CONSEIL D'ÉTAT

Etude à la demande du Premier ministre

Intelligence artificielle et action publique : construire la confiance, servir la performance

Etude adoptée
en assemblée générale plénière
du 31/03/2022

SOMMAIRE

Synthèse 5

Introduction 15

1. Construire un langage commun et intelligible de l'intelligence artificielle 21

1.1. Comprendre ce qu'est (et ce que n'est pas) l'IA pour bâtir un langage partagé 21

1.1.1. L'IA, variations sur un thème.....22

1.1.2. Les systèmes d'intelligence artificielle relèvent d'approches et de techniques différentes 30

1.1.3. Les systèmes d'IA peuvent remplir les fonctions les plus variées.....40

1.2. Promouvoir une compréhension partagée et rationnelle de l'IA..... 45

1.2.1. Écarter le fantasme de la « singularité ».....45

1.2.2 Permettre aux citoyens d'acquérir une culture des concepts et enjeux de l'IA 47

1.2.3 Œuvrer à des concepts juridiques communs aux niveaux européen et mondial.....50

2. Accélérer le déploiement des systèmes d'IA publics pour en exploiter pleinement le potentiel 59

2.1. Le recensement des cas d'usage de l'IA publique révèle le remarquable potentiel des SIA..... 61

2.1.1. Panorama général des cas d'usage de l'IA publique.....61

2.1.2. Les bénéfices attendus des systèmes d'IA dans le secteur public 71

2.2. Le potentiel des SIA reste encore à exploiter au sein de la sphère publique 81

2.3. Conduire une stratégie volontariste et lucide..... 84

2.3.1. Le secteur public ne doit pas attendre le moment, mais le créer 84

2.3.2. Les points de vigilance stratégique.....87

3. Définir et mettre en œuvre les principes et méthodes de l'IA publique de confiance.....93

3.1. Définir les principes généraux de l'IA publique de confiance..... 98

3.1.1. Primauté humaine 98

3.1.2. Performance.....108

3.1.3. Equité et non-discrimination.....113

3.1.4. Transparence.....118

3.1.5. Sûreté (cybersécurité).....126

3.1.6. Soutenabilité environnementale.....127

3.1.7. Autonomie stratégique 130

3.2. La mise en œuvre normative de l'IA publique de confiance 134
3.2.1. Le règlement IA et les espaces d'expression du droit national 134
3.2.2. L'apprentissage de la liberté : les lignes directrices de l'IA publique de confiance 138
3.3. Le recours juridictionnel, facteur-clé de confiance dans l'IA publique 144
3.3.1. Les voies de recours contre le système et ses effets 144
3.3.2. La mise en jeu de la responsabilité de l'administration devant le juge administratif 148
3.3.3. La responsabilité pénale du fait des SIA publics 153
4. Doter la France des ressources et de la gouvernance adaptées à l'ambition.....157
4.1. Doter les administrations des ressources nécessaires 157
4.1.1. Faire ou acheter ?.....157
4.1.2. Piloter une stratégie RH de l'IA publique 162
4.1.3. Libérer le potentiel de valeur de la donnée, carburant des systèmes d'IA.....169
4.1.4. Se doter des ressources techniques et financières adaptées.....186
4.1.5. Diffuser des repères méthodologiques communs.....190
4.2. Construire une gouvernance tournée vers l'innovation et la confiance 193
4.2.1. Assurer le pilotage de la stratégie de l'IA publique 193
4.2.2. Favoriser la mise en œuvre opérationnelle des SIA.....198
4.2.3. Confier le pilotage de la régulation des SIA à une CNIL transformée 200
4.2.4. Structurer la réflexion sur l'éthique et l'implication civique 203
Conclusion.....207
Annexes 209
Annexe 1 : Lettre de mission..... 211
Annexe 2 : Composition du groupe de travail..... 213
Annexe 3 : Liste des auditions conduites par la mission 214
Annexe 4 : Glossaire..... 222
Annexe 5 : Quelques dates-clés sur l'intelligence artificielle..... 237
Annexe 6 : Synthèse du projet de règlement européen sur l'IA 241
Annexe 7 : Méthodologie et bilan du questionnaire adressé aux administrations 252
Annexe 8 : Présentation des acteurs de l'écosystème de l'IA publique..... 262
Annexe 9 : Cartographie des cas d'usage des systèmes d'IA dans l'action publique..... 267
Annexe 10 : Note sur le cadre juridique actuel des systèmes d'IA dans la sphère publique..... 347

Synthèse

Domaine de recherche effervescent et enjeu économique, social et de souveraineté majeur, l'« intelligence artificielle » (IA) a, comparativement, suscité peu de littérature en ce qui concerne son utilisation dans le service public. Cette étude, commandée par le Premier ministre au Conseil d'État, entend alimenter la réflexion sur les concepts, les usages, les enjeux juridiques et éthiques et, plus largement, les conditions d'un déploiement pertinent et réussi des outils s'y rattachant au sein de la sphère publique, tant de l'État que des collectivités territoriales et des établissements publics, au service des citoyens et des agents publics. Loin d'être un exercice académique, elle se conçoit comme une **contribution à une stratégie de l'IA publique** qui reste largement à structurer et à formaliser par les pouvoirs publics, dans le sillage de la stratégie nationale pour l'intelligence artificielle lancée en 2018.

*

Sortir de la science-fiction et parler la même langue

L'IA suscite autant d'espoirs que de fantasmes. Les avancées technologiques récentes de l'apprentissage automatique, illustrées au premier chef par les « réseaux de neurones » employés pour « l'apprentissage profond » (« *deep learning* »), ont abouti à des réalisations spectaculaires, notamment dans les domaines de la vision par ordinateur (notamment la reconnaissance de personnes, d'objets ou de formes sur des images et la création automatique de contenus vidéos) et du traitement du langage (par exemple l'analyse sémantique, consistant à identifier le sens d'un texte par-delà les mots utilisés et à générer une réponse pertinente, comme le fait une enceinte connectée). Elles ont, dans le même temps, exacerbé les craintes, largement exagérées, de l'asservissement de l'humain par la machine, voire d'une prise de pouvoir par cette dernière, de la manipulation des comportements et de la surveillance de masse. Elles ont également alimenté des discours lénifiants et des stratégies commerciales abusives, promettant des « prédictions » d'ordre quasi-divinatoire là où se déployaient de simples analyses statistiques.

La très forte charge symbolique de l'expression « intelligence artificielle », ainsi que **l'absence de définition partagée et de consensus sur le contenu même de la notion**, contribuent puissamment à la confusion et compliquent l'examen rationnel des avantages et des inconvénients de ce qui est, d'abord et avant tout, un ensemble d'outils numériques au service de l'humain, qu'on peut regrouper sous le vocable de « systèmes d'IA » (SIA).

Entre deux conceptions extrêmes, l'une assimilant tout système d'information ou application numérique à l'IA, l'autre n'y voyant qu'un horizon de recherche ou un simple point de passage de l'innovation dont sont exclus les systèmes à mesure qu'ils sont créés et maîtrisés, coexistent de multiples conceptions intermédiaires qui appellent un choix.

Dans une acception étroite très répandue, les SIA sont limités à l'apprentissage automatique (dite « IA connexionniste »), approche qui consiste, en substance, à nourrir la machine d'exemples (par exemple, des millions d'images de chat) afin qu'elle en déduise les règles pertinentes (les traits caractéristiques d'un chat) pour résoudre un problème (identifier un chat sur une nouvelle image). Dans une acception plus large, qui est privilégiée par le projet de règlement européen en cours de négociation et par cette étude, les SIA incluent des systèmes dont les règles de fonctionnement sont paramétrées explicitement par l'homme (IA dite « symbolique ») mais qui disposent d'une certaine latitude pour déterminer la solution satisfaisante ou optimale d'un problème complexe (exemple des systèmes-experts).

Prendre conscience du potentiel de performance des systèmes d'IA et de sa sous-exploitation par les collectivités publiques

En l'absence de recensement préexistant des cas d'usage des SIA publics, en particulier ceux qui relèvent de l'apprentissage automatique, le Conseil d'État s'est livré à un exercice de cartographie illustrative (annexe 9) dont se dégagent trois conclusions :

En premier lieu, aucun domaine de l'action publique n'est imperméable à ces systèmes et n'a, *a priori*, vocation à l'être. A titre d'exemples non hiérarchisés, des SIA sont d'ores et déjà opérationnels dans les domaines les plus divers :

- La gestion des territoires : circulation automobile, entretien de la voirie, gestion des déchets, de l'eau, de l'éclairage public, du nettoyage urbain, transport public par véhicule dit « autonome »… ;
- La défense et la sécurité : détection de forces militaires sur des images aériennes et satellite, prévention des attaques informatiques, détection de la désinformation d'origine étrangère avec Viginum, lecture automatisée de plaque d'immatriculation, anticipation des catastrophes naturelles par les services de secours, reconnaissance faciale de suspects ou de victimes par la police judiciaire… ;
- Les activités de contrôle et de lutte contre la fraude : ciblage des contrôles fiscaux et douaniers, contrôle aux frontières, détection de constructions non autorisées sur des images satellite… ;
- La justice : pseudonymisation des jugements, recherche documentaire, évaluation des préjudices en cas de dommage corporel… ;
- La politique de l'emploi : appariement entre offre et demande d'emplois, personnalisation de l'accompagnement… ;
- L'éducation : prévention du décrochage scolaire, affectation en première année de l'enseignement supérieur… ;

- La protection sociale : liquidation des prestations, identification du non-recours aux droits… ;
- La santé : aide au diagnostic et à la prescription médicale, alertes sanitaires, robotique médicale…

En deuxième lieu, à l'instar de ses voisins européens, la France ne vit pas une révolution de l'IA publique, mais connaît un déploiement très progressif des SIA dans les services publics, très inégal selon les administrations et souvent expérimental.

Déjà largement déployés en soutien des activités de contrôle et de lutte contre les infractions (profilage des fraudeurs sur la base des manquements déjà observés par le passé et définition des stratégies de contrôle…), ils sont aussi régulièrement utilisés pour la fourniture de renseignements aux citoyens (« robots conversationnels », dits *chatbots*, par exemple dans la sphère « sécurité sociale ») et, dans une moindre mesure, pour l'automatisation de tâches répétitives et fastidieuses. Ils sont encore assez peu exploités dans l'assistance à la gestion des ressources humaines (personnalisation des plans de formation, gestion des affectations, recrutements, robots conversationnels internes…). La prise de décision entièrement automatisée reste très minoritaire, les SIA assurant, le plus souvent, une fonction d'aide à la décision (prévision des pannes, cartographie de la délinquance, appui aux professeurs des écoles pour la personnalisation des exercices donnés aux élèves en fonction de leur niveau, détection d'une pathologie sur une radio…) ou la prise en charge d'une tâche périphérique (par exemple, la traduction automatique d'un texte ou sa synthèse).

En troisième lieu, les objectifs poursuivis et les bénéfices attendus du déploiement des SIA dans l'action publique sont nombreux. Ils ont vocation à permettre d'**améliorer la qualité du service public** dans ses multiples dimensions, tant en ce qui concerne la pertinence des décisions et prestations délivrées (par exemple via l'identification précoce des entreprises en difficulté pour déclencher au plus tôt des actions de soutien), la continuité du service public (guichet numérique 24/7) et la réduction du temps d'instruction des demandes (de prestations sociales, de permis de construire…), l'accomplissement de tâches matériellement impossibles à accomplir avec les ressources humaines disponibles (identifier une personne à partir des milliers de photos enregistrées dans un fichier, retrouver l'ensemble des informations pertinentes sur un sujet donné au sein de giga-octets de documents), une meilleure égalité de traitement, l' « égalité des armes » entre les autorités de contrôle et les opérateurs contrôlés, ou encore la neutralisation ou la réduction, pour l'usager, de la complexité administrative.

Les SIA, s'ils sont bien conçus et efficacement déployés, permettent d'**optimiser l'emploi des ressources publiques**, matérielles comme humaines. A la faveur de l'automatisation d'une activité, les agents publics peuvent être redéployés sur d'autres missions, sur les cas complexes ou pour apporter une assistance personnalisée à des publics en difficulté. Si le remplacement massif des agents reste purement théorique en l'état des techniques et largement hypothétique, il n'est pas contestable que ces outils auront un impact croissant sur l'emploi public et

impliqueront, dans le cadre de la gestion prévisionnelle des emplois et des compétences, d'accompagner ceux occupant des postes ayant vocation soit à se transformer, soit à disparaître : le choix de réduire ou redéployer les effectifs appartient aux pouvoirs publics, pas aux SIA.

Enfin, l'engagement des administrations dans le développement de SIA constitue un **levier de compétitivité économique**, tant par le canal de la recherche que par celui de la commande publique.

Conduire une stratégie volontariste et lucide de déploiement de l'IA publique de confiance

Ces promesses attachées aux SIA publics, qui, lucidement évalués, n'ont rien d'un effet de mode, plaident pour la **conduite d'une stratégie de conception et de déploiement résolument volontariste**, au service de la performance publique et, à travers elle, de l'intérêt général, en réponse aux attentes croissantes des citoyens et en appui des femmes et des hommes qui les servent. La France ne doit pas attendre passivement le moment, mais le créer. Elle n'en tirera toutefois des bénéfices, et tous les bénéfices, qu'à trois conditions.

Créer les conditions de la confiance

La première d'entre elles est de **créer les conditions de la confiance**. L'acceptabilité sociale des SIA publics n'est pas définitivement acquise et ne le sera sans doute jamais entièrement, du point de vue tant des citoyens que des agents publics. Elle constitue aujourd'hui, tant au sein de l'État que des collectivités locales, et plus encore peut être au niveau de ces dernières (ainsi qu'en témoigne l'adoption préventive de chartes par certaines d'entre elles), un frein sérieux au déploiement des SIA, dans un contexte général de défiance croissante à l'égard des autorités publiques et alors que pèse sur l'administration un devoir particulier d'exemplarité.

La construction de cette « IA publique de confiance » comporte plusieurs enjeux :

- Il est impératif et urgent de **rehausser le niveau de compréhension des citoyens comme des agents publics** sur ce qu'est et ce que n'est pas l'intelligence artificielle, sur ce que ces systèmes permettent de faire ou d'espérer et ce qui leur est inaccessible, sur leur potentiel et les risques qu'ils comportent. Assis sur des concepts clarifiés et partagés et des illustrations concrètes tirées de la vie quotidienne, ce travail d'acculturation et de sensibilisation doit être une priorité pour faire pièce au « mythe de la singularité », qui marque l'avènement du règne d'une machine ayant définitivement surclassé l'homme en tous domaines, et mesurer ce qui sépare les capacités actuelles, limitées, des SIA, de « l'IA générale », capable d'accomplir l'ensemble des tâches habituellement dévolues aux humains et qui est aujourd'hui de l'ordre de la science-fiction. Cela implique notamment, au-delà de la diffusion d'une culture du numérique au sein de la société, d'associer, concrètement, les citoyens usagers, les partenaires sociaux et les représentants de la société civile à la conception et au déploiement de la stratégie de l'IA publique.

- Les pouvoirs publics doivent également définir une **doctrine administrative de l'IA de confiance, reposant sur un ensemble de principes fondamentaux**, déclinés en exigences opérationnelles et mis en œuvre par des mesures juridiques, organisationnelles, techniques, de pédagogie, de formation et de gouvernance… L'étude propose à cet égard sept principes de l'IA publique de confiance, qui doivent être conciliés entre eux et appellent des arbitrages, et à la mise en œuvre desquels il importe d'associer autant que possible l'ensemble des parties prenantes, en particulier les partenaires sociaux, et les représentants des usagers-citoyens.

Sept principes de l'IA publique de confiance

1. La primauté humaine. Les SIA publics se conçoivent comme des outils au service de l'humain, ce qui suppose qu'ils répondent à une finalité d'intérêt général et que l'ingérence dans les droits et libertés fondamentaux qui résulte de leur mise en service ne soit pas disproportionnée au regard des bénéfices qui en sont attendus. En outre, l'humain doit se porter garant du bon fonctionnement du SIA en le supervisant (grâce à des mesures techniques, juridiques, de formation et de gouvernance), y compris en cas de recours à un outil d'aide à la décision, l'humain étant en général prompt à entériner les résultats proposés par la machine (biais d'automatisation). Enfin, l'humain doit anticiper le risque d'un dysfonctionnement du système, en limitant sa dépendance, et en assumer les conséquences, l'erreur de la machine n'étant, indirectement, qu'une erreur humaine.

2. La performance. La dégradation de la qualité d'un service en raison de son automatisation est un des facteurs les plus destructeurs de la confiance dans les outils numériques, en particulier lorsque, s'agissant du service public, les usagers n'ont pas la possibilité de se tourner vers un concurrent. Les administrations doivent donc identifier les indicateurs de la performance du système (exactitude, robustesse technique, temps de réponse, etc.), et définir, au regard des conséquences de l'erreur, le niveau de performance acceptable, en veillant à ne pas détériorer la qualité du service rendu.

3. L'équité et la non-discrimination. Les concepteurs des SIA doivent choisir, parmi les différentes conceptions de l'équité, celle qui guidera le fonctionnement des systèmes et formaliser ce choix, dans le respect du principe d'égalité. Ils doivent en outre veiller à prévenir les discriminations involontaires, enjeu particulièrement prégnant pour les SIA d'aide à la décision fondés sur l'apprentissage machine. Entraînés sur de vastes jeux de données susceptibles de renfermer des biais, ces systèmes peuvent les reproduire et produire des résultats pénalisants pour certaines catégories de personnes. Ce principe implique la mise en place d'un système de gestion des risques comportant une analyse critique à toute étape (entraînement et déploiement), une sensibilisation des agents chargés des SIA à cette problématique spécifique voire une plus grande représentativité sociale des équipes de conception.

4. La transparence. Ce principe comporte, à tout le moins, le droit d'accès à la documentation du système, une exigence de loyauté consistant à informer les personnes de l'utilisation d'un SIA à leur égard, l'auditabilité du système par les

autorités compétentes ainsi que la garantie d'explicabilité. La complexité technique de certains SIA, en particulier ceux qui reposent sur l'apprentissage profond, et la difficulté ou l'incapacité de formaliser le raisonnement ayant conduit au résultat produit, risquent d'accentuer le sentiment de défiance si les personnes sur lesquelles ce résultat a une incidence ne peuvent obtenir, dans un langage simple, une explication sur les principaux ressorts de la décision ou de la recommandation formulée par le système.

5. La sûreté (cybersécurité). Tout SIA doit intégrer l'enjeu de sûreté, c'est-à-dire la prévention des attaques informatiques et la résolution de leurs conséquences. En même temps qu'ils peuvent contribuer à la détection et à la résolution des menaces en la matière, les SIA présentent des vulnérabilités particulières (l'empoisonnement des données d'apprentissage, le leurre du système ou le vol des données) qu'il est impératif d'anticiper et de contrer.

6. La soutenabilité environnementale. Les SIA n'ont pas d'impact écologique de nature différente de celui de l'ensemble des technologies numériques, mais leur généralisation concourrait, toutes choses égales par ailleurs, à une aggravation significative de la crise environnementale en cours, du fait de l'accroissement du besoin en terres rares, de l'artificialisation des sols et, surtout, de la consommation d'électricité qu'ils induisent. L'impact environnemental des SIA doit donc être pris en compte dans la stratégie de l'IA publique en général, autour d'un principe de neutralité globale de l'IA, comme dans la conception de chaque système, en mettant en regard le surcroît de performance permis par une puissance de calcul supérieure avec son empreinte écologique.

7. L'autonomie stratégique. Dès lors que les SIA concourent de façon croissante aux fonctions essentielles de la puissance publique, ils doivent être conçus de manière à garantir l'autonomie de la Nation. Si l'autarcie numérique serait un objectif illusoire et contre-productif, la France doit se doter des ressources nécessaires, en matière de compétences, de structures de recherche, d'infrastructures et de données, pour réduire et choisir ses dépendances.

- Les responsables publics doivent veiller, dans la stratégie de déploiement, à un **équilibre des usages**. L'acceptabilité par les agents publics suppose de ne pas négliger les **usages internes** des SIA, au service de la qualité de vie au travail et de l'appui à la gestion des carrières. Surtout, le risque existe que le regard des citoyens sur les SIA ne soit structuré par la place qu'occupent, dans l'action administrative, les systèmes dédiés au contrôle et à la surveillance ; il est donc crucial d'investir dans les mêmes proportions dans les **SIA de « service »** et de les valoriser, idéalement au sein d'une même politique publique (à titre d'exemple, les SIA servent aussi bien à l'identification des cas de fraude aux prestations sociales qu'à ceux du non-recours à ces prestations par des personnes y ayant droit, de même que la reconnaissance faciale dans les lieux publics peut permettre de confondre un terroriste comme de retrouver un enfant enlevé ou perdu dans une foule).

- Afin d'anticiper sur l'éventuelle entrée en application du règlement européen et de guider les choix des administrations, les pouvoirs publics devraient, de préférence à l'adoption d'une législation-cadre qui risque de s'avérer trop rigide et rapidement dépassée, élaborer des **lignes directrices de l'IA publique de confiance**, qui formaliseraient tout à la fois la stratégie, la doctrine d'emploi et la méthodologie pratique de conception, de déploiement et d'utilisation des SIA au sein de la sphère publique. Ces lignes directrices procèderaient à une uniformisation terminologique, indiqueraient les conditions de recours à certains SIA sensibles (notamment pour le recours à la prise de décision automatisée) et rappelleraient les règles de droit d'ores et déjà applicables en veillant à éclairer les administrations sur l'articulation des différents régimes existants. Leur élaboration devrait associer le Parlement et l'ensemble des parties prenantes (partenaires sociaux, agents, représentants des usagers, secteur privé…).

- Il ne peut y avoir d'IA publique de confiance sans **droit au recours effectif**. D'ores et déjà, des voies de droit existent pour contester les décisions administratives produites par un SIA ou avec son appui, et pour saisir les traitements de données à caractère personnel que constituent les systèmes qui ont recours à ces dernières. Le régime juridique des décisions de recourir à un SIA sera défini progressivement par la jurisprudence, le cas échéant en distinguant selon que le système affecte ou non des droits et libertés fondamentaux. Les conditions d'engagement de la **responsabilité de l'administration** à raison d'un système défaillant seront également éclairées par le juge, sans qu'il soit utile ni pertinent de conférer la personnalité morale au système lui-même : la victime d'un dommage causé par un SIA dont l'administration est la gardienne ne devrait pas être contrainte de rechercher la responsabilité d'un autre acteur ayant participé à la conception ou au développement du système. Le régime actuel des infractions non intentionnelles apparaît quant à lui suffisant pour discipliner les autorités publiques sur le terrain pénal.

Faire preuve de lucidité et de vigilance dans le déploiement

La deuxième condition de succès d'une stratégie volontariste est la **lucidité et la vigilance dans le déploiement**. Le volontarisme ne saurait se confondre avec le « solutionnisme », alors que les SIA ne sont pas toujours une réponse pertinente, ou la meilleure réponse, au problème posé. Ils ne sont qu'un moyen parmi de nombreux autres qui composent la boîte à outils de l'action publique. Recourir à un SIA ne saurait donc constituer une fin en soi. Ces systèmes ne résoudront pas à eux seuls les immenses défis auxquels la France est confrontée pas plus qu'ils ne doivent être le prétexte à la banalisation d'une complexité administrative et juridique « gérée » par la machine.

En outre, le potentiel considérable de ces systèmes ne doit pas faire perdre de vue leur complexité et la technicité de leur conception, qui ne s'improvise pas. L'administration ne doit pas rêver d'emblée d'un grand soir de l'IA, transformant les politiques publiques, les organisations et la façon même de rendre le service public, mais s'éveiller au petit matin de l'apprentissage automatique, en cheminant par étapes et en intégrant la logique de l'innovation fondée sur l'essai-erreur.

Doter la France des ressources et de la gouvernance adaptées à l'ambition

Les ressources nécessaires à la conception, à l'adaptation et à l'utilisation des systèmes d'IA publics sont à la fois humaines, techniques et financières, organisationnelles et juridiques. Leur calibrage dépend de multiples paramètres, notamment de **l'arbitrage entre le développement en interne (« faire ») ou le recours à un prestataire (« acheter »)**.

Cet arbitrage ne peut être fait abstraitement. Il doit notamment tenir compte de l'enjeu de montée en compétence des administrations, du particularisme des systèmes d'IA dont l'apprentissage se poursuit en phase d'utilisation, et de la disponibilité, de la maturité et de l'adaptabilité des solutions du marché. L'administration doit avoir conscience que l'externalisation suppose de disposer de ressources internes de pilotage et qu'elle n'est pas nécessairement moins chère que la conception interne, surtout sur le long terme, et qu'elle suppose de respecter les règles de la commande publique et de faire usage des souplesses qu'elles ménagent, auxquelles les administrations n'ont pas assez recours. La mutualisation de briques logicielles entre administrations est une voie prometteuse, dans le respect des règles de concurrence et de la propriété intellectuelle, de même que la diffusion de **recommandations méthodologiques** à destination des administrations qui ne sont pas encore rompues à l'exercice.

S'agissant des ressources humaines, l'enjeu premier est, au-delà de l'indispensable acculturation des agents publics, de **former les dirigeants publics** dont l'engagement personnel et le soutien constituent des facteurs-clés de succès des projets de SIA, là où, aujourd'hui, la faible culture des responsables publics peut être la source d'un désintérêt ou d'une défiance.

En ce qui concerne les **experts des données**, l'administration, relativement attractive pour les profils peu expérimentés, doit poursuivre les efforts déjà engagés pour diversifier les recrutements, les fidéliser et exploiter davantage l'exceptionnel potentiel dont elle dispose en la matière au sein des services statistiques. Des facilités particulières devraient leur être reconnues au regard des plafonds d'emplois et des budgets de personnel pour recruter de tels experts dans des équipes disposant d'une taille critique, compte tenu du retour sur investissement, alors que les programmes de subvention encouragent au contraire à privilégier l'externalisation. La fidélisation des experts, comme le recrutement de profils plus expérimentés, reste un point faible pour des raisons touchant à la rémunération, au manque de perspectives professionnelles et aux pesanteurs du fonctionnement administratif.

La disponibilité de **données** nombreuses et de qualité constitue également un déterminant essentiel de la capacité de l'administration à concevoir des systèmes d'IA basés sur l'apprentissage automatique. Les données d'entraînement sont le carburant des modèles algorithmiques.

En dépit de nombreux progrès réalisés ces dernières années, sous l'impulsion d'Etalab, en particulier dans le cadre de la politique de diffusion publique des données (« *open data* »), subsistent encore d'importants **freins culturels** au partage interne à la sphère publique, aggravés par la difficulté dans laquelle se trouvent les

administrations productrices pour tirer de leurs données une valeur souvent captée par le secteur privé. S'y ajoutent des **freins techniques** liés, notamment, au recours à des logiciels non interopérables et au manque d'interfaces de partage (API). Le principal enjeu est toutefois de nature **juridique**.

Un assouplissement du **cadre juridique du partage intra-public des données** devrait être sérieusement examiné, alors que les administrations ne disposent pas, à l'heure actuelle, de plus de droits que les citoyens, sauf dispositions particulières. En outre, les contraintes découlant du RGPD pour l'utilisation des **données à caractère personnel** à des fins de conception des systèmes d'IA restent incertaines et souvent exagérées, alors que le règlement offre de nombreuses souplesses, y compris pour les traitements à d'autres fins que celles pour lesquelles les données ont été initialement collectées. L'étude propose à cet égard plusieurs clarifications.

Sauf à prendre un retard impossible à rattraper, la France doit en outre se doter des **ressources techniques** adaptées à la conception de SIA de pointe, en particulier pour ce qui concerne la puissance de calcul, ce qui constitue tout à la fois un enjeu de performance et de souveraineté. Cette course à la compétitivité technologique ne doit pas pour autant faire perdre de vue que de nombreux agents ne disposent pas de matériels informatiques de base leur permettant de travailler dans les meilleures conditions et que les systèmes d'information métier traditionnels sont trop souvent inadaptés et obsolètes.

Sur le plan organisationnel, une réflexion devrait être conduite sur les conditions de **pilotage de la stratégie de l'IA publique**. Celui-ci incombe aujourd'hui à une multiplicité d'acteurs engagés et dynamiques mais qui n'ont pas la taille critique pour exercer un effet de levier à l'échelle de la sphère publique, dont l'action respective ne fait pas l'objet d'une articulation formalisée, et qui peinent à valoriser les réalisations françaises à l'échelle européenne et internationale.

Sans verser dans l'hypertrophie centralisatrice ni introduire de rupture nette entre le développement de l'IA publique et le soutien à l'IA privée, un renforcement voire un repositionnement d'Etalab et du coordonnateur national pour l'intelligence artificielle devraient être étudiés, en lien avec l'intervention de l'Agence nationale de la cohésion des territoires, afin notamment de positionner l'État comme prestataire de services et pourvoyeur de ressources, y compris humaines, pour les collectivités territoriales.

Enfin, l'étude propose que la fonction d'**autorité de contrôle nationale responsable de la régulation des systèmes d'IA**, notamment publics, que le projet de règlement européen envisage de consacrer, soit confiée à une **CNIL profondément transformée** pour incarner et internaliser le double enjeu de la protection des droits et libertés fondamentaux, d'une part, et de l'innovation et de la performance publique, d'autre part.

*

Les orientations que dessinent l'étude visent à conjurer le double risque que le secteur public soit le spectateur passif et le régulateur distant du développement des systèmes d'intelligence artificielle par d'autres, ou un apprenti-sorcier numérique oublieux des exigences fondamentales de l'éthique de l'action publique et, en particulier, du primat de l'humain sur la technique. Depuis la loi dite « informatique et libertés » de 1978, la France est l'inspiratrice des bons usages des nouvelles technologies numériques : ce n'est qu'en poursuivant sur ce chemin, par l'engagement résolu des administrations publiques dans l'exploitation du formidable potentiel de ces systèmes au service du bien commun et de l'épanouissement humain, dans un cadre conjuguant le respect des droits et libertés et la souplesse nécessaire à l'innovation et à la performance publiques, qu'elle le restera.

Introduction

« *La machine d'arithmétique fait des effets qui approchent plus de la pensée que tout ce que font les animaux mais elle ne fait rien qui puisse faire dire qu'elle a de la volonté, comme les animaux.* »

Pascal, *Pensées*, CXLIX

« *Il ne connaît pas la pitié, ni les remords, ni la peur et rien au monde ne peut l'arrêter, personne.* »

Terminator, dir. James Cameron (1984)

Il n'est guère de sujet plus en vogue que celui de l'intelligence artificielle. L'effervescence scientifique dont elle fait l'objet, les enjeux économiques qui l'entourent et sa visibilité médiatique alimentent une production vertigineuse d'articles[1], d'actes de colloques[2], de rapports, de stratégies et de plans d'actions.

Omniprésente, l'« IA » l'est aussi, et sans qu'on le sache toujours, dans la vie de tous les jours, aussi imperceptible qu'indispensable pour des millions de Français. Après avoir déverrouillé son smartphone par reconnaissance faciale, consulté sa messagerie électronique d'où les courriels indésirables ont été automatiquement expurgés et pris connaissance d'une publicité pour une destination touristique à laquelle elle s'était intéressée la veille sur internet, une personne qui envoie un message écrit à son conjoint à l'aide d'une fonctionnalité de saisie intuitive (suggestion de mots) et un message vocal à un ami pour le prévenir de l'heure de son arrivée avant de prendre la route par l'itinéraire présenté comme le meilleur par l'application de navigation, à une vitesse régulée par sa voiture en fonction des panneaux de signalisation qui bordent la route et en se garant à l'aide de l'assistance embarquée, a utilisé ou subi, en l'espace de quelques minutes, plus d'une demi-douzaine d'outils recourant à l'intelligence artificielle, sans même en avoir conscience, le plus souvent.

[1] Le nombre de publications scientifiques (revues par les pairs) sur l'intelligence artificielle a été multiplié par douze entre 2000 et 2019, avec une très forte accélération ces dernières années (AI100, *AI Index Report 2021*). Ce thème représente près de 4% de l'ensemble de ces publications.

[2] La participation à la conférence de référence NeurIPS a augmenté de 800% depuis 2012.

Réalité quotidienne, l'intelligence artificielle a pourtant ceci de paradoxal qu'elle semble constamment échapper au temps présent, pour nous projeter dans un futur qui alimente, chez les uns, le rêve d'un monde parfait ou l'espoir d'une vie meilleure, chez d'autres, le cauchemar de l'asservissement de l'humanité par la machine ou la crainte de la dérive et de la manipulation mentale. Le halo de mystère voire de magie qui nimbe ce concept se nourrit assurément de la complexité technique des modèles mathématiques qui en relèvent, dont la compréhension semble devoir rester inaccessible au commun des mortels, voire aux spécialistes en raison de l'effet « boîte noire » de certains modèles algorithmiques, et des prouesses techniques qui lui sont attachées, parfois survendues par des acteurs peu scrupuleux. Il est aussi le fruit d'un **imaginaire forgé par des œuvres et récits de science-fiction** qui versent plus volontiers dans la dystopie dramatique que dans l'utopie rassurante, et en alimentent une conception déformée, voire caricaturale. La trace mythologique que laisse le fantasme d'une créature quasi humaine plus ou moins asservie est longue et explique sans doute beaucoup des réactions spontanées d'effroi ou de fascination.

La notion même d'intelligence artificielle, comme le sigle « IA », **y contribuent puissamment**, tant par la symbolique des termes que par la généricité de la formule : à défaut de pouvoir se départir de ce concept fourre-tout et de son sigle, tant ils sont désormais entrés dans les usages et ancrés dans les esprits et les discours, ces termes mériteraient à tout le moins d'être placés entre guillemets. La suite de cette étude ne s'en dispensera que pour les besoins du confort visuel du lecteur.

Si l'IA fait couler beaucoup d'encre, plus ou moins sympathique, **la question de son utilisation dans la sphère publique demeure, quant à elle, assez peu étudiée**. Le nombre modeste de rapports et d'articles spécifiquement consacrés au déploiement des outils basés sur l'intelligence artificielle dans l'action publique contraste singulièrement avec le battage médiatico-scientifique général que suscite ce thème. Comme si l'administration était vouée à en rester un acteur de second plan, agissant dans l'ombre des géants privés du numérique, au risque d'ailleurs de s'exposer au soupçon et au procès d'intention, qu'il s'agisse du remplacement subreptice de l'agent public par la machine ou des velléités de surveillance de masse de la population. Quoique le propos mériterait d'être nuancé et que la réalité est sans doute plus complexe, il faut bien admettre que **le secteur public ne bénéficie pas, dans le domaine de l'innovation numérique, du même capital confiance que les entreprises**, à tout le moins que les acteurs de référence du secteur privé.

La crise sanitaire de la covid-19 n'aura pas modifié la donne. Elle a certes servi de catalyseur à de nombreux projets et laissé entrevoir au grand public l'importance de la contribution des systèmes d'IA à la gestion des pandémies, qu'il s'agisse de la compréhension du virus et de sa dynamique de propagation, du traçage de contacts (*contact tracing*), de l'optimisation des chaînes logistiques, du contrôle du port de masque dans les lieux publics (sans évoquer les robots chinois capables de contrôler simultanément la température corporelle de dix personnes dans un rayon de cinq mètres), de l'identification des personnes vulnérables présentant un risque de développer une forme grave d'infection, ou encore de la lutte contre la désinformation sanitaire en ligne. Mais elle a aussi mis en exergue, parfois de façon exagérée, les risques qui pouvaient découler d'un mésusage ou d'un détournement

de ces technologies. De fait, le regard des Français sur l'IA se serait légèrement amélioré avec la pandémie, mais il n'aurait pas radicalement changé[3].

Si tout ou presque, le meilleur peut-être moins que le pire, semble avoir été dit sur l'intelligence artificielle en général, l'essentiel reste ainsi à écrire et, surtout, à faire pour ce qui concerne « l'IAisation » du service public. Le Premier ministre a souhaité que le Conseil d'État y apporte sa contribution, en livrant son éclairage sur le **concept** d'intelligence artificielle et ceux qui lui sont liés, sur l'état des lieux de ses **usages** dans le secteur public de notre pays et ses perspectives de développement, et sur les **garde-fous juridiques et éthiques** qui lui paraîtraient appropriés.

Conformément au cadre fixé dans la lettre de mission (annexe 1), l'analyse se concentre sur les **activités de service public menées au sein de l'ensemble des collectivités publiques** – l'État, ses établissements publics, les organismes de sécurité sociale ainsi que les collectivités territoriales. Sans éluder la question de la régulation publique de l'IA, c'est d'abord et avant tout en tant que conceptrices et utilisatrices de systèmes d'IA que les administrations seront étudiées.

Il y a toutefois lieu de préciser d'emblée que nombre de constats dressés et de préconisations formulées dans la présente étude n'ont rien de spécifique au secteur public : les techniques à l'œuvre dans le public et dans le privé sont les mêmes ; de nombreux cas d'usage peuvent être déployés indifféremment dans les deux sphères, par exemple dans le domaine des fonctions support (gestion des ressources humaines, achat, comptabilité…) ; il existe une **grande et nécessaire porosité entre les initiatives privées et publiques**, qu'il s'agisse de recherche, de partenariat, de commande publique ou de réflexion sur l'usage éthique de l'IA. L'étude s'efforce, au-delà, d'intégrer en permanence la préoccupation de n'énoncer aucune orientation qui ne tienne compte de l'ensemble des intérêts sociaux et économiques en jeu, aussi bien en termes de coûts qu'en termes de généralisation des pratiques. C'est dire aussi que le rôle traditionnel que joue l'État en France – démonstrateur, expérimentateur, innovateur – trouve ici plus encore que d'ordinaire à se déployer : **réussir la généralisation d'un usage éthique, performant, contrôlé de l'IA publique sera un puissant facteur de réussite de son déploiement dans toute la société.**

Même circonscrit aux collectivités publiques, **le champ de l'étude n'en reste pas moins d'une étendue considérable**.

L'IA est un sujet omnidisciplinaire, à la confluence des mathématiques, de l'informatique, des neurosciences, de l'éthique, de la philosophie, de la sociologie, de l'économie ou encore du droit. Elle irrigue, de fait ou potentiellement, l'ensemble des activités humaines et affecte le fonctionnement même de la société et de l'économie, ce qui lui vaut classiquement d'être rangée dans les technologies à usage général (« *general purpose technology* »), aux côtés de l'électricité et de l'Internet.

[3] Selon un sondage IFOP (RB/EB n° 117705, décembre 2020, « Notoriété et image de l'Intelligence Artificielle auprès des Français et des salariés »), plus des deux tiers des personnes interrogées ont indiqué que leur opinion sur l'IA n'avait pas évolué depuis le début de la crise sanitaire. Si 21% d'entre elles déclarent que leur regard a évolué en bien, 12% font état d'une dégradation de leur opinion.

L'utilisation de ces systèmes pour une politique publique donnée (sécurité, justice, aménagement du territoire, santé, éducation…) mériterait à elle seule la rédaction de plusieurs études.

De même, **l'impressionnante rapidité des évolutions scientifiques et technologiques** – ce qui paraissait un obstacle infranchissable hier devenant chose banale le lendemain – fait courir le risque d'une obsolescence immédiate de la réflexion : elle a conduit à ne pas formuler d'hypothèses sur les capacités des systèmes, qui peuvent se développer à un rythme inattendu, et sur la vitesse de leur déploiement dans notre quotidien qui, lui, est en revanche toujours plus lent qu'on ne l'imagine[4]. On ne se demandera donc pas si et quand des véhicules dits « autonomes » seront mis au point puis circuleront sur la route que nous empruntons, ou si un système d'IA sera capable, au XXI^{e} siècle, de produire une étude solide sur le déploiement des systèmes d'IA dans le secteur public. L'exercice de prospective est d'autant plus périlleux que l'histoire de l'IA est émaillée de périodes « hivernales » au cours desquelles les investissements s'en sont détournés en l'absence de résultats suffisamment probants et rentables. On se gardera donc d'identifier ou de prophétiser une « révolution », en s'en tenant à la prudente certitude qu'une puissante dynamique est à l'œuvre et que, au demeurant, l'état de l'art offre déjà de très intéressantes perspectives.

Le cadre juridique reste lui aussi marqué, à la date de rédaction de cette étude, par de fortes incertitudes : la proposition de règlement européen sur l'intelligence artificielle, présentée il y a moins d'un an par la Commission européenne, est en cours de négociation et a déjà fait l'objet d'un projet modifié, encore susceptible de subir des inflexions majeures. En tout état de cause, ce règlement laisserait aux Etats membres de larges marges pour sa mise en œuvre. Cette proposition de « règlement IA », résumée en annexe 10, a néanmoins constitué un point central de la réflexion juridique conduite dans le cadre de la présente étude, qui a pris tant ce projet que le RGPD comme des points de départ, sans chercher à envisager des cadres juridiques complètement alternatifs.

Il est également crucial de préciser que **la question des voies et moyens d'un développement réussi de l'IA publique ne peut être dissociée du contexte global de la numérisation de l'action publique**. D'une part, il importe peu, en définitive, que l'administration réponde à un besoin en déployant un système d'IA innovant ou en recourant à une application informatique « traditionnelle », pourvu qu'elle satisfasse effectivement ce besoin, au juste coût. Elle doit seulement songer que sa boîte s'est enrichie d'un nouvel outil, dont la puissance potentielle n'en modifie pas la nature : celle d'un moyen parmi d'autres, à mettre au service de l'intérêt général. D'autre part, le Conseil d'État est parfaitement conscient du décalage qui peut être ressenti par nombre d'administrations, notamment dans les collectivités territoriales

[4] Il ne suffit pas qu'une technologie soit disponible pour qu'elle soit instantanément adoptée et diffusée, et pour qu'elle remplace des usages reposant sur des technologies plus anciennes, comme le montre le fait que le courrier électronique n'a pas instantanément fait disparaître le courrier papier. Cette inertie s'explique par de nombreux facteurs d'ordre économique, technique, organisationnel et culturel.

de petite taille, mais aussi dans les administrations déconcentrées ou dans des administrations centrales dont les budgets de systèmes d'information ont parfois été négligés, entre l'ambition affichée d'un déploiement large et rapide de l'IA publique et la réalité quotidienne d'agents confrontés au dénuement informatique. Il va de soi que **la remise à niveau des équipements et des systèmes doit constituer une priorité**.

Ces préalables étant posés, la présente étude du Conseil d'État se conçoit, pour le tout, comme une proposition : celle de **poser des points de repère** pour la construction itérative d'une stratégie de l'IA publique faite d'une succession de choix majeurs – politiques, opérationnels, éthiques, juridiques… – que d'autres études devront préalablement éclairer en fonction de l'état de l'art au moment où elles seront réalisées. Le lecteur n'y trouvera pas la planification à dix ou trente ans des cas d'usage de l'IA publique, non plus qu'un catalogue de mesures assorti d'échéanciers colorés, de dispositions législatives pléthoriques, de structures nouvelles et de crédits supplémentaires, mais un espace commun de réflexion pour la sphère publique, débarrassé des mystifications qui accompagnent trop souvent les discours sur l'IA, offrant un éclairage sur les potentialités qu'elle offre et les conditions requises pour sa pleine exploitation, dessinant les précautions à prendre pour renforcer ou, à tout le moins, ne pas rompre la confiance avec les parties prenantes, en particulier les citoyens et les agents, et ambitionnant, modestement, de donner envie aux administrations de s'emparer de ces outils, d'apprivoiser ces techniques et d'imaginer avec elles le service public de demain. **L'étude assume à cet égard d'aborder le sujet par les opportunités avant d'en exposer les risques**, à rebours d'un réflexe souvent constaté dans la littérature, en particulier celle qui émane de la communauté juridique.

Il n'y a pas lieu pour autant de céder à une technophilie béate qui prendrait le moyen pour la fin. Car les systèmes d'intelligence artificielle ne sont qu'un **moyen, au service des collectivités publiques** pour les aider à maîtriser la complexité de leurs univers d'action et accroître la performance de leur action au bénéfice des citoyens et des entreprises, **et au service des agents publics**, pour les assister dans l'exercice de leurs missions, dans des conditions de travail de qualité. Le paradoxe n'est qu'apparent : le large déploiement des systèmes d'intelligence artificielle peut ouvrir la voie à une **réhumanisation** du service public partout où il en est besoin et à une plus grande individualisation des décisions, sans impliquer la moindre renonciation aux exigences de l'éthique publique ni à aucune des ambitions que la société demande à la puissance publique de porter. L'ambition de cette étude est d'en convaincre l'ensemble des parties prenantes, et notamment les autorités publiques, et de leur fournir quelques repères pour avancer dans cette direction.

Celle-ci aborde successivement les quatre axes stratégiques d'une « IA publique » de confiance, au service de la performance publique :

I – Adopter et utiliser des concepts communs et compris par le plus grand nombre, appuyés sur des définitions claires, rompant avec les fantasmes, la confusion et l'imprécision terminologiques qui entourent l'« IA ».

II – S'engager résolument et sans tarder dans le déploiement des systèmes d'IA pour en exploiter tout le potentiel et les externalités positives, en amplifiant les initiatives déjà engagées.

III – Définir et mettre en œuvre opérationnellement une doctrine de l'IA publique de confiance, qui suppose de respecter certains principes et exigences permettant de maîtriser les risques de toute nature (individuels et collectifs) afin de garantir en toutes circonstances l'utilité publique des systèmes d'IA.

IV – Garantir la disponibilité des ressources indispensables à la conception, au déploiement et à la maintenance des systèmes d'IA (ressources humaines, données, moyens financiers, infrastructures…) **et mettre en place une gouvernance adaptée à la nécessaire ambition de l'IA publique.**

1. Construire un langage commun et intelligible de l'intelligence artificielle

Le premier enjeu de la réflexion et de l'action en matière d'intelligence artificielle consiste à comprendre de quoi il est concrètement question. Il ne s'agit pas ici de se livrer à un exercice académique de définition, mais d'un préalable nécessaire à toute stratégie efficace de l'IA publique.

La démystification de la notion est une condition de l'acceptation citoyenne du déploiement de services qui y font appel. Sa bonne compréhension est indispensable à son appropriation par les dirigeants publics et par l'ensemble des administrations. La construction d'un vocabulaire et d'une grammaire communs est nécessaire à l'efficacité des échanges et des travaux sur ce thème, comme à la pertinence des choix des acheteurs publics confrontés à une offre profuse et parfois confuse, volontairement ou non. Enfin, l'élaboration d'un cadre juridique propre à l'IA, là où il est nécessaire, est impossible ou contre-productif en l'absence de définitions claires et non équivoques.

1.1. Comprendre ce qu'est (et ce que n'est pas) l'IA pour bâtir un langage partagé

Dans le domaine de l'intelligence artificielle, l'exercice de terminologie et de définition constitue en soi un **défi** et un **enjeu**.

Un **défi**, car l'absence de définition universellement admise de l'intelligence artificielle, la diversité des approches et des réalités rattachées à ce vocable selon les interlocuteurs, ainsi que la créativité terminologique et le foisonnement des concepts qui s'y rattachent, souvent d'origine anglo-saxonne (V. sur ce point le glossaire en annexe 4), créent une forte incertitude sur les contours exacts de l'IA et du champ lexical qui lui est associé. S'y ajoute le degré de complexité parfois élevé des modèles algorithmiques qui ne permet pas toujours de se représenter mentalement le fonctionnement du système et rend difficile son appréhension par des profanes.

Un **enjeu**, car la forte charge symbolique de la notion d'intelligence artificielle dans l'imaginaire collectif, la valeur commerciale attachée à la notion et les velléités d'instrumentalisation et de dévoiement à des fins de marketing (parfois désignées comme des pratiques d' « *IA-washing* », par analogie avec le « *greenwashing* », ou verdissement) ou encore l'importance de disposer d'un langage commun maîtrisé par la multiplicité des acteurs (scientifiques, ingénieurs, agents du « métier », juristes…) impliqués dans les politiques et projets qui s'y rapportent, appellent, en la matière plus que dans bien d'autres, un effort de clarification, de convergence et de normalisation terminologiques.

1.1.1. L'IA, variations sur un thème

Il n'existe à ce jour aucune définition de l'intelligence artificielle qui ferait consensus parmi l'ensemble des acteurs, qu'ils soient scientifiques, politiques, économiques ou administratifs, en dépit de nombreux travaux académiques menés pour recenser, comparer et synthétiser les dizaines de définitions proposées au fil du temps[5].

Parmi les multiples conceptions de cette notion, deux sont particulièrement utiles à sa compréhension : l'IA peut aussi bien se concevoir, sur le plan fonctionnel, comme un (ensemble de) produit(s) ou de service(s) que, sur le plan académique, comme une discipline.

1. L'IA comme produit ou service

Le sigle « IA », voire le terme abusivement personnifiant « une IA », **est souvent utilisé par métonymie pour désigner le produit ou le service qui utilise un ou des programmes informatiques relevant de l'intelligence artificielle**.

Ce produit ou ce service peut être décrit par référence à sa **finalité**, aussi dénommée « destination », qui correspond à la fonction qu'il remplit (« une IA de météo » pour prévoir le temps qu'il fera demain, « une IA de délivrance de carte grise » pour l'automatisation de l'octroi du certificat d'immatriculation, « une IA de *scoring* financier » pour l'octroi d'un crédit immobilier compte tenu du profil de risque du client, « une IA de traduction automatique » pour la traduction en français d'un document en anglais…). Le système peut également être désigné en référence à d'autres caractéristiques, comme le **type de modèle utilisé** ou « approche » (« une IA d'apprentissage automatique », un « système-expert » utilisant des règles prédéfinies par l'homme…)[6] ou sa fonctionnalité générique (« une IA de reconnaissance d'image » pour la vision par ordinateur, une « IA de génération de texte » en matière de traitement du langage naturel, « une IA prédictive » pour la formulation d'une prévision de résultat à partir d'une situation de fait)[7].

---

[5] V. par ex. ce travail comparatif mené sur 55 documents émanant d'organismes publics ou parapublics : S. Samoili, M. López Cobo, E. Gómez, G. De Prato, F. Martínez-Plumed et B. Delipetrev, *AI Watch : Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*, Publications Office of the European Union, 2020.

[6] V. 1.1.2. sur la description des différentes approches.

[7] V. 1.1.3. pour la présentation synthétique des principales briques fonctionnelles.

Par analogie avec le terme usuel de « systèmes d'information »[8], la présente étude recourra à la notion, également utilisée au niveau européen, de système d'intelligence artificielle (SIA). La proposition de règlement européen de la Commission européenne désigne un tel système comme un **logiciel** (c'est-à-dire un ensemble d'instructions permettant la réalisation d'une tâche, qu'il soit fourni sous forme de produit ou de service[9], composé d'un ou plusieurs programmes informatiques qui permettent d'accomplir une tâche donnée), mais cette assimilation ne fait pas consensus. Il est certain qu'un système d'IA comprend nécessairement au moins une brique logicielle, qui peut constituer à elle seule le produit ou le service (comme une application informatique). Mais elle peut aussi n'être qu'une composante d'un produit dans lequel elle est physiquement intégrée ou dont elle sert la fonctionnalité (comme un bras robotisé, une voiture « autonome », le système d'exploitation du smartphone…)[10]. Il semble donc préférable de désigner le SIA comme le produit ou le service dans son ensemble, utilisant un ou plusieurs logiciels conçus selon l'une des techniques reconnues comme relevant du champ de l'intelligence artificielle[11], et qui seront présentées au point 1.1.2.

Dans son acception la plus large, le système d'IA est présenté comme un « traitement algorithmique », notion qui figure dans la loi sans être définie[12], ou réduit plus simplement encore à **l'algorithme**, qui n'en est toutefois qu'une composante.

L'algorithme est classiquement défini comme une suite d'instructions et d'opérations permettant de résoudre une catégorie de problèmes. Il s'agit de la définition des étapes qui permettent de convertir des données d'entrée (une image, un son, un texte, des chiffres…) en données de sortie (une réponse oui ou non à la question posée, un montant, une recommandation, une prévision, un contenu – image, son, texte…, une décision, le mouvement d'un robot…). Cette notion peut être comparée à une recette de cuisine décrivant la façon de transformer les ingrédients (données d'entrée) en plat (résultat), à la notice de montage d'un

[8] L'article 1er du décret n° 2019-1088 du 25 octobre 2019 relatif au système d'information et de communication de l'État et à la direction interministérielle du numérique définit « *le* » système d'information et de communication de l'État par ses composantes, à savoir « *l'ensemble des infrastructures et services logiciels informatiques permettant de collecter, traiter, transmettre et stocker sous forme numérique les données qui concourent aux missions des services de l'État et des organismes placés sous sa tutelle* ». On conviendra que le singulier utilisé est assez… singulier au regard de la réalité de l'informatique et de l'État et des usages. L'article 2 du décret emploie du reste lui-même le pluriel « systèmes d'information ».

[9] On parle dans ce cas de SAS, *Software As Service* : la nature du produit ne change pas, mais il est délivré sous forme de prestation par un tiers extérieur, et non remis comme un outil.

[10] Dans une acception un peu différente, le terme « logiciel intégré » ou « embarqué » (« *embedded* ») désigne le logiciel qui assure les fonctions de base de l'appareil dans lequel il est incorporé (par ex., le système d'exploitation du smartphone ou le logiciel de numérisation dans un scanner), par opposition à celui qui, bien que téléchargé sur l'appareil, se borne à offrir des services additionnels (l'application « boussole » du smartphone ou l'outil de reconnaissance d'objets sur un image scannée).

[11] V. 1.1.2. sur la description de ces techniques.

[12] V. par ex. l'article L. 311-3-1 du code des relations entre le public et l'administration.

meuble (résultat) à partir des pièces détachées (données d'entrées), ou au cheminement intellectuel qui conduit le juge, à partir des données du litige qui lui sont fournies (données d'entrée) à rendre un jugement dans un sens déterminé (résultat) – le syllogisme judiciaire est une forme d'algorithme.

Lorsque les données de sortie sont des résultats chiffrés (par exemple, le prix d'un bien immobilier évalué en fonction de la surface, de la localisation, de l'agencement du bien ; le score d'un candidat à l'embauche déterminé en fonction de ses diplômes, son expérience professionnelle, ses recommandations ; la probabilité qu'un délit soit commis à tel endroit à tel moment à partir des statistiques de la délinquance…), on parle d'algorithme de régression ; lorsque le résultat du système est une catégorie (oui / non ; animal / végétal ; courriel normal / courriel indésirable ; pêche / pomme / poire / abricot ; décision d'octroi d'une prestation / décision de rejet…), il s'agit d'un algorithme de classification[13].

Le terme d'« algorithme » est souvent utilisé pour désigner les catégories d'algorithmes qui font partie de la boîte à outils de l'ingénieur IA[14], dénomination la plus proche de celle, couramment utilisé, de *data scientist*[15]. On parlera plutôt de **modèle** (algorithmique) pour désigner l'algorithme conçu par l'ingénieur IA pour accomplir une tâche spécifique, en utilisant l'un de ces outils[16]. Le modèle est ainsi le cœur du système d'IA[17] : c'est la description complète de l'« usine » qui transformera la matière première (les données d'entrée) en produit fini (les résultats).

Afin d'être utilisable opérationnellement, le modèle est transcrit par un programmeur ou un développeur (voire par un système d'IA[18]…) dans un langage de programmation (R, Python…) sous la forme d'un **code-source**. Ce programme informatique est, en quelque sorte, la matérialisation du modèle constituant le système. C'est le texte décrivant la recette de cuisine, la notice de montage du meuble ou le guide complet du raisonnement juridictionnel (qui n'existe pas… encore), exécutable par un ordinateur.

---

[13] Pour une présentation précise des différents algorithmes utilisés, le lecteur est invité à se référer au glossaire en annexe 4.

[14] Il existe de très nombreuses catégories d'algorithmes, fruit d'une recherche ancienne et plus effervescente que jamais. On peut citer notamment la régression linéaire et la régression logistique, l'arbre de décision, les forêts aléatoires, les algorithmes de « boosting » (AdaBoost, XGBoost, GradientBoost…) ou la classification naïve bayesienne.

[15] La traduction littérale de « scientifique de la donnée » n'est pas usitée.

[16] L'ingénieur IA dira : « *j'ai utilisé un algo(rithme) de forêt aléatoire pour construire mon modèle de prédiction de la maladie de Crohn à partir des symptômes constatés chez un patient* ».

[17] Il conviendrait d'ailleurs de parler des modèles constituant le système, car un logiciel est en général constitué de nombreux algorithmes différents permettant d'accomplir des fonctions aussi diverses qu'afficher des informations, permettre la saisie par un utilisateur, produire des résultats, transmettre des données… La valeur ajoutée du système tient d'ailleurs souvent à l'originalité, à la pertinence et à la performance de l'assemblage de différentes briques logicielles.

[18] Il existe désormais des SIA capables, à partir d'une consigne formulée en langage courant, d'écrire le programme informatique correspondant (V. par ex. l'outil Codex d'OpenAI).

La construction du modèle algorithmique requiert **deux ressources-clés**.

La première ressource est constituée des **données**. Pour définir le cheminement qui permet d'aboutir au résultat voulu à partir d'une situation de départ donnée, il est indispensable de disposer de données associant la seconde au premier. Ces données peuvent consister en des connaissances techniques détenues par des experts du métier, par exemple une relation de cause à effet connue des femmes et hommes de l'art grâce à la formation ou à l'expérience acquise. Un juriste tire de son expertise technique l'idée qu'une sanction administrative dépourvue de toute motivation est illégale ; et que « SI » cette sanction est attaquée devant le juge compétent par le mis en cause « ET SI » les conditions de recevabilité du recours sont remplies « ET SI » le moyen tiré du défaut de motivation de la décision est soulevé, « ALORS » la sanction encourt l'annulation[19]. Ces données se confondent ainsi avec des règles existantes, qui peuvent être articulées pour construire un modèle permettant de répondre à la question : telle sanction administrative doit-elle être annulée pour défaut de motivation ? (réponse : oui/non).

Les données peuvent aussi être « brutes », et représenter les situations elles-mêmes et les résultats correspondants. Des images ou des sons auxquels sont associées des descriptions (photographie d'une voiture, aboiement d'un chien…) sont également des données reliant un ensemble de pixels (image) ou d'échantillons (son) à un objet (voiture) ou à un animal (chien) : de telles données peuvent être utilisées pour concevoir un modèle qui pourra identifier une voiture ou un chien à partir d'une photo ou d'un son.

Les opérations de collecte, de sélection, de nettoyage[20] et de qualification des données en vue de concevoir un système d'IA sont coordonnées par les ingénieurs IA[21]. Il importe de rappeler que toute donnée peut aussi être exploitée indépendamment de la construction d'un système d'IA, pour en tirer des **connaissances** utiles, représentées le cas échéant sous des formes didactiques (schémas, graphiques, tableaux de bord…)[22] à des fins de pilotage et de suivi de l'activité administrative – c'est le métier des analystes des données (« *data analysts* ») qui recoupe très largement la **statistique classique**.

La seconde ressource-clé pour la réalisation d'un SIA est l'**infrastructure technique**, comprenant des processeurs (qui déterminent la puissance de calcul disponible), de la mémoire, de la capacité de stockage et des services associés. Seule une puissance

---

[19] Il s'agit ici d'un raccourci car il faut encore que le requérant ne se désiste pas, qu'il n'y ait pas non-lieu, que le moyen soit lui-même recevable (ce qui suppose qu'un moyen de légalité externe ait été présenté dans le délai de recours) … On mesure à quel point un problème extrêmement simple en apparence peut requérir la construction d'un modèle relativement élaboré afin d'embrasser l'ensemble du champ des possibles.

[20] Pour débarrasser les données des éléments inutiles ou parasites (pour un son, un bruit de fond ou un phénomène acoustique, par ex.).

[21] Il s'agit d'ailleurs de l'activité la plus consommatrice de temps dans les projets de systèmes d'IA (le chiffre de 65% des heures d'un projet d'apprentissage machine est parfois avancé : Cognilytica Research, *Data Engineering, Preparation and Labeling for AI*, 2019).

[22] Discipline connue sous le terme de *visualisation de données* (« *dataviz* »).

de calcul élevée, générée par des infrastructures de type « super-calculateurs », permet de concevoir des modèles d'une grande complexité. Ces ressources techniques peuvent être sous le contrôle direct du concepteur de l'outil, s'il dispose de ses propres serveurs et centres de données, ou disponibles *via* un *cloud* : il s'agit alors de « louer », en tant que service, de la capacité de traitement à des entreprises ayant elles-mêmes investi dans les infrastructures physiques nécessaires.

2. L'IA comme discipline

L'IA **est aussi, et avant tout, une discipline**[23] ou, selon la commission d'enrichissement de la langue française, un « *champ interdisciplinaire théorique et pratique* »[24] et, en particulier, un **champ de recherche**, non seulement en mathématiques et en informatique, mais aussi en sciences cognitives, en linguistique, en philosophie, en sociologie, en économie ou encore en droit et en éthique.

Parce qu'elle évoque une forme d'intelligence alternative à l'intelligence humaine et susceptible à terme d'atteindre un niveau de performance similaire puis de rivaliser avec elle ou de s'y substituer et, ainsi, une démarche repoussant sans cesse les limites de la connaissance et du progrès technologique, **le contenu de la notion d'intelligence artificielle apparaît aussi très évolutif, au rythme de l'innovation**. Les percées scientifiques et techniques d'hier tendent à sortir de son périmètre ou à être rejetées à ses confins à mesure qu'elles deviennent des acquis matures voire triviaux, ne recelant plus aucune part de « mystère » ou de « magie »[25]. Ce phénomène, parfois baptisé « effet IA » et résumé par l'adage prêté à Larry Tesler « *l'IA est tout ce qui n'a pas encore été fait* »[26], est notamment illustré par l'expression répandue dans la communauté scientifique de « *good old-fashioned AI* » (GOFAI)[27] pour désigner notamment les systèmes fondés sur l'exécution d'instructions entièrement prédéterminées par l'homme (systèmes dits « basés sur les règles » – V. *infra* 1.1.2.), qui faisaient pourtant figure, il y a quelques décennies, d'innovations technologiques majeures et prometteuses. L'IA est donc un horizon, qui s'éloigne à mesure qu'on croit s'en rapprocher, et non une réalité figée. Il est pourtant nécessaire de tenter de la saisir.

---

[23] Dans son étude annuelle de 2017 consacrée à la puissance publique et aux plateformes numériques, le Conseil d'État définissait l'IA comme une « *science dont le but est de faire accomplir par une machine des tâches qui requièrent traditionnellement l'intelligence humaine ou animale* ».

[24] Avis publié au Journal officiel de la République française le 9 décembre 2018 (NOR : CTNR1832601K) qui définit l'intelligence artificielle comme le « *champ interdisciplinaire théorique et pratique qui a pour objet la compréhension de mécanismes de la cognition et de la réflexion, et leur imitation par un dispositif matériel et logiciel, à des fins d'assistance ou de substitution à des activités humaines* ». L'avis précédent, publié le 22 septembre 2000, mentionnait cette notion mais se gardait d'en proposer une définition.

[25] L'analogie avec la magie est d'ailleurs intéressante en ceci que ce terme peut tout aussi bien désigner la prestidigitation, c'est-à-dire des pratiques répondant aux lois connues de la physique et couramment pratiquées, que la pratique de la sorcellerie et l'œuvre des mages, qui relèvent du domaine du fantastique ou de la croyance religieuse.

[26] La citation originale de l'intéressé est en réalité : « *l'intelligence est ce que les machines n'ont pas encore fait* ».

[27] Expression que l'on peut traduire par : « la bonne vieille IA à l'ancienne / démodée ».

La difficulté principale de l'exercice de définition ne se loge pas dans le terme « artificielle ». Selon l'inventeur, sinon du concept, du moins du terme, en 1956, le professeur en science informatique à Stanford John Mc Carthy, l'intelligence artificielle est « *la science et l'ingénierie de la fabrication de machines intelligentes* »[28]. Cette définition évasive a le mérite de rattacher d'emblée l'IA à la machine, ou plus spécifiquement à un **ensemble d'entités automatisées** que l'homme fabrique ou programme. Le caractère « artificiel » de l'IA tient à ce qu'elle est due « *à la technique de l'homme, par opposition à ce qui a été créé et se développe naturellement* »[29].

C'est l'autre terme de la notion, celui d'intelligence, qui soulève des difficultés, à tel point que son explicitation est parfois absente de la définition proposée, au profit d'une formule périphrastique[30]. L'administration britannique définit ainsi l'IA comme « *l'utilisation de la technologie numérique pour créer des systèmes capables d'effectuer des tâches dont on pense généralement qu'elles nécessitent de l'intelligence* »[31]. Nombre de définitions procèdent par simple comparaison avec les capacités de l'humain, un système d'IA se distinguant par son aptitude à reproduire certaines facultés humaines[32].

L'**approche « imitative » de l'intelligence de l'IA** renvoie au test de Turing, aux termes duquel une machine est « intelligente » si elle est capable, aux yeux d'un humain, de se faire passer pour un autre humain et non pour une machine (il s'agit donc davantage de simuler l'humain plutôt que de le reproduire, et de faire illusion pendant une courte période de temps). Mais comme l'ont indiqué Stuart Russel et Peter Norvig dans leur ouvrage de référence *Intelligence artificielle - une approche moderne*[33], « *les chercheurs ont consacré peu d'efforts à la réussite du test de Turing, estimant qu'il est plus important d'étudier les principes sous-jacents de l'intelligence que de reproduire un modèle* ».

Étymologiquement, l'intelligence, du latin *intellego*, signifie « discerner, démêler, remarquer, se rendre compte, reconnaître », mais également « comprendre, entendre, saisir ». Chacun de ces termes se prêterait lui-même à un intéressant exercice de définition et d'explicitation.

---

[28] J. McCarthy, *What is artificial intelligence*, novembre 2007.

[29] Dictionnaire de l'Académie française.

[30] La difficulté est d'autant plus grande pour ce qui concerne la notion française qu'elle procède d'une traduction littérale du terme anglais « intelligence », qui désigne une réalité différente.

[31] *A guide to using artificial intelligence in the public sector*, juin 2019.

[32] L'Académie française définit l'IA comme « *un ensemble de propriétés rapprochant du cerveau humain certains systèmes informatiques très évolués* ». Pour Yann Le Cun, responsable de la recherche en IA de Facebook, l'IA permet à « *des machines d'accomplir des tâches et de résoudre des problèmes normalement réservés aux humains et à certains animaux* » (Y. Le Cun, *Qu'est-ce que l'intelligence artificielle ?*, Collège de France). Dans le même esprit, l'étude de la Commission européenne sur l'IA dans les services publics retient la définition suivante : « *forme spéciale de technologie de l'information et de la communication apte à reproduire un comportement intelligent et à accomplir des tâches couramment considérées comme requérant une intelligence humaine* ».

[33] Pearson, 4e édition.

Pour les pionniers de l'IA, « *l'intelligence était essentiellement celle des raisonnements logiques* »[34], c'est-à-dire la capacité à atteindre un objectif par une succession d'actions obéissant à des règles précises et formalisables. L'intelligence se définit en rapport avec ce qui lui est demandé (comme l'indiquait John McCarthy, « *l'intelligence est la partie computationnelle de la capacité à atteindre des objectifs dans le monde* »). Nombre d'auteurs mettent davantage l'accent sur les capacités d'apprentissage de la machine. Pour le chercheur François Chollet, l'intelligence d'un système se mesure à son efficacité à acquérir des compétences sur un périmètre de tâches, au regard des acquis (réflexes, connaissances préalables…), de l'expérience et de la difficulté à généraliser[35].

L'intelligence de l'IA renvoie à la modélisation de certaines fonctions cognitives humaines et à leur mise en œuvre par une machine. Parmi les **fonctions cognitives couvertes par les capacités actuelles de l'IA**, on peut citer la capacité de mémoriser et de traiter de l'information, de raisonner, d'apprendre, de parler, d'agir, de prendre une décision, de planifier, de percevoir et d'interagir avec son environnement… Les progrès scientifiques permettent également d'envisager une IA capable d'une forme (réduite) d'intelligence émotionnelle ou, plus rigoureusement capable de réagir aux émotions de son interlocuteur humain.

Pour autant, un système d'IA n'est **doté ni de bon sens, ni de conscience de lui-même, ni de capacité à se corriger de façon entièrement autonome, sans règle ou nouvel intrant lui ayant indiqué qu'il s'est trompé**. Il se borne à faire ce que l'humain l'a programmé à faire, et ce même si le chemin qu'il emprunte pour le faire n'est pas toujours entièrement intelligible pour l'homme, y compris son programmateur, et même s'il ne parvient pas nécessairement à expliquer les raisons du résultat produit. Ainsi, si un SIA de reconnaissance d'objet ou d'entité peut *discerner* un chat parmi un ensemble d'animaux, dans le sens où l'outil peut les distinguer au sein d'une image, il le fait évidemment sans *discernement*.

L'existence d'une « IA générale » qui pourrait combiner l'ensemble des fonctions cognitives et développer des capacités autonomes d'imitation complète de l'intelligence humaine relève encore (d'aucuns pensent pour toujours) de la science-fiction.

IA générale et IA étroite

« *Un **système d'IA générale** est censé être un système capable d'effectuer la plupart des activités que les humains peuvent faire. Les **systèmes d'IA étroite** sont au contraire des systèmes qui peuvent effectuer une ou quelques tâches spécifiques. Les systèmes d'IA actuellement déployés sont des exemples d'IA étroite.* »[36]

[34] V. Rialle, *L'intelligence artificielle et sa place dans les sciences de la cognition*, Bulletin de l'AFIA n° 26, 1996.

[35] F. Chollet, *On the Measure of Intelligence*, 2019, arXiv:1911.01547v2, traduction libre.

[36] Groupe d'expert de haut niveau sur l'IA, Commission européenne, *Une définition de l'IA : principales capacités et disciplines scientifiques*, 2018.

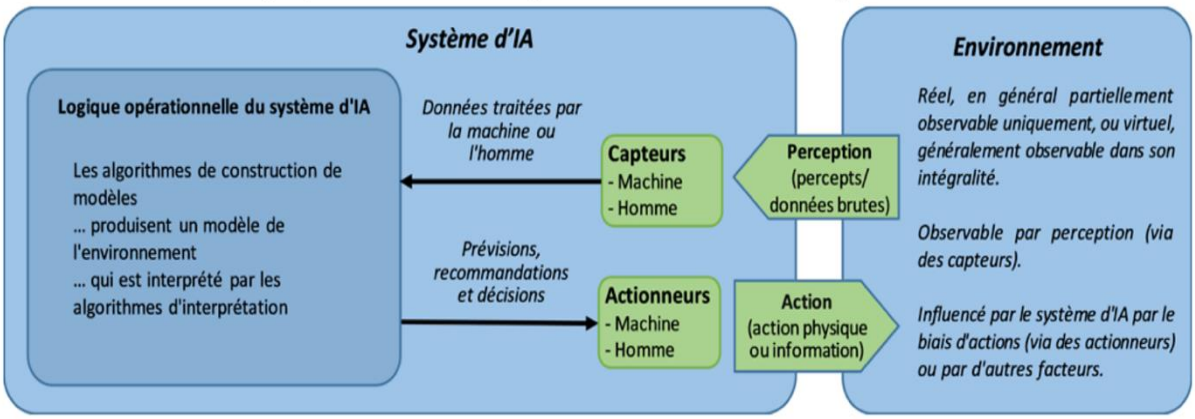
Le groupe d'experts de haut niveau sur l'IA de la Commission européenne a retenu une définition très complète qu'il est utile de reproduire :

« *L'intelligence artificielle désigne des systèmes conçus par des humains qui, étant donné un objectif complexe, agissent dans le monde physique ou numérique en percevant leur environnement, en interprétant les données structurées ou non structurées collectées, en raisonnant sur les connaissances dérivées de ces données et en décidant de la ou des meilleures actions à entreprendre (selon des paramètres prédéfinis) pour atteindre l'objectif donné. Les systèmes d'IA peuvent également être conçus pour apprendre à adapter leur comportement en analysant comment l'environnement est affecté par leurs actions précédentes.*

En tant que discipline scientifique, l'IA comprend plusieurs approches et techniques, telles que l'apprentissage automatique (dont l'apprentissage profond et l'apprentissage par renforcement sont des exemples spécifiques), le raisonnement automatique (qui comprend la planification, l'ordonnancement, la représentation et le raisonnement des connaissances, la recherche et l'optimisation) et la robotique (qui comprend le contrôle, la perception, les capteurs et les actionneurs, ainsi que l'intégration de toutes les autres techniques dans des systèmes cyber-physiques). »

Le schéma suivant[37] illustre le principe de fonctionnement d'un SIA ainsi défini :

Vision conceptuelle d'un SIA



La notion d'« **agent intelligent** » est parfois utilisée[38] pour désigner ce principe de fonctionnement comprenant la perception de l'environnement par des capteurs, l'analyse des données et la sélection d'un résultat, et le déclenchement d'une action rationnelle (c'est-à-dire l'action appropriée à la situation). Un système d'IA combine en général différents agents pour remplir la fonction qui lui est assignée (notion de « système multi-agents »).

[37] V. l'étude de l'OCDE, *L'intelligence artificielle dans la société*, 2019.

[38] V. en particulier l'ouvrage de référence de S. Russel et P. Norvig, *Intelligence artificielle : une approche moderne*, Pearson, 4e édition, et l'entrée « agent intelligent » dans le glossaire en annexe 4.

1.1.2. Les systèmes d'intelligence artificielle relèvent d'approches et de techniques différentes

On distingue classiquement deux approches de l'intelligence artificielle, historiquement en compétition l'une avec l'autre : d'une part, l'IA symbolique ou basée sur les règles, courant dominant des années 1950 au début des années 2000 ; d'autre part, l'IA connexionniste ou apprenante (apprentissage machine), dont la genèse est plus ancienne mais dont le succès, spectaculaire, est récent (une dizaine d'années), au point de nourrir une forme de monopole de l'appellation « IA ». S'y ajoutent les techniques statistiques « classiques » qui sont souvent rangées dans l'apprentissage machine mais distinguées dans certains travaux et dans la proposition de règlement européen sur l'IA[39].

1. L'IA symbolique ou déterministe, une IA fondée sur des règles

Les SIA basés sur les règles, dits IA symboliques, reproduisent un **raisonnement logique de type déductif** (« si x et y, alors z »). Ces systèmes exécutent des règles programmées, qui reflètent la connaissance que nous avons des liens entre deux données : on sait en quoi consiste une addition, on peut donc aisément concevoir une calculatrice qui effectue une telle opération. Un tel SIA n'a qu'une capacité « créative » limitée : les opérations qu'il exécute sont prédéterminées dans un modèle explicite. Les règles qui représentent le savoir dans un domaine d'application sont stockées dans une base de connaissances. Un moteur d'inférence sélectionne dans cette base les règles pertinentes et les applique pour résoudre le problème qui est posé, en fonction des faits qui lui sont soumis (par exemple, si un usager demande une autorisation administrative ou une prestation en fournissant les éléments justificatifs, le système peut vérifier que les conditions posées par les textes sont remplies et recommander une décision ou l'édicter lui-même : « si âge supérieur à X ans, nombre d'enfants supérieur à Y et ressources inférieures à Z euros, alors droit à l'allocation » ; « si droit à l'allocation, alors versement de la somme sur le compte bancaire du demandeur » ; etc.).

L'ambition historique de l'IA symbolique est d'identifier et de traduire en programmes informatiques les règles qui « gouvernent le monde ». Ainsi, avant qu'elle soit supplantée par les réseaux de neurones pour la reconnaissance d'objets sur des images, d'innombrables publications scientifiques ont rivalisé d'imagination pour définir les caractéristiques des choses (« extraction de caractéristiques ») et associer un résultat à un ensemble de caractéristiques. Des formules mathématiques expriment le « raisonnement » à l'œuvre pour déduire, par exemple, que la « chose » qui se trouve sur une photo, avec ses contours généraux, ses éléments constitutifs (une oreille, des moustaches…), ses points d'intérêts particuliers, ses couleurs… est un chien, et non un chat. Dans ce paradigme, les données, soigneusement nettoyées et fiabilisées pour éviter les erreurs qui pourraient conduire à de mauvaises déductions, sont utilisées pour comprendre le cheminement logique conduisant d'un ensemble de données d'entrée (les pixels composant l'image du chien) à une donnée de sortie (la machine indique que c'est

[39] V. 1.2.3. sur la définition retenue par le règlement européen.

un chien). On voit que, en fonction de la tâche à accomplir, les arbres décisionnels peuvent être très simples (octroi d'une prestation si la condition de ressources est remplie) ou d'une très grande complexité (il est très difficile d'extraire l'ensemble des caractéristiques d'un éléphant pour permettre sa reconnaissance sur une image, alors que sa trompe peut être partiellement masquée, l'image inversée ou altérée par des perturbations visuelles liées, par exemple, aux conditions météorologiques, à la luminosité…)[40].

Les **systèmes experts** constituent l'application la plus connue de cette approche. Il s'agit de systèmes mis en œuvre pour résoudre des problèmes qui requièrent un haut niveau d'expertise, dans des domaines spécialisés. Les spécialistes de la matière sont amenés à formaliser dans une base de connaissances les savoirs qui gouvernent l'activité ou le process dont ils ont la charge, et qui seront mobilisés par la machine.

Un programme basique de jeu d'échecs constitue un SIA symbolique. La machine est paramétrée pour respecter les règles du jeu et adopter des stratégies pré-programmées en fonction des mouvements du joueur humain. Naturellement, l'utilisation des systèmes-experts n'est pas cantonnée à ce qu'on appelle les « problèmes-jouets ». Ils sont largement déployés dans l'industrie et les services, en particulier pour l'accomplissement de tâches qui n'admettent pas (ou très peu) l'erreur et qui se prêtent donc moins à une approche statistique.

2. L'IA statistique

Les méthodes statistiques sont utilisées de très longue date pour éclairer l'action publique. On distingue schématiquement deux types de statistiques : les statistiques « fréquentistes », les plus intuitives, visent à évaluer la probabilité qu'une théorie ou une règle prédéterminée produise un résultat X ou Y ; les statistiques bayésiennes (estimation bayésienne[41]) ont pour objet, à l'inverse, d'évaluer la probabilité qu'une théorie ou une règle soit vraie compte tenu des observations réalisées.

Ces modèles mathématiques sont très utilisés pour identifier, classer et pondérer l'influence d'un ensemble limité de variables explicatives d'un phénomène (probabilité d'un défaut de paiement en cas de crédit bancaire, probabilité d'une fraude étant donné le profil d'une entreprise…). Dans le cadre d'un outil d'aide au diagnostic médical, le modèle statistique peut calculer la probabilité que le patient soit atteint de telle ou telle maladie en fonction des différentes variables causales avérées ou suspectées (qui modélisent les connaissances médicales sur les différents facteurs de maladie et les différents symptômes possibles).

---

[40] Bien qu'elle soit moins utilisée pour les systèmes basés sur les règles, la notion d'apprentissage automatique n'y est pas étrangère. Ainsi, la programmation inductive logique, que la proposition de règlement européen classe dans les « approches fondées sur la logique et les connaissances », est classiquement décrite comme une approche logique de l'apprentissage automatique. Elle consiste à classer les données disponibles en exemples positifs et en exemples négatifs, et à induire un programme logique qui confirme tous les positifs et infirme tous les négatifs.

[41] V. le glossaire en annexe 4, entrée « estimation bayésienne ».

Cette logique probabiliste distingue clairement cette approche de celle qui est fondée sur les règles, qui ne laisse en principe pas de place à l'aléa. En revanche, cette approche statistique se distingue mal de l'apprentissage machine dont il sera question ensuite. La proposition de règlement IA de la Commission européenne propose pourtant d'en faire deux catégories distinctes[42]. La plupart des personnes auditionnées par la mission n'identifiaient ainsi pas de différence de nature entre les deux. En général, on parle de technique statistique pour des modèles « simples », alors que l'apprentissage automatique est davantage utilisé pour résoudre des problèmes complexes, notamment pour traiter des données comportant un très grand nombre de dimensions (comme les pixels d'une image, qui ont chacun une forme, une couleur… et qui doivent être mis en relation les uns avec les autres pour identifier l'objet sur l'image). La confusion peut en outre être entretenue par des techniques commerciales valorisant la dimension « apprenante » de modèles statistiques « à l'ancienne ».

3. L'apprentissage machine et l'IA « connexionniste » (ou « subsymbolique »)

L'apprentissage machine ou apprentissage automatique (*machine-learning*) se distingue de l'approche symbolique (basée sur les règles) en ceci que le modèle ne résulte pas de la pré-détermination d'instructions logiques assignant un résultat à des données d'entrée, mais est défini par la machine elle-même à partir d'un grand nombre de données qui lui ont permis d'identifier des relations, des récurrences, des corrélations, des liens de proximité entre elles, donc des règles. Très schématiquement, l'IA connexionniste déduit les règles à partir des exemples qu'on lui fournit, alors que l'IA symbolique produit des résultats à partir des règles programmées. Trivialement, cela revient à « apprendre le métier » par l'observation de la pratique plutôt qu'en suivant des cours théoriques. On peut apprendre les règles du football en lisant le règlement de la FIFA ou en écoutant les instructions de l'entraîneur ; on peut aussi les déduire du visionnage de milliers de matchs, permettant de comprendre qu'il s'agit d'un sport qui se joue (au début au moins…) à onze contre onze, dont l'objectif consiste à mettre le ballon dans le but adverse plus de fois que le contraire, qui n'admet l'utilisation des mains que par le gardien de but ou pour les touches, etc.

A titre d'exemple (fictif), pour construire un système d'IA d'aide à l'examen de la recevabilité d'un recours contentieux, on peut soit transcrire en instructions informatiques les dispositions pertinentes du code de justice administrative et la jurisprudence pour que le programme les applique à un dossier donné (approche symbolique), soit construire un modèle qui va déduire ces règles à partir de l'exploitation d'un très grand nombre de jugements – il constatera, « à l'expérience », que ce délai est de deux mois la plupart du temps, sauf dans certaines matières ; qu'il n'est pas déclenché en l'absence de mention des voies et délais de recours dans la décision attaquée, etc. (approche connexionniste).

La notion d'« apprentissage » ne doit pas induire en erreur sur les capacités réelles des systèmes d'IA : en l'état des techniques, elle consiste simplement à construire

[42] En particulier, elle classe l'estimation bayésienne dans l'approche statistique alors que celle-ci est à la base de la plupart des modèles d'apprentissage machine.

une fonction mathématique, parfois simple, parfois d'une complexité vertigineuse (comportant un nombre très élevé de paramètres, sous la forme d'un espace vectoriel à multiples dimensions), à partir de données existantes.

A la différence du SIA symbolique qui est explicitement programmé par des experts du métier, le SIA connexionniste est entraîné[43]. La **phase d'entraînement** (ou d'apprentissage) à partir d'un jeu de données d'entraînement permet de construire le modèle. Elle est suivie d'une **phase de validation** sur un jeu de données de validation distinct, permettant d'apprécier les performances du modèle au regard de l'objectif fixé et, plus largement, le « comportement » du système.

Dans le cas d'un **SIA statique** (ou figé), le modèle n'évolue pas une fois validé et mis en production. A l'inverse, un **SIA dynamique ou apprenant** sera mis à jour :

- soit de façon ponctuelle ou périodique : c'est **l'apprentissage incrémental[44]**, qui peut aller jusqu'à ré-entraîner un modèle sur un jeu de données différent afin d'en construire un nouveau, plus adapté aux besoins. A titre d'exemple, un outil de traitement du langage naturel entraîné sur des articles de l'encyclopédie en ligne Wikipédia devra être ré-entraîné sur des décisions de justice si on souhaite l'utiliser pour rechercher automatiquement une décision dans une base de jurisprudence à partir d'une requête formulée en langage courant). L'**apprentissage par transfert** désigne la technique qui consiste, à partir d'un modèle pré-entraîné et « générique », à utiliser un jeu réduit de données labellisées correspondant précisément à la tâche qu'on souhaite faire accomplir au système pour l'ajuster finement.
- soit de façon permanente : c'est **l'apprentissage continu**, qui passe en général par une **boucle de rétroaction** permettant de tenir compte en temps réel des résultats produits par l'outil et de l'approbation ou de la désapprobation exprimée par l'utilisateur et d'adapter les paramètres du modèle en conséquence.

La notion parfois utilisée de SIA ou d'algorithme « auto-apprenant » est en revanche déconseillée en l'état des techniques car elle donne, à tort, l'illusion que la machine est capable d'apprendre d'elle-même de façon automatique – alors que ce réapprentissage, même en continu, n'est que le prolongement de la fonctionnalité d'origine de l'algorithme d'apprentissage machine. Des recherches sont consacrées au *lifelong learning* (apprentissage tout au long de la vie), notion qui renvoie à des systèmes capables d'apprendre en permanence, d'accumuler les connaissances au fil du temps et de les utiliser ou de les adapter pour résoudre de nouveaux problèmes ou faire évoluer leurs propres capacités d'apprentissage (et, à terme, de devenir une IA générale au sens précédemment mentionné) ou encore, plus modestement, au *méta-learning*, qui désigne la capacité de la machine à « apprendre à apprendre ».

---

[43] Fr. Chollet, *L'apprentissage profond*, p. 7.

[44] Selon la définition donnée par le référentiel de certification du LNE, l'apprentissage incrémental est l'apprentissage automatique réalisé sur des données regroupées en lots (batchs), les lots étant renouvelés périodiquement, au fur et à mesure de l'accumulation de nouvelles données tout au long du cycle de vie de la fonctionnalité d'IA.

Le cycle de vie d'un système d'IA

Ce cycle de vie comporte schématiquement les phases suivantes, qui sont itératives[45]:

1/ La conception

Cette première phase consiste à transformer une expression de besoin en spécifications fonctionnelles. Elle couvre d'abord l'évaluation des besoins, la définition des objectifs du SIA, l'élaboration d'un cahier des charges, l'évaluation des risques (y compris liés à la conduite du changement) et l'examen des conditions de conformité. Elle comprend également la collecte et le traitement des données : identification des données nécessaires, proprification des données afin de les rendre exploitables par le SIA, vérification de leur qualité (absence de biais, exhaustivité…). Enfin, il s'agit de construire le modèle, par le choix des algorithmes utilisés notamment (d'où l'importance d'une veille technologique), et de l'entraîner.

2/ Le développement et la certification

La phase de développement consiste à traduire les spécifications fonctionnelles en une version opérationnelle de la fonctionnalité d'IA, vérifiée et validée. Elle passe généralement par l'élaboration d'une **preuve de concept** (dit « POC », pour « proof of concept ») afin de déterminer sa viabilité et la possibilité d'atteindre l'objectif fixé (et, le cas échéant, de lever des fonds destinés à financer sa mise en production), et, éventuellement, d'un prototype expérimenté en conditions réelles (cas de la robotique en particulier). Le développement peut s'achever par un processus d'évaluation consistant à vérifier et valider la conformité du système aux spécifications définies avant son déploiement, en utilisant des **données de test**, différentes des données d'entraînement et de validation. Cette évaluation peut être purement interne, ou être réalisée par un organisme d'évaluation de la conformité qui pourra délivrer une **certification**.

3/ Le déploiement

La mise en production du SIA (« le passage à l'échelle ») implique notamment la gestion des changements organisationnels ou l'évaluation de l'expérience utilisateurs et, le cas échéant, une nouvelle mise en conformité réglementaire.

4/ L'exploitation et le maintien en conditions opérationnelles

L'exploitation du SIA conduit à poursuivre son évaluation au regard de ses effets et de l'écart entre les résultats produits et les résultats attendus. Le SIA est ajusté en fonction de ces écarts afin d'assurer la conformité de sa fonctionnalité aux spécificités définies. Le SIA peut être abandonné si les résultats ne sont pas satisfaisants.

[45] V. l'étude de l'OCDE sur *L'intelligence artificielle dans la société* précitée, qui se fonde sur les travaux par l'AIGO (groupe d'experts sur l'IA de l'OCDE) et le référentiel de certification de processus pour l'IA du Laboratoire national de métrologie et d'essais (LNE/DEC/CITI/CH ; LNE/DEC/IA/GA, n° 2.0, 12 juillet 2021).

Plusieurs méthodes d'apprentissage peuvent être distinguées.

a/ On parle d'**apprentissage supervisé** lorsque le modèle est alimenté par des données dites labellisées, étiquetées, annotées ou qualifiées, c'est-à-dire auxquelles on a assigné une valeur ou qu'on a rangé dans une catégorie, de manière à permettre à la machine de différencier ce qui est une bonne et une mauvaise réponse : telle photo représente une chaise, tel caractère (présenté à l'endroit, à l'envers, en police Times New Roman ou écrit d'une main ferme ou tremblante…) est le chiffre 1, tel terme est un nom propre qui doit être pseudonymisé dans une décision de justice… Les paramètres « manuels » (non entraînables[46]) du modèle sont ajustés de telle sorte que la prévision qu'effectue la machine (inférence) à partir de données d'entrée (n'importe quelle photo, n'importe quel caractère ou n'importe quelle décision de justice) soit conforme ou la plus proche possible de la valeur attendue (l'outil devra indiquer qu'une chaise se trouve sur la photo si tel est bien le cas et donner le résultat inverse dans le cas contraire ; il devra reconnaître correctement le chiffre 1 sur un papier, une enveloppe ou dans une image ; il devra proposer ou procéder à la pseudonymisation d'un nom propre dans une décision s'il s'agit bien d'un nom propre).

L'apprentissage supervisé repose aujourd'hui très massivement sur l'étiquetage de données élémentaires par des humains : pour reconnaître un chat d'un chien, des dizaines de milliers d'images de chien et de chat, qui vont constituer le jeu de données d'apprentissage, doivent être labellisées comme telle par l'humain. Ce travail d'étiquetage peut être subreptice : les « captcha » utilisés pour déterminer si l'internaute est un robot ou non (« cochez les cases où apparaît un avion ») collectent les réponses des personnes interrogées (qui sont en général exactes) pour alimenter un modèle d'apprentissage machine. Les tests A/B (*A/B testing*), dans lesquels l'utilisateur indique ce qui vous convient le mieux (par exemple, l'ergonomie du site internet expérimental A est meilleure que celle du site B), constituent une autre technique d'étiquetage (si A obtient un meilleur score que B, le site empruntera davantage à la structure et à la présentation de A).

L'objectif de l'apprentissage supervisé est de **minimiser l'erreur**, c'est-à-dire l'écart entre la bonne réponse et le résultat produit par la machine, en choisissant le bon algorithme, en en optimisant les paramètres (en « réglant » la machine) et, si besoin, en l'alimentant avec un jeu de données plus riche ou de meilleure qualité. Schématiquement, plus le jeu de données d'apprentissage est volumineux et de bonne qualité, plus le modèle algorithmique sera performant dans l'accomplissement de la tâche qui lui est assignée. A l'inverse, un volume ou une qualité insuffisants de données ne permet pas un bon apprentissage en raison de son caractère statistique – ce qui différencie encore l'apprentissage machine et l'apprentissage humain, un enfant n'ayant pas besoin de centaines d'images de chaise pour en reconnaître une avec un degré de fiabilité extrêmement élevé. Le choix de l'algorithme[47], qui est l'une des fonctions de l'ingénieur en IA, influence également la performance du système d'IA.

[46] On parle parfois d'hyper-paramètres, par opposition avec les paramètres qui sont calculés automatiquement lors de l'entraînement du modèle.

[47] Des exemples d'algorithmes sont donnés dans le glossaire en annexe 4.

Le principal enjeu de l'apprentissage machine consiste à définir un modèle algorithmique capable de **généraliser** dans son domaine de fonctionnement, c'est-à-dire de **produire un résultat correct à partir de situations nouvelles qu'il n'a jamais rencontrées lors de son entraînement**. Il ne faut pas que le modèle apprenne les données d'entraînement « par cœur » mais qu'il en identifie les caractères distinctifs pour pouvoir « extrapoler » à d'autres situations. Si l'on souhaite concevoir un système apte à distinguer un homme d'une femme, il n'est guère utile de construire un modèle qui « mémorise » chacune des dix-mille images d'hommes et dix-mille images de femmes du jeu de données d'entraînement sans en tirer le moindre enseignement quant aux traits caractéristiques (apparents) d'un homme et d'une femme. Un tel modèle reconnaîtra certes immédiatement l'une des vingt-mille images composant le jeu de données d'entraînement et y associera le bon qualificatif (homme ou femme) ; mais il sera incapable de prédire correctement le bon résultat (homme ou femme) à partir d'une nouvelle donnée d'entrée (une nouvelle photo d'homme ou de femme qu'il n'avait jamais « vu(e) » auparavant). Ainsi, on aura beau lui montrer la photo d'un autre homme ressemblant fortement à certains spécimens du jeu d'entraînement, il sera incapable de dire avec une probabilité satisfaisante qu'il s'agit d'un homme, c'est-à-dire d'inférer correctement. En d'autres termes, pour un modèle ainsi conçu, un homme est nécessairement l'une des dix-mille personnes figurant sur les images d'homme. On parle alors de surajustement (*overfitting*) ou de surentraînement[48].

b/ Lorsque les données d'entraînement ne sont pas annotées, c'est-à-dire qu'il n'est pas fourni à la machine de « bonnes réponses » qui lui permettraient de les distinguer des « mauvaises » et de construire sur cette base le « mode d'emploi » correct de la tâche qui lui est soumise, il s'agit d'**apprentissage non-supervisé** : le rôle du SIA se limite alors à structurer le jeu de données en les regroupant – on dit souvent en les partitionnant ou en les segmentant (*clustering*) – par catégories homogènes, afin d'identifier des différences, des récurrences ou des anomalies. Le système est incapable de qualifier chacune de ces catégories, mais il en identifie l'existence et un certain nombre de traits distinctifs.

Un système basé sur l'apprentissage non supervisé peut, par exemple, dresser une typologie des clients d'un magasin en fonction de leur profil de consommation, ou les plantes d'une même espèce selon leurs caractéristiques visuelles. Dans ce dernier exemple, la machine ne reconnaît pas la plante sur la base de caractéristiques prédéfinies par l'homme – ce qui serait un système-expert – ou sur la base d'une annotation des images (« marquage » d'un grand nombre de photos d'orchidées comme appartenant à la catégorie « orchidée » par exemple), ce qui constituerait de l'apprentissage supervisé, mais elle se borne à distinguer (sous forme de probabilités le cas échéant) un ensemble d'images A (qui s'avèrent être des orchidées) et un

[48] Pour utiliser une métaphore picturale, un modèle de reconnaissance de formes verra dans l'œuvre *Un dimanche après-midi à l'île de la Grande Jatte* de Seurat, s'il est surentraîné, la juxtaposition de points d'une forme et d'une couleur précises ; s'il est sous-entraîné une simple image colorée avec des formes indéterminées ; et un modèle robuste identifiera des personnages, des animaux, des bateaux… c'est-à-dire capturera, au bon niveau d'abstraction, les formes élémentaires pertinentes.

ensemble d'images B (qui sont des roses) en raison des similitudes au sein du groupe A et au sein du groupe B, et de la dissemblance entre les spécimens de A, d'une part, et de B, d'autre part.

c/ L'**apprentissage auto-supervisé**, concept situé à mi-chemin des deux précédents, est une méthode très prometteuse pour surmonter le problème de la rareté des données labellisées ou du coût de la labellisation, qui constituent la principale limite de l'apprentissage supervisé. Elle consiste à produire de façon automatisée (et non par labellisation humaine) des données d'entraînement. La technique généralement utilisée à cette fin consiste à altérer un jeu de données et à entraîner le modèle afin qu'il retrouve la « version d'origine » grâce à l'identification de caractéristiques. Il peut s'agir, par exemple, de tronquer des phrases (en enlevant un mot) pour les besoins de l'entraînement, le modèle devant « deviner » le mot manquant (si la prévision n'est pas la bonne, c'est-à-dire si le mot proposé par la machine ne correspond pas à celui qui a été escamoté dans la phrase initialement complète dont on disposait, le modèle est ajusté pour améliorer la prévision) ; ou encore d'occulter une partie d'une image, d'en modifier l'angle, d'ajouter du « bruit »... Une fois les caractéristiques pertinentes apprises grâce à ces « tâches prétexte », le modèle peut être plus aisément ré-entraîné pour accomplir efficacement les tâches souhaitées, à l'aide d'un minimum de données annotées manuellement correspondant à ces tâches (« apprentissage par transfert »).

d/ Enfin, l'**apprentissage par renforcement** consiste à définir un objectif dont l'atteinte entraînera une « récompense »[49] pour le système (il « saura » s'il a gagné ou perdu). Ce dernier identifie sur cette base les stratégies gagnantes et les meilleures actions à réaliser dans telle ou telle situation à laquelle il est confronté. Autrement dit, le système crée ses propres données d'entraînement à partir de la qualification d'un résultat. C'est ainsi qu'a été entraîné AlphaGo Zero, version avancée du programme AlphaGo initialement basé sur l'apprentissage supervisé et qui a battu pour la première fois le champion du monde du jeu de go. Contrairement à l'apprentissage supervisé, aucune donnée historique (issue de parties jouées par des champions humains) n'est utilisée, et l'humain n'intervient pas dans l'apprentissage, autrement qu'en implémentant les règles du jeu et le système de récompense, et en ordonnant au système de jouer contre lui-même pendant une durée déterminée jusqu'à ce qu'il atteigne des performances maximales ou satisfaisantes (définies par l'homme).

4. Réseaux de neurones et apprentissage profond

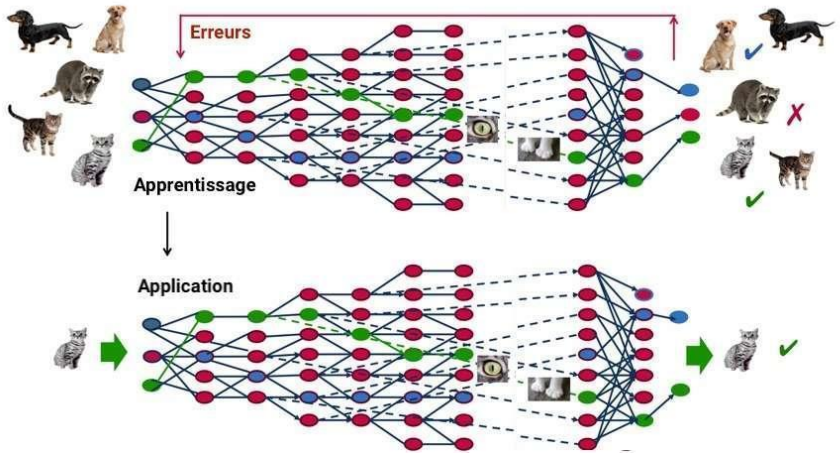
Parmi les algorithmes utilisés dans le domaine de l'apprentissage machine, celui qui suscite le plus d'engouement, voire de fantasmes, consiste à utiliser des réseaux de neurones artificiels, d'où provient l'appellation IA « connexionniste » (en référence aux connexions neuronales).

Ces réseaux sont composés d'au moins trois couches de neurones : une couche d'entrée qui reçoit des données brutes ; reliée à une couche cachée qui traite ces

[49] Nouvel exemple du caractère trompeur du vocabulaire employé appelant une grande prudence dans son maniement. Bien entendu le système n'est pas sensible à la récompense ou au châtiment ; il s'agit seulement de coder sa réponse par une indication binaire lui indiquant qu'il est, ou non, sur le bon chemin de raisonnement.

données ; elle-même reliée à une couche de sortie qui produit le résultat. Lorsque le réseau comporte plus d'une seule couche cachée, on parle d'apprentissage profond.

Illustration du fonctionnement d'un réseau de neurones



Le réseau de neurones, dans sa phase d'apprentissage, identifie les différentes caractéristiques permettant d'attribuer à une image la qualification de chat, de chien ou de raton-laveur (dans l'exemple ci-dessus). Dans sa phase d'application, il utilise ces différentes caractéristiques préalablement « apprises » pour identifier les chats.

Les neurones interagissent entre eux pour résoudre des tâches complexes, comme la reconnaissance d'images ou la catégorisation. A chaque couche revient une tâche d'abstraction : par exemple, pour identifier un visage, la première couche identifiera des formes élémentaires, la seconde les associera à un élément du visage, la troisième identifiera leur taille, etc. Chaque neurone pondère la valeur des autres neurones : si le neurone A reçoit une bonne information du neurone B, mais une mauvaise information du neurone C (ce que confirme le programme d'entraînement), alors le neurone A pondèrera à la hausse la valeur des informations du neurone B et minorera celle du neurone C. Si le nombre de couches et de nœuds est défini par l'homme (hyperparamètre), le « poids » de chaque neurone est quant à lui ajusté automatiquement pour optimiser le résultat.

Si les modèles connexionnistes ont été théorisés de très longue date, leur supériorité opérationnelle dans certaines tâches, en particulier la reconnaissance d'images, n'est définitivement démontrée, sur le plan empirique, que depuis une dizaine d'années[50].

[50] La conférence ECCV d'octobre 2012 où a été présenté un modèle algorithmique basé sur l'apprentissage profond développé sous la houlette de Geoffrey Hinton, constitue à cet égard un point de bascule (V. pour le récit pittoresque de cet épisode et une mise en perspective historique : D. Cardon, J.-P. Cointet et A. Mazières, *La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle*, Réseaux 2018/5, n° 211, p. 173-220.

L'essor de cette technique s'explique par une forme d'« *alignement des astres* » : la disponibilité accrue des données d'entraînement grâce à la numérisation croissante et accélérée de l'activité humaine ; la disponibilité d'une puissance de calcul sans précédent, avec le développement de processeurs incomparablement plus performants qu'auparavant (en particulier les GPU[51], que l'on trouve dans les cartes graphiques d'ordinateurs et prisés des amateurs de jeux vidéo) et le perfectionnement rapide des algorithmes grâce à l'effervescence de la recherche et au principe du partage public des codes sources (*open source*).

Les réseaux de neurones sont utilisés aussi bien en apprentissage supervisé (les poids des neurones sont ajustés pour que le résultat produit soit le plus proche possible de la réponse attendue telle qu'elle ressort de la donnée annotée par l'humain) qu'en apprentissage non supervisé, par exemple par les réseaux génératifs antagonistes (*generative adversarial networks*) qui consistent à opposer un réseau de neurones, dit « générateur », chargé de créer des contenus nouveaux à partir d'exemples réels (par exemple, des photos de personnes fictives) et un réseau de neurones, dit « discriminateur », chargé de distinguer les contenus originaux et les contenus créés par le générateur[52].

5. IA symbolique contre apprentissage machine : match nul ?

L'IA symbolique et l'IA statistique sont généralement présentées comme concurrentes et il est vrai que, **ne reposant pas sur les mêmes techniques, elles présentent des avantages et des inconvénients différents**. Schématiquement et sous réserve d'exceptions tenant à la complexité des modèles, l'IA symbolique a en principe le mérite d'être explicable, dès lors qu'elle repose sur des règles définies par l'humain, à l'inverse de l'apprentissage profond, dont les modèles échappent davantage à la compréhension humaine. Toutefois, l'IA symbolique ne peut incorporer un nombre extrêmement important de règles et se prête plutôt à la résolution de problèmes spécifiques dans un environnement suffisamment normé. En outre, la mise à jour de ces systèmes basés sur les règles peut être complexe et extrêmement coûteuse, alors que les réseaux neuronaux, par exemple, apprennent à réaliser une tâche sur la base de nouvelles données qui leur sont fournies.

L'exceptionnel engouement pour les techniques d'apprentissage automatique ne doit pas occulter que l'immense majorité des systèmes d'information en fonctionnement reposent sur une logique symbolique. En outre, ces deux approches peuvent se révéler **complémentaires**.

[51] *Graphics Processing Unit*.

[52] En pratique, le réseau discriminateur est entraîné à « comprendre » les données originales (par ex., la photo de personnes réelles) ; de son côté, le générateur apprend à produire des données nouvelles, contrefaites. Le discriminateur doit alors indiquer quelles sont les photos authentiques et quelles sont les photos inventées par le générateur. Puis, sur la base des résultats que lui renvoie le discriminateur, le générateur adapte son modèle pour perfectionner son mode de fabrication et, ainsi, réduire la proportion des cas dans lesquels le discriminateur a, à juste titre, identifié une photo comme ayant été artificiellement créée.

D'une part, de nombreux services les combinent. Tel est le cas, par exemple, des robots conversationnels (*chatbots*) qui utilisent des algorithmes d'apprentissage machine se rattachant au traitement automatique du langage naturel (TALN) pour « comprendre » la demande entrée par l'utilisateur, puis font appel à une base de connaissances programmée pour choisir et restituer la réponse qu'ils estiment correcte. Dans ce cas, le SIA, dans son versant « statistique », convertit du texte dans un langage exploitable et, dans son versant symbolique, lui applique des règles logiques. Il est tout aussi fréquent que soient associées une brique logicielle utilisant des réseaux de neurones pour la vision par ordinateur et une autre brique symbolique : tel est le cas, par exemple, d'un SIA capable de détecter la présence d'un bâtiment sur une image satellite et de le qualifier (maison, silo, usine…), à l'aide d'un modèle d'apprentissage machine, puis, par l'application de simples règles, de déterminer sa date de construction en comparant les données cartographiques à différentes dates, d'apprécier si un bâtiment de cette nature est susceptible d'appeler la délivrance d'une autorisation d'urbanisme (eu égard à sa surface, à la proximité d'un monument historique…) et, le cas échéant, de vérifier, avec l'aide d'un fichier des permis de construire, si un tel bâtiment a été autorisé, avant de notifier à l'agent humain une alerte en cas de suspicion de construction irrégulière.

D'autre part, l'hybridation des deux approches dans la construction d'un même modèle algorithmique, afin de tirer le « meilleur des deux mondes », constitue un champ de recherche actif.

1.1.3. Les systèmes d'IA peuvent remplir les fonctions les plus variées

Les fonctionnalités des systèmes d'IA se rattachent aux fonctions cognitives de l'humain : la perception (voir, entendre, toucher…), la mémoire (stocker et mettre à disposition des informations), le raisonnement (trier et classer des données, déduire des connaissances à partir des données, convertir des données...), la planification (définir et mettre en œuvre des plans d'actions en réponse à des situations et des problèmes), l'apprentissage (prendre en compte le passé pour adapter le comportement futur), l'expression et la création (parler, écrire, dessiner…) ou encore la motricité (se déplacer, accomplir un geste…).

Les SIA sont ainsi susceptibles de s'acquitter de tâches innombrables, dont certaines sont à la portée des humains et d'autres non. **Il n'existe pas de corrélation directe entre la simplicité de la tâche pour un humain, d'une part, et la faisabilité par la machine, d'autre part**. Un geste trivial pour un humain – serrer la main d'une autre personne ou saisir un verre posé sur une table – est très difficile et coûteux à faire accomplir par un robot ; à l'inverse, retrouver un mot unique ou une information dans un document de 500 000 pages en une heure n'est pas à la portée d'un humain, mais l'est sans difficulté d'un programme informatique, sans même qu'il s'agisse d'un système d'IA, d'ailleurs. De même, si le système *DeepMind* surclasse les meilleurs joueurs d'échecs humains, il s'est révélé incapable d'obtenir la moyenne à

un examen de mathématiques de niveau lycée (faute notamment de bien comprendre les énoncés), ou encore a répondu que 1+1+1+1+1+1+1=6[53].

Il convient ainsi de se garder du biais cognitif qui tend à déduire la prétendue supériorité générale de la machine sur l'humain de la seule circonstance que la première est capable d'accomplir certaines fonctions spécifiques avec une performance supérieure au second. Il faut encore et toujours rappeler qu'au regard de l'éventail des actions qu'un humain est capable de réaliser et des conditions dans lesquelles il le fait, y compris en utilisant son bon sens, sa capacité d'adaptation, ses émotions, l'humour ou d'autres qualités relationnelles, les systèmes d'IA font toujours pâle figure.

Outre la disponibilité des ressources nécessaires, notamment en données, c'est la facilité ou la difficulté de modéliser la tâche, c'est-à-dire de définir humainement (dans les systèmes basés sur les règles) ou de générer automatiquement (dans l'apprentissage machine) les opérations à accomplir (le modèle algorithmique) pour obtenir le résultat recherché (données de sortie) à partir d'une situation initiale (données d'entrée), qui constitue le facteur déterminant.

Certaines tâches sont automatisées de longue date, par des systèmes basés sur les règles. Il y a bien longtemps qu'il est possible de jouer à un jeu vidéo « contre l'ordinateur », ce dernier suivant des instructions programmées, lui prescrivant de réagir de telle ou telle manière dans telle ou telle situation afin de battre le joueur humain ou un autre ordinateur. Les systèmes d'information de l'administration sont en général basés sur le paradigme symbolique, en ce qu'ils produisent mécaniquement des résultats par l'application d'instructions simples.

De même, les systèmes simples fondés sur les statistiques sont maîtrisés depuis longtemps et continuent d'être utilisés avec grand profit pour comprendre le monde et prévoir des évolutions.

Grâce au dynamisme de la recherche, des fonctionnalités avancées des systèmes d'IA basées sur l'apprentissage machine ont aujourd'hui acquis une très grande maturité technique. Sans prétention à l'exhaustivité, les développements qui suivent présentent les principales d'entre elles, en s'abstenant à ce stade de toute appréciation sur l'intérêt, le bien fondé, la dangerosité ou l'acceptabilité de l'emploi de ces techniques, notamment par les pouvoirs publics : il s'agit d'en comprendre les capacités, pour réfléchir, ensuite, à l'opportunité de leur usage.

1. La vision par ordinateur

Introduite à partir des années 1960, la **vision par ordinateur**, ou *computer vision*, est le domaine qui a connu les progrès les plus spectaculaires depuis dix ans, avec l'utilisation des techniques d'apprentissage profond, en particulier des réseaux de neurones convolutifs, appliquées à des jeux de données toujours plus riches. Elle permet d'analyser des images fixes, des vidéos ou des objets afin d'identifier et de

[53] D. Saxton et a., *Analysing Mathematical Reasoning Abilities of Neural Models*, Conference paper ICLR 2019.

catégoriser des personnes et des choses à partir de leurs caractéristiques physiques (formes, proportions, couleurs, motifs…). La précision des modèles tend désormais à dépasser celle de l'humain[54] et le temps nécessaire à l'apprentissage ne cesse de se réduire[55].

L'**identification (et la distinction) de personnes** à partir de métriques physiques (reconnaissance biométrique), notamment du visage (reconnaissance faciale), utilisée aussi bien dans des produits et services commerciaux (pour « débloquer » son smartphone ou pour classer les clients dans l'ordre d'arrivée au comptoir d'un pub anglais...) que dans des services publics comme la sécurité (identification d'un suspect ou d'une victime à partir d'une photo ou d'une vidéo) ne constitue que l'une des très nombreuses illustrations des potentialités de cette fonctionnalité. Les expressions du visage peuvent être analysées pour reconnaître des **émotions** (des spectateurs d'un événement, d'une personne interrogée...) ou le degré d'attention d'une personne (du conducteur d'un véhicule, des élèves en classe...).

La vision par ordinateur permet aussi d'**identifier des choses** (des caractères manuscrits ou typographiques sur un document, comme une adresse postale sur une enveloppe, ou un objet, comme la plaque d'immatriculation d'un véhicule, une chaise sur une photo, une installation nucléaire en construction sur une image satellite, un panneau de limitation de vitesse au bord de la route, un masque sur le visage d'une personne...), de les **localiser sur l'image et de les détourer** (présence d'objets différents sur une même image, détection d'un objet dans une vidéo) et de **dénombrer et classer des entités par catégories** (comptabiliser le nombre de voitures sur une route ou de rhinocéros dans un parc africain, rattacher un champignon à une variété vénéneuse ou comestible, identifier la race d'un chien sur une image, trier des photos de monuments historiques ou l'ensemble des images représentant une casquette...).

Cette fonctionnalité est ainsi très utilisée pour **déceler des anomalies par rapport à un état de référence** (présence d'une tumeur cancéreuse sur les radios, pièce d'identité ou billet de banque frauduleux, analyse des dégâts causés aux bâtiments par des intempéries ou un sinistre, détection de la chute d'une personne, d'un mouvement de foules ou de la fuite d'un animal en captivité, défaut d'une pièce sur une ligne de production dans le cadre du processus de contrôle de la qualité…).

---

[54] Le taux d'erreur sur la reconnaissance d'un objet sur une image (défi ImageNet) s'est effondré : en 2013, le système de reconnaissance d'objet le plus performant réussissait, dans 63% des cas, à identifier un objet sur une image ; ce taux est passé à 90% en 2021. Lorsqu'on s'intéresse à la question de savoir si la réponse correcte figure parmi les cinq premières « prédictions » du système, le taux de précision est passé de 85% à 98% sur la même période, une performance supérieure à celle d'un humain moyen (environ 95%). Les performances tendent toutefois à plafonner depuis quatre ans. Source : AI100, *AI Index Report 2021*.

[55] Sur le défi ImageNet, la durée nécessaire à l'entraînement du modèle est passée de 6,2 minutes en 2018 à 47 secondes en 2020 (même source). Le coût de l'entraînement correspondant (pour un taux de précision de 93%), qui était d'environ 2000 dollars en 2017, était de l'ordre de 7 dollars en 2019.

La vision par ordinateur permet aussi de générer des contenus visuels, incluant aussi bien des « œuvres d'art » (transformer la photo d'un paysage en peinture impressionniste) que des hyper-trucages (*deep fakes*), vidéos à des fins d'usurpation d'identité.

2. Le traitement automatique des sons

Les technologies de détection des sons permettent d'identifier, à partir de bases de données sonores, tous types de sons (voix, instruments de musique, signaux sonores non encore classifiés…) et de les exploiter. Elles sont notamment utilisées par les logiciels de **dictée vocale** (*speech-to-text*) ou de **commande vocale** (comme Siri d'Apple ou Alexa d'Amazon, couplées à des briques de traitement automatique du langage), mais aussi dans le domaine de la sécurité (détection de bruits suspects dans un bâtiment sécurisé), du divertissement (clonage de voix) ou de l'environnement (détection de mammifères pour éviter les collisions avec les bateaux).

Là encore, des progrès continus sont accomplis, en particulier en matière de conversion de paroles en texte (sur le jeu de données LibriSpeech, le taux d'erreur sur les mots est passé de plus de 5 % en 2016 à 1,4 % en 2020) et de reconnaissance du locuteur (sur le jeu de données VoxCeleb, la machine la plus performante se trompait sur l'identité de la personne qui parle dans près de 8% des cas en 2017 ; ce taux est passé à 0,6% en 2020).

3. Le traitement automatique du langage (naturel)

Le traitement automatique des langues (TAL), ou traitement automatique du langage naturel (TALN) compte parmi les briques fonctionnelles de l'IA les plus développées à ce jour, grâce à des progrès réalisés ces toutes dernières années[56].

Il permet l'**analyse sémantique de textes** et l'identification de leur contenu (ex. les entités nommées à anonymiser dans une décision de justice ; le traitement du courrier pour l'orienter automatiquement vers le service chargé d'y répondre), le sens (repérage des fausses informations par exemple), de les résumer ou de les reformuler, et même d'en identifier la tonalité émotionnelle. A ce titre, les modèles de TALN décuplent les possibilités de recherche d'informations précises au sein d'importants *corpus* de textes et la collecte massive de données de textes (*data mining*).

Le TALN est également utilisé par les logiciels de **traduction**, de **génération automatique de textes** (agents conversationnels ou *chatbots*, réponses types à des mails, description sommaire des établissements dans Google Maps, rapports de présentation de comptes annuels[57]), de **catégorisation** (classement automatique

[56] La performance moyenne des meilleurs systèmes dans la compréhension de textes en anglais, évalué dans le cadre de l'initiative SuperGLUE, est passée de moins de 50% en 2019 à environ 90% en 2021, un score légèrement supérieur à celui d'un humain (AI100, *AI Index Report 2021*).

[57] L'un des usages controversés de l'IA générative en matière de texte consiste à modéliser les idées, le vécu, les expressions et les tournures de phrase d'une personne de son vivant, en analysant la masse de ses productions écrites (notamment ses messages) et orales (vidéos), afin de construire un système capable, après sa mort, de « converser » avec ses proches.

d'un mail dans un dossier spam). Sur la base d'un historique de consultation ou d'achats, le TALN sert à des applications de recommandations d'achats (ex. sites de vente en ligne). Associé au traitement de la voix, il contribue aux applications de synthèse vocale (*text-to-speech*), tels les *voicebots* ou les agents conversationnels (par exemple, dans les enceintes connectées).

En général, les modèles de TALN utilisent la vectorisation de mots et l'« intégration lexicale » (*word embedding*), qui permet notamment de déduire un mot du contexte dans lequel il est utilisé (en le rapprochant des autres mots de la phrase dans laquelle il est inséré), donc, par exemple, de générer une réponse pertinente à la question posée.

4. L'acquisition, la gestion et la représentation de la connaissance

Cette brique fonctionnelle a pour objet d'explorer et d'analyser de très grands volumes de données afin de **qualifier ces données** (reconnaissance de leurs caractéristiques, ou *pattern recognition*), de les ordonner, de **trouver des corrélations entre elles**, qui pourront ensuite être qualifiées ou non de liens de causalité, ou encore de les représenter. Elle permet de transformer des jeux de données « brutes » (non ou peu structurées) en informations intelligibles et en connaissances opérationnellement exploitables pour l'humain.

Parmi les utilisations classiques de cette brique fonctionnelle, on trouve la « visualisation des données » (« *data viz* »), qui consiste à représenter les données traitées sous une forme compréhensible (par des graphiques, des cartographies, des environnements virtuels…) et l'analyse dite « prédictive » qui consiste à tirer d'une masse de données historiques les variables explicatives d'un phénomène (météo, sanitaire, délinquant…) afin d'identifier la probabilité de ses futures manifestations. Cette seconde technique est notamment utilisée dans le cadre de la recherche de fraudes (pour identifier les écarts à la norme de comportement), dans le marketing (identification des comportements d'achats des consommateurs en fonction d'une série de critères), dans la santé (identification des facteurs de risques pour le développement d'une pathologie à partir d'une masse de données cliniques…) et dans l'industrie (maintenance prédictive permettant d'identifier les signes précurseurs de pannes à partir d'une série de signaux).

5. La robotique

Le terme robot évoque intuitivement la figure humanoïde popularisée par de nombreuses œuvres littéraires et cinématographiques de science-fiction, et par les réalisations expérimentales plus ou moins convaincantes dont la presse se fait régulièrement l'écho. Cette notion est en réalité plus polysémique. On peut distinguer schématiquement la robotique matérielle et la robotique immatérielle.

La première, qui est la plus intuitive, renvoie à l'ensemble des **systèmes automatisés interagissant directement avec l'environnement physique**, quelle que soit la forme physique dans laquelle ils sont embarqués. Les cas d'application sont innombrables, des drones aux appareils ménagers et de jardin dits « intelligents » (aspirateurs, tondeuses…) en passant par le traditionnel bras articulé sur une chaîne de production

de produits industriels (voitures par exemple), les véhicules assistés et « autonomes » ou encore les robots de gardiennage, chargés du déminage ou du démantèlement d'installations nucléaires (« Spot Mini »). La **cobotique** est le champ de recherche et d'application portant sur la coopération entre l'humain et la machine qui l'assiste, afin d'en exploiter la complémentarité (les robots médicaux par exemple).

La seconde, connue sous l'appellation anglo-saxonne de « bot », renvoie à un logiciel **simulant le comportement humain dans l'environnement numérique**. Il inclut notamment l'agent conversationnel (chatbot) et l'assistant vocal (voicebot), ainsi que le domaine de l'automatisation de process par robot (RPA, pour robotic process automation), consistant à confier à un agent machine des tâches numériques souvent répétitives (collecte de données dans divers système d'information, saisie de données, traitement de courriers électroniques…) qu'il accomplit comme le fait l'humain, le cas échéant par un apprentissage fondé sur l'observation du comportement de l'homme (quoique plus rapidement et sans erreur liée à la fatigue par exemple).

Pour conclure sur ce point, il convient de préciser que **ces différentes briques fonctionnelles peuvent se combiner au sein d'un même produit ou service**. C'est même très souvent le cas. Une enceinte connectée est l'archétype du produit composite : la reconnaissance de la voix de l'utilisateur (déclenchement de l'enceinte et captation des propos), la conversion de la question énoncée en texte exploitable par la machine, la construction de la réponse et la transformation de celle-ci en voix (de la machine) font appel, chacune, à l'intelligence artificielle. A l'inverse, un produit ou une prestation donné peut ne faire appel à l'IA que pour une toute petite partie de ses fonctions.

Le recensement des cas d'usage de l'IA dans le secteur public, synthétisé dans la deuxième partie de l'étude et présenté dans l'annexe 9, donne de multiples illustrations de la mise en œuvre de ces briques fonctionnelles.

1.2. Promouvoir une compréhension partagée et rationnelle de l'IA

Avant d'expliciter les motifs qui paraissent établir la nécessité de s'engager résolument dans le déploiement de SIA publics, la discussion de cet impératif doit s'engager sur des bases communes. Celles-ci ne pourront être édifiées qu'en déblayant le terrain du débat des fantasmes qui l'encombrent (le mythe de la singularité les synthétisant tous), en acquérant une culture collective commune qui exige de l'État un effort important de pédagogie, et en partageant des concepts juridiques de base que le droit européen, en première analyse, est à même de fournir.

1.2.1. Écarter le fantasme de la « singularité »

La réflexion sur l'intelligence artificielle est souvent parasitée par la focalisation excessive sur l'intelligence artificielle générale[58] et le « mythe de la singularité », qui définit le point de bascule d'un monde dans lequel la machine intelligente surpasse

[58] V. sa définition dans l'encadré ci-dessus.

et domine les humains. Des récits de science-fiction aux nombreux ouvrages de « vulgarisation » sur l'IA naît l'idée, largement admise outre-Atlantique, qu'il est possible d'inventer une telle machine, que cette échéance est comprise entre après-demain et une cinquantaine d'années, que rien ne garantit qu'elle s'assignera des fins et recourra à des moyens acceptables pour les humains, à commencer par leur compatibilité avec la survie de l'espèce humaine[59] et qu'il est nécessaire d'imaginer des « lois » (comme les lois de la robotique d'Isaac Asimov) et de veiller au bon « alignement » des systèmes, concept réducteur issu de la culture des jeux de rôle d'*Héroïc Fantasy*[60] qui invite à fonder entièrement l'éthique sur un choix des « bonnes valeurs » pour les programmer dans le système[61].

Ce fantasme, onirique pour les uns, cauchemardesque pour les autres, présente l'avantage d'attirer l'attention sur de nombreux problèmes réels, notamment le préoccupant silence des philosophes, à de rares exceptions près, la significative et parfois inquiétante prévalence de biais culturels dans les conceptions qui s'affrontent, et l'absence de traitement historique de la question. Pour autant, l'hypothèse de la survenance de la singularité, qui suscite une imposante littérature, des colloques, et des prises de position publique passionnées, a été battue en brèche par de nombreux auteurs, en particulier au sein de la communauté scientifique.

Elle ne pourrait être un outil conceptuel puissant pour inviter à réfléchir sur les règles qu'en refusant les raccourcis méthodologiques qui conduisent ses partisans à bondir à la conclusion de l'omnipotence de la machine sans expliquer les étapes intermédiaires, en laissant à l'état d'hypothèses implicites plusieurs sauts critiques, en considérant que pouvoir comprend vouloir, en confondant intelligence humaine et intelligence artificielle toutes deux réduites à des phénomènes physiques dont l'équivalence technique amènerait à celle des comportements, croyant que la capacité potentielle d'une intelligence lui permettrait de créer aussi les moyens matériels de l'exercice physique de ses potentialités ou encore négligeant toute capacité humaine à prendre la moindre précaution.

---

[59] On se bornera ici à souligner que l'énergie nécessaire au fonctionnement de la machine peut s'avérer non soutenable d'un point de vue environnemental pour l'espèce humaine et à renvoyer, sur ce point précis, aux développements en 3.1.6.

[60] En règle générale, l'alignement d'un personnage est défini selon deux dimensions : l'objectif qu'il poursuit à l'égard des autres et du monde (« bon », « mauvais », « neutre ») et la fidélité à des règles de conduite (qui oppose les personnes « loyales » aux personnes « chaotiques », en passant par les personnages « neutres »). La combinaison des deux caractéristiques (de « loyal bon » à « chaotique mauvais ») donne une base éthique dont on conviendra qu'elle est un peu en retrait de l'éthique à Nicomaque ou de la critique de la raison pure...

[61] Ce que veut dire suivre les bonnes valeurs, en dépit de l'abondance de la littérature, demeure d'ailleurs assez mystérieux. Sur ce point, des exemples dans : *Possible minds, 25 ways of looking at AI*, ed John Brockman, Penguin 2020, B. Christian, *The alignment problem*, Atlantic books, 2020 ; S. Russel, *Human Compatible*, Penguin 2020 ; M. Gibert, *Faire la morale aux robots*, Flammarion 2021. Eléments historiques du débat dans l'étude fondatrice de N. Wiener, *The human use of human beings*, Da capo Perseus 1950 ; contra : A. Daub, *What the tech calls thinking,* Farar Straus Giroux, 2020 ; J. Gabriel Ganascia, *Le mythe de la singularité*, Seuil, 2020.

En outre, cette approche d'ordre eschatologique de la question occulte le fait que, d'ores et déjà, des systèmes d'IA exercent, sur le fonctionnement des sociétés, une influence majeure, mais plus discrète que l'avènement du « règne des machines », selon des finalités (bonnes ou mauvaises) définies par leurs concepteurs ou avec des conséquences qui leur échappent.

Il est indispensable que la stratégie de l'IA publique fasse pièce à ces fantasmes. La réalité est que les systèmes d'IA ne sont que des applications informatiques, parfois basiques, parfois puissantes, mais qui ne disposent en aucun cas d'une autonomie qui leur permettrait de s'affranchir de l'humain. **Le risque n'est pas que la machine prenne le pouvoir mais que l'humain en perde le contrôle** en se défaussant de ses responsabilités. C'est la raison pour laquelle l'IA publique de confiance doit notamment être fondée sur le **principe de primauté humaine**[62].

Il convient en outre de ne pas verser dans la caricature qui tendrait à assigner aux systèmes d'IA une vocation exclusive à remplacer l'humain. S'ils peuvent assumer une partie des fonctions aujourd'hui dévolues aux femmes et aux hommes qui composent l'administration, ils en sont le plus souvent complémentaires, la machine pouvant permettre à l'humain d'exprimer tout son potentiel ou de le consacrer aux tâches qui requièrent son intelligence singulière (on parle parfois d'humain « augmenté » ou, pour décrire l'interaction entre l'humain et le système d'« interface homme-machine »).

1.2.2 Permettre aux citoyens d'acquérir une culture des concepts et enjeux de l'IA

L'ignorance suscite plus souvent la défiance, voire l'hostilité, que la curiosité ou l'intérêt. L'IA est particulièrement exposée à ce phénomène en raison de sa technicité, qui est réelle, et des pouvoirs magiques qui lui sont parfois prêtés, qui n'existent pas. Le risque d'un rejet en bloc des SIA est d'autant plus grand que, à l'instar de l'informatisation par le passé, l'automatisation est souvent associée, dans l'imaginaire collectif, à la destruction d'emplois et à la déshumanisation de la société. Les SIA publics y sont particulièrement vulnérables en raison de la méfiance croissante que peut susciter l'action publique au sein d'une partie du corps social. Les récents débats publics sur les nanotechnologies, le déploiement des réseaux de communications électroniques de la 5e génération ou la vaccination obligatoire montrent la sensibilité de l'opinion à l'introduction de technologies nouvelles ou de produits nouveaux et le risque que la vulgarisation intéressée ou le complotisme remplacent la pédagogie et le contrôle civique.

En sens inverse, l'ignorance peut susciter un excès de confiance et conduire à négliger les risques associés à ces outils. Cette ignorance peut elle-même être déniée et laisser la place à des certitudes infondées. Elle peut encore être flattée par des opérateurs dénués de scrupules préférant laisser dans l'ombre la nature de leurs pratiques.

[62] V. 3.1.1. sur la primauté humaine comme principe de l'IA publique de confiance.

A l'heure actuelle, il n'existe pas d'aversion sociale généralisée, ni même majoritaire, aux SIA, au contraire. Selon le sondage annuel de l'IFOP « Notoriété et image de l'intelligence artificielle auprès des Français et des salariés »[63], 77% des personnes interrogées déclarent avoir une « bonne image » de l'IA, et 64% déclarent lui faire confiance, dont 73% parmi les 18-24 ans. Elles sont toutefois moins de 10% à en avoir une « très bonne image ». La concertation citoyenne organisée par la CNIL sur les « enjeux éthiques liés à la place des algorithmes dans notre vie quotidienne » en 2017 a confirmé que, si les algorithmes ne suscitaient aucune hostilité *a priori*, les citoyens du panel attachaient une très grande importance aux garanties entourant leur développement. Le corps social semble donc manifester, en majorité, une forme de bienveillance vigilante à l'égard de ces techniques. La **conduite d'une plus large enquête, centrée sur les systèmes d'IA publics**, permettrait d'appréhender de manière plus fiable et plus fine l'état des mentalités.

Ces éléments sur l'état de l'opinion publique doivent cependant être mis en regard du degré de connaissance de l'IA. À cet égard, on ne peut qu'être frappé de constater que, selon le même sondage de l'IFOP, environ 50% des personnes interrogées sur l'IA disent « *voir précisément de quoi il s'agit* ». Le sondage commandé par la Commission européenne en 2020 a montré qu'en moyenne, 8 entreprises européennes sur 10 déclaraient savoir ce qu'est l'IA. Selon l'enquête de l'IFOP « Notoriété et attentes vis-à-vis des algorithmes », commandée par la CNIL et publiée en janvier 2017, 31 % des personnes interrogées voient précisément ce que sont les algorithmes. Or il est certain que la proportion des personnes ayant objectivement une vision claire et techniquement étayée de ce qu'est l'IA est très sensiblement inférieure. Il semble que l'omniprésence médiatique du sujet, sous un prisme souvent sensationnaliste, la vulgarisation de discours pseudo-scientifiques et l'exploitation cinématographique de l'IA, aient créé une redoutable illusion de connaissance qu'il faut s'employer à dissiper.

Le rehaussement du niveau général de compréhension du fonctionnement des systèmes d'IA doit constituer un axe prioritaire de la stratégie pour le déploiement de SIA publics, et poursuivre deux objectifs.

Le premier est de favoriser une compréhension minimale de ce qu'est l'IA, de ce qu'elle n'est pas, ce qu'elle permet de faire et ce qu'elle ne permet pas de faire, en remettant le sujet à sa juste place, sans se retrancher derrière la complexité technique d'un sujet qui ne saurait rester confiné aux cénacles scientifiques.

En termes de méthode, la compréhension des enjeux de l'IA et l'acceptation de son utilisation dans la sphère publique **ne peuvent résulter d'une démarche descendante et assertive**. Il ne suffira pas d'asséner que les SIA publics sont utiles pour en convaincre le citoyen. Au catéchisme de l'IA, il faut en préférer la pédagogie, exemples à l'appui. La prise de conscience de ce que le quotidien des Français est d'ores et déjà largement structuré par ces outils est une étape clé[64].

---

[63] RB/EB n° 117705, décembre 2020.

[64] Selon le sondage IFOP « Notoriété et image de l'intelligence artificielle auprès des Français et des salariés » de décembre 2020, précité, seuls 27% des Français pensent avoir déjà eu

Il existe d'ores et déjà une offre gratuite de cours en ligne qui pourrait utilement être adaptée et déployée, rapidement et pour un coût limité, dans les écoles, en commençant dès le niveau élémentaire[65], comme dans les administrations et les entreprises. A titre d'exemple, l'Université d'Helsinki a conçu et diffusé en 2018 un cours en ligne (« *Elements of AI* »), disponible dans la plupart des langues des États membres de l'Union, notamment en français[66], et permettant à tout un chacun de se familiariser avec les concepts de l'IA. Cette approche générique doit être complétée par la description non-technique et très concrète du fonctionnement d'objets de la vie courante et de services administratifs recourant à ces techniques. Il faut expliquer qu'un SIA ne peut, par exemple, pas (encore ?) assurer de bout en bout la collecte des ordures ménagères, mais peut optimiser les circuits de collecte au bénéfice premier des agents de collecte, tout en réduisant nuisances et coûts ; un SIA ne délivre pas (encore ?) de permis de construire, mais il peut aider à l'instruction du dossier en automatisant certains points de contrôle…

Naturellement, viser l'émergence d'une culture commune de l'IA est vain s'il n'existe pas, plus largement, une **culture commune du numérique**. A ce titre, la sensibilisation au numérique dès le plus jeune âge est primordiale, aussi bien pour promouvoir l'usage légal et raisonnable des outils (notamment des réseaux sociaux) que pour la compréhension de leur fonctionnement. **Les langages de programmation sont d'authentiques langues vivantes**, au même titre que l'anglais ou l'espagnol. Ils mériteraient d'être enseignés de la même façon.

En outre, le développement des SIA publics ne peut qu'aller de pair avec le déploiement d'actions spécifiques en direction des personnes qui rencontrent des difficultés avec les outils numériques. Par rapport à la moyenne de l'UE (25% de faibles compétences numériques), la littératie numérique en France demeure médiocre ; à titre d'illustration, elle est très inférieure à un pays comme les Pays-Bas (4% de non-raccordés à l'internet contre 12% en France, et 16% de faibles compétences numériques seulement contre 29% en France).

Selon l'INSEE[67], **l'illectronisme** (ou illettrisme numérique) touche 17% de la population : une personne sur six n'utilise pas Internet, une personne sur quatre ne sait pas s'informer *via* Internet, et une personne sur cinq est incapable de communiquer sur ce réseau. L'illectronisme touche essentiellement les personnes de 75 ans ou plus (8,81% de cette tranche d'âge), mais également, dans une moindre mesure les personnes sans diplôme ou n'ayant qu'un CEP (4,02%). Ainsi, l'âge est

recours à des produits ou services faisant appel à l'IA dans leur vie personnelle, et 16% dans leur vie professionnelle, ce qui est manifestement bien inférieur à la réalité.

[65] Dans son rapport déposé le 15 mars 2017 (*Pour une intelligence artificielle maîtrisée, utile et démystifiée*, C. de Ganay et D. Gillot), l'Office parlementaire d'évaluation des choix scientifiques et technologiques préconisait déjà d'organiser des formations à l'informatique dans l'enseignement primaire et secondaire faisant une place à l'intelligence artificielle (proposition n° 12).

[66] Il a été adapté par une équipe de Sorbonne Université et est disponible à l'adresse suivante : https://course.elementsofai.com/fr/

[67] INSEE Première, n° 1780, octobre 2019.

facteur discriminant, mais, contrairement aux idées souvent véhiculées, le manque de compétences numériques touche aussi largement les jeunes, loin d'être tous *digital natives* : une personne entre 15 et 29 ans sur cinq souffre d'une incapacité numérique, ce qui constitue un frein majeur à la recherche d'un emploi, à l'apprentissage tout au long de la vie, et plus généralement à l'insertion sociale. Par rapport au développement des pratiques sociales numériques, cette situation entraîne un sentiment de déclassement et une contestation de la pertinence de l'essor de la société numérique, plus accentués encore chez les jeunes concernés.

Les politiques en faveur de l'inclusion numérique, et notamment le Plan national pour un numérique inclusif (septembre 2018) doivent donc intégrer la spécificité des SIA, en particulier en termes d'explicabilité (formation des aidants et des conseillers numériques France Service par exemple). Cette acculturation est de nature non seulement à assurer une meilleure acceptabilité des SIA, mais à susciter la demande de SIA[68] voire l'envie de participer à leur conception et à leur amélioration, mais aussi à leur contrôle.

Le second objectif de l'acculturation consiste à cet égard à **aiguiser le sens critique des citoyens sans alimenter la défiance**. L'information sur l'IA ne doit pas relever de la propagande, mais décrire loyalement et de façon équilibrée les grands principes de fonctionnement des SIA, les avantages et inconvénients qu'ils présentent, les potentialités et les limites, les risques et les garanties qui les entourent (V. sur ce point la troisième partie de l'étude). L'**association des acteurs de la société civile aux décisions stratégiques en matière d'intelligence artificielle** est indispensable.

Cela emporte une conséquence : ces acteurs doivent eux-mêmes être sensibilisés et formés, et se départir de postures dogmatiques fustigeant par principe « l'IA », parfois en réaction aux promesses excessives et aux discours lénifiants. De la même manière qu'une association de patients a besoin d'expertise médicale pour jouer un rôle utile, qu'une association environnementale a besoin d'expertise pour n'être pas instrumentalisée, qu'un syndicat bénéficie de crédits publics pour former ses membres afin d'égaliser les conditions de compétence, il faut développer une offre de formation au bénéfice du tissu associatif – militant, syndical, consumériste – pour assurer l'égalité des armes avec les promoteurs des SIA. Donner les garanties d'un contrôle des représentants des usagers sera vécu comme une farce ou une illusion, si ces représentants ne sont pas formés, de manière adéquate et indépendante, à leur tâche.

1.2.3. Œuvrer à des concepts juridiques communs aux niveaux européen et mondial

Un concept aussi hétérogène et évolutif que l'intelligence artificielle se prête particulièrement mal à l'adoption d'une définition juridique. Celle-ci est toutefois nécessaire dès l'instant qu'on envisage d'encadrer les systèmes d'IA par un régime juridique dédié, là où le cadre juridique actuel, synthétisé en annexe 10, ne permet pas de répondre entièrement au besoin d'encadrement des produits et services

[68] V. 2.3.1. sur l'éventualité de voir émerger la revendication d'un droit à l'IA.

recourant aux techniques précédemment mentionnées, compte tenu des risques qu'ils présentent et qui sont décrits dans la troisième partie de cette étude.

1. L'enjeu terminologique du règlement européen sur l'IA

La Commission européenne a pris l'initiative d'une proposition de règlement établissant des règles harmonisées concernant l'intelligence artificielle[69] (dite législation sur l'intelligence artificielle ou *AI Act)* qui constituerait le premier cadre juridique transversal et spécifique à l'IA dans le monde. Le contenu de cette proposition, incluant les modifications proposées par la présidence slovène en décembre 2021, est présenté en annexe 6 et résumé ci-après.

Aperçu du projet de règlement européen sur l'IA

Les principales caractéristiques de cette initiative sont les suivantes :

- Un **champ d'application large**, tant au regard de la définition des systèmes d'IA (V. ci-dessous) que sur le plan organique (applicabilité aux SIA publics comme privés), géographique (le droit européen s'appliquerait non seulement aux systèmes mis en service et utilisés dans l'Union, mais aussi à ceux qui produiraient des résultats utilisés dans l'Union) et fonctionnel (exclusions sectorielles limitées) ;

- Une **approche par les risques** : contrebalançant le champ d'application matériel large, la législation européenne proportionnerait les contraintes aux risques présentés par chaque système. Schématiquement, elle comporterait quatre étages : 1° quelques interdictions de principe (SIA à des fins de manipulation mentale, d'abus de faiblesse, de « notation sociale » et, sauf exceptions, d'identification biométrique dans l'espace public à des fins répressives) ; 2° un régime juridique complet pour les SIA dits « à haut risque » (par exemple, ceux qui sont utilisés pour l'admission à l'université, la délivrance de prestations sociales ou les contrôles aux frontières), incluant des exigences sur les données, la transparence, le contrôle humain, l'exactitude, la robustesse et la cybersécurité… ; 3° des obligations de transparence minimales pour certains SIA ; et 4°, pour le surplus des systèmes (présentant un risque faible), une simple incitation à l'adoption de codes de conduite ;

- Une **approche « produit »** : l'essentiel des obligations pèserait sur le fournisseur, entendu comme l'opérateur qui met le système sur le marché (ou en service). Il devrait, pour les SIA à haut risque, mettre en place un système de gestion de la qualité incluant un système de gestion des risques et élaborer une documentation technique, et respecter les exigences du règlement et les normes harmonisées qu'il prévoit (ou les spécifications communes). Ces SIA feraient l'objet d'une certification préalable à leur mise sur le marché et d'un marquage CE. Des obligations de surveillance après commercialisation sont prévues et un rôle important est dévolu aux autorités de surveillance des marchés.

[69] COM (2021) 206 final, 2021/0106 du 21 avril 2021.

- Un **texte évolutif** : afin d'éviter l'obsolescence prématurée, le règlement comporterait de nombreuses souplesses par le biais de renvois à des actes délégués de la Commission européenne (notamment pour l'énumération des SIA à haut risque) et à des normes harmonisées.

- Un **corps de règles autonome** : pour l'essentiel, les obligations posées par le règlement s'ajouteraient aux règles existantes, sans s'y substituer. Le principe posé est que la conformité à ce règlement ne vaudrait pas conformité au RGPD, s'agissant des SIA constituant aussi des traitements de données à caractère personnel. Le texte comporte quelques rares souplesses dans la mise en œuvre du RGPD pour la conception et l'utilisation des SIA (« dé-biaisement » des modèles par l'utilisation de données sensibles, « bacs à sable réglementaires », prise en compte de la notice d'utilisation du fournisseur par l'utilisateur en tant que responsable du traitement de données à caractère personnel…).

Afin de ne pas introduire de désordre terminologique dans un univers sémantique déjà complexe, **il importe que le droit national et la réflexion interne reprennent autant que possible les principaux concepts et les définitions associées** tels qu'ils résulteront du droit de l'Union. Un soin particulier doit donc être apporté à leur élaboration. On se bornera ci-après à évoquer les principaux concepts sur lesquels repose le projet de règlement.

a/ S'agissant des systèmes d'IA, l'article 3 de la proposition initiale de la Commission européenne a retenu une définition largement inspirée de celle de l'Organisation de coopération et de développement économiques (OCDE) : un SIA est défini comme un « *logiciel qui est développé au moyen d'une ou plusieurs des techniques et approches énumérées à l'annexe I et qui peut, pour un ensemble donné d'objectifs définis par l'homme, générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels il interagit* ». Le SIA se caractériserait donc par sa **nature** (un logiciel), les **modalités de sa conception** (l'une des approches énumérées à l'annexe I de la proposition, qui sont celles évoquées précédemment[70] : approches d'apprentissage automatique[71], approches fondées sur la logique et les connaissances[72] – ce qui renvoie aux systèmes basés sur les règles, c'est-à-dire l'IA symbolique, et approches statistiques[73]), et une **fonction** (générer des résultats sur la base d'objectifs définis par l'homme).

[70] Cette annexe pourrait être complétée par la Commission pour tenir compte des évolutions technologiques.

[71] Y compris l'apprentissage supervisé, non supervisé et par renforcement, utilisant une grande variété de méthodes, y compris l'apprentissage profond.

[72] Y compris la représentation des connaissances, la programmation inductive (logique), les bases de connaissances, les moteurs d'inférence et de déduction, le raisonnement (symbolique) et les systèmes experts

[73] La proposition évoque à ce titre l'estimation bayésienne ainsi que les « méthodes d'exploration et d'optimisation ».

Cette formulation très large, technologiquement neutre, permet de ne pas introduire de distorsions de concurrence entre les fournisseurs de systèmes selon la technique utilisée. Elle présente en revanche le risque d'attraire dans le champ d'application du règlement l'ensemble des logiciels informatiques et des produits les utilisant, alors même qu'ils ne sont pas communément considérés comme relevant de l'IA. En effet, un simple tableur informatique (comme Excel) génère des résultats de calculs à partir de données d'entrée en recourant à un modèle algorithmique basé sur les règles, de même qu'un logiciel de traitement de texte propose des corrections orthographiques ou des synonymes selon une approche comparable. Or l'objectif n'est manifestement pas d'encadrer la fourniture et l'utilisation de tels produits.

C'est la raison pour laquelle la présidence slovène a proposé au Conseil Transports, télécommunications et énergie (Télécommunications) du 3 décembre 2021 de resserrer le champ d'application du texte en utilisant la définition suivante : « *un système d'IA est un système qui : / (i) reçoit des données et des entrées provenant de machines et/ou de l'homme, / (ii) déduit comment atteindre un ensemble donné d'objectifs définis par l'homme en utilisant l'apprentissage, le raisonnement ou la modélisation mis en œuvre avec les techniques et les approches énumérées à l'annexe I, et / (iii) génère des sorties sous forme de contenus (systèmes d'IA générative), prédictions, recommandations ou décisions, qui influencent les environnements avec lesquels il interagit* ».

Le SIA ne serait donc plus caractérisé par sa nature (logiciel) mais par son mode d'alimentation (i), sa capacité à déduire lui-même une méthodologie d'atteinte des objectifs fixés par l'homme à l'aide d'une des approches (ii) et la nature de sa production (iii). Outre que la distinction entre « données » et « entrées » au (i) n'apparaît pas clairement, cette caractéristique n'est pas de nature à restreindre le champ d'application, puisqu'elle est partagée par des outils informatiques courants (pour fournir un résultat, la calculatrice est alimentée par une opération entrée par l'utilisateur). Le (iii) reprend la logique de la proposition initiale.

La caractéristique exposée au (ii) semble quant à elle se référer autant à la conception du système qu'à son utilisation : un SIA recourt à l'une des approches énumérées pour déduire la façon d'atteindre les objectifs qui lui sont assignés. Cette formulation ne se laisse pas aisément appréhender. Elle fait écho à l'idée, de maniement délicat, d'« autonomie », évoquée dans les considérants[74], et de complexité, à travers la multiplicité des objectifs et l'existence de plusieurs méthodes pour les atteindre, parmi lesquelles la machine choisit. Elle couvre sans difficulté l'ensemble des SIA fondés sur l'apprentissage machine voire l'approche statistique, qui, d'une certaine manière, construisent leur propre méthodologie de résolution du problème posé à partir des données qui leur sont fournies en phase de conception (et, s'agissant des modèles dont

[74] Avant la formalisation de la proposition de la Commission, le Parlement européen a proposé d'introduire l'autonomie dans la définition des systèmes d'IA, en la définissant comme la caractéristique d'un système qui ne se borne pas à exécuter des instructions prédéterminées (Résolution du Parlement européen du 20 octobre 2020 contenant des recommandations à la Commission concernant un cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes (2020/2012(INL)).

l'entraînement se poursuit en phase d'inférence, en phase d'utilisation). S'agissant des systèmes basés sur les règles, l'idée pourrait être d'exclure ceux dont les résultats procéderaient de l'exécution d'instructions simples et entièrement prédéterminées par l'homme[75], à la différence de modèles plus complexes où, eu égard aux caractéristiques du moteur d'inférence, le système disposerait d'une certaine latitude pour choisir celles qui produisent les meilleurs résultats, sans que l'humain soit toujours en mesure de les anticiper précisément[76].

A supposer que cette lecture soit confirmée, la ligne de partage entre les SIA symboliques relevant ou non du règlement resterait cependant assez incertaine et mériterait, à tout le moins, un éclairage dans les considérants du texte. Quoiqu'il en soit, cette voie semble devoir être privilégiée pour exclure du champ d'application du texte les logiciels basiques et d'usage courant, qui ne présentent aucun risque particulier et qui ne méritent pas même de faire l'objet des simples « codes de conduite » envisagés par le règlement pour les SIA qui ne sont pas « à haut risque »[77]. On ne peut exclure, au final, que cette définition soit interprétée en partie de façon finaliste, afin d'y inclure des systèmes qui méritent, en raison de leur mode de fonctionnement et des risques qu'ils comportent, de faire l'objet de l'encadrement prévu par le texte[78].

b/ S'agissant des « *opérateurs* » des systèmes, le projet de texte distingue notamment :

- le « *fournisseur* », qui est la personne, l'organisme ou l'entité qui « *développe ou fait développer un système d'IA et le met sur le marché ou en service sous son propre nom ou sa propre marque, à titre onéreux ou gratuit* » : il s'agit, en d'autres termes, du fabricant du système, qui est celui qui en définit la « destination » et qui est responsable de la mise à disposition du système auprès des utilisateurs ;

---

[75] Une application « calculatrice », un tableur ou un logiciel de traitement de texte n'entreraient donc pas dans le champ. A titre d'exemple d'outils utilisés par la juridiction administrative, la base de jurisprudence Ariane, dont le moteur de recherche identifie simplement les décisions, analyses ou conclusions comportant le mot-clé entré par l'utilisateur, ne serait pas davantage considérée comme un SIA. Il en va de même de l'application Télérecours en ce qu'elle permet le dépôt de documents, affiche des informations au gré de la navigation et procède à des notifications automatiques pour des évènements entièrement prédéterminés.

[76] En reprenant la typologie proposée par S. Russel et P. Norvig (V. en annexe 4, l'entrée « agent intelligent »), la définition proposée pourrait à tout le moins recouvrir les agents intelligents « fondés sur des buts », les agents « fondés sur l'utilité » et les agents « apprenants ».

[77] On peut observer que la définition de l'IA que donne l'article 5002 du *National Artificial Intelligence Initiative Act* américain de 2020 est très similaire, à ceci près que l'un de ses éléments constitutifs est la capacité du système à « *transformer les informations perçues en modèles par une analyse automatisée* », ce qui se réfère plus naturellement à l'apprentissage automatique.

[78] A titre d'exemple, l'outil Parcoursup, qui ne relève pas de l'apprentissage machine, utilise un modèle algorithmique d'appariement (algorithme de Gale-Shapley) qui consiste à trouver la solution optimale parmi la multitude de couples étudiant-formation. Il semble pouvoir entrer dans la définition posée. Cette inclusion est renforcée par l'importance des enjeux de fonctionnement de ce type d'outil, qui explique que les systèmes d'admission dans les universités soient classés parmi les SIA à haut risque (annexe III).

- et l'« *utilisateur* », qui est la personne, l'organisme ou l'entité « *utilisant sous sa propre autorité un système d'IA* ».

Selon le cas, l'administration pourra avoir la qualité de fournisseur du système – si elle le conçoit elle-même pour les besoins d'un tiers, celle de simple utilisateur, si elle se borne à recourir à un système existant conçu par un tiers (achat « sur étagère »), ou les deux, si elle décide de concevoir et développer le système elle-même, pour ses propres besoins, ou de confier cette mission à un tiers pour son compte. Hormis le cas, assez rares, dans lequel une administration commercialiserait un système d'IA, celle-ci procèdera à une « mise en service » au sens du règlement (« *fourniture d'un SIA directement à l'utilisateur en vue d'une première utilisation ou pour usage propre dans l'Union, conformément à la destination du système* »).

Les citoyens, entreprises, associations et autres personnes formant le « public », qu'ils soient simplement destinataires d'une décision prise par ou avec l'aide d'un système d'IA ou qu'ils interagissent eux-mêmes avec le système (par exemple, l'usager d'une téléprocédure recourant à l'IA, le bachelier formulant ses vœux sur la plateforme Parcoursup, l'internaute interrogeant un robot conversationnel …), n'en sont pas des « utilisateurs » au sens de la proposition de règlement, mais des « *personnes à l'égard desquelles le système d'IA est utilisé* », des personnes qui « *interagissent* » avec le système ou encore celles qui y sont « *exposées* », notions distinctes utilisées par la proposition pour désigner les personnes sur lesquelles le système aura un impact.

c/ Chaque SIA comporte une (ou plusieurs) « **destination**(s) », celle-ci étant définie comme « *l'utilisation à laquelle un système d'IA est destiné par le fournisseur, y compris le contexte et les conditions spécifiques d'utilisation* ». Il s'agit, en d'autres termes, de sa ou de ses finalités, notion-clé du droit des traitements de données à caractère personnel et évoquée dans l'exposé des motifs de la proposition, mais qui n'a pas été reprise par la Commission dans le texte lui-même, le cas échéant pour éviter toute confusion entre les deux *corpus* juridiques.

2. Le droit conventionnel européen

Le Conseil de l'Europe s'est très tôt préoccupé des incidences du développement des systèmes d'IA sur les droits fondamentaux, l'État de droit, la démocratie et la justice (avec l'adoption en 2018 de la charte éthique européenne sur l'utilisation de l'IA dans les systèmes judiciaires, dans le cadre des travaux de la Commission européenne pour l'efficacité de la justice - CEPEJ). Il a mis en place en 2019 un Comité *ad hoc* sur l'intelligence artificielle (CAHAI) chargé d'examiner la faisabilité et le contenu potentiel d'un cadre juridique contraignant[79], telle qu'une convention, dont le principe a été proposé par la commission permanente de l'Assemblée Parlementaire du Conseil de l'Europe au Comité des Ministres.

Il est crucial pour la lisibilité du droit de l'intelligence artificielle que les notions qui seraient utilisées dans une telle convention du Conseil de l'Europe soient convergentes avec celles du règlement IA, de la même façon que les obligations

[79] V. l'étude de faisabilité du 17 décembre 2020, CAHAI (2020) 23.

qu'elle prévoirait, en termes nécessairement généraux, ne devraient pas excéder celles qui résultent, pour les Etats membres, de l'application de ce règlement. A titre de comparaison, la cohabitation entre le RGPD et la convention n° 108 du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel n'a, à ce jour, pas soulevé de difficulté.

3. L'enjeu de la normalisation

La proposition de règlement IA applique les principes du « nouveau cadre législatif » (ou *New Legislative Framework* - NLF), selon lequel la règlementation détermine les « exigences essentielles » et renvoie, pour la définition des exigences techniques, à des normes européennes dites « harmonisées » (et, à défaut de publication de la norme au Journal officiel de l'UE, à des spécifications techniques communes arrêtées par la Commission) dont le respect permettra de présumer celui du règlement s'agissant des SIA à haut risque. En outre, de nombreux travaux ont été engagés, au niveau international, pour faire converger les concepts et les règles de référence en matière d'intelligence artificielle (normes ISO, IEEE…).

Il existe d'ores et déjà plusieurs normes ISO relatives à l'intelligence artificielle, élaborées sous la responsabilité du comité technique ISO/IEC JTC 1/SC 42, dédié à la normalisation de l'IA, et cette production devrait considérablement s'accélérer dans les années à venir.

On peut notamment signaler la norme sur la fiabilité en matière d'IA[80], qui, selon ses propres termes, porte sur la confiance dans les SIA et « examine les approches visant à établir la confiance dans les systèmes d'IA par la transparence, l'explicabilité et la contrôlabilité » ainsi que celles permettant d'évaluer et de réaliser des SIA résilients, fiables, précis, sûrs et confidentiels. A également été publiée, en novembre 2021, une norme sur les biais dans les SIA d'assistance à la prise de décision, tant sur le plan de la collecte des données que sur ceux de l'apprentissage, de l'évaluation et de l'utilisation du SIA[81].

Il faut enfin noter que **plusieurs normes ISO ont donné une première définition de l'intelligence artificielle**, sans que cette définition ne constitue le cœur des normalisations en cause. La norme sur la fiabilité précitée définit ainsi l'IA comme la « capacité d'un système d'ingénierie à acquérir, traiter et appliquer des connaissances et des compétences ». A ce jour, la norme ISO relative au vocabulaire des technologies de l'information[82] donne trois définitions de l'IA, dont une seule se rapporte aux fonctionnalités attendues d'un SIA :

- « *discipline qui traite des systèmes informatiques capables d'exécuter des fonctions généralement associées à l'intelligence humaine, telles que le raisonnement, l'apprentissage et l'auto-amélioration* » ;
- « *domaine interdisciplinaire, communément considéré comme branche de l'informatique, consacré au développement de modèles et de systèmes capables*

[80] ISO/IEC TR 24028:2020, publiée en mai 2020.

[81] ISO/IEC TR 24027:2021, publiée en novembre 2021.

[82] ISO/IEC 2382:2015.

d'exécuter des fonctions généralement associées à l'intelligence humaine, telles que le raisonnement et l'apprentissage » ;

- « *capacité d'une unité fonctionnelle à exécuter des fonctions généralement associées à l'intelligence humaine, telles que le raisonnement et l'apprentissage* ».

D'autres normes ISO sont en cours de discussion, dont une dédiée à la terminologie de l'IA, ce qui atteste du caractère sensible de l'harmonisation conceptuelle en la matière.

Il convient d'ailleurs de noter que le puissant organisme américain de normalisation, le *National Institute of Standards et Technologies* (NIST) a engagé des travaux sur l'élaboration d'un « cadre de management des risques de l'IA », dont il est prévu qu'ils aboutissent à l'hiver 2022/2023[83].

Au niveau français, il semble qu'aucune norme dédiée à l'IA n'ait été publiée par l'AFNOR, qui, à raison, s'investit en priorité dans les instances internationales de normalisation. Ainsi, l'AFNOR participe activement au comité technique CEN-CLC/JTC 21 *Artificial Intelligence* créé au printemps 2021. Il a vocation à traiter des questions éthiques, de sécurité et de respect de la vie privée, de la qualité et de l'ingénierie des systèmes d'IA, des données pour les systèmes d'IA ou encore de la responsabilité.

L'implication des acteurs français, publics comme privés, dans ces travaux de normalisation doit en effet constituer une priorité stratégique en raison du caractère structurant de ces normes dans la compétitivité de notre industrie comme dans la fluidité des échanges internationaux, par-delà les fortes divergences culturelles et de valeurs. Si notre pays apparaît bien positionné dans la gouvernance des instances de normalisation, la présence des industriels et des administrations dans les réunions de travail reste souvent trop modeste en comparaison d'autres pays comme l'Allemagne, sans même évoquer la place hypertrophiée des grandes compagnies internationales du numérique. La France a la culture de la législation, non de la normalisation. Or il serait regrettable que, dans un renversement de paradigme normatif, nos administrations soient tenues d'observer des règles qui auront été conçues par, voire pour, le secteur privé.

[83] Selon la présentation qui en est faite sur le site du NIST, consulté en mars 2022, ce cadre devrait définir « *les caractéristiques de la fiabilité, notamment l'exactitude, l'explicabilité et l'interprétabilité, la fiabilité, la vie privée, la robustesse, la sûreté, la sécurité (résilience) et l'atténuation des biais involontaires et/ou nuisibles, ainsi que des utilisations nuisibles* » et prendre en compte des « *principes tels que la transparence, l'équité et la responsabilité lors de la conception, du déploiement, de l'utilisation et de l'évaluation des technologies et des systèmes d'IA* ».

2. Accélérer le déploiement des systèmes d'IA publics pour en exploiter pleinement le potentiel

Comme la plupart des autres pays de l'OCDE[84], la France n'a pas formalisé de stratégie nationale propre aux systèmes d'IA (SIA) de l'administration. Le déploiement de ces systèmes au sein des collectivités publiques s'inscrit dans le cadre plus général et transversal de la **Stratégie nationale pour l'intelligence artificielle** (SNIA), baptisée « *AI for humanity – L'intelligence artificielle au service de l'humain* », dont la France s'est dotée en 2018, à la suite du rapport de la mission présidée par le député Cédric Villani.

Au-delà des quatre secteurs prioritaires pour l'investissement dans les systèmes d'intelligence artificielle identifiés dans la SNIA (santé, transport, environnement et défense/sécurité), qui concernent aussi bien la sphère publique que la sphère privée, l'une des ambitions affichées est, de manière plus transverse, de promouvoir la transformation publique en accompagnant les administrations dans l'expérimentation, le développement et le déploiement de l'IA pour améliorer l'efficacité de leur action, en mettant en réseau les experts de la donnée de l'État et en développant un pôle de compétences mutualisé.

Cette stratégie repose ainsi sur la création d'un écosystème favorable aux initiatives publiques « locales », et non sur une planification du déploiement ni même, au niveau national et interministériel, sur la définition de priorités générales de conception, assises sur une analyse des principaux gisements de valeur publique des systèmes. Ce choix pragmatique, cohérent avec l'idée que l'innovation requiert de laisser aux acteurs de terrain une certaine liberté d'imagination et une grande capacité d'adaptation, explique aussi **l'absence de démarche structurée de recensement des cas d'usage existants** et **la faible visibilité, à l'échelon national, sur les cas d'usage à venir**.

[84] Selon l'OCDE, l'Italie, la Finlande et l'Uruguay se distingueraient par la mise en place d'une stratégie dédiée au secteur public (étude de l'OCDE, *Hello world, Artificial intelligence and its use in the public sector*, 2019).

Les services d'administration centrale en charge de la mise en œuvre de la SNIA[85] (V. la description des acteurs en *Annexe 8* et le schéma au point 4.2.1) ne disposent pas d'une vision d'ensemble formalisée des cas d'usage dans la sphère publique, comme a entrepris de le faire le Centre commun de recherche (JRC) de la Commission européenne à travers l'initiative *AI Watch*[86] à l'échelle de l'Union européenne[87]. Ils ont essentiellement connaissance des projets bénéficiant d'une subvention ou d'un dispositif d'appui, de ceux qui ressortent des échanges, notamment informels, qu'ils organisent ou auxquels ils participent, et de ceux qui ont été rendus publics par les porteurs de projets. Au niveau des ministères, certaines **feuilles de route** des administrateurs ministériels des données, des algorithmes et des codes sources (AMDAC) évoquent quelques actions portant sur des SIA, mais n'en fournissent que rarement une vue globale, laquelle n'est en général pas davantage disponible sur les sites institutionnels. La connaissance qu'ont les ministères des SIA développés par les opérateurs relevant de leur tutelle est également très inégale. Enfin, en dépit des efforts de leurs associations, les collectivités territoriales manquent à l'évidence d'informations sur les initiatives prises par leurs homologues sur le territoire national, que l'Agence nationale de la cohésion des territoires (ANCT) n'est pas, en l'état, en capacité de leur fournir.

Il est certes matériellement impossible de recenser de manière exhaustive l'ensemble des cas d'usage actuels de l'IA dans l'action publique nationale. Un tel travail encyclopédique serait nécessairement incomplet et obsolète dès sa publication et appellerait un lourd travail d'actualisation. En retenant une définition large des SIA comme celle qu'a proposée la Commission européenne, il ne présenterait du reste qu'un intérêt limité puisqu'il conduirait peu ou prou à répertorier la totalité des systèmes d'information publics. En revanche, la réalisation d'une **cartographie publique des cas d'usage innovants**, en particulier ceux qui reposent sur l'apprentissage automatique ou qui combinent les différentes approches de l'IA, répondrait à un besoin exprimé par nombre d'administrations désireuses de mieux comprendre l'intérêt et l'utilité pratique de ces systèmes, favoriserait l'émergence de projets nouveaux, la mutualisation et la duplication tout en prévenant les initiatives en doublon, et fournirait aux citoyens une vue d'ensemble des utilisations de l'IA et des informations générales sur les systèmes qui les affectent.

---

[85] Coordonnateur national de la stratégie IA, direction interministérielle du numérique (DINUM), direction interministérielle de la transformation publique et mission « Numérique » de l'Agence nationale de la cohésion des territoires (ANCT) pour les collectivités territoriales.

[86] Cette entité a conçu et mis en ligne un tableau recensant, à date, environ 140 cas d'usage à travers l'Union européenne, sommairement décrits, avec un renvoi vers le site du porteur de projet ou d'autres contenus explicatifs. En outre, l'étude de référence de 2020 sur l'IA dans le secteur public, pilotée par la Commission, examine 230 cas d'usage dans les différents domaines de l'action publique.

[87] Le projet de règlement européen prévoit en outre que les SIA à haut risque de l'annexe III, qualifiés d'« autonomes » (c'est-à-dire autres que ceux qui sont utilisés comme composants de sécurité de produits soumis à une législation d'harmonisation ou qui constituent de tels produits), lorsqu'ils sont « autonomes » (c'est-à-dire qu'ils ne sont pas incorporés dans un produit) sont répertoriés dans une base de données tenue par la Commission et mise à la disposition du public.

A défaut d'être en capacité de livrer une cartographie complète de cette nature, le Conseil d'État propose, en réponse à la lettre de mission, un **panorama illustratif des usages**, réalisé à l'aide des travaux du groupe de travail, des recherches documentaires, des auditions réalisées et des contributions reçues. Cet échantillon, présenté dans l'annexe 9 et synthétisé ci-dessous, montre que l'ensemble des domaines de l'action publique sont concernés, l'ensemble des catégories d'acteurs impliquées et l'ensemble des briques logicielles relevant de l'IA, telles qu'elles ont été présentées dans la première partie, mobilisées. Il révèle aussi le remarquable potentiel des SIA dans la sphère publique, lequel reste encore insuffisamment exploité. Ces constats justifient que les pouvoirs publics s'engagent dans une démarche résolument volontariste, qui exige dans le même temps de faire preuve de la plus grande vigilance dans les choix stratégiques (quelle ambition assigner aux SIA ? quels cas d'usage privilégier ? quel rythme de déploiement ?) et de ne jamais perdre de vue l'enjeu crucial, développé dans la troisième partie dont la lecture est indissociable de celle-ci, de la construction d'une IA publique de confiance, répondant aux attentes et aux intérêts des humains qu'elle sert.

2.1. Le recensement des cas d'usage de l'IA publique révèle le remarquable potentiel des SIA

2.1.1. Panorama général des cas d'usage de l'IA publique

Les travaux consacrés à l'IA, notamment dans l'action publique, proposent différentes typologies d'usages qui se heurtent en général à la même difficulté de cohérence, en ce qu'elles juxtaposent, au lieu de les imbriquer, quatre niveaux d'analyse distincts :

- **l'approche** d'IA utilisée (ex. : système-expert, apprentissage profond…) ;
- la nature de la **brique fonctionnelle** que cette approche permet de mettre en œuvre, au sens du point 1.1.3 (ex. : vision par ordinateur, traitement du son, traitement du langage naturel, automatisation de process…) ;
- la **fonctionnalité** que cette brique fonctionnelle permet de servir (ex. : identification de personnes à l'aide de la vision par ordinateur, agent conversationnel et traduction linguistique grâce au traitement du langage naturel, « analyse prédictive » par l'utilisation d'un modèle de régression linéaire…) ;
- et, enfin, le **cas d'usage** proprement dit, c'est-à-dire la mise en œuvre de cette fonctionnalité (ou de plusieurs fonctionnalités combinées) par une entité ou une catégorie d'entités déterminées, dans un produit ou un service ayant une destination précise (ex. : système d'identification biométrique mis en œuvre par la police à des fins de recherche d'auteurs d'infractions, robot conversationnel déployé par un organisme de sécurité sociale pour informer les allocataires de leurs droits, traducteur automatique du récit d'un demandeur d'asile devant un officier de protection, maintenance prédictive du réseau routier d'une collectivité territoriale…)[88].

[88] A titre d'exemple, l'étude de référence de la Commission européenne (en cours d'actualisation) sur l'IA dans le secteur public propose une typologie en dix catégories, dont

Loin d'être académique, cette distinction est susceptible d'avoir des incidences juridiques : selon le projet de règlement européen résultant des travaux de la présidence slovène, seraient exclus de son champ les **SIA à « usage général »** (*general purpose AI systems*), notion qui semble recouvrir la brique fonctionnelle et la fonctionnalité qu'elle sert. Un modèle générique de reconnaissance faciale, susceptible d'être utilisé dans les contextes les plus variés, relèverait de cette exclusion. C'est seulement au stade du cas d'usage, lorsque la destination du système est clairement définie et située, que les obligations du règlement s'appliqueraient. Au regard de l'enjeu d'encadrement des systèmes en fonction des risques qu'ils présentent, cette construction apparaît tout à fait justifiée : les incidences sur les libertés publiques d'un système d'identification biométrique des personnes par les forces de l'ordre dans l'espace public sont sans commune mesure avec les systèmes d'authentification à l'entrée d'un local réservé au personnel ou, comme le dispositif ALICEM, d'une application numérique en vue d'y accomplir des démarches administratives, alors que tous reposent sur la vision par ordinateur et, plus précisément, la reconnaissance faciale.

L'annexe 9, à laquelle le lecteur est invité à se reporter pour mieux appréhender les utilisations actuelles des SIA dans l'action publique, recense, par grand secteur d'activité, les principales initiatives connues en France, voire à l'étranger, mettant en œuvre de tels systèmes, fondés sur l'apprentissage automatique ou l'approche statistique. Les développements qui suivent présentent de façon synthétique les principales familles de cas d'usage qui s'en dégagent.

1. L'automatisation des tâches répétitives ou fastidieuses

De longue date, des systèmes basés sur les règles ont été déployés pour substituer l'intervention d'un logiciel à un traitement humain de tâches administratives simples et routinières. Beaucoup d'entre eux ne sont d'ailleurs plus considérés comme relevant de l'IA.

Dans la sphère publique, ces systèmes sont largement utilisés dans le domaine social, s'agissant de la liquidation automatique des prestations ou d'un segment seulement du processus de liquidation (ainsi du renseignement de certaines données nécessaires à cette liquidation *via* un robot transcrivant ces informations d'un système d'information à l'autre, en lieu et place d'un agent humain). Dans le domaine éducatif, l'appariement automatisé de l'offre et de la demande pour

« *Apprentissage automatique et apprentissage profond* », « *Systèmes-experts et systèmes basés sur les règles* », « *Vision par ordinateur et identification* », « *Traitement du son* », « *Traitement du langage naturel, Fouille de textes et analyse de discours* », « *Robots conversationnels, Assistants digitaux intelligents et systèmes de recommandations* » et « *Robotique cognitive, Automatisation de process et Véhicules connectés et autonomes* », alors que l'apprentissage profond (réseaux de neurones) est un ensemble de techniques utilisées pour la vision par ordinateur, le traitement du son et le traitement du langage naturel, que la vision par ordinateur est mobilisée dans les véhicules autonomes, de même que les robots conversationnels reposent à la fois sur des systèmes basés sur les règles et sur le traitement du langage naturel, ou encore que l'analyse de discours recourt à la brique fonctionnelle « traitement du son ».

l'admission en première année d'une formation de l'enseignement supérieur, avec le service Admission Post-Bac, devenu Parcoursup, relève également de cette logique (sans recours à un algorithme d'apprentissage automatique).

L'apprentissage automatique ouvre de nouvelles perspectives d'utilisation en permettant l'automatisation de tâches répétitives légèrement plus complexes, en ce qu'elles requièrent une analyse – d'image, de son et surtout de texte – et une opération de qualification, et non pas seulement l'exécution d'une action déclenchée par un constat simple et purement factuel (comme la question de savoir si les ressources du demandeur d'une prestation sont supérieures, inférieures ou égales à un plafond). Tel est le cas des SIA de tri et d'orientation des courriers (à la Caisse des dépôts et consignations) ou des courriels (des systèmes classiques de détection des messages indésirables à des outils plus sophistiqués qui détectent le niveau de priorité des courriels voire qui suggèrent une réponse, comme chez Pôle Emploi). Dans le champ de la justice, où les SIA publics sont encore rares, les outils de **pseudonymisation des décisions juridictionnelles**, développés à partir d'algorithmes de traitement du langage naturel, visent à automatiser la suppression des éléments permettant l'identification des personnes mentionnées dans les décisions et, ainsi, à faciliter leur mise à disposition du public, prévue par l'article 24 de la loi pour une République numérique en 2016, puis précisé par l'article 33 de la loi du 23 mars 2019 de programmation 2018-2022 et de réforme de la justice[89]. Etalab a développé à cette fin un outil pour le Conseil d'Etat. La Cour de cassation a conçu un outil propre décrit dans l'annexe 9 (fiche n° 4).

Le constat factuel du potentiel de réalisation des tâches répétitives ou massives est évidemment dressé sans préjudice de la nécessaire réflexion sur le choix des règles (dont Parcoursup a donné une illustration, toujours controversée), sur le rôle, à redéfinir, des agents auxquels ces tâches sont épargnées, et sur la maîtrise de processus complexes dont des pans entiers sont confiés à des SIA, autant de points critiques abordés en troisième partie.*2. L'automatisation de la relation à l'usager ou à l'agent public*

Une deuxième famille de cas d'usages, connexe à la précédente, concerne l'objectif d'**amélioration de l'information et de la réponse aux usagers du service public, essentiellement fondée sur le traitement automatique du langage**.

A l'instar des entreprises dans le cadre de leur relation clients, plusieurs administrations ont développé des robots conversationnels (*chatbots*) ou des agents vocaux (*voicebots*) qui constituent un guichet d'information supplémentaire, accessible tous les jours et à toute heure, depuis chez soi et sans temps d'attente.

Ce mouvement est particulièrement développé au sein des institutions, pour lesquelles la relation au contribuable (pour la DGFIP), à l'allocataire ou au cotisant (pour les caisses de sécurité sociale : CNAV, CNAF et URSSAF Caisse nationale) est un

[89] Et dont la mise en œuvre a été définie par le décret du 29 juin 2020 relatif à la mise à disposition du public des décisions des juridictions judiciaires et administratives.

enjeu majeur de la qualité de service attendue[90]. Pour la plupart de ces outils, il s'agit, pour la partie technique, d'un produit « sur étagère », qui est ensuite adapté à la terminologie propre à l'administration utilisatrice et nourri, s'agissant du fond des réponses, d'une base de connaissance métier qui lui est propre, et, le cas échéant, enrichi au fur et à mesure des interactions avec les usagers.

Ces outils peuvent également être mis au service de la **politique de ressources humaines de l'administration**, notamment pour automatiser le conseil aux agents publics, sur leurs droits et leurs obligations, qu'il s'agisse de questions relatives à leur carrière, aux avantages sociaux ou aux formations qui leur sont accessibles. Si l'on voit poindre, à l'étranger, des initiatives originales visant à associer un robot aux procédures de recrutement dans la fonction publique (V. par exemple le projet *Tengai Unbiased* en Suède, évoqué *infra* en 3.1.3.), les outils déployés ou en développement en France demeurent rares : on peut citer le projet du rectorat de Versailles, financé par la Banque publique d'investissement, de développer un SIA de personnalisation des offres de formation et les projets de la gendarmerie nationale et des douanes.

Le volet RH du programme « Valorisation des données » de la direction générale des douanes et des droits indirects

Les douanes ont engagé en 2019 une démarche de développement de nouvelles fonctionnalités numériques basées sur une meilleure valorisation de la donnée, avec le soutien financier du fonds de transformation de l'action publique (FTAP). Outre les cas d'usage relatifs aux contrôles douaniers sur le terrain (V. ci-dessous et l'annexe 9, fiche n° 3), un simulateur de mutation pour les agents a été mis en service afin de les éclairer sur leurs chances d'obtenir telle ou telle affectation eu égard à leur profil, sur la base d'une analyse automatique des déroulés de carrière passés, et un « *chatbot* RH » est en cours de déploiement pour répondre aux questions de premier niveau et décharger les agents des services des ressources humaines.

Certaines administrations ont, en outre, recours à l'intelligence artificielle pour **faciliter l'accès du public aux sources d'information**. Le code du travail numérique représente à cet égard un projet phare. Ses objectifs ont été définis par l'article 1er de l'ordonnance n° 2017-1387 du 22 septembre 2017 relative à la prévisibilité et la sécurisation des relations de travail, qui dispose qu'il vise « *en réponse à une*

[90] On recense également, à l'échelle européenne, de très nombreux cas d'usage des robots conversationnels : Mona en Autriche (questions sur la crise sanitaire) ; Projet VDAB en Belgique (demandeurs d'emploi) ; Concierge urbain dans la ville de Plovdic en Bulgarie (informations touristiques) ; Kiri à Frederiksberg au Danemark (services locaux aux citoyens) ; Iti en Estonie (informations sur les statistiques du pays) ; Kamu en Finlande (informations sur l'entrée et le séjour des étrangers) ; Voicebot irlandais et Toms en Lettonie (informations sur la situation fiscale)… L'un des robots conversationnels les plus développés est Kommune-Kari, en Norvège, disponible pour les habitants de 80 communes représentant plus de 30% de la population, dont la base de connaissances est l'une des plus larges au monde dans la sphère publique (6000 entrées), et qu'il est envisagé de déployer en Suède, en Finlande et au Danemark.

demande d'un employeur ou d'un salarié sur sa situation juridique, l'accès aux dispositions législatives et réglementaires ainsi qu'aux stipulations conventionnelles, en particulier de branche, d'entreprise et d'établissement, sous réserve de leur publication, qui lui sont applicables ». Mérite également d'être signalé le projet, actuellement en développement, de traduction automatique en langue étrangère et de vocalisation des fiches techniques du site *service-public.fr*, piloté par la DILA. La consolidation du droit sur le site *legifrance.gouv.fr* constitue également un domaine d'expérimentation[91]. Enfin, Pôle emploi a développé son propre moteur de recherche, dont la performance très satisfaisante a permis de réduire le nombre de sollicitations par courriel des conseillers[92].

3. Les systèmes d'aide à la décision publique

De nombreuses collectivités publiques ont recours aux SIA pour tirer des données dont elles disposent des informations, des connaissances et des analyses exploitées par les agents pour prendre les meilleures décisions. Cette famille de cas d'usages recouvre des réalités extrêmement diverses, de la caractérisation d'une anomalie sur une image à la compréhension, l'anticipation (avec l'analyse dite « prédictive ») et la gestion de phénomènes économiques, sociaux, environnementaux ou culturels complexes.

a/ La **gestion des territoires et des réseaux** (V. la fiche n° 1 en Annexe 9) constitue à cet égard un champ de déploiement extrêmement intéressant et prometteur, qu'il s'agisse de transport et de mobilité (identification des dégradations de la voirie en vue d'orienter les choix d'entretien, aide à la gestion des flux routiers et optimisation des plans de circulation, anticipation des accidents de la route afin de pré-positionner des moyens de secours…), de collecte des déchets (conteneurs « intelligents » qui détectent les types de déchets déposés ou le taux de remplissage des bennes et permettent d'adapter les tournées en conséquence), de gestion de l'eau (anticipation et prévention des fuites ou de la dégradation de la qualité de l'eau, compteurs d'eau « intelligents »), d'entretien des espaces verts (arrosage des parcs en fonction de la pluviométrie et des prévisions météorologiques) ou de prévention des risques naturels (simulation de l'évolution du niveau des nappes phréatiques en tension par la région Occitanie, par exemple) et des sinistres.

[91] Les SIA ouvrent également la voie à une forme de « codification numérique » permettant, grâce à l'analyse sémantique, de réunir l'ensemble des dispositions pertinentes sur un même thème, indépendamment des codes créés par la loi ou le règlement, et à la qualification juridique de la donnée (quelles règles juridiques s'appliquent à telle ou telle donnée ou situation de fait ?).

[92] Ce type de cas d'usage est aussi illustré par le réseau de neurones utilisé par l'agence du numérique de l'État britannique (*Government Digital Service*) pour assigner automatiquement à chaque page du site de référence *GOV.UK* (qui s'apparente à *servicepublic.fr*) des mots-clés en fonction du contenu (96 % des 100 000 pages ont été traitées de cette façon, en six mois).

PredictOps, un SIA d'aide à la décision des secours

Le service départemental d'incendie et de secours du Doubs a développé à partir de 2018, en partenariat avec un laboratoire de recherche (Femto-ST, IUT de Belfort-Montbéliard), et mis en service en février 2021, un outil baptisé PrédictOps, qui utilise 1200 variables (données météorologiques, éphéméride, données de trafic routier, qualité de l'air, niveau des cours d'eau, données épidémiologiques…) et un modèle entraîné sur les données issues des 40 000 interventions effectuées par les sapeurs-pompiers de ce département chaque année, afin de prévoir aussi bien les périodes de crue que les pics d'accidents de la circulation ou de suicides, ou la survenance d'arrêts cardiaques et, en conséquence, de déterminer le niveau de sollicitation du SDIS à une échéance comprise entre une heure et 72 heures, selon le type de sinistre. Les moyens humains et matériels peuvent être adaptés en conséquence, en mobilisant les personnes sous astreinte, les pompiers volontaires, voire les équipes du département le plus proche. Les résultats obtenus en milieu urbain semblent probants.

On peut citer ici l'exploitation à grande échelle des données aériennes, tant par les armées que par l'Institut national de l'information géographique et forestière, ou encore l'ambitieux projet « Gwadastreet » de la région Guadeloupe qui, par la mise en place d'un système de photographies automatiques à 360° de l'ensemble des routes de Guadeloupe[93], vise à améliorer la perception de la redevance d'occupation du domaine public et la gestion de la voirie en planifiant les travaux d'élagage.

Les SIA présentent aussi l'intérêt de simuler les conséquences des décisions publiques.

Certaines localités ont entrepris de se doter de « jumeaux numériques », qui constituent des représentations visuelles et statistiques très avancées de la réalité urbaine (ville virtuelle), intégrant en temps réel un vaste ensemble de paramètres (les constructions, le trafic routier, les flux de populations, la qualité de l'air, les nuisances sonores, la biodiversité…) et qui offrent la possibilité de simuler les multiples incidences directes et indirectes, à court ou moyen terme, qu'aurait telle ou telle décision publique, afin d'enrichir l'éclairage des choix à opérer.

La modélisation économique au sein de l'Autorité de régulation des transports

L'Autorité de régulation des transports (ART) exerce, entre autres, des missions de suivi économique des marchés de transports qu'elle régule, qui nécessitent des traitements statistiques complexes, que des SIA viennent faciliter. Le premier de ces systèmes, fondé sur une combinaison des apprentissages supervisé et non-supervisé, permet de réaliser une segmentation du marché du transport ferroviaire à grande vitesse de voyageurs en France, afin d'analyser les comportements de la demande finale et de mesurer les élasticités-prix associées

[93] L'application *Google Street View*, qui se concentre essentiellement sur les sites les plus touristiques de l'île, ne donne pas accès à l'ensemble du territoire guadeloupéen.

et, à terme, de contre-expertiser la tarification de l'accès à l'infrastructure. Le second est un modèle de prévision de la fiabilité des circulations de train, ayant recours à un apprentissage supervisé (identification de facteurs de segmentation de la ponctualité des circulations ferroviaires).

b/ Les SIA d'aide à la décision sont très largement déployés dans le domaine de la **santé**, avec le développement des **outils d'aide au diagnostic** en imagerie médicale (détection de cancers, analyse de radiographies) et d'aide à la prescription et à la décision thérapeutique sur la base de dossiers médicaux résolus (complémentaires des outils d'automatisation du contrôle des ordonnances dans le cadre de la pharmaco-vigilance). Sur un plan plus organisationnel, les SIA peuvent aider au pilotage des services, à l'image de l'outil en déploiement au centre hospitalier de Valenciennes, qui a pour objectif d'anticiper le nombre de patients qui se rendront aux urgences à un jour, cinq jours et dix jours, en fonction des flux passés, des actualités ou encore des prévisions météorologiques, avec pour objectif, à terme, d'ajuster le dimensionnement et la nature du dispositif de prise en charge aux besoins.

c/ Ces outils sont aussi utilisés dans le domaine éducatif, où des **SIA d'aide à l'apprentissage des savoirs fondamentaux** (français et mathématiques), en phase de déploiement, assistent les professeurs des écoles en proposant des exercices personnalisés adaptés au niveau et à la progression des élèves. Le ministère chargé de l'Éducation nationale a développé six de ces SIA, dans le cadre du Partenariat d'innovation en intelligence artificielle (P2IA), en ayant recours à un marché public innovant. D'autres SIA en déploiement peuvent être mentionnés, notamment des assistants vocaux pour l'apprentissage des langues ou, dans le cadre du dispositif « Devoirs Faits », des robots conversationnels pour le soutien aux collégiens scolarisés au CNED.

Les SIA éducatifs

La solution « MathIA » propose des exercices personnalisés aux élèves, et adapte le niveau des questions aux résultats de chacun. Le professeur peut suivre sur un tableau de bord les activités de ses élèves et de leur progression, constituer des groupes de niveaux et choisir des parcours « clés en main » proposés par l'outil.

Kaligo combine lui deux principales applications : la lecture d'une part, l'outil analysant si la lecture de l'enfant est correcte avant de le corriger s'il y a lieu, et la dictée d'autre part, grâce à l'analyse de la graphie et la correction des fautes. Un tableau de bord permet à l'enseignant de suivre le niveau de ses élèves et, éventuellement, de mettre en place des stratégies de remédiation adaptées.

d/ Dans le champ **juridictionnel**, deux projets d'outils d'aide à la décision, encore au stade de la preuve de concept, méritent d'être signalés : la détection des séries contentieuses, projet conduit par le Conseil d'Etat, vise à permettre de coordonner voire d'unifier le traitement contentieux de flux de requêtes posant des questions similaires et appelant des réponses analogues, par exemple en désignant rapidement un tribunal administratif pilote ; le projet de détection des divergences d'interprétation

de la loi a pour objet d'éclairer la Cour de cassation sur l'existence de jurisprudences contradictoires, pouvant appeler une prise de position de la juridiction suprême de l'ordre judiciaire ou une clarification des textes.

e/ Il existe de nombreux autres cas d'usage des SIA pour l'aide à la décision, dans différents domaines. Il peut s'agir, par exemple, de tirer une information précise d'une masse importante de données[94] afin d'évaluer une politique publique.

La mesure de la représentation des femmes dans les médias

L'ex-Conseil supérieur de l'audiovisuel (CSA), aujourd'hui l'ARCOM, exploite depuis 2019 un outil développé par l'Institut national de l'audiovisuel (INA) permettant, grâce à une algorithme d'apprentissage supervisé, l'identification du genre du locuteur. Il lui permet de mesurer la présence des femmes dans les médias audiovisuels dans la perspective de l'établissement des rapports qu'il publie sur le sujet, dans le cadre de sa mission de veiller à une juste représentation des femmes et des hommes dans les programmes des services de communication audiovisuelle (article 3-1 de la loi du 30 septembre 1986 relative à la liberté de communication).

A une échelle plus locale, les SIA peuvent aider à anticiper sur des besoins des entreprises et à orienter les demandeurs d'emploi en conséquence. Telle est l'ambition de « *La bonne boîte* », développé par Pôle Emploi, dont le modèle algorithmique détecte, sur un secteur géographique donné, les entreprises qui sont susceptibles d'embaucher à brève échéance, afin d'inciter les demandeurs d'emploi à présenter des candidatures spontanées.

L'usage des systèmes dits prédictifs dans le domaine pénal, en particulier pour l'évaluation du risque de récidive ou d'évasion d'une personne ou celle du niveau de criminalité ou de délinquance d'un secteur géographique (afin d'orienter les contrôles), est à la fois plus connu et plus controversé. De fait, ces systèmes n'occupent pas, à ce stade, une place centrale dans la boîte à outils des services de police et de gendarmerie.

4. Les activités de contrôle, terrain de déploiement privilégié des SIA

A la confluence de plusieurs familles de cas d'usage, les activités de contrôle méritent assurément une mention particulière dans cette présentation générale. Les SIA occupent en effet, depuis plusieurs années et avec une dynamique remarquable, une place importante dans les **politiques de lutte contre les infractions et les fraudes** (V.

[94] On peut citer, à titre d'exemple, le modèle GRID3 de « recensement à distance » réalisé dans plusieurs pays en développement par le département du développement international britannique sur la base de l'analyse automatisée (vision par ordinateur) des images satellite, permettant d'évaluer la densité de population à partir des sites de peuplement, des bâtiments, des réseaux de transport, de l'éclairage public, ces données étant croisées avec celles de recensements locaux partiels. Il est envisagé d'utiliser les données de téléphonie mobile à cette fin. L'objectif est de suivre l'évolution démographique d'un secteur lorsqu'il n'est pas possible de conduire un recensement en raison de conflits ou d'une situation d'insécurité.

annexe 9, fiche n° 3)[95]. Or, ces systèmes soulèvent des questionnements particuliers en termes de maîtrise, d'équité et de non-discrimination et, au-delà, d'acceptabilité par les citoyens.

Les SIA sont, d'abord, largement utilisés dans la **définition des stratégies de contrôle**, sur la base d'un profilage des comportements délinquants et d'une cotation du risque de fraude. Les usages les plus avancés concernent la sphère économique et sont illustrés par les initiatives de Tracfin, des douanes et de l'administration fiscale.

Le ciblage du contrôle fiscal par la Direction générale des finances publiques

La DGFIP a très tôt (dès 2014) identifié l'IA comme complémentaire des agents pour identifier des facteurs de risques : à ce jour, 30 des 180 facteurs ont été identifiés par un SIA à partir des contrôles passés et de données externes. Demeure toutefois à la main des agents la définition des stratégies locales de contrôle, qui s'appuient sur les dossiers hiérarchisés en fonction de ces facteurs. L'utilisation de l'IA est aujourd'hui totalement intégrée à la stratégie de la DGFIP, dont le taux de contrôles programmés à l'aide de l'IA est devenu un indicateur de performance. Cette dynamique se poursuit avec le lancement du projet « Foncier innovant » qui vise l'identification des bâtiments ou des aménagements (ex. piscines) non déclarés.

Le ciblage du contrôle des fraudes sociales est également un cas d'usage assez développé (V. en particulier les initiatives de la CNAF, annexe 9, fiche n° 7).

Un projet pilote, s'appuyant sur un algorithme d'exploration et de traitement des commentaires des consommateurs sur Internet (Trip Advisor, Google…), est développé par le ministère chargé de l'Agriculture, afin d'identifier les établissements à risque sanitaire potentiel et d'appuyer les services vétérinaires dans la programmation de leurs contrôles.

Outils d'orientation et de planification des contrôles, les SIA sont aussi mobilisés directement pour **détecter les menaces et les infractions elles-mêmes**.

A ce titre, ils peuvent être déployés **en ligne** afin d'identifier des comportements suspects par l'exploitation massive de flux de données échangés publiquement : détection de cyberattaques sur des systèmes d'information publics (Agence nationale de sécurité des systèmes d'information), de menaces terroristes (technique dite de l'algorithme des services de renseignement), de diffusion de fausses informations par ou pour le compte de puissances étrangères (Viginum), de certaines infractions fiscales révélées par des contenus publiés sur les réseaux sociaux (outil de la DGFIP autorisé à titre expérimental par l'article 154 de la loi n° 2019-1479 du 28 décembre 2019 de finances pour 2020), de fonctionnement biaisé de plateformes numériques comme des comparateurs d'offres (outils

[95] Ce constat n'est pas propre à la France. V. par ex. aux Pays-Bas : A. van Veenstra, F. Grommé et S. Djafari, *The use of public sector data analytics in the Netherlands*, octobre 2020, Emerald Publishing Limited.

développés par le pôle d'expertise de la régulation économique – PEReN), d'offres d'emploi illicites (Pôle Emploi), d'offres de formation frauduleuses dans le cadre du compte personnel formation (Caisse des dépôts et consignations), ou encore de contenus pédopornographiques (Land de Rhénanie du Nord Westphalie). Dans la sphère culturelle, les SIA sont aussi susceptibles d'être utilisés, en matière de protection des droits d'auteur, pour la détection d'œuvres de contrefaçons[96].

Les SIA peuvent aussi servir à la détection d'infractions **dans le monde réel**. La vision par ordinateur est ainsi couramment utilisée en matière de contrôle des infractions routières et du paiement du stationnement payant, avec la lecture automatisée des plaques d'immatriculation qui permet, grâce à la reconnaissance optique de caractères, d'extraire leur numéro à partir d'une image captée par un radar automatique ou un véhicule équipé, avant d'identifier dans les bases de données le titulaire du certificat d'immatriculation et de procéder à l'envoi d'un procès-verbal – cette dernière étape n'étant pas intégralement automatisée s'agissant des contraventions routières, dont la constatation fait intervenir des officiers de police judiciaire. L'usage le plus controversé dans ce domaine est celui de la prévention des troubles à l'ordre public et l'identification de suspects dans l'espace public par la reconnaissance biométrique. On peut également citer, dans le domaine du sport, la mise en œuvre d'un « algorithme de statistique prédictive » par l'unité de gestion du passeport d'un athlète, qui consiste à détecter des variations suspectes de certains paramètres biologiques afin, le cas échéant, de déclencher un contrôle anti-dopage classique consistant à déceler directement une substance prohibée dans l'organisme ou de recourir à une expertise[97].

L'Autorité des marchés financiers (AMF) a déployé plusieurs SIA dans le but d'améliorer **la protection des investisseurs particuliers et la détection des escroqueries**. Trois outils ont été déployés : FIS.H identifie les sites Internet susceptibles de proposer des investissements frauduleux ; Spade identifie des propositions d'investisseurs *via* le classement automatique de spams (par exemple, des campagnes de démarchage par des acteurs non autorisés à le faire) ; Wetrend vise à mettre en évidence les nouvelles tendances d'arnaques aux investissements.

Le ministère de l'écologie a également mis l'accent sur les activités de contrôle, dans le domaine de la fraude aux biocarburants (projet Carbur), des installations classées pour la protection de l'environnement ou de la sûreté nucléaire (projet SIANCE déployé par l'Autorité de sûreté nucléaire, qui, *via* un outil de traitement automatique du langage, analyse les lettres d'inspection – 22 000 courriers – et permet d'identifier, *via* la consolidation des données analysées, les problèmes récurrents sur un type d'installation, et d'améliorer l'efficacité et le suivi des contrôles).

---

[96] V. sur ce sujet le rapport de la mission conjointe du Conseil supérieur de la propriété littéraire et artistique, du Centre national du cinéma et de l'image animée et de la Haute Autorité pour la diffusion des œuvres et la protection des droits sur Internet, *Vers une application effective du droit d'auteur sur les plateformes numériques de partage : état de l'art et propositions sur les outils de reconnaissance des contenus*, 29 janvier 2020.

[97] Art. R. 232-67-9 et R. 232-67-10 du code du sport.

Un dernier exemple est celui du « traitement automatisé d'analyse prédictive destiné à assister les comptables publics assignataires de l'État dans la mise en œuvre des modalités de contrôle de la dépense », géré par l'Agence pour l'informatique financière de l'État, qui détermine la liste des dépenses présentant des risques d'irrégularité au vu des résultats des contrôles effectués par le passé (données relatives aux fournisseurs, à l'organisation de la chaîne de la dépense et aux imputations budgétaires et comptables)[98].

5. La robotique matérielle appliquée à l'action publique

Les prestations de service public délivrées dans le monde physique sont susceptibles d'avoir recours à des SIA pour interagir avec leur environnement et, au premier chef, avec les personnes (dont les usagers).

Certains cas d'usage sont d'ores et déjà expérimentés ou en cours de développement. Dans le domaine de la sécurité, on peut citer aussi bien les drones que les systèmes d'arme. Le **véhicule dit « autonome »**, qu'il conviendrait plutôt de dénommer « véhicule à conduite automatisée » (comme le propose le comité national pilote d'éthique du numérique) ou, comme le fait l'ordonnance n° 2021-443 du 14 avril 2021, « véhicule à délégation de conduite », ouvre également d'intéressantes perspectives pour les activités de service public recourant à des flottes de véhicules conduits par des humains. Il y a bien longtemps que des métros circulent sans conducteur, ce qui est possible sans recourir à l'apprentissage automatique, dès lors qu'il s'agit d'un transport guidé dont le trajet et les arrêts sont parfaitement normés (en conditions normales d'utilisation). Il est nettement plus difficile de faire circuler en toute sécurité des bus ou des navettes routières, même sur un trajet prédéfini, alors que les obstacles sont multiples, les conditions de trafic très changeantes et les conditions météorologiques très variables. Des expérimentations sur le terrain sont en cours dans plusieurs villes, en France et dans le monde.

Les robots physiques peuvent être utilisés pour des tâches aussi diverses que des opérations chirurgicales, le déminage, les interventions en milieu dangereux ou périlleux (centrales nucléaires, sites industriels pollués…), le nettoyage urbain et le déneigeage, ou encore des entretiens d'embauche.

2.1.2. Les bénéfices attendus des systèmes d'IA dans le secteur public

Le propos peut sembler trivial mais il mérite d'être rappelé dans le cadre de la présente étude : les SIA déployés dans le secteur public ont d'abord et avant tout vocation à être créateurs (nets) de valeur. Les cas d'usages recensés attestent de la diversité et de l'ampleur des bénéfices qui peuvent être attendus de systèmes correctement conçus, avec pour ambition de répondre très directement à une demande du corps social : celle d'une administration de service, plus réactive, plus affûtée, plus présente là où se manifestent les besoins, et globalement moins coûteuse.

[98] Arrêté du 29 janvier 2019 portant création d'un traitement automatisé d'analyse prédictive relatif au contrôle de la dépense de l'État.

1. Franchir un cap dans l'amélioration de la qualité du service public

Les SIA peuvent être un levier majeur d'optimisation de la qualité du service public, dans ses différentes dimensions.

a/ Leur intérêt premier est sans doute d'améliorer l'adéquation et la pertinence des décisions, des actions et des prestations délivrées. Grâce à la collecte et l'exploitation massives de données, matériellement hors de portée de l'humain, et sous réserve que les précautions requises soient prises pour garantir, notamment, l'intégrité et l'objectivité de ces données, les administrations peuvent disposer d'une connaissance incomparablement plus approfondie des besoins du public, des activités et situations économiques ou sociales, de l'environnement dans lequel se déploient les politiques publiques et des conséquences potentielles des décisions qui se présentent au choix de la puissance publique. En complément de l'expertise métier et dans le prolongement de l'analyse statistique traditionnelle, l'apprentissage automatique bien conduit ouvre des perspectives inégalées pour la détection et la qualification d'évènements et de phénomènes inconnus, dont l'existence relevait d'une vague intuition humaine ou était masquée par un préjugé, ainsi que pour la compréhension de leur origine, en particulier par l'identification de leurs causes et des facteurs d'influence, et de leur dynamique, par le recours à des simulations dans un environnement contrôlé de plus en proche de la réalité (« jumeau numérique »). Elles offrent également de nouvelles opportunités pour l'évaluation des politiques publiques.

L'un des usages les plus classiques des SIA consiste à déterminer les variables explicatives d'une situation et à en pondérer l'importance, ouvrant la voie à un meilleur ciblage des interventions publiques. On peut citer, à titre d'exemple, l'outil « Signaux faibles », expérimenté en Bourgogne-Franche-Comté à compter de 2016 et étendu en 2019 au niveau national, dans le cadre d'un partenariat entre la direction générale des entreprises, la délégation générale à l'emploi et à la formation professionnelle, l'URSSAF Caisse nationale, la Banque de France et la DINUM. Ce SIA, qui recourt à un algorithme d'apprentissage automatique entraîné sur les données financières, économiques et d'activité détenues par les administrations précitées, vise à détecter plus précocement les entreprises de plus de 10 salariés en difficulté afin de déclencher au plus tôt des actions de soutien et d'accompagnement.

Des SIA performants peuvent aussi améliorer la **qualité des prestations**, affranchie de la disparité des compétences de ceux qui les délivrent. Dans certains domaines et en l'état des techniques, la **précision de l'analyse automatique** ne surclasse pas nécessairement celle des meilleurs experts (par exemple pour la détection de tumeurs cancéreuses dans les radiographies par un modèle de vision par ordinateur), mais elle dépasse sans nul doute la moyenne. Pour d'autres tâches, sa performance est inégalable par l'humain, en particulier lorsqu'elle est couplée avec des capteurs puissants de données – par exemple, l'identification univoque de personnes à plusieurs kilomètres grâce à la combinaison de caméras haute définition (par exemple, dans les « boules optroniques » aéroportées utilisées par la gendarmerie nationale) et d'un algorithme de reconnaissance faciale adossé à une large base de données.

b/ La qualité de service associée aux SIA présente en outre l'intérêt d'être à la fois **disponible en permanence et constante dans le temps**, sous réserve d'une maintenance appropriée du modèle. Une administration n'est en général accessible par téléphone ou au guichet qu'une partie de la journée, alors qu'un robot conversationnel peut intervenir à tout moment du jour et de la nuit[99]. A titre d'illustration, le *chatbot* thématique « de crise » déployé par l'URSSAF Caisse nationale dans le cadre de la crise sanitaire a répondu à un million de questions simples, ce qui aurait été impossible en y affectant des agents humains aux horaires ouvrables. L'automatisation offre ainsi de grands avantages du point de vue de la **continuité du service public**, à condition bien sûr d'être pensée avec les agents afin d'articuler convenablement la continuité des systèmes avec l'organisation des horaires de travail de ces derniers.

c/ Du fait de l'accélération du temps social et politique et de la réduction du temps d'attente jugé raisonnable par les usagers, le **temps public**, qu'il soit administratif ou juridictionnel, reste aujourd'hui un point faible de la qualité du service public auquel les SIA peuvent, dans certains cas, puissamment contribuer à remédier grâce à **l'automatisation totale ou partielle des tâches**. Il en va aussi bien de l'exécution mécanique de tâches répétitives répondant à un schéma logique aisément modélisable (par exemple, générer automatiquement des courriers à partir d'un modèle-type et de quelques données personnalisées) que de l'instruction des demandes plus ou moins complexes adressées à l'administration (demandes d'autorisation, d'allocations…) : les SIA peuvent être mobilisés afin d'analyser et d'exploiter les documents présentés (contrôle de l'authenticité, extraction des informations pertinentes…), vérifier la satisfaction de conditions purement objectives, notamment chiffrées (seuil, plafond…), classer des demandes par ordre de mérite ou de priorité, ou encore recommander une position au vu de la doctrine passée telle qu'elle se dégage de la multitude des décisions analysées…

On voit bien, par exemple, tout l'intérêt que les intéressés comme l'administration peuvent retirer de l'**automatisation de la prise de décisions « positives »**, c'est-à-dire ayant des effets favorables pour leurs destinataires, sans porter préjudice aux tiers ni une atteinte excessive à l'intérêt public, et en tenant compte, à titre de soupape, des possibilités de retrait si la décision s'avère illégale (dans un délai de quatre mois ou sans délai en cas de fraude), ainsi que des infractions de faux prévues aux articles 441-1 et suivants du code pénal, qui permettent de réprimer la « fraude au SIA » par fausse déclaration[100]. Loin des fantasmes sur la prise du contrôle du social par une intelligence artificielle, qui refuserait que, demain, l'aide personnalisée au logement soit allouée en quelques jours plutôt qu'en quelques semaines ?

La rapidité d'exécution propre à la capacité de calcul de la machine, incomparablement supérieure à celle du cerveau humain, peut non seulement faciliter, mais même rendre possible la réalisation de certaines tâches. Tel est le cas, par exemple, de la **recherche**

[99] Ce qui peut présenter un intérêt, notamment, pour les personnes travaillant en horaires décalés ou les touristes (V. le projet de conciergerie urbaine de la ville de Plovdic en Bulgarie, qui propose aux visiteurs un agent conversationnel pour optimiser leur séjour).

[100] V. en particulier l'art. 441-6 du code pénal.

d'informations dans un très vaste *corpus* documentaire non structuré : un système d'IA peut identifier l'information pertinente en quelques secondes (comme c'est déjà le cas, par exemple, d'une simple recherche par mot-clé dans un ouvrage encyclopédique) là où plusieurs personnes à temps plein devraient être mobilisées sur un longue période pour y parvenir.

d/ Les SIA peuvent également apporter un élément de réponse à l'aspiration profonde des usagers à **l'égalité de traitement**, à deux égards.

D'une part, lorsqu'ils sont dénués de préjugés (c'est-à-dire sous réserve des biais qui peuvent eux-mêmes les affecter[101]), les systèmes d'IA sont totalement indifférents aux caractéristiques non pertinentes de leur interlocuteur : son accent, sa mise, son attitude sont insusceptibles d'affecter la réponse donnée. C'est évidemment un grave défaut toutes les fois où la relation humaine doit capter, par l'empathie, les facteurs clefs d'une situation ; mais dans tous ceux, nombreux, où un SIA peut sans dommage suppléer provisoirement ou de manière limitée à la relation humaine, il s'agit d'un atout essentiel.

D'autre part, les SIA peuvent aussi contribuer, lorsque cela est nécessaire, à la convergence des positions, par-delà la fragmentation institutionnelle, accentuée par la déconcentration et la décentralisation, de la décision publique, qu'elle soit administrative ou juridictionnelle.

A l'heure actuelle, aucun mécanisme ne garantit de façon systématique et complète l'égalité devant la loi ou devant la justice à l'échelle nationale. Le principe d'égalité, qui suppose une comparabilité minimale et est ainsi tenu en échec lorsqu'un requérant demande au juge de comparer l'incomparable[102], est impuissant à corriger ces discordances. Il n'est certes pas opportun de se saisir du levier des SIA pour promouvoir une vision absolutiste et mécanique du principe d'égalité, au risque de condamner la diversité et l'émulation, qui sont de puissants stimulateurs d'idées et de progrès de l'action publique. Et si des divergences, injustifiées, peuvent résulter d'écarts de compétence, de rigueur d'analyse ou de temps consacré, d'autres peuvent être parfaitement acceptables, en l'absence même de différences de situation objectivables, lorsqu'elles résultent, par exemple, de choix politiques émanant d'autorités démocratiquement élues ou s'inscrivent dans des contextes économiques, sociaux ou culturels légitimant une différence d'approche.

Pour autant, on ne peut ignorer que ces différences sont susceptibles d'être vécues comme des injustices ou, à tout le moins, d'être plus sérieusement contestées à mesure que les citoyens, les associations et les entreprises disposeront d'outils puissants de comparaison, recourant eux-mêmes à des SIA, leur permettant de déchirer le voile d'ignorance actuel. Qu'il s'agisse de l'attribution de places de crèches, des décisions de la Cour nationale du droit d'asile ou de la doctrine des Caisses d'allocations familiales en matière de RSA, les autorités publiques, confrontées les unes aux autres, seront davantage invitées à se justifier ou à

[101] V. 3.1.3. sur les biais discriminatoires qui peuvent affecter les SIA.

[102] Le juge refuse ainsi de comparer les règles statutaires applicables aux fonctionnaires appartenant à des corps différents.

s'aligner. De ce point de vue, les SIA publics peuvent apporter une contribution décisive, en anticipant cette revendication en tant que de besoin. Toutes les fois où rien ne justifie des discordances de vue ou de doctrine, l'utilisation d'un modèle unique permet d'homogénéiser la réponse publique à l'échelon pertinent (qui n'est pas nécessairement le niveau national) ou, à tout le moins, de favoriser la convergence des positions. La conception des SIA elle-même, en ce qu'elle implique souvent d'expliciter des ressorts de la décision publique jusqu'alors cachés ou diffus[103], sert aussi le rapprochement des modèles des différentes collectivités publiques en charge d'une même mission de service public.

e/ Le recours au SIA peut aussi être un élément de **réponse à la crise de la complexité publique**. En réponse aux attentes d'une partie des usagers des services publics, d'innombrables lois se sont donné l'objectif d'un « choc de simplification » (des textes, des procédures, des normes, des critères…), avec un insuccès notoire, rappelé à plusieurs reprises par les rapports du Conseil d'État. C'est qu'on ne peut à la fois satisfaire aux exigences d'individualisation de la décision et prendre en compte les subtilités et variations infinies du monde, sans produire de la complexité normative et procédurale.

Cette complexité largement irréductible semble appeler un choix binaire : ou bien elle est supportée par l'usager, entraînant souvent privation de droits élémentaires, frustration, désocialisation, parmi les plus modestes ou les plus faibles, et contentieux ; ou bien elle est assumée et internalisée par l'administration, avec le risque que les agents publics eux-mêmes ne soient pas en capacité d'y faire face, créant des dysfonctionnements, des inégalités de traitement, des risques de tous ordres, notamment contentieux et psychosociaux.

Les SIA offrent une opportunité de dépasser ce dilemme : d'abord, en proposant au public des interfaces simples et en rejetant la complexité dans les coulisses de l'action publique, car elle ne saurait, en démocratie, peser à titre principal sur les citoyens et les entreprises ; ensuite, en assistant les agents publics dans la mise en œuvre des textes et des dispositifs (par exemple, en pré-qualifiant, en orientant et en classant les demandes, ou encore en automatisant certains points de contrôle fastidieux), voire en les dispensant de toute intervention directe, au profit d'une simple supervision du système et d'un traitement manuel des anomalies. Ainsi, c'est un système d'IA basé sur les règles, le système d'immatriculation des véhicules (SIV) qui traite les demandes de carte grise, croise les informations que l'administration détient, apprécie la conformité de la demande aux exigences textuelles, et procède à la délivrance du certificat. Seul son refus peut faire l'objet d'une décision humaine, vérifiant les obstacles que le SIA a repérés. On peut également citer les outils de liquidation

[103] La réalisation d'un SIA peut impliquer de pousser à son terme la formalisation des critères d'appréciation et de leur pondération, qui ne ressort pas nécessairement des textes ou de la pratique administrative, ou, s'agissant d'un système d'apprentissage automatique, de fournir aux usagers une explication sur les principaux déterminants des décisions prises (V. 3.1.4. sur cet enjeu d'explicabilité), qui n'est pas forcément exigée d'un humain. Cet enjeu soulève d'ailleurs des questionnements éthiques lourds (exemple : quelle priorisation des patients se présentant aux urgences ou pour la distribution de vaccins ?).

automatique des droits sociaux, chargés de réaliser les contrôles réglementaires et d'enclencher le versement de la prestation, y compris en allant chercher dans un fichier particulier les informations manquantes à une telle liquidation.

Ce sont des exemples simples d'internalisation de la complexité, dont on devine à quel point elle pourrait améliorer la perception de l'action administrative comme sa conduite par les agents, si on en recherchait la généralisation pour des processus autrement compliqués, et en veillant à prévenir les dysfonctionnements, tels que ceux que le SIV a connus à son lancement.

Il ne faut naturellement pas perdre de vue que les facilités qu'offre le numérique, dont les SIA, peuvent aussi avoir pour effet de banaliser la complexité et aiguiser la tentation du raffinement et de la sophistication, justifiée par la maîtrise supposée qu'en donnera le système. Les SIA ne doivent en aucun cas en fournir le prétexte : une complexité inévitable peut être maîtrisée par un tel système ; l'existence d'un SIA ne saurait jamais justifier une complexité complaisamment acceptée. Dans la méthodologie de recours à ces outils, esquissée au point 4.1.5 ci-dessous, la recherche de la suppression de la complexité qu'on prétend dominer avec son aide devrait être un préalable systématique. Et dans tous les cas où un SIA est retenu comme outil de compréhension ou de gestion de la complexité, il convient de garder à l'esprit que cette dernière demeure.

f/ Le dernier enjeu de performance porte sur la **prévention du risque de décrochage technique des administrations**. Le recours aux SIA dans le secteur public peut s'imposer parce que le monde que le service public sert, anime, contrôle ou régule, a déjà lui-même largement recours à ces outils. L'enjeu est double.

Vis-à-vis de l'usager, l'administration perd en **crédibilité** si elle ne peut pas, pour des services comparables à ceux que propose le secteur privé, offrir une qualité de prestation comparable ou de même gamme. Ainsi – exemple parmi mille autres – banques et assurances proposent toutes, désormais, des applications recourant à l'IA au service de leurs clients, dont le premier contact avec l'entreprise est le plus souvent un *chatbot*, qui facilite les démarches, supprime les déplacements, prépare les entretiens sans s'y substituer… Le service public, qui est un ensemble de services, doit se fixer pour objectif d'atteindre sinon de dépasser ces standards partout où cela apparaît pertinent et adéquat.

Le second enjeu est celui du **contrôle**. Sans doute une autorité de régulation ou une juridiction n'est-elle pas structurellement dans l'incapacité de s'assurer de la légalité ou de la loyauté d'un SIA au seul motif qu'elle n'en utilise pas elle-même. Mais il est certain que l'utilisation récurrente d'outils recourant à l'apprentissage machine pour ses besoins propres, voire l'association de certains agents à leur conception, lui permettent de mieux en comprendre les principes de fonctionnement. Outre le risque de méconnaissance, c'est celui de **l'asymétrie des moyens technologiques** déployés respectivement par les contrevenants et par les contrôleurs qu'il faut redouter : nombre d'infractions ne sont détectables que par le recours aux SIA, à tout le moins dans des proportions garantissant une répression crédible et dissuasive. Comme le relevait l'étude adoptée par l'assemblée générale du Conseil d'État du 15 avril 2021 sur les pouvoirs d'enquête de l'administration, « *il*

est indispensable que les contrôleurs ne se trouvent pas en infériorité technique par rapport à certains acteurs qui ont massivement investi le champ du numérique pour le développement de pratiques contraires à la loi ».

L'ensemble des avantages potentiels que peuvent présenter des SIA bien conçus, que l'on vient de résumer, concernent l'ensemble de la sphère publique. Ils énoncent d'eux-mêmes l'attrait particulier de ces outils pour les collectivités territoriales : en charge de l'essentiel des fonctions de contact avec les usagers de la plupart des grands services publics tout au long de la vie ; compétentes pour la majeure partie des usages du sol et de l'espace, des flux, de l'aménagement et du développement, il est peu de compétences qu'elles gèrent qui ne se prêteraient à l'usage de SIA.

2. Optimiser l'emploi des ressources publiques

Il n'est pas contestable que les systèmes d'IA opportunément choisis, bien conçus et efficacement déployés fournissent d'intéressants leviers d'efficience de l'action publique, en sus du surcroît de recettes qu'ils peuvent générer, tant par leur utilisation dans la lutte contre la fraude et le recouvrement des créances publiques, fiscales ou non, que, plus indirectement, par le potentiel de croissance économique qu'ils renferment. La raréfaction des ressources publiques disponibles et la difficulté structurelle des pouvoirs publics à assumer la suppression de missions ou d'activités administratives entraînent même souvent une sur-focalisation sur les économies budgétaires que ces systèmes permettraient de réaliser. Ces économies sont de deux ordres.

Le premier gisement, trop souvent oublié, réside dans l'**optimisation des moyens matériels** de l'administration et des prestations délivrées, permettant d'éviter des surcoûts inutiles, le cas échéant en épargnant à la planète l'empreinte écologique correspondante. La performance énergétique des bâtiments publics et la modulation de la consommation en fonction des besoins réels, notamment par l'exploitation des données de consommation passées, en constituent un bon exemple. De la même façon, ne déclencher l'éclairage public nocturne que lorsqu'une personne ou un véhicule s'engage dans une voie, en restant insensible à l'errance d'un animal (ce qui suppose de détecter et qualifier les formes, même sommairement) permettrait de réduire la facture d'électricité des collectivités, tout en maintenant la sécurité que l'option de l'extinction complète des lampadaires ne permet pas d'assurer.

La seconde source d'économies, qui accapare l'attention et suscite interrogations et crispations, porte sur les **ressources humaines**. De même que la singularité reste un mythe à ce jour, **le remplacement massif des agents publics par l'IA et l'avènement d'une « IAcratie »**, répondant à ses propres consignes ou même à celles de dirigeants humains, ne résiste pas à l'épreuve de l'analyse des SIA déployés aujourd'hui dans les administrations publiques, même si la concomitance d'une démarche de promotion de l'IA et d'un sentiment de démotivation nourri par la contraction de l'emploi public dans certains secteurs n'offre pas les meilleures conditions pour une réflexion apaisée et une ambition partagée en ce domaine.

Il ne saurait pour autant être occulté qu'un certain nombre d'emplois ont vocation soit à se transformer, soit à disparaître à mesure que l'automatisation se diffuse et se perfectionne. Il n'est pas contestable que les SIA constituent, aussi, un gisement

d'économies. L'ampleur des emplois concernés reste toutefois difficile à déterminer, non seulement en raison des incertitudes sur le rythme de déploiement et les capacités des systèmes d'IA, mais aussi des choix politiques qui seront faits par les pouvoirs publics.

D'abord, comme on l'a vu, de nombreux SIA ne se substituent pas à l'intervention humaine mais en démultiplient l'efficacité. Le choix peut être fait, selon le cas, non pas de maintenir la qualité du service à effectifs réduits, mais de l'améliorer à moyens constants.

Ensuite, les ressources humaines libérées des tâches automatisées peuvent être **redéployées,** non seulement pour assurer l'entraînement (la complémentarité entre le travail humain d'annotation des données et le SIA est manifeste, par exemple, dans le déploiement de l'outil de pseudonymisation des décisions de justice de la Cour de cassation), la supervision et la maintenance des systèmes, mais aussi et surtout pour assurer des prestations qui ne le sont pas actuellement, ou qui le sont mal, faute de moyens humains suffisants. Tel est le cas du traitement de situations ou dossiers complexes, sensibles ou sortant de l'ordinaire, exigeant concentration, discernement et temps, et qui ne trouvent pas de réponse satisfaisante dans le fonctionnement actuel d'une administration embolisée par le « tout-venant ». L'automatisation des décisions faisant droit à des demandes (décisions favorables) permet de consacrer davantage de ressources à l'examen de celles pour lesquelles le système a recommandé un rejet. On voit tout l'intérêt que présente l'automatisation partielle, en concentrant les moyens humains sur la petite minorité des dossiers non standards nécessitant une prestation sur-mesure que seul un agent public peut aujourd'hui offrir.

Enfin, les SIA étant prioritairement affectés, en l'état des techniques, à l'accomplissement de tâches matérielles et d'activités de support, les personnels qui en avaient la charge peuvent être déployés sur le terrain, au service de missions impliquant une forte dimension humaine et, à ce titre, insusceptibles d'être prises en charge par la machine, notamment dans le secteur sanitaire, social et médico-social. De même, le temps qu'un policier ou un gendarme économise sur la rédaction des procédures grâce à un SIA de transcription des propos tenus par les personnes entendues peut être consacré aux enquêtes ou à la prise en charge psychologique des victimes. Il n'est pas chassé de son emploi par le SIA : il y est, au contraire, réinstallé.

Il y a lieu d'ajouter que, eu égard aux capacités actuelles des SIA, ce sont d'abord des **tâches simples, répétitives et rébarbatives** qui ont vocation à être automatisées. Or celles-ci peuvent être à l'origine d'une lassitude, d'une démotivation et d'un sentiment de déclassement des agents, voire de risques psycho-sociaux. Un agent d'accueil téléphonique dont un *chatbot* décharge la ligne pour ne lui transmettre que les appels que le SIA – à raison de leur complexité ou de leur originalité, ou de la nature du besoin exprimé – l'aide à sélectionner est ainsi mieux disposé pour fournir une réponse personnalisée et de qualité. L'automatisation peut donc aussi constituer une opportunité du point de vue de la qualité de vie au travail.

Cette présentation mérite évidemment d'être tempérée et confrontée au réel.

D'abord, elle fait fi d'**effets pervers liés aux potentialités et facilités offertes par le numérique**, qui, loin de simplifier la vie des agents, peut au contraire s'avérer synonyme d'alourdissement de la charge de travail ou de sur-sollicitation, en sus des irritants liés aux dysfonctionnements informatiques. L'usage intensif des messageries électroniques, jusqu'à l'embolie parfois, en fournit le meilleur exemple. C'est la raison pour laquelle il est indispensable de prévoir des garde-fous contre le mésusage numérique, à l'instar du droit à la déconnexion permettant de préserver la vie privée des agents en dépit de l'effacement des frontières entre la vie « de bureau » et la vie « à la maison », exacerbé par le télétravail. Il y a là un champ neuf de négociation collective pour définir de nouveaux droits et de nouvelles garanties, afin que les agents s'emparent d'auxiliaires utiles, et ne soient pas submergés par des systèmes invasifs qui accréditeraient le sentiment d'une submersion.

Il ne faut pas non plus ignorer le phénomène de « **(sur)charge mentale »** de l'agent public recentré sur la gestion de la complexité, de la pathologie et de l'atypie. Ce dernier peut aussi trouver intérêt, dans une certaine mesure, à ménager une respiration dans son plan de charge, en s'acquittant occasionnellement de tâches simples, tout en conservant une compétence qui lui confère à la fois la capacité de demeurer critique envers le système et de gérer le service en cas de défaillance.

Il ne doit pas davantage être perdu de vue que la conception et le développement des modèles d'apprentissage automatique ne font pas nécessairement appel, à titre exclusif, à des compétences de haut niveau. L'annotation des données peut, dans certains cas, ne requérir que du bon sens ou un niveau d'instruction minimal, et éventuellement une très brève formation. Si l'entraînement d'un modèle performant de détection de pathologies dans des images médicales requiert la mobilisation de professionnels de santé, les données d'apprentissage d'un SIA de pseudonymisation de documents supposent, en substance, de savoir distinguer le nom propre d'une personne physique d'un nom commun. Les progrès de la recherche scientifique en algorithmique, en particulier les apprentissages auto-supervisés et par renforcement, tendent certes à réduire le besoin d'intervention humaine pour la labellisation des données, mais ils ne le feront pas disparaître à brève ou moyenne échéance. Et les administrations pourront trouver intérêt à y associer la multitude des agents du service plutôt que des prestataires externes.

Enfin, l'exercice de « redéploiement » des agents restera parfaitement vain et artificiel, voire, dans certains cas, source de souffrance au travail, en l'absence d'une **gestion prévisionnelle des compétences et des carrières** adaptée à ce « choc numérique », établie en concertation avec les organisations syndicales, et d'un effort massif d'accompagnement et de formation des agents. La reconversion ne s'improvise pas, elle s'anticipe et se prépare. La question du devenir de fonctionnaires inaptes à toute autre fonction que celle qu'ils exercent parfois de très longue date appellera, selon l'ampleur du phénomène, des dispositifs particuliers de gestion de la transition professionnelle et de réinsertion.

3. Soutenir la compétitivité économique du pays

On sait à quel point la performance des administrations publiques d'un pays est un facteur clé dans les choix d'investissements. La simplicité des démarches, la rapidité de leur traitement et la fiabilité des réponses auxquelles contribuent les SIA sont autant de facteurs d'attractivité économique que la France aurait tort de négliger. Dans l'espace européen, qui stimule la concurrence entre Etats membres, l'avantage compétitif que confère le bon usage des SIA au service public, donc à l'économie qu'il sert, ne saurait être ignoré.

L'investissement public dans les SIA, à travers la **commande publique** et les **partenariats public-privé**, est également de nature à stimuler notre économie et à favoriser le dynamisme d'un écosystème d'entreprises compétitives, des jeunes pousses aux grands intégrateurs. Le marché français de la « govtech », qui correspond aux prestations numériques commandées par les collectivités publiques, est évalué à une quinzaine de milliards d'euros par an. Il excède 100 milliards aux Etats-Unis[104]. La part des systèmes d'IA dans ce potentiel est difficile à évaluer, d'autant qu'il n'existe pas, à l'heure actuelle, de recensement des marchés passés par les administrations françaises portant spécifiquement sur ces prestations. Mais il n'est pas douteux qu'il est amené à occuper une part croissante et significative, notamment dans des secteurs où les cas d'usage sont exclusivement ou principalement publics, comme dans le domaine de la défense ou de la sécurité.

La sphère publique joue également un grand rôle dans l'attractivité de la France pour les talents du secteur de l'intelligence artificielle, particulièrement convoités. Notre pays ne peut espérer conserver et attirer les meilleurs et, ainsi, déclencher une dynamique vertueuse (la compétence attirant la compétence), si le premier employeur de France qu'est la fonction publique reste en marge du développement des SIA.

Il importe ainsi de mesurer à quel point le déploiement de systèmes d'intelligence artificielle dans le service public peut jouer un rôle majeur dans la croissance économique du pays et, partant, dans l'amélioration du bien-être collectif. Il ne s'agit pas de concevoir un nouveau plan calcul et de promouvoir ses champions industriels : il s'agit de partager en Europe des choix (de sous-traitants, de soutien aux entreprises naissantes, de financement judicieux de la recherche et du développement) qui fasse jouer au déploiement des systèmes d'intelligence artificielle dans le service public le rôle d'un accélérateur, d'un incubateur, d'un stimulateur de la généralisation à l'économie et à la société de l'évolution en cours dans ce qu'elle a de plus productif, plutôt que de laisser des opérateurs privés, souvent étrangers, en capturer les fruits, sans retour pour la collectivité nationale.

[104] M-B. Girard et G. Fonlladosa, PUBLIC, *Govtech en France - Etat des lieux et perspectives*. Le secteur de l'éducation (« EdTech ») représenterait 20% de ce marché global, de même que les « villes intelligentes », suivis de la mobilité (17%) et de la santé (15%), de la sécurité et de la défense (10%) et de la participation citoyenne (« CivicTech » - 6%).

2.2. Le potentiel des SIA reste encore à exploiter au sein de la sphère publique

L'exercice de recensement des cas d'usage auquel le Conseil d'État s'est livré révèle que le potentiel des SIA tel qu'il vient d'être décrit reste encore très largement sous-exploité dans la sphère publique.

Certes, les administrations intègrent sans grande difficulté les fonctionnalités les plus matures, et qui sont d'ailleurs déjà largement utilisées par les acteurs privés, à l'exemple des robots conversationnels et de la reconnaissance d'objets. Elles sont capables d'extraire de plus en plus de connaissances des données qu'elles produisent ou se procurent, notamment pour comprendre les facteurs explicatifs d'un problème. Mais elles restent, pour l'essentiel, dans une **phase de découverte et d'expérimentation**, plus ou moins concluante, et n'ont pas enclenché, autour de l'IA, un mouvement structurel et massif d'automatisation ni, *a fortiori*, de redéfinition des politiques publiques, de leurs missions et de leurs principes de fonctionnement.

Cet état de fait n'est pas propre à la France. Ce constat a été dressé à l'échelle de l'Union européenne par l'étude de la Commission européenne (*AI Watch*) déjà mentionnée sur l'IA dans le secteur public. Celle-ci observe un déséquilibre entre le potentiel de transformation de l'IA et l'utilisation effective de solutions y recourant par les administrations, et relève qu'il existe peu de preuves d'impacts socio-économiques significatifs de cette technologie dans le secteur public. La plupart des 230 cas d'usage étudiés portent sur des outils relativement frustes, dont les effets ont essentiellement été testés dans des environnements contrôlés et non en situation réelle, et dont l'impact sur le changement est, la plupart du temps, qualifié d'« incrémental ». Seuls trois cas d'usage sur ces 230, dont la consistance n'est d'ailleurs pas précisée, sont répertoriés comme s'inscrivant dans un changement qualifié de « radical » dans la définition des politiques publiques ou du modèle de service public. L'étude conclut que le secteur public joue avant tout un rôle de régulateur et de facilitateur (« *gouvernance de l'IA* ») plutôt que d'utilisateur ou de bénéficiaire direct (« *gouvernance par l'IA* »).

La France n'accuse donc pas de retard au regard des pratiques constatées dans les administrations des autres Etats membres de l'Union européenne, mais elle ne bénéficie pas non plus d'une quelconque avance, sous réserve, éventuellement, d'un secteur comme celui de la défense, dans lequel l'État investit traditionnellement davantage que d'autres pays. La France ne représentait d'ailleurs que 5% des cas d'usage recensés par *AI Watch* dans son étude, et 8% des 376 cas répertoriés à la fin septembre 2021.

Au-delà de ce tableau général, on constate surtout une **très grande hétérogénéité** dans la maturité des administrations, le degré d'avancement des réflexions et des projets et l'ampleur des investissements qui y sont consacrés. Alors que les activités de contrôle, d'enquête et de sanction, dans lesquelles le « retour sur investissement » est bien identifié, sont assez dynamiques, le déploiement de SIA innovants est plus balbutiant dans les activités de service à l'usager, et embryonnaire pour ce qui concerne les fonctions support, notamment dans le domaine des

ressources humaines. Les SIA sont souvent conçus comme un moyen d'accroître directement la performance opérationnelle des agents, en leur fournissant des outils supplémentaires, voire comme un outil de substitution, et peu comme un levier d'amélioration de leurs conditions de travail et de la qualité de vie au travail.

La disparité constatée au niveau des administrations de l'État est plus flagrante encore au niveau local. Elle tient évidemment, en partie, à la très grande hétérogénéité des collectivités territoriales, en termes de taille et de moyen. Mais pour des collectivités comparables, l'ambition et les réalisations peuvent être très éloignées, selon le plus ou moins grand volontarisme politique en matière numérique et l'investissement personnel d'élus ou d'agents publics versés dans ces questions.

Les freins au déploiement de SIA, tels qu'ils ressortent des auditions et des réponses au questionnaire[105] qui a été transmis à tous types d'administrations, tiennent, d'une part, à l'absence des ressources de toute nature nécessaires à ce déploiement, et, d'autre part, à une aversion à une série de risques auxquels les SIA sont associés.

S'agissant des ressources, le premier obstacle identifié est le manque de données, ou **l'insuffisante qualité des jeux de données**. La plupart des administrations ont ainsi souligné les difficultés à trouver le temps nécessaire pour nettoyer et faire monter en qualité les jeux de données nécessaires à la construction d'un SIA. Or, les services, dont certains ont fait état d'une exigence de résultat peu compatible avec le principe de « l'essai-erreur » qui préside à l'élaboration d'un SIA, sont peu enclins à se lancer dans un tel chantier, dont l'issue est, par définition, incertaine. Plus de la moitié des 45 administrations ayant répondu au questionnaire identifient d'ailleurs parmi les obstacles rencontrés l'utilité non avérée des projets de SIA, ce qui peut traduire tant l'échec d'un projet passé et abandonné que l'aversion au risque d'un potentiel échec futur.

Le second obstacle tient au **manque de moyens**. Il s'agit à la fois d'un manque de temps et d'un manque de crédits budgétaires. Un certain nombre d'administrations n'identifient pas clairement auprès de quels services elles peuvent trouver un appui méthodologique (trois quarts des administrations déplorent un manque d'accompagnement et de formation) et font état, pour la grande majorité d'entre elles, des difficultés à fidéliser certaines compétences techniques qu'elles réussissent pourtant, mieux que par le passé, à recruter. Le manque de temps révèle plus largement l'atrophie de certaines équipes numériques au regard des chantiers qu'elles ont à mener par ailleurs, reléguant au second rang de leurs priorités l'engagement d'une démarche IA : la contraction des budgets de fonctionnement et la priorisation des dépenses d'investissement sur le maintien en conditions opérationnelles des systèmes d'information existants laissent peu de marge de manœuvre pour le financement de projets innovants dont le retour sur investissement est incertain, y compris dans les plus grandes collectivités.

[105] La synthèse des réponses à ce questionnaire figure en annexe 7.

S'agissant des obstacles tenant aux risques, réels ou supposés, associés au déploiement d'un SIA, ils sont de trois ordres :

- Le **risque juridique** est le plus partagé, ce que deux facteurs tout à fait liés peuvent expliquer. En premier lieu, la conformité au RGPD est anticipée comme une difficulté majeure, annihilatrice des enthousiasmes initiaux, ce qui traduit plus largement une méconnaissance du cadre juridique applicable depuis son entrée en vigueur. Les difficultés à répondre aux demandes (effectives ou escomptées) du régulateur ont également été largement citées. Les administrations anticipent notamment les obstacles à la mise en relation de traitements de données à caractère personnel dont les finalités diffèrent, mais dont le rapprochement ou l'interconnexion permettraient de mieux évaluer les politiques publiques (par exemple, la réalisation d'une étude sur les freins sociaux à l'utilisation des bibliothèques suppose de rapprocher les fichiers des services sociaux et des usagers des bibliothèques municipales). En second lieu, il faut souligner une crainte partagée en termes de responsabilité, ou plus précisément d'imputabilité, et d'explicabilité des décisions prises avec l'appui d'un SIA, rejoignant ici un questionnement éthique largement répandu au sein des administrations. C'est ce qui explique que la grande majorité d'entre elles ne trouve pas d'utilité ou qu'une utilité faible au développement de SIA décisionnels sans supervision humaine.

- Le **défaut d'acceptabilité** est également un risque identifié comme majeur, qu'il s'agisse des usagers du service ou des agents publics. Lié à l'acculturation des parties prenantes sur les enjeux et potentialités offerts par l'intelligence artificielle, ce risque tend à se réaliser lorsque la méthode de construction du SIA est exclusivement descendante, ne répondant pas à un besoin des métiers clairement exprimé par les agents. Il ne faut, toutefois, pas nier le poids de certaines habitudes : qu'il s'agisse de la rigidité de l'organisation administrative ou de l'aversion psychologique au risque et au changement, les administrations, même locales, peuvent manquer de l'agilité nécessaire pour concevoir des SIA réellement transformants, ou ne peuvent le faire qu'à un horizon temporel éloigné.

- Le troisième risque tient aux garanties de **sécurisation de l'outil**, dans un contexte de vigilance accrue en matière de cybersécurité. Les administrations sont en effet très conscientes que les bases de données nettoyées constituent une nouvelle richesse, nécessitant une forte protection.

Enfin, il faut souligner que **les obstacles rencontrés par les collectivités ne sont pas fondamentalement différents de ceux auxquels font face les administrations de l'État**. Toutefois, certaines spécificités doivent être mentionnées.

D'une part, elles sont davantage confrontées au manque de compétences disponibles. Selon la taille de la collectivité, il est soit impossible financièrement, soit très difficile compte tenu des conditions de rémunération et des perspectives de carrière proposées, de recruter des experts de la donnée susceptibles de concevoir des SIA ou d'en piloter la réalisation par des prestataires privés. Cela est notamment le cas hors des grandes métropoles, et lorsque la collectivité se situe dans un territoire sans écosystème local dynamique ni grand pôle de formation numérique.

D'autre part, s'agissant des contraintes de la commande publique[106], certaines collectivités regrettent de ne pas disposer d'une plus grande liberté dans le recours à des prestataires privés, lesquels peuvent être dissuadés de candidater à des consultations longues et gourmandes en ressources, surtout les petites entreprises de leur écosystème territorial de la « govtech » qu'elles souhaitent pourtant, et légitimement, privilégier.

2.3. Conduire une stratégie volontariste et lucide

L'ensemble des entités de la sphère publique sont susceptibles de tirer le plus grand profit des SIA et des bénéfices précédemment mentionnés.

Les collectivités territoriales le sont peut-être plus que toute autre. Le potentiel de connaissances que renferment les très nombreuses données qu'elles collectent ou produisent sur les caractéristiques sociales et économiques de leur territoire et de leur population, l'agilité dont elles bénéficient et la confiance que leur accordent plus spontanément les citoyens sont autant d'atouts qui peuvent en faire un formidable laboratoire pour le déploiement de systèmes d'IA au service de leurs nombreuses compétences, qu'elles relèvent de la prestation de service aux usager ou du contrôle, et répondre à une demande sociale toujours plus forte dans un contexte budgétaire toujours plus contraint. Les SIA judicieusement déployés sont de nature à restaurer une autonomie autrement menacée par la nécessité de recourir à des expertises extérieures, de l'État, d'autres collectivités ou de prestataires.

Ces bénéfices ne pourront être engrangés qu'au prix d'une démarche volontariste impliquant l'ensemble de la sphère publique.

2.3.1. Le secteur public ne doit pas attendre le moment, mais le créer

La plupart des administrations ont la conviction ou l'intuition diffuse – selon leur degré d'acculturation au sujet – que les systèmes d'IA, loin d'être un effet de mode, structureront de façon croissante leur fonctionnement à l'avenir, comme en leur temps l'informatisation puis la généralisation d'Internet. L'inéluctabilité apparente de cette évolution ne doit pas pour autant justifier des postures passives d'acquiescement ou de résignation face à un phénomène subi, ni un calendrier qui serait dicté par d'autres.

Sans doute la pression citoyenne à la performance publique et les contraintes croissantes de ressources du secteur public l'amèneront-ils naturellement à se tourner plus massivement, tôt ou tard, vers cet ensemble de solutions innovantes. On ne peut pas non plus totalement exclure que les performances élevées et croissantes des SIA pour l'accomplissement d'un certain nombre de tâches, au point de dépasser dans certains cas les facultés humaines, suscitent l'émergence de revendications directes d'un « **droit à l'IA** » par des citoyens ou des entreprises qui soutiendraient que leur demande ou leur situation serait ou aurait été mieux traitée en faisant appel à un tel système, ou par des agents publics ou des organisations syndicales qui considèreraient que certaines tâches sources d'accidents du travail ou

[106] V. 4.1.1. sur ces contraintes.

de risques psycho-sociaux, notamment les plus répétitives, devraient être automatisées. Mais l'hypothèse reste peu probable compte tenu de la défiance ou de la vigilance qu'inspirent ces systèmes dans une partie importante du corps social. Et, en tout état de cause, le débouché juridictionnel apparaît étroit.

Existe-t-il un droit à l'IA ?

L'administration dispose en principe du **choix des moyens** pour s'acquitter d'une obligation pesant sur elle. L'exigence de mutabilité du service public, dont l'une des facettes est son adaptation aux évolutions techniques, constitue davantage une construction doctrinale qu'un principe général dégagé par la jurisprudence. En tout état de cause, celle-ci n'a jamais tiré de cette exigence un droit à l'innovation publique et à la « contemporanéité technologique » de l'administration, mais seulement la faculté reconnue à l'administration de se défaire de contraintes, notamment contractuelles, qui entravent l'amélioration du service public.

En l'état de la jurisprudence, **aucune norme supérieure ne paraît susceptible de fonder directement un « droit au SIA »**, pas plus qu'une disposition législative, à l'instar de ce qui peut exister aux Etats-Unis à travers le principe de performance[107]. Les normes constitutionnelles touchant à la qualité et à la performance de l'action publique sont, pour la plupart, des objectifs de valeur constitutionnelle insusceptibles d'être directement invoqués par le justiciable, quoiqu'ils s'imposent au législateur (objectif de bon usage des deniers publics, objectif de bonne administration de la justice, lutte contre la fraude…). Le « *droit à une bonne administration* » consacré par l'article 41 de la Charte des droits fondamentaux de l'Union européenne, dont la portée est circonscrite au droit de voir ses affaires traitées impartialement, équitablement et dans un délai raisonnable et aux garanties énumérées aux paragraphes 2 à 4, n'est opposable qu'aux institutions et organes de l'Union, et non aux autorités nationales, y compris lorsqu'elles mettent en œuvre le droit de l'Union[108].

De même qu'il est possible d'attaquer devant le juge de l'excès de pouvoir le refus de l'administration de créer un service public ou d'en modifier les conditions d'organisation et de fonctionnement afin d'en assurer un

[107] Un décret présidentiel du 3 décembre 2020 (*Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*) prévoit que les agences doivent notamment respecter un « *principe de performance* » qui leur fait obligation de recourir à un SIA toutes les fois où les avantages qu'ils présentent l'emportent « *largement* » sur les risques et que ces derniers peuvent être évalués et gérés. On relèvera aussi que, dans un arrêt n° 2270 du 8 avril 2019, le Conseil d'État italien a jugé que la prise de décision administrative automatisée constituait une « *application cohérente* » du principe constitutionnel de bonne action administrative garanti par l'article 97 de la Constitution, qui oblige l'administration à « *atteindre ses objectifs avec le moins de moyens et de ressources possibles et par la rationalisation et l'accélération des procédures* ». Mais cette décision ne dégage pas un droit à l'automatisation des décisions.

[108] CJUE, 17 juillet 2014, *YS e. a.*, C-141/12 et C-372/12, pts 67-69. Cet arrêt précise toutefois que l'article 41 de la Charte « *reflète un principe général du droit de l'Union* », laissant entrevoir la possibilité d'en tirer d'autres exigences que celles qui sont posées par ce texte.

fonctionnement normal, et de soutenir utilement que ce refus repose sur une erreur manifeste d'appréciation[109], un usager pourrait certes s'aviser de contester sous cet angle un refus d'automatiser une procédure administrative ou de recourir à un système d'assistance numérique à la décision. Mais, à défaut de dispositions particulières garantissant un certain niveau de qualité du service public[110], il est permis de douter qu'une telle contestation puisse vraisemblablement prospérer en l'absence d'un décalage flagrant entre le fonctionnement du service en cause et « l'état de l'art », créant un écart inacceptable de qualité entre le service effectivement rendu et celui qui est raisonnablement exigible de l'administration, eu égard aux ressources dont elle dispose et aux besoins des usagers. Enfin, des actions en responsabilité, invoquant une perte de chance liée à l'absence d'utilisation de SIA, sont également envisageables. À cet égard, il appartiendra au juge de déterminer à partir de quel stade de maturité et de généralisation d'un SIA l'absence de son utilisation par une administration est susceptible d'ouvrir droit à réparation du préjudice subi par la victime de ce choix ou de cette inertie technologique.

Au total, si les chances de succès d'une action en carence fondée sur le non-recours à un SIA paraissent limitées à l'heure actuelle, même si elles s'accroîtront à mesure que ces techniques deviendront d'usage courant, il n'en serait pas moins regrettable que les collectivités publiques soient contraintes par le juge de les utiliser.

Il est donc certain que le rythme de diffusion de SIA dans la sphère publique et les effets qu'ils y produiront, dépendront d'abord et avant tout des initiatives que prendront les autorités compétentes, et de l'engagement des dirigeants publics.

De l'ensemble des auditions conduites, il ressort en effet clairement que **l'engagement et le soutien au plus haut niveau**, politique comme administratif, constituent une condition indispensable au succès de tout projet de SIA. Si les décideurs politiques et l'ensemble de la chaîne hiérarchique administrative ne sont pas convaincus du bien-fondé d'un recours plus fréquent et plus ambitieux à ces outils, le pays accumulera retard et déconvenues. A l'inverse, la volonté politique peut enclencher une **dynamique auto-entretenue de l'IA publique** exploitant les **effets de système et d'échelle** propres à ces technologies innovantes. D'un écosystème public résolument engagé dans le développement de ces solutions naîtraient une émulation et un partage d'expérience, d'idées, de données et de modèles algorithmiques, qui transcenderaient

[109] CE, 19 juillet 1991, *Fédération nationale des associations d'usagers des transports*, n° 115294, Rec. ; CE, 10 novembre 1997, *P.*, n° 173137, Rec.

[110] A titre d'exemple, l'art. L. 1110-5 du code de la santé publique garantit à toute personne, compte tenu de son état de santé et de l'urgence des interventions que celui-ci requiert, « *le droit de recevoir, sur l'ensemble du territoire, les traitements et les soins les plus appropriés et de bénéficier des thérapeutiques dont l'efficacité est reconnue et qui garantissent la meilleure sécurité sanitaire et le meilleur apaisement possible de la souffrance au regard des connaissances médicales avérées* ». De telles dispositions sont susceptibles de motiver des actions tendant à ce que le praticien ait recours des SIA voire des actions en responsabilité fondées sur la perte de chance d'éviter un accident médical en raison de l'absence d'utilisation d'un tel système.

d'ailleurs le clivage public/privé. A défaut, se formera, dans le prolongement de la situation actuelle, un archipel de l'IA publique composé d'une poignée de services en pointe mais qui n'auront ni la capacité d'entraînement, ni même la volonté, pour faire monter en compétence les autres entités publiques.

L'implication résolue et précoce des dirigeants publics est d'autant plus cruciale que le développement des SIA est un **chemin exigeant**, qui suppose d'investir tôt et dans la durée pour espérer en récolter des fruits. Cet investissement porte naturellement sur les ressources nécessaires à leur conception, développées dans la quatrième partie de l'étude, qu'il s'agisse du recrutement et de la formation des experts de la donnée, très convoités sur le marché du travail, des infrastructures coûteuses, ou encore de la politique des données dont les ressorts sont autant techniques et juridiques que culturels. Il consiste aussi dans l'apprentissage, qui n'a rien d'automatique pour les administrations, de règles nouvelles ou appliquées à un objet nouveau et de bonnes pratiques révélées par l'expérience, y compris par les erreurs.

2.3.2. Les points de vigilance stratégique

1. Le volontarisme n'est pas le solutionnisme

L'approche volontariste peut rapidement entraîner une dérive solutionniste, qui verrait dans la technologie numérique la solution à tous les problèmes que rencontrent les administrations, et qu'il est impératif de conjurer d'emblée.

Le premier écueil consisterait à regarder les SIA comme des fins et non comme des moyens. Or, ces systèmes sont des outils au service des politiques publiques, des auxiliaires d'administration et de justice, et ne sont que cela. Pas plus que l'électricité n'a déterminé la géographie, les horaires d'ouverture ou les modalités de l'action publique (mais a pu faciliter ici son implantation, là son amplitude d'ouverture, et ailleurs sa performance), le déploiement de SIA ne peut ni ne doit être l'objectif même du service public. Il ne s'agit donc pas de céder à l'élan lénifiant de l'effet de mode ou du marketing politico-administratif.

Le deuxième écueil consisterait à **regarder les SIA, par principe, comme le meilleur moyen d'accomplir une tâche et comme une garantie absolue de performance administrative**. Tout en cultivant le « réflexe IA », au sens où, confrontée à un problème, la collectivité publique doit avoir conscience qu'elle a, entre autres, cet outil à sa disposition, elle ne saurait s'affranchir d'une réflexion au cas par cas sur la meilleure méthode à retenir pour le résoudre, en tenant compte des coûts de toute nature de l'innovation numérique (coût humain, coût écologique, coût financier, ingérences dans les droits et libertés fondamentaux…). Il n'y a pas lieu d'entreprendre la construction d'un réseau de neurones à vingt ou cent couches quand un système-expert simple ou une régression logique suffiraient à répondre au besoin de façon satisfaisante. Et il est, somme toute, assez fréquent qu'une tâche puisse être réalisée efficacement en recourant à un tableur basique, sans qu'il soit besoin de concevoir à grands frais un système d'information complexe reposant sur l'IA. De manière générale, l'approche la plus économe technologiquement (« *low-tech* ») pour atteindre l'objectif mérite d'être privilégiée.

Il serait également **illusoire de considérer les SIA comme un substitut systématique aux efforts de bonne gouvernance et de rationalisation des organisations et des normes**. Si, dans une certaine mesure, ces systèmes permettent de gérer la complexité voire le désordre (qu'on songe, par exemple, à la recherche d'informations dans des bases de données non structurées), ils pâtiront tout autant que les agents humains des silos qui entravent la circulation des données et des modifications incessantes du cadre juridique et des dispositifs administratifs, qui les rendront à plus ou moins brève échéance obsolètes. Et ils resteront longtemps impuissants à assister l'administration dans la mise en œuvre de lois et règlements « de niche » (qui ne permettront pas de constituer un jeu de données suffisant pour bâtir un modèle d'apprentissage automatique) ou truffés d'exceptions (ce qui complique considérablement la conception de systèmes basés sur les règles), de formules vagues et de critères subjectifs qui ne se prêteront pas à l'automatisation. Si la règle de droit est confuse ou imprécise, aucune machine n'en assurera une application satisfaisante.

Plus largement, jamais un SIA ne transformera le plomb d'une mauvaise politique publique en or socio-économique. S'ils peuvent y contribuer, ces systèmes ne désendetteront pas la France, n'éradiqueront pas la pauvreté et n'assureront pas la cohésion sociale et nationale, pas plus qu'ils ne mettront fin à la criminalité et à la délinquance.

2. *Cheminer par étapes*

L'administration ne doit pas rêver d'emblée d'un grand soir de l'IA mais s'éveiller au petit matin de l'apprentissage automatique. Les prérequis et les chausse-trappes sont nombreux, notamment sur le plan méthodologique[111] ; la démarche peut s'avérer potentiellement ingrate, tant les résultats peuvent se révéler modestes au regard des efforts consentis et des ressources mobilisées ; la sous-performance d'un modèle développé trop rapidement ou prématurément peut fragiliser la confiance dans les systèmes (V. 3.1.2. sur ce point) ; et le risque d'échec, illustré par plusieurs projets récents comme l'outil DataJust, n'est jamais à écarter. A ce titre, le lancement d'un processus de déploiement de SIA publics doit être conjugué avec une nouvelle culture de l'échec au sein des administrations : la culture entrepreneuriale (« échouer souvent, échouer rapidement ») est souvent louée comme un fondement des innovations, et le secteur public pourrait utilement s'en inspirer, en distinguant les échecs inacceptables au regard des investissements consentis et les échecs « normaux » au regard d'une dynamique d'innovation numérique qui implique nécessairement un processus d'apprentissage non linéaire.

[111] V. 4.1.5. pour quelques points de repères méthodologiques.

DataJust

Le décret n° 2020-356 du 27 mars 2020 a autorisé le ministère de la justice, à titre expérimental, à développer un outil permettant d'extraire et d'exploiter de façon automatisée les données relatives aux montants des demandes indemnitaires en réparation des préjudices résultant de dommages corporels, par poste de préjudice, des offres indemnitaires des personnes mises en cause, des évaluations effectuées dans le cadre de procédures de règlement amiable et des condamnations prononcées à ce titre, qui figurent dans les décisions rendues par les juridictions d'appel, de l'ordre administratif et de l'ordre judiciaire.

Un tel outil pourrait fournir aux victimes une information supplémentaire leur permettant de fixer leurs prétentions de manière éclairée et, une fois la décision rendue, d'apprécier l'opportunité de la contester ou non. Il permettrait en outre au juge de statuer en pleine connaissance des pratiques « habituelles » et de favoriser une convergence des doctrines d'indemnisation entre les deux ordres, sans pour autant créer un barème contraignant dont l'utilisation ne pourrait être autorisée que par la loi (V. CCass. 1e civ., 23 octobre 2013, n° 12-25301). A terme, le ministère de la justice envisageait de mettre en ligne un « référentiel indicatif d'indemnisation » à destination du grand public.

Le Conseil d'État, statuant au contentieux, a confirmé la licéité d'un tel traitement de données à caractère personnel au regard du RGPD (CE, 30 décembre 2021, *Sté Gerbi Avocat Victimes et préjudices et autres*, n° 440376, 440976, 442327, 442361, 442935).

La complexité technique du projet et la nécessité de ressources conséquentes pour atteindre un niveau de performance acceptable, conjuguées aux importantes réserves suscitées par une partie importante de la communauté juridique, ont convaincu la Chancellerie de l'abandonner au terme de deux ans d'expérimentation. Cette expérience est sans doute une utile leçon dont toutes les composantes (marchés, compétences, objectifs, pilotage…) devraient être examinées, non dans un esprit de recherche de responsables ou de stigmatisation de l'échec, mais pour mieux identifier les facteurs de succès et les points de vigilance, et y sensibiliser l'ensemble des administrations.

Dans un premier temps, l'ambition pourrait se limiter à deux actions :

- d'une part, développer dans chaque collectivité un ou deux systèmes modestes à fort enjeu symbolique afin d'éprouver et de démontrer le potentiel de ces techniques (« victoires rapides ») ;

- d'autre part, rechercher les possibilités de conversion numérique du fonctionnement de l'administration, à objectifs et cadre d'organisation inchangés. Il s'agirait, par exemple, de demander à chaque service de consacrer quelques heures à un passage en revue des processus administratifs qu'ils conduisent, en les décomposant finement et en cotant les tâches qui les composent au regard de la **faisabilité (et de la facilité) d'automatisation et de son bilan avantages/inconvénients**, avant de définir sur la base de cette

cartographie quelques projets prioritaires à fort enjeu opérationnel. L'un des prérequis de cet exercice est la bonne compréhension des principes de fonctionnement des SIA et des bénéfices qui peuvent en être attendus, sans surestimation (au risque de partir dans de mauvaises directions, à fonds perdus) ni sous-estimation (au risque de manquer des opportunités d'amélioration).

Forte de la maturité acquise par la conduite de quelques projets, notamment dans la compréhension des capacités et des limites des SIA, et des résultats de leur évaluation critique, notamment par la mesure de la satisfaction des usagers comme des agents utilisateurs, une administration peut se projeter sur le moyen/long terme, et entreprendre de repenser les politiques publiques, les modes d'organisation du service public et le contenu même des prestations publiques. Elle peut alors s'affranchir de l'existant et du simple réflexe de « traduction numérique » du fonctionnement physique, et développer sa capacité à imaginer des services qui n'existent pas et ne peuvent pas exister à l'heure actuelle, parce qu'ils requerraient des ressources humaines qui ne sont pas disponibles ou ne peuvent pas l'être.

3. Veiller à l'équilibre des usages

Quand bien même n'obéit-elle pas à une planification nationale pluriannuelle, la dynamique de déploiement des SIA publics devrait respecter un équilibre des usages, à au moins deux égards.

D'une part, si le service du public est nécessairement prioritaire, il convient de **ne pas négliger l'investissement dans les SIA au service des agents eux-mêmes** (par exemple, dans le domaine de la gestion des ressources humaines, l'automatisation des réponses aux questions les plus fréquentes qu'ils se posent en droit de la fonction publique). Ces usages-support présentent de nombreux intérêts : ils favorisent une bonne appropriation des concepts et des outils, contribuent à dissiper des idées fausses sur ces systèmes, et permettent d'aiguiser le sens critique voire de susciter des vocations et de bonnes pratiques pour le développement des SIA à usage « externe ». Ils peuvent en outre permettre un redéploiement des moyens humains du support interne vers les prestations au public.

D'autre part, il importe de veiller à un déploiement équilibré des usages au regard de leur **perception sociale par le public** – schématiquement, entre les systèmes dédiés au « service », vus comme des usages « positifs » (allocation d'aides, fourniture d'informations, accélération de l'instruction des dossiers, amélioration de la prise en charge thérapeutique…), et ceux qui le sont au « contrôle », perçus comme des usages « négatifs » (surveillance par les forces de l'ordre, ciblage des contrôles fiscaux et sociaux…)[112]. **Les effets des indispensables efforts de pédagogie, rappelés en première partie, peuvent être anéantis par le sentiment que les SIA**

[112] A titre de repère grossier, l'étude d'AI Watch fait état de la distribution suivante des 230 cas d'usage répertoriés : le contrôle représente 20%, l'aide à la définition des politique publiques, 17%, les SIA consacrés à la délivrance des prestations, 5%, les SIA de relation avec les usagers, de participation citoyenne et de service général, 38%, et les SIA de fonctionnement interne, 20%.

sont d'abord et avant tout des instruments de coercition ou, qu'à tout le moins, c'est ainsi que les pouvoirs publics les conçoivent.

Or on a vu que les SIA de contrôle se prêtaient particulièrement bien à l'utilisation de l'apprentissage automatique, ce qui, avec la perspective d'un « retour sur investissement » à court ou moyen terme, explique qu'ils figurent parmi les projets les plus aboutis. Mais le risque est réel qu'ils structurent le regard que les citoyens portent globalement sur l'IA dans la sphère publique. La crispation suscitée par la proposition de règlement de la Commission européenne en ce qui concerne l'utilisation des SIA pour la reconnaissance biométrique en temps réel dans les lieux publics, dont l'avis du comité européen de la protection des données porte la trace[113], atteste de l'extrême sensibilité des usages répressifs et coercitifs, reléguant à l'arrière-plan de l'attention démocratique des cas de recours qui pourraient s'avérer bien plus problématiques sur le plan sociétal.

La stratégie des usages de l'IA publique doit impérativement tenir compte de cet enjeu. Pour y répondre, il convient de veiller à ne pas déployer prématurément des systèmes à finalité répressive dont l'acceptabilité par les parties prenantes – usagers comme agents – n'est pas suffisamment garantie, quand bien même servent-ils l'intérêt général et collectif bien compris. Mais pour ne pas freiner à l'excès ce déploiement, il est tout aussi crucial d'investir dans des projets utilisant l'IA à des fins d'assistance et de secours, d'amélioration du service des prestations, d'accessibilité du service public ou d'inclusion sociale et dans leur valorisation.

Cet équilibre devrait être recherché non seulement **globalement**, mais aussi, idéalement, **à l'échelle d'un secteur d'activité ou d'une politique publique**. L'acceptabilité du recours à un SIA pour la détection de la fraude sociale sera d'autant mieux garantie qu'elle s'accompagnera d'un autre système d'IA permettant d'identifier les cas de non-recours aux prestations par des personnes qui en remplissent pourtant les conditions et de les leur servir. De la même façon, on peut parfaitement imaginer qu'un SIA de contrôle fiscal permette aussi de détecter les cas dans lesquels le contribuable s'est acquitté d'un impôt d'un montant supérieur à celui qu'il devait, faute d'avoir demandé le bénéfice d'un crédit d'impôt ou d'une réduction d'impôt. Dans le champ de la sécurité, il y a lieu de mettre en exergue l'intérêt que présentent les SIA tant pour la détection d'infractions pénales dans certains lieux publics particulièrement exposés à un risque de sécurité que pour l'analyse dite « prédictive » et la détection précoce des sinistres afin d'améliorer la réactivité et l'efficacité du déploiement des moyens d'intervention de la sécurité civile. Au sein même des usages policiers, il ne faut pas perdre de vue que les SIA peuvent être utilisés aussi bien pour arrêter les auteurs d'infractions que pour disculper des suspects (l'exploitation de leurs traces numériques pouvant étayer un alibi qu'ils n'étaient pas en mesure de fournir eux-mêmes), ou retrouver et porter assistance aux victimes.

---

[113] Cet avis sur la proposition de règlement, rendu en juin 2021, appelle à « *une interdiction générale de toute utilisation de l'IA pour la reconnaissance automatique des caractéristiques humaines dans les espaces accessibles au public, telle que la reconnaissance des visages, de la démarche, des empreintes digitales, de l'ADN, de la voix, des frappes au clavier et d'autres signaux biométriques ou comportementaux, quel que soit le contexte* ».

Il est important de préciser qu'il ne s'agit en aucun cas, ce faisant, de se livrer à une forme de manipulation destinée à contenir artificiellement la contestation contre les usages répressifs des SIA, mais de rendre compte fidèlement de la diversité des cas d'usage et des apports de l'IA et, ainsi, de lutter contre l'image déformée que peut donner l'investissement particulier consenti dans les technologies numériques de sécurité et les systèmes de détection et de répression des infractions et manquements de toute nature.

Aux points de vigilance qui viennent d'être évoqués, portant sur le calibrage de l'ambition générale, le choix des cas d'usage et le rythme de déploiement, s'en ajoute un dernier, cardinal en ce qu'il conditionne purement et simplement le succès de cette stratégie. Il porte sur la capacité du secteur public à maîtriser les risques que comportent les systèmes d'IA et à inspirer confiance par le respect de standards éthiques élevés, qui se traduisent par des principes fondamentaux auxquels la partie suivante est consacrée.

3. Définir et mettre en œuvre les principes et méthodes de l'IA publique de confiance

On a vu dans les deux parties précédentes que, pour répondre à la méfiance naturelle qu'inspire le recours à des machines et à des technologies souvent mal comprises et favoriser un déploiement fluide des systèmes d'IA dans l'action publique, il était nécessaire, tout à la fois, d'élever le niveau de familiarité des citoyens avec les concepts eux-mêmes et les principes de fonctionnement de ces systèmes, exemples à l'appui, de valoriser les bénéfices que la collectivité et chacun de ces membres peut (espérer) en retirer, et de veiller à un développement équilibré des usages, en particulier entre ceux dédiés au « service » et ceux qui le sont au « contrôle ».

Le défi prioritaire de l'acceptabilité sociétale des systèmes d'IA ne peut toutefois être relevé sans une identification assumée des **risques et limites** qu'ils présentent, notamment au regard des droits fondamentaux et des libertés publiques. Il existe à cet égard, sous l'appellation générique et quelque peu galvaudée d'« éthique de l'IA », une intense réflexion scientifique et doctrinale, ainsi qu'une très abondante littérature, qu'il est vain de recenser, qui documente ces multiples inconvénients. Certains travaux s'en prévalent, illustrations pittoresques[114] ou glaçantes[115] à l'appui, pour disqualifier par principe le recours à ces outils : ils n'ont d'intérêt que diagnostique. D'autres, plus constructifs, explorent les voies et moyens d'une maîtrise de ces risques à travers des solutions de nature politique, organisationnelle, technique ou encore juridique. Tel est l'enjeu de l'**IA digne de confiance** (« *trustworthy AI* »)[116], concept-pivot des institutions européennes qui a donné lieu à plusieurs productions[117], lesquelles ont inspiré en

[114] En 2020, le club de football écossais d'Inverness a remplacé le cameraman par un SIA utilisant la vision par ordinateur, lequel a passé une grande partie de la rencontre à filmer un arbitre de touche chauve dont la tête lui rappelait le ballon…

[115] Un robot conversationnel médical utilisant le modèle GPT-3 de traitement du langage naturel (contre les mises en garde de ses concepteurs) a par exemple suggéré à un (faux) patient de se suicider...

[116] Celle, étymologiquement proche, d'IA « fiable » est en revanche d'usage moins courant et renvoie davantage à l'idée d'exactitude des résultats fournis par le système.

[117] V. notamment les lignes directrices en matière d'éthique pour une IA de confiance du groupe d'experts indépendants de haut niveau sur l'intelligence artificielle de 2019 ; le livre blanc de la Commission européenne du 19 février 2020 (*Intelligence artificielle - Une approche*

partie les obligations juridiques que comporte la proposition de règlement européen sur les systèmes d'IA : définir les garanties propres à assurer que les systèmes d'IA accroissent le bien-être collectif et fonctionnent au bénéfice du plus grand nombre et à rassurer les parties prenantes à cet égard.

Les pouvoirs publics ont tout particulièrement la responsabilité de définir les conditions et garanties de l'IA de confiance dans le secteur public – qui s'imposeront tant aux acteurs publics qu'aux acteurs privés en tant qu'ils fourniront des SIA aux administrations. Pèse en effet sur l'administration un **devoir particulier d'exemplarité**, qui va au-delà du concept de responsabilité sociale appliqué aux entreprises. Les citoyens exigent plus de l'administration et lui pardonnent moins, tant parce qu'ils entretiennent avec elle un lien particulier forgé par l'histoire, d'ordre affectif, que parce qu'ils se perçoivent à la fois comme créanciers (ils la financent très majoritairement à travers des prélèvements obligatoires – contraints – plutôt que par le paiement du prix de la prestation[118]) et comme captifs à son égard : ils ne la choisissent pas, car elle est le plus souvent en situation de monopole, et ils ne peuvent s'en dispenser car elle offre souvent des prestations essentielles. L'erreur ou la dégradation de la qualité du service rendu à raison d'un SIA défectueux ou mal maîtrisé se paiera souvent plus cher dans la sphère publique en termes de confiance, pénalisant l'ensemble des démarches entreprises en la matière et, partant, l'intérêt général que cet outil a vocation à servir.

Or cette pression à l'exemplarité, qui n'a rien de nouveau, s'inscrit dans une tendance générale plus récente, loin d'être propre à la France, de **dégradation de la confiance dans la capacité des autorités publiques à régler les problèmes des citoyens**, voire, dans une partie du corps social, de défiance à l'endroit d'institutions suspectées d'avantager les puissants, de corseter les libertés, sinon de conspirer contre le peuple. Parce qu'elle véhicule aussi un imaginaire angoissant voire apocalyptique, dont il a déjà été question, l'intelligence artificielle s'expose, plus encore que bien d'autres instruments de l'action publique, à des contestations et des remises en cause potentiellement irrationnelles voire violentes, et donne prise à des thèses complotistes sur le thème de la surveillance et de la manipulation généralisées. Sans l'exagérer, cette réalité sociologique et psychologique doit impérativement être prise en compte dans la définition et la mise en œuvre de la stratégie de l'IA publique de confiance, qu'elle rend plus nécessaire encore.

Pour autant – et c'est toute la difficulté de l'exercice –, il convient aussi, en sens inverse, de se garder d'initiatives qui **entraveraient excessivement le déploiement des SIA publics**, dont on a précédemment exposé les avantages. A exagérer les

européenne axée sur l'excellence et la confiance) ; la résolution du Parlement européen du 20 octobre 2020 relative au *Cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes*) ; le rapport de l'Agence des droits fondamentaux de l'Union européenne de 2020, *Intelligence artificielle et droits fondamentaux*...

[118] Ce biais est d'ordre largement psychologique. Il est aujourd'hui très difficile de se passer d'un abonnement à une offre de téléphonie mobile, donc de ne pas s'acquitter du montant du forfait. Mais le paiement de la facture en contrepartie directe de la prestation fournie est mieux accepté que l'acquittement d'un impôt versé au « pot commun », sans contrepartie directe, conformément au principe de non-affectation budgétaire.

risques et à trop en faire, on finit, paradoxalement, par susciter la défiance et par créer l'anomalie là où on devrait simplement encadrer la normalité. D'une certaine manière, la construction du droit des traitements de données à caractère personnel en France à partir de 1978, dans le sillage de l'affaire Safari[119], s'inscrit dans ce biais cognitif qui explique, encore aujourd'hui, la connotation péjorative des termes « fichiers » et, plus encore, « fichage » et la vision des traitements comme un « mal nécessaire » plutôt que comme un simple besoin à encadrer.

Ce régime juridique des données à caractère personnel illustre un autre écueil pour la définition des contraintes pesant sur les SIA publics. Il tient à la **tradition légaliste des administrations françaises**. L'immense majorité d'entre elles sont mues, culturellement et en raison des multiples contrôles[120] dont elles font l'objet, par la préoccupation de respecter scrupuleusement les règles de droit qui leur sont applicables, reléguant les manquements délibérés à un phénomène quantitativement marginal. Le fonctionnement des entreprises repose quant à lui, le plus souvent, sur une analyse du risque juridique incluant l'occurrence d'une action contentieuse ou d'un contrôle et la gravité de ses conséquences (élevée pour le risque pénal, le plus souvent faible pour le risque civil). Ainsi, pour un petit responsable privé de traitement de données à caractère personnel, le risque de sanction résultant d'une violation du RGPD apparaît très limité en l'état des moyens de contrôle dont dispose la CNIL, et dépend essentiellement de l'introduction d'une plainte d'une personne concernée s'estimant lésée. On ne peut qu'être frappé du sentiment récurrent des administrations, exprimé lors des auditions menées pour la rédaction de l'étude, d'une forme d'« inégalité des armes » avec le secteur privé : **à obligation formellement équivalente, la contrainte juridique réelle sera sensiblement supérieure, ou au moins vécue comme telle, par les administrations**. Cette considération doit impérativement rester à l'esprit des concepteurs du cadre juridique et déontologique de l'IA publique[121], d'autant que les SIA publics font déjà l'objet d'un encadrement juridique (présenté en annexe 10 et résumé ci-après), issu de la sédimentation de multiples règles dont la cohérence d'ensemble reste à bâtir.

[119] Le projet de Système automatisé pour les fichiers administratifs et répertoire des individus (SAFARI), jamais déployé, visait à interconnecter plusieurs fichiers nominatifs.

[120] Ces contrôles sont aussi bien internes (contrôle hiérarchique, contrôle des autorités politiques elles-mêmes exposées au risque médiatique de l'illégalité, contrôle d'un organe délibérant, rôle des commissaires du gouvernement et de la tutelle…) qu'externes (contrôle du régulateur, le cas échéant dans le cadre d'obligations consultatives, contrôle des juridictions administratives, notamment en référé, et des juridictions financières, contrôle du Parlement…).

[121] Précisons à cet égard qu'il n'y a pas lieu de s'émouvoir, comme on le lit parfois (V. par ex. l'étude d'*AI Watch* sur l'IA dans le secteur public), d'un « conflit d'intérêts » entre l'État régulateur de l'IA et l'État utilisateur de l'IA. Il appartient à l'État, dans sa fonction normative, de définir un cadre permettant de concilier au mieux l'enjeu de l'innovation (publique et privée) et celui de la protection des droits et libertés. La même configuration se présente lorsque l'État définit le droit de la fonction publique ou de l'électricité.

Bref aperçu du cadre juridique actuel des SIA publics

Le droit applicable aux SIA publics résulte :

- Des règles, issues de la loi du 6 janvier 1978 et, le cas échéant, du RGPD, encadrant les **traitements de données à caractère personnel**, dès lors qu'ils utilisent de telles données.

S'agissant de la **prise de décision automatisée**, l'article 47 de la loi du 6 janvier 1978 interdit qu'une décision qui produit des effets juridiques à l'égard d'une personne ou l'affecte de manière significative soit prise sur le seul fondement d'un traitement automatisé, sauf, notamment, s'il s'agit d'une décision administrative individuelle et si, entre autres conditions, elle mentionne qu'elle a été fondée sur un traitement algorithmique. Une telle possibilité est en outre exclue pour les décisions entrant dans le champ de la directive « police-justice », les décisions de justice impliquant une appréciation sur le comportement d'une personne, ou encore les services de conciliation, de médiation judiciaire ou d'arbitrage. Enfin, l'article L. 4001-3 du code de la santé publique encadre strictement le recours aux SIA dans le champ de la prévention, du diagnostic et du soin.

- Des règles relatives au **droit à l'information** des usagers sur les traitements algorithmiques. Le doit à la communication des documents administratifs est consacré par l'article 15 de la Déclaration des droits de l'Homme et du citoyen de 1789 et précisé par les dispositions du code des relations entre le public et l'administration (CRPA).

L'article L. 311-2 du CRPA prévoit notamment le droit d'obtenir communication de ces documents, par exemple les codes-sources, sauf si est en jeu un secret protégé par la loi ou la protection de la vie privée, ce qui pourrait faire échec à la communication de tout document permettant de retrouver des données personnelles. S'agissant des destinataires des décisions individuelles, le CRPA prévoit une information renforcée : ces décisions doivent mentionner qu'elles ont été prises sur le fondement d'un traitement algorithmique et, si l'intéressé en fait la demande, l'administration doit lui communiquer les règles définissant ce traitement et les principales caractéristiques de sa mise en œuvre, les informations communiquées devant l'être sous une forme intelligible. Ces règles générales n'excluent pas des règles spéciales plus restrictives applicables à certains traitements.

- Des **normes supérieures, constitutionnelles ou conventionnelles**. S'agissant du bloc de constitutionnalité, la jurisprudence du Conseil constitutionnel est peu fournie, mais mobilise, ce qui est peu étonnant, le droit au respect de la vie privée ou encore la liberté d'expression et de communication. S'agissant des exigences conventionnelles, le droit positif repose essentiellement sur la CEDH et la Charte des droits fondamentaux de l'Union européenne, et la jurisprudence des cours européennes ne porte que sur des traitements de données personnelles. Comme le Conseil constitutionnel, la CEDH opère un contrôle de nécessité et de proportionnalité des atteintes aux droits fondamentaux en cause.

La dernière spécificité importante des SIA publics à prendre en compte lors de la définition des obligations pesant sur l'administration tient à la **contrainte de ressources**. Là où un opérateur privé peut financer les SIA qu'il déploie en anticipant sur les recettes qu'ils sont susceptibles de générer, l'administration, elle, est plus encline à anticiper les coupes budgétaires qu'elle risque de subir. Elle n'est d'ailleurs guère incitée par les raisonnements budgétaires classiques à développer des SIA permettant de dégager des économies, car il y a toutes les chances que son budget soit revu à la baisse en conséquence. Les contraintes d'ordre juridique ou déontologique, y compris sur un plan purement procédural ou documentaire (à l'instar des obligations prévues par la proposition de règlement européen pour les SIA à haut risque), peuvent ainsi, selon le cas, dissuader les administrations ou les placer dans l'incapacité matérielle de lancer un SIA, ou les inciter, faute de ressources humaines suffisantes, à externaliser plus que l'intérêt général ne le commanderait[122].

Ces préalables étant posés, la construction de l'IA publique de confiance, qui ne peut être qu'une démarche évolutive et itérative compte tenu de la rapidité des progrès techniques et de l'émergence de nouvelles problématiques sociales, doit reposer sur le triptyque suivant.

En premier lieu, il convient, sur la base des risques recensés et documentés, de définir un **nombre limité de principes généraux** qui doivent structurer la réflexion stratégique (nationale, ministérielle, locale…) sur la place que les SIA doivent occuper dans l'action publique et sur leurs effets systémiques sur la société et les institutions (au-delà des effets directs sur les individus), et dont découlent des **exigences opérationnelles** devant guider la décision de principe de recourir à un système d'IA donné pour accomplir une tâche déterminée et les choix de conception et de déploiement de ce système.

En deuxième lieu, il y a lieu de déterminer les exigences, de fond et procédurales, qui méritent une **consécration juridique pleine et entière** (« droit dur ») en raison de leur importance dans la protection des droits et libertés et de la force des garanties que les pouvoirs publics entendent donner aux parties prenantes. Ces règles de droit doivent être respectées par l'administration. Les autres exigences constituent des **repères d'ordre déontologique**, dont les administrations doivent s'efforcer de tenir compte pour renforcer l'acceptabilité sociale des systèmes, de façon proportionnée à l'ampleur des risques et en arbitrant les contradictions qui peuvent surgir entre elles. Ils peuvent utilement donner lieu à l'édiction d'actes de droit souple : le Conseil d'État recommande à cet égard d'édicter des lignes directrices de l'IA publique de confiance.

Enfin, l'effectivité des garanties posées suppose l'existence d'un **droit au recours effectif**, portant aussi bien sur la légalité des décisions relatives aux SIA que sur la mise en jeu de la responsabilité de l'administration.

La présente partie développe chacun de ces trois points.

[122] V. 4.1.1. sur la problématique de l'externalisation.

3.1. Définir les principes généraux de l'IA publique de confiance

Au regard des risques connus et documentés à date, il est proposé de retenir sept principes structurants de l'IA publique de confiance :

- La primauté humaine
- La performance
- L'équité et la non-discrimination
- La transparence
- La sûreté (cybersécurité)
- La soutenabilité environnementale
- L'autonomie stratégique

3.1.1. Primauté humaine

Le principe de primauté humaine renvoie à l'idée qu'un humain doit veiller à ce que les systèmes d'IA fonctionnent à son bénéfice, se porter garant de leur bon fonctionnement et répondre des conséquences de ses dysfonctionnements. Derrière la machine, il y a toujours l'homme, et d'abord lui. Pas plus qu'il n'est une fin en soi, mais seulement un moyen au service de l'action humaine, un SIA n'est pas un processus extérieur autonome, mais toujours un élément du service, sous la responsabilité de celui-ci et des agents compétents.

Ce principe suppose aussi que chacun puisse tirer bénéfice de l'IA, sur tout le territoire et dans toutes les situations. À cet égard, la mise en œuvre des systèmes d'IA, par les moyens dégagés, doit permettre d'apporter une aide aux publics les plus éloignés du numérique et de leur offrir l'accès à celui-ci.

Ce principe implique enfin une exigence de bénéfice humain (1), une supervision du système par l'humain (2), une gestion de la dépendance humaine à l'égard du système et de la réversibilité du processus d'automatisation (3) et l'assimilation des dysfonctionnements du système à l'erreur humaine (4). Le corollaire de ce principe est la possibilité de mise en jeu de la responsabilité de la personne (physique ou morale) qui maîtrise le SIA ayant causé un dommage, et non celle du système en tant que tel[123].

1. Bénéfice humain

L'administration doit être en capacité de démontrer que son choix de recourir à un système d'IA et les modalités de fonctionnement qu'elle a définies sont guidés par la préoccupation d'apporter un bénéfice à la collectivité humaine et, dans l'idéal, le plus grand bénéfice possible – ce qui a aussi pour conséquence pour l'administration de devoir garantir son accessibilité à tous[124].

a/ Cette exigence n'est respectée que si le SIA public répond à une **finalité d'intérêt général avérée** et que **l'ingérence dans les droits et libertés fondamentaux qui**

[123] V. 3.3.2. au sujet de la mise en jeu de la responsabilité de l'administration.

[124] V. 3.1.3. sur l'accessibilité et l'universalité des SIA.

résulte de sa mise en service n'est pas disproportionnée au regard des bénéfices qui en sont attendus. Encore faut-il déterminer le champ de ce qui est socialement acceptable.

Le principe même d'une liste de SIA interdits par principe, telle qu'elle figure à l'article 5 de la proposition de règlement, ne relève pas de l'évidence. Ce parti rompt avec la logique du droit européen de la protection des données à caractère personnel, qui repose pour l'essentiel sur une poignée de grands principes, notamment celui qui exige que le traitement de telles données repose sur une finalité « légitime », et qui invite pour le surplus les responsables de traitement et les autorités nationales de contrôle, sous le contrôle des juridictions, à procéder à la balance des intérêts en présence et à des tests de proportionnalité (sur l'ampleur de la collecte, sur la durée de conservation…), sans interdire de manière générale et absolue certains traitements en particulier. Autrement dit, dans cette matière, le champ de l'interdiction se déduit de l'application des principes généraux et non de l'identification *a priori*, par le législateur, de pratiques prohibées.

De manière générale, il convient d'**accueillir avec prudence l'idée d'interdictions propres aux SIA**, au regard de l'objectif de neutralité technologique, de la préoccupation de ne pas « diaboliser » ces systèmes et de la nécessité de ménager l'avenir. Il ne faut pas perdre de vue, en outre, que les règles de droit commun permettent déjà d'interdire des usages socialement inadmissibles des SIA.

A titre d'exemple, on peut sérieusement douter que le RGPD autorise le déploiement des systèmes de « notation sociale », visés par l'interdiction figurant au c) du paragraphe 1 de l'article 5 de la proposition de règlement sur l'IA, qui supposent des traitements de données à caractère personnel. La prohibition résultant du RGPD serait d'ailleurs plus large puisqu'elle s'appliquerait aussi à des dispositifs, automatisés ou non, ne recourant pas à un SIA et qui, s'ils ne permettraient pas un contrôle social aussi poussé que celui qu'on observe en Chine, aurait néanmoins des effets dévastateurs sur les libertés publiques. De même, il n'est pas certain qu'il soit nécessaire d'interdire explicitement le recours aux SIA pour servir certaines pratiques (manipulation subliminale et abus de faiblesse) comportant un risque de préjudice physique ou psychologique et qui sont d'ores et déjà pénalement sanctionnés dans les Etats membres. Quant à la reconnaissance biométrique en temps réel dans les lieux publics, il est sûr qu'elle appelle un encadrement des plus stricts, national ou européen ; mais il est plus douteux qu'il faille faire de leur interdiction la règle, et de leur déploiement l'exception, et, en tous les cas, il apparaît inopportun, au regard des enjeux d'intérêt général (retrouver un enfant enlevé ou perdu, identifier un terroriste notoire dans la foule d'un grand évènement…), de les interdire purement et simplement, comme le suggère le comité européen et le contrôleur européen de la protection des données[125]. Une telle interdiction ne saurait en tous les cas reposer sur les errements bien connus de systèmes de reconnaissance faciale, notamment aux Etats-Unis, imputables à une maturité insuffisante des technologies, à un déploiement hâtif ou à un usage dévoyé.

[125] Avis conjoint 5/2021 du 18 juin 2021.

Il importe en outre que ces interdictions de principe ne brident pas la **recherche scientifique** et les activités tendant au développement de systèmes permettant précisément de détecter et de combattre ces usages dévoyés de l'intelligence artificielle (ce qui peut supposer d'en concevoir à des fins de test). Le « texte de compromis » en cours de discussion prévoit utilement d'exclure du champ du texte les systèmes spécifiquement développés et mis en service pour les seuls besoins de la recherche scientifique et du développement, et la sanctuarisation des activités de recherche et développement en amont de toute mise sur le marché ou mise en service.

Au-delà des interdictions de principe que le règlement pourra comporter, c'est au cas par cas qu'il appartiendra à l'administration, sous le contrôle du juge, d'apprécier si les atteintes portées aux droits et libertés sont justifiées et proportionnées, en tenant dûment compte des spécificités des SIA publics. En effet, on ne peut pas, dans cet exercice de balance, traiter un SIA déployé pour les besoins d'un service public administratif de la même façon qu'un système poursuivant une finalité commerciale. Le premier poursuit un but d'intérêt général, qui fait le plus souvent écho à des exigences constitutionnelles, y compris, de façon transverse, l'objectif de bon usage des deniers publics qui découle des articles 14 et 15 de la Déclaration de 1789. Son déploiement présente donc une légitimité particulière qui peut justifier une ingérence plus forte dans les droits et libertés.

b/ Une autre implication de l'exigence de bénéfice humain, souvent oubliée, consiste à veiller à minimiser les atteintes que sa conception même peut causer aux droits et libertés fondamentaux, à la santé et à la sécurité. Si le **devoir de vigilance** qui s'impose aux grandes sociétés[126] n'est pas juridiquement applicable aux administrations, il est certain qu'elles ne peuvent se désintéresser des conditions de fonctionnement des prestataires extérieurs, des sous-traitants et des fournisseurs.

À cet égard, l'une des problématiques propres aux SIA **porte sur l'annotation des données** dans le cadre de l'apprentissage supervisé, qui peut être externalisé auprès d'entreprises susceptibles de recourir elles-mêmes à la sous-traitance, le cas échéant en cascade, jusqu'à des opérateurs étrangers qui mobiliseront une main d'œuvre peu qualifiée pour exécuter une tâche de labellisation répétitive dans des conditions de travail parfois très éloignées des standards nationaux et mettant en péril la santé ou la sécurité des salariés. Les collectivités publiques doivent faire preuve de discernement dans le choix des opérateurs, de rigueur dans l'encadrement du recours à la sous-traitance, et d'intransigeance quant aux conditions dans lesquelles ces tâches sont effectivement accomplies, y compris en recourant au réseau diplomatique pour qu'il soit procédé par les autorités compétentes à des contrôles inopinés.

c/ Enfin, l'exigence de bénéfice humain suppose de s'extraire de l'immédiateté pour analyser les **conséquences à long terme du système**.

---

[126] Art. L. 225-10-4 du code de commerce résultant de la loi n° 2017-399 du 27 mars 2017. Le devoir de vigilance désigne l'obligation de prendre les mesures raisonnables propres à identifier les risques et prévenir les atteintes graves envers les droits humains et les libertés fondamentales, la santé et la sécurité des personnes ainsi que l'environnement, résultant des activités d'une entité. Cette obligation est actuellement circonscrite aux grandes sociétés et ne concerne donc, au sein du secteur public, que certaines entreprises publiques.

Comme toutes les grandes révolutions technologiques, le numérique n'est pas seulement un moyen permettant d'effectuer certaines tâches ; environnement quotidien de la très grande majorité des citoyens, il façonne une nouvelle société. Dans ce contexte, une question fondamentale posée par l'avènement d'un monde assisté par l'intelligence artificielle est celle de la dépossession des citoyens de leur capacité d'agir. La garantie d'une IA inclusive en est une condition, mais elle ne résout pas le défi de transformer l'usage de SIA en pratiques capacitantes, fondées sur une culture numérique partagée et la possibilité pour l'humain d'analyser le changement à l'œuvre, pour, le cas échéant, le contester.

En réalité, le boom de l'IA pose la question de l'utilisation de l'intelligence humaine, et du **risque d'abaissement collectif de notre capacité à comprendre les règles qui régissent une partie de notre quotidien**. Or dès lors qu'une partie d'entre nous (par manque d'intérêt, de temps, de croyance dans la possibilité que cela soit utile) s'abandonne à la technologie sans la questionner, le risque est grand que se creuse davantage le fossé entre « ceux qui comprennent » et « ceux qui ne comprennent pas » (ou, peut-être plus dangereux encore, entre décideurs et ceux qui ne le sont pas).

De même, les réflexions sur l'impact de l'IA sur le travail montrent que ces technologies, que l'on pense qu'elles peuvent contribuer à l'avènement d'une société du temps libéré ou qu'elles conduiront inexorablement à l'effondrement du contrat social, bouleverseront les équilibres nationaux comme internationaux. A ce titre, il faut souligner le lancement en novembre 2021 d'un programme pluriannuel de 5 ans « LaborIA » au sein de l'Inria, financé par le ministère du travail, et en partie lié à la participation de la France au Partenariat mondial pour l'IA, qui vise à explorer les conséquences potentielles de l'IA sur « l'avenir du travail » en raison de l'automatisation d'une série de tâches.

L'enjeu est donc, comme pour toute forme de délégation d'action, non pas le recours accru aux technologies, mais la **capacité individuelle de chacun, et plus encore la capacité collective du corps social, à les penser et à en maîtriser le processus** ; autrement dit, à réfléchir.

Il faut dès lors souligner que l'emploi de l'IA par les administrations publiques commande, à l'échelle « macro-démocratique », l'émergence d'une culture numérique partagée et, à un niveau « micro-démocratique », doit se faire en **coproduction avec les usagers** (V. la quatrième partie de l'étude) qui doivent être associés à la fois au choix du recours à l'IA mais aussi (et surtout) au contrôle de l'utilisation de l'IA, dans un cadre leur permettant d'être non seulement informés mais également consultés (tant dans la phase de déploiement que, le cas échéant, en cas d'incident).

2. Supervision humaine

L'idée selon laquelle un SIA ne saurait complètement évincer l'humain et fonctionner en roue libre est l'une des plus évidentes et des plus consensuelles dans la réflexion éthique et déontologique sur ces systèmes. L'IA de confiance ne saurait être celle d'une confiance aveugle dans la machine.

En ce sens, la proposition de règlement européen entend soumettre les SIA à haut risque au principe du « contrôle humain » (*human oversight*), avec pour objectif de prévenir ou réduire au minimum les risques pour la santé, la sécurité ou les droits fondamentaux d'un SIA utilisé conformément à sa destination ou dans des conditions de mauvaise utilisation raisonnablement prévisible. L'humain doit notamment être en capacité d'appréhender complètement les capacité et limites du système, d'interpréter correctement les résultats qu'il produit, de surveiller le fonctionnement et de détecter les anomalies, et de décider d'interrompre son fonctionnement (« bouton d'arrêt ») – ce qui soulève la question de la réversibilité (V. *infra* 3.).

Mais la place respective qu'il y a lieu de reconnaître concrètement à la machine et à l'humain est plus indécise et débattue. On ne peut s'en tenir à l'équation simple selon laquelle davantage d'humain équivaudrait à davantage de garantie. La fiabilité de l'intervention humaine peut en effet s'avérer plus faible que celle de la machine pour l'exécution d'une tâche donnée. Sous l'effet des progrès accomplis dans ce domaine, le « regard de la machine » tend de plus en plus à surclasser celui du regard humain. La détection de pathologies dans des images issues d'examens radiologiques[127] ou l'identification par reconnaissance faciale sur la base d'un gabarit biométrique en fournissent de bonnes illustrations. Potentiellement contre-productive, l'intervention de l'humain peut aussi, selon les cas, donner l'illusion de la maîtrise et une fausse assurance aux usagers.

La question des modalités d'association de l'humain aux systèmes d'IA se pose à deux niveaux complémentaires :

- dans la réflexion stratégique sur le principe et le degré d'automatisation des activités d'une administration (est-il possible et souhaitable de confier telle tâche à la machine et dans quelle mesure ?) ;
- puis dans la conception de chaque SIA (quel rôle opérationnel pour l'humain dans l'utilisation du système ?)

Il existe un dégradé subtil dans le degré d'automatisation possible des tâches, celle-ci pouvant être **partielle** (sur une partie du processus ou certaines sous-tâches seulement), **conditionnelle** (doctrine administrative laissant le choix ou faisant obligation à l'agent de recourir au système ; validation systématique des décisions à partir d'un certain seuil de criticité ou possibilité de passer outre la décision de la machine ; obligation de confier

[127] Depuis 2019, les systèmes les plus performants ont dépassé les performances humaines dans la détection du cancer de la peau, du cancer de la prostate, en repliement de protéines et en diagnostic de la rétinopathie diabétique (S. Russel et P. Norvig, *Intelligence artificielle - une approche moderne*, Pearson, 4e édition, p. 25). V. sur ce sujet les cas d'usage en annexe 9, fiche n° 8.

à un humain le traitement des recours administratifs contre les décisions automatisées…) ou **supervisée**, soit **par l'humain**, pendant le fonctionnement ou au vu des résultats produits (identification et correction des erreurs par réentraînement du modèle, redéfinition des cas dans lesquels la décision est automatisée…)[128], soit par une **autre machine** programmée à cette fin par l'humain (le système d'IA pouvant aussi servir à vérifier la cohérence ou la rectitude de projets de décisions élaborés par un humain avant validation définitive par ce dernier). Le choix du degré et des modalités adéquates d'automatisation est toujours affaire d'espèce.

La distinction la plus classique est celle qui oppose la prise de décision automatisée et l'outil d'aide à la décision (ou la prise de décision « assistée »).

Dans le premier cas, la machine prend elle-même la décision. Si elle soulève peu de difficulté lorsque cette décision procède de l'application mécanique de règles simples à des faits objectifs (par exemple, l'attribution d'une aide si les ressources du demandeur sont inférieures à X euros, qui suppose de comparer les ressources déclarés au plafond réglementaire), ce qui rejoint largement le champ de la compétence liée de l'administration, il en va très différemment lorsqu'il y a lieu de porter une appréciation sur la satisfaction d'une condition plus « subjective », de mettre en œuvre un faisceau de critères ou d'indices à pondérer, et, *a fortiori*, lorsque l'administration dispose d'un pouvoir discrétionnaire. L'automatisation de la prise de décision reste techniquement envisageable et (sauf exceptions) juridiquement possible sous certaines conditions, mais elle suppose, d'une part, d'être en capacité de formaliser clairement les ressorts de la décision c'est-à-dire, en quelque sorte, de traduire la doctrine administrative en langage de programmation[129] et, d'autre part, d'être conscient de deux effets pervers : d'une part, un **effet de toise**, consistant à ramener l'ensemble des situations individuelles à des catégories prédéterminées nécessairement imparfaites, au risque de négliger des situations particulières qui justifieraient une application adaptée de la règle ; d'autre part, un effet de **fossilisation de la doctrine administrative**, si les paramètres du système-expert ne sont pas actualisés ou, en apprentissage machine, si l'on se

[128] Les lignes directrices du HLEG distinguent schématiquement le modèle « humain dans la boucle » (*human in the loop*), dans lequel l'humain intervient dans chaque cycle de décision du système (lequel ne peut pas prendre la décision de façon totalement autonome) ; le modèle « humain sur la boucle » (*human on the loop*), qui se réfère à la capacité d'intervention humaine pendant le cycle de conception du système et la surveillance de son fonctionnement (la machine peut prendre les décisions elles-mêmes mais l'humain peut intervenir pour s'y opposer) ; et le modèle « humain aux commandes » (*human in command*), dans lequel l'humain supervise l'activité globale du SIA et ses impacts et peut décider quand et comment utiliser ou ne pas utiliser le système dans une situation donnée. Cette distinction n'est pas reprise par la proposition d'AI Act. On observera que l'expression « *human in the loop* » est plus souvent utilisée par la communauté de l'IA pour désigner le rôle de l'humain dans l'évaluation de la performance du modèle algorithmique en apprentissage machine et son amélioration par une boucle de rétroaction.

[129] Comment créer un modèle algorithmique qui déciderait qu'une construction faisant l'objet d'une demande de permis de construire porterait atteinte à l'intérêt ou au caractère des lieux avoisinants ?

contente d'entraîner le modèle avec des données historiques, ce qui conduit mécaniquement à perpétuer l'existant.

Dans le cas de la prise de décision assistée, les informations et connaissances produites grâce au SIA sont ou peuvent être prises en compte par l'agent chargé de prendre une décision (au sens large du terme : il peut s'agir aussi bien de l'édiction d'un acte juridique que de l'émission d'une position, l'adoption d'un comportement, d'un choix de politique publique…). La machine peut intervenir en amont (par exemple en fournissant des informations utiles à la prise de décision publique, en formulant des suggestions ou en émettant une alerte donnant lieu à une levée de doute humaine) et/ou en aval (en effectuant un « contrôle qualité » des positions qu'envisage de prendre l'agent, en l'avertissant d'une potentielle erreur).

Claire sur le papier, la dichotomie entre ces deux modes d'intervention du SIA est plus complexe à appréhender en pratique.

Il importe en effet de ne pas s'arrêter au fonctionnement nominal du SIA et à l'existence formelle d'une étape de validation humaine, mais de tenir compte de deux biais connexes qui peuvent subrepticement muer un outil d'aide à la décision en outil de décision automatique : d'une part, le **biais d'ancrage**, qui se manifeste par l'incapacité ou la difficulté de l'agent à se départir de la « première impression » que lui fournit le système, s'il intervient en amont ; d'autre part, et surtout, le **biais d'automatisation**, qui se définit comme la tendance de l'humain à faire davantage confiance aux résultats produits par la machine qu'à son propre jugement et, ainsi, à avaliser de façon systématique ou excessive les recommandations qu'elle formule. Les utilisateurs peuvent croire qu'ils n'ont pas besoin d'autres éléments d'appréciation de la réalité que ceux que produit le système. Ce dernier peut apporter l'illusion d'une représentation complète et intelligible du monde. S'il formule des recommandations ou des propositions, il peut donner le faux sentiment d'une démarche rationnelle par nature (alors qu'elle ne l'est qu'autant que les concepteurs du système l'ont été) et infaillible (alors qu'elle ne l'est jamais). Ces biais conduisent naturellement à un enfermement cognitif, que les études sur les réseaux sociaux ou les moteurs de recherche ont amplement confirmé. Le complotiste recevant des suggestions de lecture indexée sur ses propres convictions a chaque jour la confirmation qu'un complot est à l'œuvre. Il en va de même pour un SIA, par exemple, aiguillant les contrôles, qui confirme les catégories de fraudeurs en suspectant naturellement ceux déjà convaincus de fraude, en reproduisant le passé et en supprimant toute (laborieuse) curiosité. La capacité à détecter les erreurs en devient anesthésiée et le système prend en quelque sorte le contrôle de ceux qui sont censés l'utiliser.

L'un des principaux indices du biais d'automatisation est, évidemment, le taux d'approbation humaine des résultats de la machine. Un taux proche de 100% pourra laisser suspecter qu'en réalité, l'outil ne fournit pas un simple éclairage à l'agent et que ce dernier n'intervient que « pour ordre »[130]. L'approche quantitative ne suffit

[130] À ce sujet, l'étude annuelle 2014 du Conseil d'État sur *le numérique et les droits fondamentaux* indiquait déjà qu'il fallait éviter que « *les systèmes présentés comme relevant*

toutefois pas : si le SIA est extrêmement performant, il est logique qu'il soit très souvent suivi, le désaveu occasionnel de la machine par l'humain pouvant révéler l'existence d'une surveillance effective, pour prévenir les erreurs les plus flagrantes ou aux conséquences les plus fâcheuses. La qualification de prise de décision automatisée suppose donc une analyse globale des conditions de la prise de décision. Elle devra être retenue toutes les fois où l'intervention humaine s'avère purement formelle, parce que l'agent ne dispose pas du temps suffisant ou des compétences adaptées pour porter une appréciation autonome et prendre lui-même la décision ou parce que l'examen des déterminants de ses prises de position révèle qu'il ne tient compte d'aucun autre facteur que de la recommandation formulée par la machine. Une signature sur des milliers de décisions ou de relevés, apposée par un agent qui n'est jamais mis en capacité de la refuser à raison d'une analyse critique des opérations conduits n'est certainement pas une décision humaine ultime, mais le paravent d'une complète démission au profit du seul SIA.

La proposition de règlement européen prescrit aux fournisseurs de SIA à haut risque d'identifier les mesures de prévention du biais d'automatisation et de garantir aux utilisateurs la possibilité de décider de ne pas utiliser le système ou de ne pas suivre le résultat produit. Cette dernière exigence ne doit pas être confondue avec la question de savoir si, au sein de l'administration utilisatrice, chaque agent public en charge d'un process doit se voir reconnaître la possibilité de ne pas recourir au système et les conditions dans lesquelles il peut, le cas échéant, se séparer de ses préconisations : il s'agit là d'un choix de fonctionnement de l'administration, selon le degré d'automatisation et de souplesse qu'elle souhaite retenir.

Sur le plan opérationnel, la proposition de règlement prévoit que l'exigence de supervision humaine doit être garantie par des « *interfaces homme-machine appropriées* », des mesures intégrées par le concepteur avant la mise en service autant que possible, et des mesures identifiées par ce dernier dans la notice d'utilisation et qu'il appartient à l'utilisateur de mettre en œuvre.

Les **mesures techniques** touchent aux choix de conception du système et, notamment, aux souplesses laissées aux utilisateurs. A titre d'exemple – qui fera écho au vécu de milliers d'agents et d'usagers exaspérés – l'agent a-t-il la possibilité technique de passer outre un blocage informatique (« vous ne pouvez pas… », « erreur système »...), parce que le cas particulier qu'il a sous les yeux le justifie ? L'arbitrage doit naturellement intégrer la problématique de l'erreur et de la fraude. La formule de la notification d'avertissement (« attention… voulez-vous continuer ? »), avec la conservation des cas de « passer outre » dans le journal des opérations, constitue un bon compromis.

Les **mesures juridiques** renvoient notamment à l'encadrement du recours à la prise de décision automatisée, présenté dans l'annexe 6. On observera que, en l'état du droit, l'essentiel des obligations porte sur les SIA impliquant des traitements de

de « l'aide à la décision » soient en réalité presque toujours suivis et commandent la décision, l'intervention humaine n'étant alors qu'apparente » et recommandait d'indiquer dans un instrument de droit souple les critères d'appréciation du caractère effectif de l'intervention humaine (proposition n° 23).

données à caractère personnel. Or la prise de décision automatisée peut emporter des conséquences majeures indépendamment de la nature – personnelle ou non – des données traitées.

Les mesures organisationnelles peuvent être de diverses natures. Deux d'entre elles méritent une attention particulière.

D'une part, la **formation** qu'il convient de dispenser aux agents publics utilisateurs[131] doit notamment consister à aiguiser leur capacité à se distancier à bon escient des résultats produits par la machine et à ne pas laisser leur discernement s'émousser au fil du temps, au bénéfice d'une confiance excessive dans le système.

D'autre part, la **gouvernance** de la supervision humaine constitue un point-clé. Pour les SIA les plus sensibles, il peut être justifié de privilégier un regard collégial, pluridisciplinaire (expert métier, expert de la donnée, juriste…) et multipartite (y compris les organisations syndicales et les représentants des usagers s'il y a lieu) sur le fonctionnement de l'outil. A titre d'exemple, l'Union française pour la santé bucco-dentaire, avec l'appui de l'association Ethik-IA, expérimente à cette fin un « collège de garantie humaine ». Composé de représentants d'établissements de santé, de patients et de la société conceptrice de l'outil de reconnaissance d'images utilisé, et bénéficiant de l'expertise de chirurgiens-dentistes indépendants, ce collège se réunit trimestriellement pour procéder à des contrôles, aléatoires ou orientés par les remontées d'événements indésirables sur les résultats produits par l'outil, et apprécier la nécessité d'actions correctrices.

3. Non-dépendance et assistance humaine

a/ Les évolutions technologiques créent une très grande dépendance des activités humaines à l'égard des applications déployées. On imagine avec une certaine angoisse les effets économiques, sanitaires, psychologiques et sociologiques d'un « confinement numérique » destiné à prévenir la propagation à grande échelle d'un virus informatique destructeur, même pour quelques jours voire quelques heures.

Il est impératif, pour garantir en toutes circonstances la continuité du service public, d'élaborer un **plan de continuité des activités** en cas de défaillance, ponctuelle ou prolongée, du SIA, en définissant les conditions de délivrance des services par les agents, le cas échéant à l'aide d'outils de secours. Il importe également, dans la conduite des projets de SIA, d'envisager l'hypothèse d'un échec ou d'une obsolescence et d'organiser la réversibilité, fut-elle progressive.

Sur le plan à la fois psychologique et technique, il est aussi nécessaire de lutter contre un phénomène d'**assistanat numérique** qui aboutirait progressivement à l'atrophie des capacités cognitives humaines, à l'instar des effets des correcteurs automatiques sur le niveau de maîtrise de l'orthographe.

[131] V. 4.1.2. sur la formation des agents publics comme pilier de la stratégie de gestion des ressources humaines qu'appelle le déploiement de SIA publics.

D'ores et déjà, l'article 16 de la loi pour une République numérique fait obligation aux administrations de veiller à « l'indépendance » de leurs systèmes d'information, terme peu évocateur mais qui semble devoir être compris comme exigeant que « *le logiciel choisi ne rende pas l'utilisateur dépendant de ses fonctionnalités* »[132].

b/ Il est également impératif que, indépendamment même de tout dysfonctionnement, les destinataires des SIA, qu'il s'agisse des citoyens et entreprises ou des agents publics pour les systèmes déployés en interne, disposent en toutes circonstances d'un **interlocuteur humain** apte à régler le problème particulier qu'ils rencontrent. Il n'est rien – ou presque – de plus destructeur de confiance dans l'automatisation que « l'enfermement numérique » d'un usager confronté à un système informatique défaillant, incapable d'appréhender son cas particulier ou proposant des solutions kafkaïennes, comme celle qui consiste à inviter un usager en panne d'internet à consulter une aide en ligne. Il s'agit en outre d'une condition première de l'acceptabilité sociale des SIA, face à la critique – pas infondée – de l'affaiblissement des interactions humaines, emportées par la numérisation généralisée d'un quotidien déshumanisé.

L'article L. 111-2 du code des relations entre le public et l'administration garantit à toute personne le droit de connaître le prénom, le nom, la qualité et l'adresse administrative de l'agent chargé d'instruire sa demande ou de traiter l'affaire qui la concerne, sous réserve des exigences de sécurité. Mais il est douteux que ces dispositions s'appliquent au cas des affaires intégralement traitées par un système d'IA. La loi, qui consacre le droit des usagers de saisir l'administration par voie électronique (art. L. 112-8), pourrait également leur garantir le droit à une assistance humaine lorsque l'usage de cette voie est rendu obligatoire. Son exercice devrait être subordonné à la condition de l'existence d'un besoin objectif (en particulier pour ceux qui ne maîtrisent pas l'outil informatique – V. ci-dessous l'exigence d'accessibilité) et à la sanction des usages abusifs. Un tel droit n'impliquerait évidemment pas celui d'exiger de manière générale la possibilité, alternativement à une solution dématérialisée, de s'adresser à un guichet physique. Il s'agirait seulement d'une soupape.

4. Acceptabilité sociale de « l'erreur machine », forme indirecte de l'erreur humaine

L'un des défis que soulève le développement des SIA a trait à **l'aversion sociale exacerbée à l'erreur informatique** : l'erreur est humaine, mais la machine est priée d'être infaillible. Cette exigence rend insupportable tout dysfonctionnement, à commencer par celui d'un logiciel de traitement de texte qui ruine plusieurs heures d'efforts rédactionnels ou du système électronique embarqué dans une voiture qui refuse de démarrer, ce qui peut discréditer des outils pourtant globalement performants, le cas échéant bien davantage qu'un humain. Il n'est pas tout à fait certain que la société serait prête à troquer la conduite automobile humaine, qui cause en moyenne plus de 3000 décès chaque année en France, pour une conduite

[132] V. l'intervention de la secrétaire d'État au numérique lors de la séance publique du Sénat du 27 avril 2016, en première lecture du projet de loi.

automatisée qui ferait trente fois moins de victimes, par exemple (à supposer même qu'on parvienne à atteindre une telle performance)[133].

Ce biais cognitif, intolérant à l'erreur informatique en même temps qu'il idéalise le « bon vieux temps » où les humains étaient aux commandes, constitue un frein sérieux et en partie irrationnel au déploiement des SIA et, partant, une source d'opportunités manquées pour l'intérêt général. **Pour l'accomplissement d'une même tâche, il ne devrait pas être exigé davantage de fiabilité de la machine que des êtres humains**. C'est seulement lorsque la machine accomplit des actions surhumaines que des exigences spécifiques peuvent se justifier, sans toutefois perdre de vue le coût d'opportunité d'une renonciation au regard des bénéfices que l'humain en tire.

Un rapport apaisé et « normalisé » aux SIA, de ce point de vue, ne se décrète pas. Il peut néanmoins être servi par un discours plus clair sur le fait que « l'erreur de la machine » n'est, en réalité, rien d'autre qu'une erreur humaine. Si l'intelligence est artificielle, c'est parce qu'elle est créée par l'homme. Dépourvu d'autonomie réelle et de volonté propre, ce système est un artefact dont les conséquences dommageables sont toujours imputables à des humains, qu'ils soient identifiés comme tels ou qu'ils œuvrent au sein de personnes morales, et qui, fut-ce avec la plus parfaite bonne foi et une compétence irréprochable, n'en ont pas moins construit un ou des modèles à l'origine d'un dommage. À cet égard, la possibilité de mettre en jeu la responsabilité d'une personne physique ou morale en cas de dommage, étudiée au point 3.3[134], contribue, dans une logique compensatoire, à l'acceptabilité sociale de la défaillance de la machine.

3.1.2. Performance

Parmi les facteurs les plus destructeurs de confiance dans les outils numériques, donc dans les SIA, figure la **sous-performance**, qui se manifeste par un service qui ne fonctionne pas ou mal, ou qui ne s'acquitte pas de la mission qui lui est dévolue avec un niveau de qualité acceptable[135]. Trivialement, l'usager fait confiance à ce qui marche.

Ce principe est d'autant plus crucial pour l'administration en l'absence de concurrent vers lequel les mécontents pourraient se tourner. On voit bien, à l'inverse, à quel point, dans la sphère commerciale, la capacité de l'application à répondre voire à

[133] V. sur cette question l'avis du comité national pilote d'éthique du numérique, « *Le véhicule autonome : enjeux d'éthique* », adopté le 7 avril 2021.

[134] V. 3.3. sur le recours juridictionnel, clé de voûte de l'IA publique de confiance.

[135] En l'état, la proposition de règlement IA se borne à définir la performance d'un SIA comme « *la capacité d'un système d'IA à remplir sa destination* » (article 3, point 18). Elle assigne en outre une exigence « d'exactitude » aux SIA à hauts risques, qu'elle ne définit pas et dont on peut penser qu'elle sera précisée dans les normes harmonisées. Selon le référentiel de certification de processus pour l'IA publié par le Laboratoire national de métrologie et d'essais, la performance se définit comme le « *degré selon lequel un système ou composant accomplit ses fonctions désignées avec un ensemble de contraintes données, telles que la vitesse, la précision ou l'utilisation de la mémoire…* ».

devancer les besoins des utilisateurs et le soin apporté à « l'expérience utilisateur », y compris en termes d'ergonomie et de convivialité de l'interface, peuvent accroître l'acceptabilité voire susciter la demande de SIA, au prix, d'ailleurs, d'une certaine insouciance sur le respect des droits et libertés.

1. Les indicateurs de la performance

La performance d'un SIA s'apprécie en évaluant différents aspects du système. Sans entrer dans un détail technique superflu dans le cadre de la présente étude, on se bornera à citer les plus importants.

a/ L'**exactitude** du système peut être définie[136] comme l'absence d'écart entre la prévision, la décision ou l'action de la machine et la « bonne réponse » (soit objectivement, soit la réponse qu'un humain considère comme étant la bonne[137]). Le SIA doit être conçu pour se tromper le moins possible. Il en existe de multiples indicateurs (appelés « métriques » par la proposition de règlement), qui sont complémentaires, selon qu'on mesure la justesse du modèle (*accuracy*)[138], sa précision (*precision*)[139], sa sensibilité (« rappel »/ *recall*)[140] ou encore, combinant précision et rappel, la capacité du modèle à donner des prédictions justes et à éviter les fausses (moyenne harmonique - *F1 score*).

Le choix de l'indicateur dépend notamment des conséquences de l'erreur, selon son sens. Un « faux positif » pour un drone armé signifie que la machine a tué ou blessé une personne qui ne devait pas l'être, ce qui est inacceptable ; pour une alarme incendie, il s'agit d'une fausse alerte, ce qui est préférable à la non-détection d'un départ de feu. Dans le premier cas, le système doit être conçu pour éliminer les faux positifs (tirs injustifiés) : le taux de précision doit tendre vers 100% (aucun « innocent » ne doit être ciblé) ; dans le second, l'enjeu est de minimiser voire éliminer les faux négatifs (départ de feu non détecté pouvant entraîner des décès) : le taux de rappel doit tendre vers 100% (tout départ de feu doit être détecté)[141].

---

[136] La proposition de règlement européen utilise cette notion sans la définir.

[137] La mesure de l'exactitude est simple en présence d'une réponse binaire (est-ce la photo d'un chien ? oui/non) et toutes les fois où il existe une vérité absolue indiscutable. Elle peut être vertigineuse et éminemment polémique lorsqu'on s'interroge sur le point de savoir si une action est la bonne du point de vue sociétal, économique, moral, environnemental…

[138] Pourcentage total de bonnes réponses : (vrais positifs + vrais négatifs) / total des réponses. Ex.: dans 97% des cas, la machine a correctement classé des courriels indésirables dans les « spams » et des courriels souhaités dans la boîte de réception

[139] Pourcentage des bonnes réponses positives sur le total des réponses positives : vrais positifs / (vrais positifs + faux positifs). Ex. : dans 98% des cas où la machine a proposé de pseudonymiser un mot dans une décision de justice, il s'agissait bien d'un mot à pseudonymiser (nom propre…).

[140] Pourcentage des prédictions positives sur le total des cas objectivement positifs : vrais positifs / (vrais positifs + faux négatifs). Ex. : 80% des mots à pseudonymiser dans le jugement l'ont été par la machine.

[141] De manière générale, lorsque le résultat « positif » a des conséquences favorables pour les personnes, il est préférable d'avoir des faux positifs plutôt que des faux négatifs (mieux vaut donner à tort un titre de séjour ou une prestation que d'en priver à tort une personne).

Cet indicateur est crucial en matière d'apprentissage machine puisque la construction du modèle algorithmique est précisément fondée sur la minimisation de l'erreur (statistique). Les systèmes basés sur les règles peuvent être à l'origine d'erreurs liées à un paramétrage humain incorrect, mais, sous cette réserve, leur fiabilité est extrêmement élevée voire, dans certains cas, absolue en conditions normales d'utilisation.

Il importe de rapporter cette exactitude au **domaine de fonctionnement**, dont l'ampleur constitue également un critère de performance. Un modèle de traitement du langage naturel entraîné essentiellement avec du texte en langue française pourra donner des résultats extrêmement fiables pour des écrits en français mais très approximatifs en québécois et inutilisables en anglais.

b/ La **robustesse technique du système,** en conditions normales d'utilisation, recouvre une multiplicité d'exigences. Elle inclut la capacité à neutraliser les anomalies et éviter que des modifications accessoires ou parasites des données d'entrée (par exemple, pour une image, la luminosité, le contraste, l'occultation d'un objet par un autre…) aboutissent à des résultats erronés voire aberrants. Le modèle algorithmique lui-même doit aussi être reproductible dans son domaine de fonctionnement, et capable de maintenir sa performance dans le temps en dépit des variations de contexte et de données d'entrée[142]. L'article 16 de la loi pour une République numérique prescrit dans cet esprit aux administrations de veiller à la « pérennité » de leurs systèmes d'information, notion qui signifie, selon les travaux préparatoires, que « *les systèmes ne sont pas obsolètes trop rapidement et qu'il faut s'interroger sur la capacité de maintenance à plus long terme* ».

La robustesse suppose, à l'échelle du système ou du produit dans lequel il est embarqué, d'assurer la **disponibilité et la stabilité de l'outil**, ainsi que la **sécurité**, entendue comme la prévention des incidents résultant d'actes non malveillants ou de cas fortuits (comme la panne d'un capteur) et la résilience à ces évènements (à distinguer de l'exigence de sûreté développée ci-après, qui porte sur les actes de malveillance).

c/ Le **temps de réponse** du système est également, à l'évidence, un paramètre déterminant de la satisfaction ou de l'exaspération de l'utilisateur.

d/ De façon générale et synthétique, la **satisfaction des utilisateurs et destinataires du système**, évaluée sur la base d'une évaluation auprès de ces derniers (tel qu'un

Inversement, lorsque le résultat positif est préjudiciable à la personne, les faux positifs sont plus problématiques que les faux négatifs (sanctionner un innocent est pire que de ne pas sanctionner un coupable).

[142] On parle plus généralement du phénomène de « dérive des données » pour désigner le décalage qui peut se créer, progressivement ou brutalement, entre les données d'entraînement et les données de production, altérant la représentativité des premières à l'égard des secondes, à l'origine d'une dégradation de la performance et, potentiellement, d'une obsolescence du modèle. L'apparition d'une pandémie mondiale, qui modifie radicalement les comportements (au moins le temps de la crise) et le fonctionnement de l'économie, en constitue un exemple extrême.

« *net promoter score* »: sur une échelle de 1 à 10, recommanderiez-vous cette application ?), constitue un indicateur subjectif synthétique de la performance du système. Cette satisfaction dépendra notamment de la capacité du système à répondre aux besoins et de l'ergonomie de l'interface.

2. La définition du niveau de performance acceptable

Au regard de ces différents critères de performance, il appartient au responsable du système de fixer le **niveau minimal de performance acceptable**, qui devra être répercuté dans les contrats que l'administration passera avec ses fournisseurs de SIA. Il n'est pas possible de fixer un seuil de performance acceptable dans l'absolu. Plusieurs considérations doivent entrer en ligne de compte.

En premier lieu, **la performance doit être d'autant plus élevée que les conséquences de l'erreur sont graves**. Il va de soi qu'un SIA dont les dysfonctionnements pourraient occasionner des décès ou des blessures doit présenter un degré de fiabilité extrêmement élevé, pouvant par exemple justifier qu'il soit recouru à un système-expert (par un paramétrage des instructions) plutôt qu'à un SIA utilisant l'apprentissage machine.

En deuxième lieu, **le système se doit d'être d'autant plus performant que l'intervention humaine est limitée**. Entre un système de prise de décision automatisée et un système d'aide à la décision dont le taux d'erreur est connu de l'opérateur humain, mis ainsi à même de se distancier des résultats, les enjeux de performance ne sont pas comparables.

En troisième lieu, un soin particulier doit être apporté aux **systèmes qui interagissent directement avec les citoyens**, dont les évaluations négatives peuvent considérablement discréditer une démarche qui ne mérite pas nécessairement l'indignité.

En quatrième et dernier lieu, il convient de tenir compte, plus largement, du niveau d'acceptabilité sociale de ces systèmes, et de mettre en regard l'investissement financier consenti avec les résultats obtenus. Une performance médiocre à un coût financier élevé est la promesse d'une polémique qui fragilisera la confiance dans les SIA en général.

De manière générale, les citoyens sont en droit d'exiger que les SIA publics soient conçus et développés de manière à maximiser leur contribution à l'intérêt général. Le principe de base devrait être celui de l'amélioration de la qualité du service rendu ou, à tout le moins (si l'objectif principal est autre, comme, par exemple, permettre le redéploiement d'agents sur d'autres tâches), la **non-régression :** le recours à un SIA ne saurait entraîner une dégradation de la qualité du service public au regard de la prestation humaine ou d'une prestation numérique antérieure ne recourant pas à l'IA ou utilisant un autre SIA.

3. Les déterminants de la performance

Enfin, pour atteindre le niveau de performance fixé, il y a lieu d'identifier et de jouer sur les déterminants de la performance. S'agissant des modèles d'apprentissage machine, si le **choix de l'algorithme** a son importance, de même que les conditions de maintenance du système, il est évident que la **qualité et la pertinence des données d'apprentissage** constituent un facteur-clé (que retranscrit l'adage « déchets en entrée, déchets en sortie », c'est-à-dire que des données d'entraînement de mauvaise qualité produiront forcément des résultats médiocres ou mauvais). Un arbitrage peut être nécessaire, non seulement au regard des ressources dont dispose le responsable du système et des délais qui lui sont impartis pour déployer le système, mais aussi des autres principes déontologiques qui pèsent sur lui. Par exemple, la performance du système pourrait être accrue par le recours à certaines données d'apprentissage, mais le RGPD fait obstacle à ce qu'elles soient utilisées, ou l'utilisation de ces données introduirait des biais discriminatoires inacceptables.

À cet égard, l'exigence posée par l'article 10 de la proposition de règlement de la Commission selon laquelle les jeux de données d'entraînement, de validation et de test doivent être non seulement pertinents et représentatifs, mais aussi « *exempts d'erreurs et complets* », apparaît à la fois peu réaliste et en décalage avec la logique même de l'approche statistique, comme l'a relevé le rapport au COREPER de la présidence slovène sur l'état des travaux[143]. D'une part, il ne pourrait s'agir que d'une obligation de moyens, conduisant à sanctionner l'utilisation, délibérée ou par négligence caractérisée, de jeux de données manifestement défectueux, et non d'une obligation de résultat, d'autant que la proposition prévoit, pour les manquements à cette obligation, une amende maximale de 30 millions d'euros ou de 6% du chiffre d'affaires de l'auteur de l'infraction. D'autre part, l'exigence de complétude apparaît d'application délicate car on en ignore la base de référence, qui est potentiellement illimitée, et elle peut même heurter le principe de minimisation des données à caractère personnel résultant du RGPD, dans le cas où elle n'est pas nécessaire à l'atteinte des objectifs de performance. Enfin, l'existence d'erreurs peut rester sans incidence sur les performances du modèle, qui les traitera comme des anomalies statistiques et les neutralisera dans le calcul.

Le principe de performance peut être rapproché de l'exigence d'explicabilité exposée ci-après (3.1.4) à deux égards. D'une part, la complexité du modèle servira souvent le premier au détriment de la seconde, ce qui peut appeler un arbitrage[144]. D'autre part, l'explicabilité du modèle et, plus précisément, son interprétabilité, permettent à son concepteur de comprendre les raisons des erreurs et de les corriger afin d'en accroître la performance.

---

[143] Rapport n° 13802/1/21, 22 novembre 2021.

[144] S'exprimant à la conférence NeurIPS de 2017, Yann Le Cun tirait de son expérience dans l'industrie le constat que, lorsqu'on présentait deux modèles à un client, l'un précis à 99 % mais non explicable, l'autre explicable mais précis à 90 % seulement, l'intéressé préférait toujours le premier.

3.1.3. Equité et non-discrimination

La littérature sur l'usage éthique de l'intelligence artificielle recourt invariablement au concept d'équité (« *fairness* »). Un SIA ne saurait aboutir à un résultat inéquitable et, en particulier, à un résultat qui introduirait une discrimination injustifiée. Consensuel dans son énoncé, ce principe peut se voir conférer une portée très différente selon le contexte culturel, les conceptions philosophiques et politiques et les sensibilités individuelles. Trois exigences s'en évincent : la nécessité pour les concepteurs du système d'opérer un choix parmi les différentes acceptions de l'équité et de le formaliser ; la nécessité de prévenir les biais discriminatoires involontaires ; et la nécessité de garantir en toutes circonstances l'accessibilité et l'universalité des systèmes vis-à-vis des usagers.

1. Le choix de l'équité

L'action publique, y compris lorsqu'elle emprunte la voie d'un SIA, est soumise au respect du principe d'égalité devant la loi et ses déclinaisons (comme l'égal accès aux emplois publics). Mais celui-ci ne commande pas absolument le choix de l'équité à assigner à un système. D'une part, il n'exige pas, en droit interne, de traiter différemment des personnes placées dans des situations différentes, contrairement au principe général du droit de l'Union de non-discrimination. D'autre part, les différences de situation et les motifs d'intérêt général qui peuvent justifier des traitements différenciés offrent, à la faveur du contrôle distant du juge administratif (qui se borne à censurer les disproportions manifestes), d'importantes latitudes aux concepteurs des systèmes. Sauf à laisser ce choix au hasard, ces derniers doivent nécessairement formaliser dans le paramétrage de l'outil la vision de l'équité qu'ils entendent promouvoir.

Or il existe de multiples conceptions de l'équité, qui relèvent de choix largement politiques et dans lesquels le projet de règlement européen ne s'immisce d'ailleurs pas explicitement. Schématiquement, on distingue l'**équité collective**, qui vise l'égalité de traitement de deux ou plusieurs groupes distincts (les femmes et les hommes ; les personnes quelle que soit leur origine ethnique ou leur orientation sexuelle…), et l'**équité individuelle**, qui vise à traiter de façon équivalente deux personnes se trouvant dans la même situation, au regard des critères objectifs et rationnels qui déterminent la décision à prendre. Le plus souvent, ces deux conceptions ne sont pas conciliables : pour atteindre l'équité de groupe, il est en général nécessaire de traiter plus avantageusement des personnes relevant du groupe le plus défavorisé, dans une logique de **discrimination positive**.

Au sein même de chacune de ces catégories, l'équité peut se concevoir de façon très différente. On s'en convaincra en prenant l'exemple d'un outil de sélection à l'embauche. L'objectif du SIA peut consister à assurer une stricte égalité de résultat (si 20 postes sont ouverts à l'embauche aux hommes et aux femmes, 10 postes devront être confiés à des hommes et 10 à des femmes, quelle que soit la répartition des candidats), à garantir que la probabilité d'embauche est la même pour les hommes et les femmes (si 60 hommes et 40 femmes postulent pour 20 postes, il faut embaucher 12 hommes et 8 femmes – soit 20% des candidats à chaque fois), à faire en sorte que,

parmi les candidats qualifiés, les chances d'être embauché soient les mêmes, qu'on soit un homme ou une femme[145] ou que, parmi les candidats non qualifiés embauchés (à tort), la proportion d'hommes et de femmes soit identique, c'est-à-dire que le taux d'erreur est le même pour les deux catégories[146], ou encore que la proportion d'hommes et de femmes embauchés soit identique tant parmi les candidats qualifiés embauchés que parmi les candidats non qualifiés embauchés[147].

L'équité individuelle peut quant à elle être recherchée de deux manières : soit, en « positif », en définissant de façon exhaustive et non discriminatoire les critères de comparaison entre les personnes (par exemple, le niveau de diplôme, l'expérience et la motivation des candidats à l'embauche) ; soit, en « négatif », en neutralisant les critères non pertinents et discriminatoires (par exemple, le genre, la couleur de peau…) de telle sorte que le modèle ne les prenne pas en compte.

Il est impératif que le système d'IA repose à cet égard sur un choix conscient, concerté, argumenté et assumé (en sus d'être transparent – V. *infra*).

S'il est ainsi nécessaire que l'objet du SIA ne soit pas inéquitable, il importe aussi qu'il n'ait pas cet effet subreptice, ce qui renvoie à l'exigence de non-discrimination algorithmique.

2. La non-discrimination algorithmique

La non-discrimination constitue une exigence centrale du principe d'équité. Elle implique **d'identifier, de prévenir et d'éliminer les biais discriminatoires prohibés** qui peuvent survenir de façon subreptice lors de la conception et la mise en service d'un SIA. Elle est largement évoquée dans les considérants du projet de règlement européen, et l'est plus discrètement dans le corps de ce dernier à travers la notion générique de « biais ».

La principale famille de risques concerne les **biais affectant les données** initialement employées. Une donnée n'est jamais un fait totalement objectif. Les données expriment non « la réalité » ou la « vérité absolue », mais viennent schématiser, de façon abstraite, la réalité telle que nous nous la représentons. Ainsi, l'existence même d'une donnée, loin d'être seulement factuelle, matérielle et objective, est le produit d'un arbitrage humain, qui peut être conditionné par des considérations culturelles, morales, politiques ou économiques – tout autant de considérations sujettes aux préjugés, conscients ou inconscients. Il en va de même du traitement de la donnée et, en particulier, de son annotation (pour les SIA basés sur l'apprentissage supervisé). Ces opérations sont tributaires du contexte, de la mentalité et des valeurs

[145] « *Equal opportunity* » (égalité des chances), qui consiste à garantir que les décisions (ou prévisions) correctes du système profitent de la même façon aux deux groupes.

[146] « *Predictive equality* » (équité prédictive ou équivalence d'erreur), qui consiste à faire en sorte que le système ne se trompe pas plus (ni moins) lorsqu'il s'agit d'une femme que lorsqu'il s'agit d'un homme.

[147] « *Equalized odds* » (équité de sort), qui permet de tenir compte aussi bien des « vrais positifs » (candidats qualifiés embauchés) que des « faux positifs » (candidats non qualifiés embauchés à tort).

des intervenants, d'une organisation sociale et des préjugés collectifs, plus ou moins conscients. Le modèle algorithmique entraîné à l'aide des données disponibles est, en quelque sorte, un reflet du passé et une reproduction synthétique de la structure mentale et des raisonnements de celles et ceux qui les ont produites.

Ce risque existe, naturellement, pour les systèmes basés sur les règles[148]. Mais **il est particulièrement aigu pour les modèles basés sur l'apprentissage machine**, où la pondération des facteurs qui déterminent la décision est fixée par la machine elle-même au regard de l'expérience humaine passée. Or, lorsque nous demandons à un SIA de travailler sur ces données, il reproduit – « naïvement », sans conscience de le faire – l'ensemble des particularités implicites ou explicites issues de ces données, et ce d'autant plus qu'il n'a pas accès au sens, à l'intention, et moins encore au doute… Tous les biais, réprimés, combattus, mais existants, dans une société sont alors reproduits et systématisés.

Deux exemples bien connus viennent illustrer comment un SIA peut, hors de toute intention de ses créateurs, aboutir à des résultats manifestement discriminatoires :

- Si un modèle de présélection de candidatures pour l'accès à un emploi est entraîné sur la base de données résultant de décisions humaines discriminatoires (en ne recrutant pas ou peu de femmes, nonobstant leurs compétences), il risque fort de produire des résultats eux-mêmes discriminatoires (en faisant du sexe du ou de la candidate un critère plus déterminant que les compétences). C'est ce qu'il s'est passé, au milieu des années 2010, avec le logiciel d'appui au recrutement utilisé par Amazon, finalement abandonné ;
- Ces biais sont également à l'œuvre dans le domaine de la reconnaissance d'images ou de reconnaissance faciale. La plupart de ces systèmes connaissent ou ont connu des dérives racistes manifestes.

Ce risque est d'autant plus redoutable qu'un facteur de discrimination peut surgir des données de façon indirecte, lorsque **le choix d'une variable cible ou d'une caractéristique *a priori* neutre peut s'avérer être une source de discrimination**. Ainsi, l'adresse postale peut fournir une certaine probabilité que la personne soit issue de l'immigration, en fonction de la concentration de cette catégorie dans le quartier de résidence ; en analysant les décisions passées, le modèle est susceptible de déceler une corrélation entre l'adresse et le sens de la décision, ce qui peut aboutir non seulement à des décisions aberrantes, mais aussi à des résultats discriminatoires au détriment des personnes issues de l'immigration, alors même que l'origine des personnes n'est pas renseignée. Il en va de même lorsqu'une variable pertinente est omise, soit parce qu'elle n'est pas disponible et est donc absente du jeu de données d'entraînement de l'algorithme, soit qu'elle n'a pas été identifiée par le programmeur, lui-même empreint de biais cognitifs (biais de confirmation, biais de *statu quo*, biais de représentativité, biais de corrélation illusoire, etc.).

[148] Étude de F. Zuiderveen Borgesius, *Discrimination, intelligence artificielle et décision algorithmiques*, Conseil de l'Europe, 2018.

Il y a donc lieu pour les pouvoirs publics de **mettre en place un système de gestion des risques** de nature à prévenir la survenance ou la reproduction de biais discriminatoires, et ce **durant toute la vie du SIA** (pendant les phases de test du modèle, lors du passage à l'échelle, dans le cadre de ré-entraînements réguliers, etc.).

En premier lieu, les risques de discrimination directe ou indirecte engendrés par le recours à un SIA doivent faire l'objet d'une **évaluation préalable**, permettant d'identifier l'ensemble des facteurs susceptibles de créer des biais, menée dans un cadre pluridisciplinaire, associant des agents métiers et des experts techniques, ainsi que des représentants de la société civile (représentatifs, notamment, des utilisateurs cibles mais aussi des groupes victimes de discriminations systémiques) et des chercheurs. L'article 11 du règlement européen et son annexe IV prévoient, à ce titre, que les SIA classés « à haut risque » sont accompagnés, avant leur mise sur le marché ou leur mise en service, d'une documentation comportant des informations détaillées sur « *les résultats imprévus prévisibles et les sources de risques pour la santé et la sécurité, les droits fondamentaux et la discrimination, compte tenu de la finalité prévue du système d'IA* ». Cette évaluation devrait, en outre, être actualisée à chaque phase de déploiement du SIA (entraînement, industrialisation, fonctionnement « de croisière ») afin de s'assurer qu'aucun risque n'a été omis et qu'aucun nouveau risque n'apparaît au fur et à mesure de ce déploiement, le cas échéant par l'étude de scenarios contrefactuels susceptibles d'identifier les variables déterminantes de l'*output* du système (ex. pour l'octroi d'un prêt bancaire, les revenus du demandeur, ce qui semble normal, mais *quid* si le SIA prend en compte le lieu de résidence ?).

En deuxième lieu, s'agissant des données, une attention particulière doit être portée à l'**analyse critique de données d'apprentissage**, en raison de leur rôle structurant sur le fonctionnement du système. Leur pertinence matérielle ne suffit pas : encore faut-il que, dans toute la mesure du possible, tout biais soit écarté, ou au moins identifié. À cet égard, deux types de correctifs peuvent être identifiés[149] :

- Les **pistes statistiques** portent sur les données servant à l'entraînement et au fonctionnement du SIA. Lorsqu'un jeu de données est incomplet, les données manquantes peuvent être reconstituées par un modèle statistique, si tant est que ce modèle n'est pas lui-même biaisé. L'expérience acquise sur les biais incite également à **privilégier un principe de correction** de ceux-ci par reprise de l'apprentissage sur des données permettant d'y remédier, plutôt que par des solutions techniques (des rustines) qui privent le SIA de son objectivité et cachent son dysfonctionnement plus qu'ils ne les corrigent, en introduisant d'autres biais. Cette opération de « débiaisement » peut aller jusqu'à introduire des données d'entraînement synthétiques (artificielles), créées pour les besoins de la cause. Elle peut justifier le traitement de données à caractère personnel sensibles (par exemple, relatives à l'origine ethnique, afin de s'assurer de la représentativité du jeu de données d'entraînement), comme le prévoit expressément le projet de règlement européen.

[149] Étude de P. Bertail, D. Bounie, S. Clémençon, P. Waelbroeck. *Algorithmes : Biais, Discrimination et Équité*, Telecom Paris Tech, 2019.

- Les **pistes algorithmiques** (ou de procédure) portent, comme leur nom l'indique, sur la programmation d'algorithmes auxquels il est explicitement demandé de ne pas discriminer certains groupes. Il s'agit donc d'une contrainte dans l'arbitrage, par le SIA, entre différentes variables, et de faire de la non-discrimination une variable de premier rang. Ce type de contrainte (la règle que l'IA doit toujours suivre) est appelée « liste blanche » et s'oppose aux « listes noires » qui sont les restrictions que le SIA ne doit jamais enfreindre. Toutefois, la possibilité de produire un algorithme totalement équitable reste discutée dès lors que les différentes conceptions de l'équité, collective et individuelle, sont rarement compatibles.

En dernier lieu, la prévention des discriminations implique que les agents chargés des SIA (décision, développement, utilisation) soient **sensibilisés** à cette problématique et aient conscience des risques potentiels qui, malgré les mesures citées ci-dessus, ne peuvent pas toujours être totalement éliminés. Le sujet des biais doit impérativement figurer parmi les modules de formation des agents en prise avec des SIA, point également abordé en quatrième partie. Se pose en outre la question de la représentativité (de genre[150], sociale, ethnique…) des équipes chargées de concevoir les SIA publics et les éventuels biais, parfaitement involontaires, qui peuvent les affecter à ce titre.

Il convient enfin de préciser que ces systèmes, s'ils sont conçus dans le respect des exigences qui précèdent, peuvent aussi, en sens inverse, constituer des outils d'atténuation voire de suppression des biais discriminatoires qui affectent le jugement humain.

Tengai Unbiased, le robot-recruteur

En Suède, la municipalité d'Upplands-Bro a expérimenté à partir de 2019 un robot-recruteur dénommé Tengai Unbiased, qui se présente comme un boîtier surmonté d'une tête humanoïde. Il effectue une première sélection sur la base d'une audition classique, au cours de laquelle il analyse le comportement du candidat à l'embauche, lui soumet des mises en situation ou des problèmes à résoudre, et évalue ses compétences, notamment techniques. Le robot est capable, dans une certaine mesure, de réagir aux réponses du candidat en lui soumettant de nouvelles questions. S'il permet d'accélérer les procédures et de réduire les ressources humaines qui leur sont consacrées, l'objectif premier du système est de supprimer les biais discriminatoires inhérents à l'intervention humaine, au tout début du processus de recrutement. Le système est ainsi conçu pour être insensible à l'âge, au sexe, à la tenue vestimentaire ou à l'esthétique des candidats. La commune a décidé de pérenniser l'expérimentation.

[150] Selon l'INSEE, les femmes ne représentent, au niveau national, que 30 % des salariés du secteur numérique et seulement 15 % sur des fonctions techniques.

3. L'accessibilité et l'universalité des SIA publics

Outre l'élévation du niveau général de connaissance et de maîtrise des outils numériques et la lutte contre l'illectronisme, dont il a déjà été question dans la première partie de l'étude, la multiplication des SIA publics implique de penser leur **accessibilité universelle, en particulier pour les personnes en situation de handicap**.

De multiples SIA développent des fonctionnalités visant à pallier un handicap. Il en va ainsi, par exemple, de la description de photos pour les personnes aveugles et malvoyantes, ou de la reconnaissance/transcription de la langue des signes. Toutefois, le déploiement de SIA par les administrations publiques ne peut être envisagé sans que soit pensée, dans le même temps, leur accessibilité. Cette accessibilité répond à deux enjeux :

- Un enjeu de conformité à la règle de droit, tout d'abord, dès lors que les personnes de droit public sont soumises à une obligation d'accessibilité des services de communication au public en ligne (sites Internet, progiciels, applications mobiles, etc.) et que l'absence de conformité peut donner lieu à une sanction administrative d'un montant pouvant aller jusqu'à 25 000 euros[151].
- Un enjeu de garantie d'accès effectif de toutes et tous aux services publics, qui dépasse largement le sujet de l'intelligence artificielle et se pose pour tout applicatif numérique. Le récent rapport de la Défenseure des droits sur la dématérialisation des services publics[152] fait ainsi état de ce que le formulaire permettant de déposer une pré-plainte en ligne demande à ce que l'utilisateur renseigne un « Captcha » en fin de processus (recopier une suite de caractères figurant sur une image), ce qui est impossible pour les personnes ayant un handicap visuel.

3.1.4. Transparence

L'opacité est destructrice de confiance. Si les citoyens ont, à tort ou à raison, le sentiment que le recours aux systèmes d'IA a pour objet ou pour effet de leur « cacher quelque chose », de dissimuler des intentions ou des décisions derrière la complexité de l'outil informatique voire l'impossibilité d'en expliquer le fonctionnement de façon simple, dans toute la mesure du possible, les systèmes d'IA deviendront rapidement la cible de critiques plus ou moins rationnelles et d'actions de désinformation qui retarderont ou compromettront leur déploiement. C'est la raison pour laquelle la proposition de règlement IA de la Commission européenne accorde une place éminente au principe de transparence.

Le principe de transparence comporte à tout le moins les exigences suivantes.

[151] Art. 47 de la loi n° 2005-102 du 11 février 2005 pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées ; décret n° 2019-768 du 24 juillet 2019 relatif à l'accessibilité pour les personnes handicapées des services de communication au public en ligne.

[152] Rapport *Dématérialisation des services publics : trois ans après, où en est-on ?*, février 2022.

1. Le droit d'accès à la documentation du système

Ce droit découle de la façon la plus intuitive du principe de transparence.

a/ Le **public** doit en principe être en mesure d'accéder aux documents qui décrivent les caractéristiques des SIA utilisés dans la sphère publique.

Dès lors que ces SIA sont déployés aux fins de l'exercice d'une mission de service public ou, à tout le, moins qu'ils présentent un lien suffisamment direct avec celle-ci, l'ensemble de la documentation s'y rapportant entre dans le champ du **droit à la communication des documents administratifs** garanti par l'article L. 311-2 du code des relations entre le public et l'administration (CRPA). Les codes-sources des systèmes sont au nombre de ces documents, même s'il convient de souligner que leur communicabilité constitue, bien souvent, une garantie de papier compte tenu de l'inexploitabilité de lignes d'écritures informatiques absconses pour le commun des mortels.

Le droit à communication, qui peut consister aussi bien à exiger une copie des documents qu'à obliger l'administration à les mettre en ligne, est toutefois subordonné à la condition que les documents aient perdu tout caractère préparatoire (ce qui sera le cas, en principe, dès lors que le SIA auquel ils se rapportent aura été déployé), qu'ils ne soient pas grevés de droits de propriété intellectuelle de tiers (ce qui dépend notamment des clauses du marché conclu par l'administration qui s'est procuré le système auprès d'un fournisseur tiers[153]), ou qu'ils ne soient pas couverts par un secret protégé par la loi. A ce titre, le droit d'accès peut être mis en échec par la protection de la vie privée, le secret des affaires[154], les exigences de sécurité des systèmes d'information, de sécurité publique ou de défense nationale, ou encore des exceptions spéciales comme le secret des délibérations des équipes pédagogiques pour la sélection des élèves admis à l'université (décision n° 2020-834 QPC du 3 avril 2020, *Union nationale des étudiants de France*). La doctrine de la CADA et la jurisprudence ne permettent pas encore d'apprécier la portée de ces exceptions appliquées aux SIA et, ainsi, leur

[153] V. sur ce point le CCAG TIC 2021, qui distingue les résultats réalisés dans le cadre du marché et financés par l'acheteur (exemple : le développement d'un SIA spécifique, sur mesure, ou une fonctionnalité spécifique d'interfaçage avec le système d'information existant), et les connaissances antérieures, réalisées et financées dans un autre cadre et qui peuvent appartenir au titulaire du marché, à des tiers ou à l'acheteur (ex : un SIA standard développé par un éditeur). Les résultats peuvent être utilisés par l'acheteur public pour les besoins exprimés dans le marché ou découlant de l'objet des prestations, sans droit de commercialisation par défaut. L'acheteur a un droit d'accès au code source développé spécifiquement pour ses besoins et peut le placer sous licence libre.

[154] Le Tribunal de première instance de l'Union européenne a rendu le 15 décembre 2021 un intéressant jugement sur la communicabilité de documents relatifs au projet de SIA « iBorderCtrl » qui vise à faciliter la détection de fausses déclarations aux frontières extérieures de l'Union (TPIUE, *Breyer c/ Agence exécutive européenne pour la recherche*, T-158/19). Il a partiellement annulé la décision de refus opposé par l'agence européenne, en écartant l'opposabilité du secret des affaires pour certains documents ou parties de documents.

degré de transparence formelle. Il importe d'en faire une interprétation stricte, non seulement en pure logique juridique (puisque le principe est celui de la liberté d'accès), mais aussi en opportunité administrative, afin de ne pas risquer de fragiliser sans justification solide l'acceptabilité du système déployé.

Il importe de préciser que **l'exigence de transparence doit se concilier avec l'impératif de confidentialité qui s'attache à certaines activités sensibles**, tout particulièrement dans le domaine de la sécurité et de la défense. En la matière, et par exception, la transparence doit être totale à l'égard de l'autorité de contrôle compétente, mais restreinte pour ce qui concerne le grand public.

Le droit à l'information du public ne peut s'exercer avec une pleine effectivité que si le fonctionnement du SIA est documenté de façon complète. En effet, le droit d'accès aux documents administratifs ne s'exerce qu'à l'égard de documents existants. Le principe de transparence impose donc la formalisation des choix et arbitrages effectués lors de la conception et des constats effectués après la mise en service.

On peut s'interroger sur l'opportunité d'assurer la **transparence de la performance** du système. En effet, ce dernier peut être en capacité de mesurer sa propre fiabilité, en indiquant un « taux de certitude » (probabilité que le résultat qu'il produit soit correct). A titre d'exemple, en matière de reconnaissance d'images, le système peut indiquer qu'il y a X% de chances que la personne photographiée soit M. ou Mme…. Toutefois, si l'existence de possibilités d'erreur doit naturellement faire l'objet d'une information des utilisateurs du système, la quantification précise de ce risque d'erreur peut, d'une part, s'avérer anxiogène pour le grand public voire source de contentieux et, d'autre part, être elle-même erronée. Son affichage constitue toutefois un puissant aiguillon pour l'amélioration de la performance du système par l'administration. *A minima*, devraient être communiqués à l'usager un ordre de grandeur du taux d'erreur, son évolution dans le temps et des explications claires sur les enjeux et les efforts déployés pour accroître la performance du système.

Outre le droit d'accès sur demande, les administrations qui emploient au moins 50 agents ou salariés (en ETP) sont tenues de publier spontanément en ligne les règles définissant les « principaux traitements algorithmiques » qu'elles utilisent, lorsqu'ils fondent des décisions individuelles (article L. 312-1-3 du CRPA). Toutefois, cette obligation n'est assortie d'aucune sanction et son effectivité est douteuse en pratique.

b/ Il va de soi que **l'administration utilisatrice d'un SIA** doit disposer, de la part de son fournisseur lorsqu'elle recourt à l'externalisation, d'une information suffisamment complète sur les caractéristiques du système. Le projet de règlement européen prévoit ainsi, pour les seuls SIA « à haut risque », la remise d'une notice d'utilisation indiquant de manière concise, complète, exacte, claire, pertinente, accessible et compréhensible une série d'informations sur les caractéristiques, les capacités et les limites de performance du système, notamment sa destination, son niveau d'exactitude, de robustesse et de cybersécurité, ses « performances » ainsi que sur les jeux de données d'entraînement. Les clauses du marché public peuvent naturellement prévoir une information plus étendue.

2. L'exigence de loyauté

Le principe de transparence doit aussi et surtout bénéficier aux personnes sur lesquelles l'utilisation de l'IA a un impact direct. A ce titre, il comporte une **exigence de loyauté** selon laquelle le responsable du système doit informer les personnes intéressées du recours à un système d'IA lorsque ce dernier a exercé une influence significative sur le sens de la décision administrative dont ils sont les destinataires ou sur la conduite d'une activité qui les concerne. Cette exigence inclut des obligations d'information passive (informations que les personnes peuvent se procurer en accomplissant une démarche) et active (« notification individuelle »).

S'agissant de l'obligation d'information « passive », outre l'application des règles générales relatives au droit de communication des documents administratifs, les administrations d'une certaine importance ont d'ores et déjà l'obligation de publier les caractéristiques des « principaux traitements algorithmiques » qu'elles utilisent (V. sur ce point la note juridique en annexe 10). L'effectivité de cette garantie pose toutefois question au regard des pratiques constatées. La notion de « principaux traitements » mériterait d'être précisée ou l'obligation recentrée sur les systèmes d'IA à haut risque[155] et ceux qui, sans être regardés comme tels par le règlement européen, présentent une certaine sensibilité, tout en encourageant les collectivités publiques à publier en ligne, sur une base volontaire, une information sommaire sur l'ensemble des SIA qu'elles utilisent.

S'agissant de l'obligation d'information « active », cette garantie existe d'ores et déjà pour des décisions individuelles prises par les administrations sur le fondement d'un traitement algorithmique[156].

Le projet de règlement européen pose la règle selon laquelle, en dehors du domaine pénal, toute personne doit être informée qu'elle interagit avec un système d'IA (qu'il soit ou non à haut risque), quand bien même ne se traduirait-il pas par l'édiction d'un acte administratif, « *sauf si cela ressort clairement des circonstances et du contexte d'utilisation* »[157]. Les conditions de cette information ne sont toutefois pas précisées. Dès lors qu'elle est matériellement possible[158] et n'exige pas des efforts disproportionnés, l'information active, qui offre davantage de garanties, devrait être privilégiée.

---

[155] Cette transparence devra inclure, le moment venu, les références de la certification d'un tel système, qui constitue un gage de confiance important.

[156] D'une part, toute décision individuelle fondée sur un tel traitement doit le mentionner (art. L. 311-3-1 du CRPA). D'autre part, si l'intéressé le demande, l'administration doit lui communiquer, sous une forme intelligible, les règles et les caractéristiques de la mise en œuvre du traitement (art. R. 311-3-1-2 du même code).

[157] En outre, et sous la même réserve, les personnes exposées à un système de reconnaissance des émotions ou de catégorisation biométrique seraient informées de son fonctionnement. Par ailleurs, les utilisateurs d'un SIA d'hyper-trucage (« *deepfake* ») préciseraient que les contenus produits ont été générés ou manipulés artificiellement, sauf exceptions.

[158] Tel n'est pas toujours le cas. Il est malaisé, par exemple, d'informer individuellement les personnes fréquentant les transports en commun de la mise en œuvre d'un système de comptage des masques en vue de l'établissement de statistiques (V. pour ce cas d'usage le

3. *L'explicabilité*

Les systèmes d'IA sont souvent associés au phénomène dit de « **boîte noire** », qui renvoie à l'incapacité de l'homme à comprendre le cheminement et les raisons ayant conduit la machine à produire tel résultat à partir de telles données d'entrée. Il est toutefois important de rappeler d'emblée que ce phénomène ne touche, sauf exceptions, que les modèles complexes, et qu'il n'est donc en rien consubstantiel à l'intelligence artificielle. Certains modèles algorithmiques simples et très utilisés (régression logistique, régression linéaire, arbre de décision…) sont en général parfaitement explicables. Il peut en aller différemment de modèles d'apprentissage machine plus complexes. Tel est souvent le cas des méthodes dites ensemblistes[159], et, plus encore, celui des réseaux de neurones, dont le poids respectif est ajusté automatiquement par la machine en fonction des données qui l'alimentent afin de maximiser l'exactitude des données de sortie, et dont le nombre de couches peut se compter en dizaines et celui des paramètres, en milliards voire en trillions[160].

Néologisme issu de l'anglais « *Explanability* », l'exigence d'explicabilité, au sens large, impose au responsable du système d'être en capacité de comprendre les opérations que la machine a réalisées pour produire ses résultats – on parle d'interprétabilité (comment le modèle est-il parvenu à ce résultat ? en recourant à quelles données ? par quels calculs ?) – et de restituer dans un langage compréhensible par toute personne (ou au moins par celle qui est concernée) les éléments-clés du « raisonnement » – c'est l'explicabilité au sens strict (pourquoi ce résultat a-t-il été produit, et pas un autre ?). Cette exigence peut se concevoir à deux niveaux.

L'explicabilité « locale » suppose d'être en mesure d'expliquer le résultat produit **pour un cas particulier** (par exemple, pourquoi ce SIA d'aide au diagnostic médical a-t-il estimé que le patient X était atteint de telle ou telle maladie, au vu d'une image ou d'un vaste ensemble de signes cliniques ?). Il s'agit, *a minima*, d'identifier les variables-clés qui ont déterminé la production du résultat par le système. Pour ce faire, il est en général fait appel à un modèle explicatif[161] qui constitue une approximation et une simplification du modèle complet (par exemple, une régression linéaire simple en lieu et place d'un réseau de neurones complexe). Par analogie, un agent public est censé pouvoir expliquer que, parmi les dizaines de critères ou d'indices du faisceau que la loi et/ou la doctrine administrative l'amènent à manier, il a pris la décision de contrôler telle entreprise ou d'accorder telle autorisation sur la base de deux ou trois critères qui lui ont paru décisifs.

---

décret n° 2021-269 du 10 mars 2021 relatif au recours à la vidéo intelligente pour mesurer le taux de port de masque dans les transports).

[159] V. le glossaire pour l'explicitation de cette notion. A titre d'exemple, les « forêts aléatoires », qui consistent à recourir à et à combiner les résultats de centaines d'arbres de décision pour la résolution d'un même problème, peuvent soulever des problèmes d'explicabilité.

[160] Le modèle Wu Dao présenté lors de la conférence de juin 2021 de l'institut de recherche en intelligence artificielle de Pékin compte 1,75 trillion de paramètres, soit dix fois plus que le modèle GPT-3 (175 milliards de paramètres) sorti en 2020.

[161] Parmi les modèles d'explication les plus utilisés figurent LIME et les Ancres.

L'explicabilité peut aussi consister à déterminer, dans une logique contrefactuelle, les modifications qui pourraient être apportées à la situation pour aboutir à un résultat différent (par exemple, un SIA chargé de la délivrance de prestations sociales ayant refusé son octroi à une personne au vu de son profil, quelles sont les caractéristiques qu'il faudrait modifier, et dans quelle mesure, pour obtenir une décision positive ?).

L'**explicabilité globale**, plus ambitieuse, vise à rendre le modèle interprétable pour l'ensemble des données, et non pour une personne ou une situation déterminée. Elle va jusqu'à la « preuve formelle » des systèmes basés sur l'apprentissage automatique, c'est-à-dire la formalisation de l'ensemble des règles d'un modèle entraîné, comme si ce dernier relevait de l'IA « basée sur les règles » (comme un système-expert).

L'explicabilité des modèles basés sur l'apprentissage machine constitue un **champ de recherche** à part entière[162]. Des progrès importants ont déjà été accomplis dans la compréhension de modèles complexes, y compris les réseaux de neurones, même si la perspective de démontrer formellement la conformité de ces derniers aux exigences apparaît très éloignée. C'est l'un des enjeux du « Grand défi » sur l'intelligence artificielle sélectionné par le Conseil de l'innovation et soutenu par le Fonds pour l'innovation et l'industrie, qui porte sur la sécurisation, la certification et la fiabilisation des systèmes fondés sur l'intelligence artificielle.

Cette exigence revêt une sensibilité particulière pour les SIA publics, qu'exprime l'article 15 de la Déclaration de 1789 : l'administration doit rendre des comptes, ce qui peut inclure l'obligation d'expliciter le fonctionnement des outils informatiques qu'elle utilise.

Le droit applicable le garantit déjà très largement pour les systèmes qui fondent en tout ou partie des décisions individuelles. Les articles L. 311-3-1 et R. 311-3-1-2 du CRPA font obligation à l'administration qui a pris une décision individuelle sur le fondement d'un traitement algorithmique de porter à la connaissance de son destinataire qui le demande « *les règles définissant ce traitement* » et les « *principales caractéristiques de sa mise en œuvre* », dont, sous une forme intelligible, « *les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé* »[163]. L'obligation de motivation à laquelle sont soumises de nombreuses décisions impose également de fournir à leur destinataire les considérations de droit et de fait qui en constituent le fondement (art. L. 211-5 du CRPA). La connaissance des déterminants de la décision est une condition d'effectivité du contrôle de l'erreur de droit, permettant de détecter une méconnaissance de la loi (par exemple, en ce qu'elle fixe une liste limitative de critères à prendre en compte) voire des biais discriminatoires (V. 3.1.3 ci-dessus).

---

[162] V. notamment le projet « *Explanable AI* » (« XAI ») de l'agence de recherche du ministère de la défense américain – DARPA.

[163] Dans le même esprit, s'agissant des traitements de données à caractère personnel, l'article 9 de la convention 108 du Conseil de l'Europe modifiée fonde le droit de toute personne « *d'obtenir, à sa demande, connaissance du raisonnement qui sous-tend le traitement de données, lorsque les résultats de ce traitement lui sont appliqués* ».

Il est légitime que les citoyens revendiquent une certaine explicabilité des SIA publics en dehors même de l'édiction automatisée ou assistée d'actes administratifs. A titre d'exemple, un SIA de ciblage des contrôles fiscaux ne conduit pas à la prise d'une décision attaquable et le fait qu'un contribuable déterminé ait fait l'objet d'un contrôle à la suite d'une recommandation d'un tel système, fût-ce dans des conditions illégales, apparaît insusceptible d'emporter la décharge d'impositions qui seraient effectivement dues par l'intéressé. Pour autant, il ne serait pas acceptable que l'administration fiscale soit dans l'incapacité d'expliquer pourquoi (c'est-à-dire au regard de quels éléments de son profil) ce contribuable a fait l'objet d'un signalement par le système. De même, si un outil de simulation – comme des jumeaux numériques – est utilisé pour prévoir les conséquences d'un choix d'aménagement du territoire communal, comme la fermeture d'un axe routier, les habitants de la commune sont en droit de connaître les raisons pour lesquelles le système a indiqué qu'il n'en résulterait aucune congestion supplémentaire ou report sur tel autre axe.

L'explicabilité pure et parfaite ne saurait toutefois constituer une condition *sine qua non* du déploiement des systèmes d'IA publics, pas plus d'ailleurs que le fonctionnement du cerveau humain dans une situation donnée n'est connu et compris de façon complète, ce dont les citoyens doivent bien s'accommoder. Trois fils conducteurs peuvent guider la réflexion :

- d'une part, il y a lieu de proportionner la portée de cette exigence à la sensibilité du système. Il n'est certes pas inutile de savoir ce qui a déterminé un SIA de comptage des animaux errants (à l'aide de la vision par ordinateur) à prendre tel chien en laisse pour un animal en divagation, mais on ne saurait contraindre l'administration à se livrer à d'intenses recherches pour le comprendre et l'expliquer (la laisse blanche était-elle « indétectable » sur le fond blanc d'un camion qui passait par là ?). De même, il n'est probablement pas nécessaire, voire contreproductif, de connaître dans le moindre détail le taux de contribution de tous les paramètres d'une décision[164] ;
- d'autre part, il convient de mettre en balance la contribution du système à la mission de service public et, notamment, l'exigence de performance qui peut justifier de recourir à un modèle complexe, avec les inconvénients qui s'attachent à la persistance de zones d'ombre et à l'impossibilité, en l'état des connaissances scientifiques, de comprendre en détail le « cheminement intellectuel » du système ;
- enfin, l'enjeu est moins de définir un standard objectif d'explicabilité que de faire en sorte que, subjectivement, la personne sur laquelle le système a eu un impact soit satisfaite des explications fournies, ce qui suppose de s'adapter à son niveau de compétences et à l'ampleur de l'impact. L'enjeu est moins d'assurer une transparence technique et normée que d'en faire la pédagogie pratique.

[164] Le bureau américain de protection des consommateurs en matière financière recommande par exemple que les décisions de refus de crédit fondés sur des SIA de notation financière des clients ne fassent pas apparaître plus de quatre raisons (https://www.consumerfinance.gov/rules-policy/regulations/1002/9/#a-3-i-B).

Il y a lieu d'observer que, hormis une allusion dans un considérant consacré aux SIA utilisés à des fins répressives, de façon d'ailleurs distincte de la transparence, le projet de règlement européen ne reprend pas formellement cette exigence. Il se borne, dans son article 13, à imposer que le fonctionnement du SIA à haut risque soit « *suffisamment transparent pour permettre aux utilisateurs d'interpréter les résultats du système et de l'utiliser de manière appropriée* ». Il s'agit donc d'une exigence d'interprétabilité, circonscrite à l'utilisateur, et non d'explicabilité au bénéfice des personnes à l'égard desquelles cet utilisateur (comme l'est l'administration) met en œuvre le système.

4. *La conception transparente et l'auditabilité*

La transparence doit également guider la conception même des SIA publics. Il importe que l'administration qui envisage de recourir à un tel outil **affiche le plus en amont possible son intention et les objectifs qu'elle poursuit**, sans dissimuler des conséquences adventices ou fatales.

En particulier, la question de l'incidence de l'automatisation d'une activité par un SI sur l'emploi public et les conditions de travail des agents publics doit faire l'objet d'un dialogue social approprié, qui ne saurait se réduire à la tenue d'un comité technique. S'il est parfaitement concevable que l'administration elle-même ne soit pas en mesure d'anticiper précisément l'ensemble des conséquences de son initiative, ces incertitudes, qui doivent être assumées par le responsable, ne sauraient fournir de prétexte à l'opacité.

Lorsqu'elle est possible, l'association de l'ensemble des parties prenantes, y compris des usagers, à la conception même de l'outil, est de nature à renforcer considérablement la confiance des citoyens à l'égard d'une administration ouverte, et de prévenir les procès d'intentions complotistes. Il est d'ailleurs tout à fait possible de partager avec les membres d'un comité des parties prenantes des informations à titre confidentiel, sans qu'un tel échange implique de les mettre sur la place publique.

Dans tous les cas, l'administration doit être en capacité d'indiquer les raisons pour lesquelles elle a décidé de recourir au système et les choix de conception qu'elle a effectués, parmi les options qui s'offraient à elle.

Ainsi, les objectifs, la démarche et les choix de conception du modèle algorithmique, les jeux de données utilisés et l'ensemble des processus (y compris de maintien en conditions opérationnelles), doivent être rigoureusement documentés afin que le système d'IA soit **auditable par les autorités compétentes**, tant en ce qui concerne le processus d'apprentissage que les résultats produits. Cette documentation doit être conservée pendant une durée suffisante pour que le système puisse être contrôlé après la cessation de son utilisation, dans une certaine mesure. Au regard des exigences du RGPD, les besoins d'un tel contrôle sont de nature à justifier la conservation des données à caractère personnel utilisées pour l'apprentissage, dans des conditions sécurisées incluant la pseudonymisation et le chiffrement, et excluant toute exploitation courante (segmentation logique).

3.1.5. Sûreté (cybersécurité)

En sus du principe de performance déjà évoqué, qui inclut la minimisation, la prévention et la résolution rapide des dysfonctionnements du système en conditions normales d'utilisation, la conception d'un système d'IA public de confiance doit impérativement intégrer l'enjeu de sûreté, c'est-à-dire la **prévention des attaques et la résolution de leurs conséquences**.

Outre les vulnérabilités que partagent l'ensemble des systèmes d'information (y compris au niveau des infrastructures physiques), les SIA présentent certains points faibles particuliers susceptibles d'être exploités à des fins malveillantes. Trois d'entre elles méritent d'être signalées :

- l'« **empoisonnement des données d'apprentissage** » est une pratique qui consiste à altérer le jeu de données utilisé pour l'entraînement du modèle afin d'en biaiser le fonctionnement et d'en vicier les résultats (par exemple, en modifiant la classification opérée par le système dans un sens favorable à l'attaquant). Des précautions particulières doivent donc être prises pour assurer l'intégrité des données, détecter les intrusions et (tentatives de) modifications, et notifier les brèches de sécurité à l'autorité compétente, voire les porter à la connaissance du public. Si le 6° de l'article 4 de la loi du 6 janvier 1978 impose au responsable de traitement de « *garantir une sécurité appropriée des données à caractère personnel, y compris la protection contre le traitement non autorisé ou illicite et contre la perte, la destruction ou les dégâts d'origine accidentelle, ou l'accès par des personnes non autorisées, à l'aide de mesures techniques ou organisationnelles appropriées* », la loi ne fixe pas d'exigence analogue pour les données ne présentant pas un caractère personnel. Or l'acuité de la problématique de la sûreté ne dépend pas de la nature de la donnée, mais de la sensibilité du SIA et des conséquences qu'entraînerait son dysfonctionnement ;
- le risque de **leurre des systèmes**, dont certains peuvent être sensibles à des variations minimes des données d'entrée. La modification, imperceptible à l'œil nu, d'un seul pixel sur une image peut ainsi entraîner une « prédiction » totalement différente et parfois aberrante[165]. Cette pratique, parfois présentée sous l'appellation d'« exemples adverses » (comme le fait la proposition de règlement IA) ou d'« attaque adversariale » (référentiel du LNE), appelle des efforts de recherche et de sécurisation des modèles ;
- le **vol de données** par rétro-ingénierie (ou ingénierie inversée) illégale : il s'agit de reconstituer le code-source utilisé (lorsqu'il n'est pas public) en vue de l'exploiter à des fins personnelles, voire de remonter aux données d'entraînement « mémorisées » par le modèle, lesquelles peuvent présenter une certaine sensibilité, notamment s'il s'agit de données à caractère personnel ou

[165] Par ex. : la présence de quelques stickers sur des panneaux de circulation ou sur le sol peut amener un véhicule autonome à confondre un stop avec une limitation de vitesse voire à rouler à contre-sens. Une simple paire de lunettes peut, selon le motif imprimé, tromper un modèle de reconnaissance faciale qui reconnaîtra à tort une personne donnée ou confondra une personne et un animal.

d'informations classifiées[166]. Là encore, les modèles doivent être conçus de manière à empêcher des pratiques de réidentification ou un risque de compromission du secret, par l'ajout de « bruit » ou de « perturbations ».

L'intrusion dans un système de traitement automatisé de données, la modification de données enregistrées ou encore le fait d'entraver ou de fausser le fonctionnement du système, sont d'ores et déjà réprimés par les articles 323-1 et suivants du code pénal. Ces dispositions sont applicables aux SIA et n'appellent aucune adaptation.

La proposition de règlement IA prescrit le recours à des solutions techniques « *adaptées aux circonstances pertinentes et aux risques* » tels qu'analysés dans le système de gestion des risques obligatoirement mis en place par le fournisseur. Ces exigences seraient réputées respectées en cas de certification pertinente au titre d'un schéma de cybersécurité. La sûreté constitue par conséquent l'un des **enjeux-clés de l'évaluation de la conformité**, qui devra recourir à des simulations d'attaque et/ou à des vérifications mathématiques pour éprouver la résistance et évaluer les vulnérabilités du système.

3.1.6. Soutenabilité environnementale

Les SIA n'ont pas d'impact écologique d'une nature différente de celui de l'ensemble des techniques et technologies du numérique. Pour rappel, l'étude générale publiée en janvier 2022 par l'ADEME et l'ARCEP établit que l'impact environnemental le plus lourd du numérique est celui lié aux terminaux, autour de 70 à 80% en moyenne, très loin devant les équipements de réseaux et le fonctionnement des centres de données. Mais l'impact écologique des SIA peut atteindre un degré d'intensité qui exige une attention et des précautions renforcées.

La question est, pourtant, souvent écartée par les tenants d'une vision utopique de l'intelligence artificielle comme constituant elle-même la solution au problème : une intelligence artificielle générale saura résoudre toutes les questions, notamment celle du réchauffement climatique, de la perte de biodiversité ou de l'épuisement des ressources. En attendant cette épiphanie, force est de constater que **la généralisation des SIA, tels qu'ils sont aujourd'hui conçus, concourrait significativement à l'aggravation de la crise environnementale**.

Ce n'est pas nier que les SIA peuvent être des auxiliaires précieux dans la lutte contre le dérèglement climatique. D'ores et déjà, on trouve de nombreux exemples de contribution de l'intelligence artificielle à l'optimisation de la consommation des ressources (par exemple, celle des véhicules à moteur) et à l'amélioration de l'efficience des sources de production d'énergie, notamment par une meilleure gestion des réseaux et des réserves. De même, les SIA optimisant les déplacements, réduisant les consommations de matières premières ou améliorant les process de fabrication, contribuent à la préservation de l'environnement. Mais cet impact,

[166] V. sur ce point E. Berthier (DGA), *Protection des données d'entraînement pour l'apprentissage statistique*, 2019, Conférence Intelligence Artificielle et Défense.

quelle qu'en soit l'importance, ne dispense en rien d'observer les conséquences négatives de la généralisation des SIA afin de les prévenir.

La multiplication du recours aux SIA a pour première conséquence d'accroître le **besoin en terres rares**. Il est difficile aujourd'hui de quantifier le facteur d'accroissement de ce besoin, ou d'imaginer de quelle manière la dépendance envers ces minerais pourrait être réduite (les ordinateurs quantiques, dont les spécialistes estiment que nous sommes encore très loin de pouvoir annoncer un fonctionnement dans des conditions qui permettraient leur généralisation, auraient pour effet certain de diminuer le besoin de ces ressources mais ne supprimeraient pas totalement le problème). Les difficultés sont ici multiples : l'exploitation s'opère dans des conditions environnementales (gaspillage énergétique, détérioration et pollution des milieux…) et dans les pays où les gisements majeurs se trouvent, qui sont loin d'être au niveau des exigences qu'une gestion raisonnable des conséquences de l'exploitation imposerait sur le territoire européen ; les ressources en cause sont naturellement finies, et leur épuisement incite à rechercher l'exploitation de sites moins riches appelant là encore des conséquences environnementales négatives significatives (plus de consommation énergétique, plus de pollution…) ; le transport des matières premières ou des produits après leur transformation, bien que maritime, a une empreinte écologique assez négative.

Le deuxième ordre de conséquences porte sur l'**artificialisation des sols**. L'illusion du stockage des données dans le nuage (*cloud*) fait oublier que celui-ci est posé au sol. Aucune donnée ne reste dans l'atmosphère, elles figurent dans des lieux de stockage qui ont une existence physique. La multiplication à l'infini du nombre de données, d'apprentissage, traitées, produites par la multiplication des SIA, exige des sites de traitement et de stockage dont la surface, en l'état de la technique, est de plusieurs dizaines d'hectares. Une planification des besoins d'espace, répondant aussi à l'exigence de protection de la souveraineté française, en laissant sur le territoire national celles des données dont l'intégrité est estimée essentielle à la défense de nos intérêts, et de ceux de la population, devrait conduire à subordonner le développement de ces lieux à l'objectif gouvernemental d'arrêt de l'artificialisation des sols. C'est donc dans la réutilisation des anciens sites, en apportant un soin particulier à l'insertion dans le tissu urbain ou rural de bâtiments dont l'attractivité architecturale et la contribution à l'animation ne sont pas les qualités premières, que réside un développement acceptable et durable des sites de stockage.

Mais la conséquence majeure de la généralisation du recours aux SIA, publics comme privés, consiste en une augmentation très importante de la **consommation d'électricité**. L'empreinte carbone de l'utilisation du système, en phase d'inférence, reste certes limitée, à l'échelle individuelle : la plupart du temps, une application basée sur l'IA, fût-elle un réseau de neurones, peut être lancée sur un smartphone ou un ordinateur classique. Mais la puissance de calcul déployée pour la construction de modèles complexes d'apprentissage machine requiert, elle, une consommation électrique considérable. Celle-ci peut être exposée en pure perte si le modèle n'est finalement pas mis en production. La course à la performance, en particulier à l'exactitude du modèle, peut conduire à une disproportion inacceptable entre son incidence environnementale et les bénéfices obtenus, lorsqu'il s'agit de réduire de

quelques dixièmes un taux d'erreur sans enjeu majeur au regard de la destination du système – même s'il convient de tenir compte des usages en cascade et, en particulier, des possibilités d'adaptation, par un réentraînement sur un ensemble limité de données et à un coût énergétique faible, de giga-modèles correctement pré-entraînés.

Comme toujours dans ce domaine, les prévisions sont approximatives, mais elles donnent un ordre de grandeur qui fait l'objet d'un assez large consensus. Une étude[167] montre ainsi que si, grâce aux mesures d'économie d'énergie prises par le secteur, la consommation imputable aux technologies de l'information et de la communication est restée stable de 2010 à 2017, elle pourrait doubler d'ici 2030. Le rapport du député Cédric Villani se faisait l'écho de projections de l'association française *The shift project*, qui estimait que selon les scénarii, le numérique représenterait de 20 à 50% des consommations électriques en 2030. A titre d'exemple, la désormais fameuse victoire d'AlphaGo sur le meilleur joueur humain a consommé l'équivalent de quatre mois de consommation d'électricité d'un ménage français moyen pour chaque partie. Quant au modèle géant GPT-3 dans le domaine du traitement du langage naturel, une étude estime que chaque session d'entraînement représente la consommation électrique annuelle de 126 ménages danois ou l'empreinte carbone d'une voiture thermique qui ferait plus de 17 fois le tour du monde à l'équateur (700 000 km)[168].

S'agissant des émissions de gaz à effet de serre, un ouvrage publié en 2020[169] estime que les technologies numériques représenteront, en 2030, 14% des émissions au monde, soit la moitié environ des émissions globales du secteur des transports (selon un rapport de l'ARCEP[170], les technologies numériques contribuent déjà à 3% des émissions mondiales de gaz à effet de serre). En France, les centres de données représenteraient, à eux seuls, 14% des émissions dues au numérique[171].

Ces chiffres, même approximatifs, appellent une réaction et l'intégration de l'enjeu environnemental au sein des principes encadrant le déploiement des SIA publics, à deux niveaux.

D'une part, la question de l'impact environnemental du recours à des systèmes d'intelligence artificielle doit impérativement être prise en compte comme une **dimension de la stratégie publique** en la matière. Outre le nécessaire développement d'infrastructures adaptées, abordé supra, les pouvoirs publics doivent d'emblée poser le **principe de la neutralité environnementale globale de l'intelligence artificielle**, au moins pour les SIA publics. C'est en effet à un niveau global que la neutralité doit être recherchée, combinant bilan positif et bilan négatif de chaque système pour une enveloppe globale neutre, au niveau par exemple d'un ministère, d'un établissement, ou d'un territoire. Cet objectif demande en conséquence une planification raisonnée

[167] HSBC, *Powering the data revolution*, mai 2019.

[168] Sachant que le modèle suivant, GPT-4, dont le développement a été annoncé en septembre 2021, compterait 100 trillions de paramètres, soit 500 fois plus que GPT-3…

[169] K. Crawford, *Atlas of AI Power, politics and the planetary cost of AI*, Yale university press.

[170] Réseau du futur, *L'empreinte carbone du numérique*, juillet 2019.

[171] Arcep, *Pour un numérique soutenable*, décembre 2020.

du déploiement des SIA, et la mise en place de mécanismes d'évaluation environnementale objectifs, sincères, et indépendants, pour à la fois dessiner les objectifs et mesurer leur atteinte. L'ADEME, mais aussi la CRE et RTE, pourraient sans doute jouer un rôle significatif, du moins pour le volet énergétique, en définissant des cadres de référence, des bonnes pratiques, et des modèles d'évaluation.

D'autre part, **la décision de recourir à un SIA public, l'arbitrage entre l'achat et la conception en régie et, selon le cas, les critères de l'achat public et les choix internes de conception, doivent pleinement intégrer cette dimension.** A-t-on réellement besoin d'un modèle totalement nouveau ou peut-on utiliser un système déjà sur le marché, et, le cas échéant, quel est l'impact d'un éventuel réentraînement ? A-t-on besoin de concevoir un modèle aussi complexe au regard de l'enjeu ou peut-on se contenter d'un modèle plus fruste, donc aussi plus « frugal » sur le plan environnemental ? Dans tous les cas, une évaluation sommaire de l'impact environnemental, tenant compte de l'intégralité du cycle de vie, devrait être menée, afin d'alimenter le bilan global ; s'agissant des modèles les plus complexes, requérant une puissance de calcul significative (réseaux de neurones comportant de très nombreuses couches et paramètres), une étude d'impact plus complète devrait être exigée.

Naturellement, le prérequis de toute démarche en la matière est la création d'**indicateurs de mesure** de l'empreinte écologique qui soient à la fois fiables et partagés. Il s'agit d'un axe de recherche actif. L'évaluation des systèmes d'IA doit intégrer celle de l'efficience énergétique et, plus largement, de la soutenabilité environnementale.

On relèvera que cet enjeu environnemental, s'il fait partie intégrante de la Charte des droits fondamentaux de l'Union européenne (article 37), est quasiment absent de la proposition de règlement IA de la Commission européenne, celle-ci se bornant, au titre des codes de conduite dont elle encourage l'élaboration, à prévoir que ces derniers peuvent inclure des engagements volontaires liés à la durabilité (ou à la viabilité) environnementale. Le Parlement européen avait quant à lui proposé d'en faire une exigence à part entière, consistant en une évaluation et l'obligation de mettre en place des mesures pour atténuer et corriger l'incidence générale des SIA sur les ressources naturelles, la consommation d'énergie, la production de déchets, l'empreinte carbone, la situation d'urgence en matière de changement climatique et la dégradation de l'environnement.

3.1.7. Autonomie stratégique

Parce qu'elle sert l'intérêt général et le bien-être de nos concitoyens, et qu'elle conditionne en partie – et le fera de façon croissante – la capacité de la puissance publique à assurer ses fonctions essentielles, la stratégie de l'IA publique doit être conçue de manière à préserver la souveraineté de la France et à garantir l'autonomie de la Nation.

Cette autonomie stratégique suppose à la fois que les collectivités publiques disposent de **la plus grande liberté de choix possible quant aux solutions qu'elles entendent déployer en matière de SIA**, et que **leur maîtrise de ces systèmes** soit telle qu'elle neutralise le risque d'une utilisation contraire aux intérêts de la Nation

et des citoyens qu'elle protège par des entités sur lesquelles elle n'exerce, directement ou indirectement, aucune forme de contrôle – au premier rang desquels les puissances étrangères, en particulier extra-européennes. Il serait naïf de croire, dans ce domaine essentiel, que les rapports entre États alliés seraient moins gouvernés par la protection de leurs intérêts respectifs que, par exemple, dans le domaine nucléaire. A ce titre, la stratégie de la France doit tenir les deux bouts de la chaîne : poursuivre l'ambition d'une autonomie stratégique européenne tout en s'assurant d'un certain degré d'indépendance nationale.

A l'évidence, l'objectif n'est pas de garantir une totale indépendance du pays, ce qui serait parfaitement illusoire et contre-productif. L'autarcie numérique heurterait du reste d'autres exigences tel que le principe de performance, qui peut justifier de recourir à des ressources contrôlées par un acteur étranger et d'en être, dans une certaine mesure, tributaire, comme l'ont révélé les débats sur le *Health Data Hub*[172]. On voit bien, par ailleurs, qu'une approche nationale peut paradoxalement faire perdre à la France toute autonomie stratégique, lorsque les moyens dont le pays dispose sont insuffisants pour rivaliser avec les compétiteurs étrangers et que seule la coopération interétatique, en particulier à l'échelon européen, peut la garantir. La France doit pouvoir choisir ses dépendances, être en capacité de s'en affranchir et maîtriser ses vulnérabilités. L'autonomie stratégique doit ainsi se concevoir par cercles concentriques, en modulant la dépendance externe en fonction de la confiance qui peut être accordée aux partenaires, et en accordant une place particulière à la coopération européenne.

La liberté de choix n'est effective que si la **disponibilité des ressources** nécessaires à la conception et au déploiement des SIA publics est suffisamment garantie. Ces ressources sont décrites plus avant dans la quatrième partie de l'étude. Du point de vue de l'autonomie stratégique, une attention particulière doit être portée aux enjeux suivants :

- les **compétences :** tant les chercheurs que les experts de la donnée, au moins les meilleurs d'entre eux, se trouvent sur un marché du travail international où la compétition entre employeurs fait rage. Des acteurs étrangers, au premier rang desquels les « GAFAM », peuvent leur offrir des niveaux de rémunération et des conditions de travail sans commune mesure avec celles qu'on trouve dans l'administration. Alors que la France se distingue par l'excellence de ses jeunes pousses, la problématique de la « fuite définitive des cerveaux » doit recevoir des réponses dans la stratégie de l'IA publique (et plus généralement, dans la stratégie numérique du pays) ;
- les **structures de recherche (et développement)** : au-delà des talents individuels, la constitution d'équipes publiques ou privées de recherche de premier ordre au niveau mondial est clé. Malgré l'excellent positionnement dans les conférences internationales en IA d'instituts de recherche comme l'Inria, et malgré la taille croissante des équipes d'experts des données et de robotique des licornes ou des meilleures jeunes pousses de la *French Tech* spécialisées en IA, la France ne semble pas encore disposer d'équipes de R&D de la puissance de celles qui

[172] V. 4.1.3. au sujet de la problématique de la plateforme nationale des données de santé.

contribuent le plus fortement au front pionnier de l'IA (non seulement comparables à celles hébergées par les GAFAM, mais aussi à celles d'organisations de taille plus modeste comme Deepmind, OpenAI, Tesla, the Allen Institute…) ;

- les **infrastructures :** la vulnérabilité générale de la France en matière de semi-conducteurs s'étend aux composants électroniques adaptés aux types de calculs requis par les SIA. Il convient donc de veiller à la disponibilité de la puissance de calcul nécessaire à la conception des SIA les plus sophistiqués (V. sur ce point la quatrième partie de l'étude), notamment dans le domaine du traitement du langage naturel, et ce en tirant parti des initiatives européennes en la matière[173]. En outre, le nébuleux « *cloud* », largement maîtrisé par des acteurs étrangers et qui offre aujourd'hui des services indispensables à l'écosystème français de l'IA est nécessairement situé sur un territoire. D'une manière générale, le développement et la maintenance en production des SIA nécessitent tout un écosystème d'infrastructures numériques de services (comme les plateformes de collecte et d'annotation de données) dont l'absence de maîtrise souveraine peut entraîner des vulnérabilités dans certains domaines sensibles. La localisation française ou, à tout le moins, européenne des serveurs constitue une première protection contre le risque d'indisponibilité lié, notamment, à des facteurs géopolitiques ;
- les **données d'apprentissage :** là encore, l'administration doit systématiquement s'interroger sur la situation de dépendance que peut créer le recours à des données dont la production est maîtrisée par des acteurs extra-européens. Ainsi qu'il a été dit, un modèle basé sur l'apprentissage automatique a vocation à être maintenu en conditions opérationnelles et adapté à l'évolution des besoins et de l'environnement, ce qui suppose de l'alimenter en données d'entraînement fraîches.

L'autonomie stratégique passe également par la maîtrise des systèmes et de leurs composants afin de prévenir tout mésusage, indépendamment des questions de sûreté déjà évoquées. L'intégrité des données et des systèmes que se procurent les administrations auprès d'acteurs contrôlés ou liés à des puissances étrangères doit être assurée aussi bien par des garanties juridiques que par des contrôles techniques. Mais c'est surtout dans la **captation abusive de données** que réside le risque. La France doit, dans toute la mesure nécessaire et sans renoncer à l'ouverture consubstantielle à toute recherche, conserver la maîtrise de ses données. Ce n'est pas seulement parce qu'elles sont la nouvelle richesse inépuisable de l'économie numérique : c'est parce

[173] La question ne se pose pas différemment en Europe, où la production de semi-conducteurs ne représente que 10% du marché mondial. Les États membres de l'UE s'en sont saisis dès 2020, avec une déclaration commune dans laquelle ils convenaient de travailler ensemble afin de renforcer la chaîne de valeur européenne de l'électronique et des systèmes embarqués. En 2021, la Commission européenne a présenté le « Chips Act » (action européenne sur les semi-conducteurs), visant à renforcer la coordination des investissements européens et nationaux et à conjuguer les capacités européennes de recherche. Un paquet législatif a été présenté en février 2022, comprenant notamment deux propositions de règlements. Il s'accompagne de 43 milliards d'euros d'investissement, dont 11 milliards de crédits publics.

que collectées, élaborées, traitées par nos soins, elles ne peuvent ni ne doivent, pour les plus essentielles, servir d'autres intérêts que ceux du pays.

Or cette maîtrise ne peut être garantie dès l'instant que les données sont collectées par des acteurs soumis, par le droit, à l'obligation de les transmettre à une puissance étrangère (ou par la collusion, à la tentation de le faire spontanément…). Il ne suffit certainement pas, à cet égard, que les données soient physiquement hébergées sur des serveurs situés en France ou dans l'Union. Comme le montre la législation américaine, la nationalité d'une entreprise (*Cloud Act* de 2018), voire le simple fait qu'elle dispose d'un établissement ou d'intérêts dans le pays susceptibles d'être utilisés comme moyen de pression (menace d'une sanction), sont de nature, par le jeu de l'extra-territorialité, à la contraindre à déférer aux réquisitions de la puissance publique étrangère, comme les services de renseignement (exemple du FISA américain). C'est précisément en raison de ce risque que la Cour de justice de l'Union européenne a, à deux reprises, annulé les instruments qui entendaient encadrer les échanges de données à caractère personnel traitées avec les Etats-Unis[174].

L'exigence de « maîtrise » des systèmes d'information, entendue comme incluant la capacité des collectivités publiques à décider de ce que peut faire le système sans s'exposer à un usage détourné des données[175], en particulier à la transmission de ces données à l'étranger, figure d'ores et déjà à l'article 16 de la loi pour une République numérique. Mais cette obligation, exprimée en termes souples (« veillent à préserver »), s'apparente largement à un vœu pieux.

Sur le plan opérationnel, outre le recours à des infrastructures propres (modèle dit « *on-premise* »), **la conduite des procédures de la commande publique mérite une attention particulière** :

- d'une part, lorsqu'il importe d'assurer une complète maîtrise des données, eu égard à leur sensibilité et à leur valeur (par exemple lorsque le SIA touche aux intérêts fondamentaux de la Nation ou lorsqu'il s'agit de secteurs particulièrement sensibles, comme celui de la santé) et que le risque d'infructuosité du marché peut être écarté, il apparaît légitime que seules les entreprises entièrement soumises au droit de l'Union soient admises à candidater. Il ne s'agit pas de discriminer des soumissionnaires parce qu'ils sont étrangers, mais de prendre acte qu'en raison de leur dépendance juridique à une puissance étrangère, ils n'offrent pas les garanties objectives requises pour être retenus ;
- d'autre part, sauf cas particuliers, les contrats devraient définir avec précision les réutilisations autorisées des données par le prestataire, ménager à l'administration un droit de contrôle sur leur usage ou leur transfert, et garantir leur destruction à l'issue de la mission.

---

[174] V. en dernier lieu CJUE, Gr. Ch., 16 juillet 2020*, Data Protection Commissioner c/ Facebook Ireland Ltd et Maximillian Schrems*, C-311/18.

[175] V. l'intervention de la secrétaire d'État au numérique lors de la séance publique au Sénat du 27 avril 2016 lors de l'examen du projet de loi en première lecture, mentionnant l'accessibilité aux services de renseignement américains de données transitant par des routeurs largement utilisés par l'administration (« *back door* »).

Enfin, la France doit veiller à ce que la future réglementation européenne des SIA garantisse la **souveraineté juridique de l'Union** et fasse pièce aux manœuvres de contournement. L'article 2 de la proposition de règlement qui en définit le champ d'application répond à cette préoccupation : un SIA est soumis aux exigences européennes dès l'instant qu'il est mis sur le marché ou en service dans l'Union, quel que soit le lieu d'établissement du fournisseur, mais également si le fournisseur et l'utilisateur se trouvent en-dehors du territoire européen mais que le résultat généré par le système est utilisé dans l'Union.

Ainsi, lorsqu'une radiographie est envoyée dans un pays tiers où un système d'IA l'interprète, avant que le compte rendu rédigé par le système soit renvoyé sur le territoire de l'Union, où le praticien qui a prescrit l'examen l'utilisera, le système entre dans le champ d'application du règlement au titre de cette clause de compétence additionnelle.

Toutefois, pour être robuste, le système n'est pas entièrement étanche. Il faut d'abord que le citoyen européen sache qu'un système d'IA est utilisé, et que cette utilisation sera prise en compte pour nourrir la relation qu'il a avec un prestataire. En théorie, la loyauté et la sincérité des relations avec sa contrepartie lui donneront cette information ; en pratique, il n'est pas sûr que ce sera toujours le cas, ou que des habiletés ne permettront pas à des opérateurs de s'exonérer de leurs obligations.

3.2. La mise en œuvre normative de l'IA publique de confiance

Autant que le fond des exigences, la portée normative des contraintes – selon qu'elles prennent la forme de dispositions législatives ou réglementaires, de lignes directrices ou autres règles de droit souple susceptibles de recours ou d'invocation devant le juge, ou encore de simples orientations générales destinées à éclairer les administrations et dépourvues de tout effet juridique et de justiciabilité – conditionne à la fois la confiance qu'inspirent les SIA et la capacité de l'administration à les développer sans obstacles excessifs.

3.2.1. Le règlement IA et les espaces d'expression du droit national

Les développements qui précèdent montrent qu'il existe d'ores et déjà des règles de droit qui font écho aux principes fondamentaux de l'IA publique de confiance tels qu'ils sont proposés. Elles touchent essentiellement aux traitements de données à caractère personnel, à la transparence des traitements algorithmiques et à de grands principes communs aux systèmes d'information publics, complétées, en surplomb, par l'ensemble des normes supérieures qui font obstacle à ce que les conditions de fonctionnement et les résultats produits par un SIA n'aboutissent à une méconnaissance des droits et libertés fondamentaux.

Les nombreux angles morts législatifs seront en grande partie comblés, le cas échéant, par le règlement européen, dont le contenu, dans son état actuel, est synthétisé en annexe 6 – sans toutefois apporter de réponse à l'ensemble des

questionnements sectoriels que fait émerger le recours accru aux SIA[176]. Ce texte fixera le premier cadre juridique transverse dédié aux SIA au monde[177], afin d'harmoniser les conditions de fourniture et d'utilisation de ces systèmes au sein de l'Union et d'offrir aux opérateurs comme au public un niveau de garantie élevé pour leur santé, leur sécurité et leurs droits fondamentaux susceptibles d'être affectés par ces systèmes. Il importe toutefois de souligner que **ce règlement ne guidera pas totalement les choix normatifs de la France, pour quatre raisons**.

a/ En premier lieu, le règlement ne couvrira pas l'intégralité du champ de l'action publique, notamment en raison de ses **exclusions sectorielles**. Contrairement à d'autres actes de droit dérivé connexes, comme la directive dite *e-privacy*[178] ou le RGPD et la directive dite « police-justice », la proposition de la Commission n'exclut pas de son champ d'application **l'ensemble des matières classiquement exclues du champ du droit de l'Union**, à savoir la sécurité nationale (y compris le renseignement), la défense et la politique étrangère. La seule exclusion porte sur les SIA « développés ou utilisés exclusivement à des fins militaires » – terminologie dont il n'est pas certain qu'elle recouvre l'ensemble du champ de la défense. Le « texte de compromis » en cours de discussion prévoit d'y ajouter l'exception de sécurité nationale. Si la notion de « développement exclusif » est claire, en ce qu'elle couvre tout SIA qui a été spécifiquement et uniquement conçu à des fins militaires ou de sécurité nationale, celle d'« utilisation exclusive » recèle une ambiguïté : exclut-elle l'ensemble des SIA utilisés par l'armée ou les services de renseignement qui trouveraient leur équivalent dans d'autres domaines de l'action publique ? Cette équivalence doit-elle être stricte (produit standard) ou s'apprécier « en substance » (une légère adaptation du système pour les besoins du client militaire n'excluant alors pas l'application du règlement) ?

b/ En deuxième lieu, conformément à l'approche par les risques qui a été retenue, le projet de règlement IA ne prétend pas imposer d'obligations à l'ensemble des SIA, y compris publics. Outre les interdictions de principe qu'il envisage, **seuls les SIA dits « à haut risque »** feraient l'objet d'un régime juridique complet incluant la certification du produit par un tiers indépendant (« organisme notifié ») avant la mise sur le marché ou la mise en service. Ainsi qu'il a été dit, certains systèmes interagissant avec les usagers seraient en outre soumis à des obligations de transparence. Mais **l'ensemble des autres SIA ne feraient l'objet d'aucune prescription spécifique**, la proposition se bornant à encourager l'élaboration de « codes de conduite ». Il convient toutefois de tempérer ce constat à deux égards :

---

[176] V. par ex. dans le domaine de la propriété littéraire et artistique, à propos de la protection du droit d'auteur : CSPLA, *Rapport de la mission Intelligence artificielle et culture*, janvier 2020.

[177] Certains pays ont commencé à se doter de législations traitant spécifiquement d'intelligence artificielle. Tel est le cas, par exemple, des États-Unis, avec le *National Artificial Intelligence Initiative Act*, même si l'approche américaine repose davantage sur la normalisation (volontaire) que sur la réglementation obligatoire.

[178] Directive 2002/58/CE du Parlement européen et du Conseil du 12 juillet 2002 concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques.

- d'une part, **une proportion importante des systèmes d'IA publics relèveraient de la catégorie des systèmes « à haut risque »**. En effet, les domaines énumérés par l'annexe III du projet de règlement recouvrent l'essentiel des activités régaliennes (police, justice, immigration), l'accès et le droit aux services publics et aux prestations sociales, l'enseignement et la formation professionnelle, la gestion et l'exploitation de réseaux publics (eau, électricité…) ou encore le recrutement, l'évaluation, la promotion et la sortie de service des agents. Si, en l'état de cette annexe, seules certaines catégories de SIA se rattachant à ces domaines relèveraient de ce régime juridique, la Commission serait autorisée à compléter cette liste selon une méthodologie qui lui laisse d'importantes marges d'appréciation. **L'entrée en vigueur du règlement, dans son équilibre actuel, entraînerait donc d'importantes conséquences pour les administrations**. En tant que conceptrices de systèmes, elles seraient astreintes aux obligations pesant sur les fournisseurs. En tant que simples utilisatrices de systèmes fournis par des tiers, leurs obligations seraient moindres et le texte leur offrirait au contraire des garanties, au prix toutefois d'un renchérissement des prestations dès lors que les fournisseurs répercuteront sur les clients publics les coûts liés à la conformité des systèmes ;
- d'autre part, la simple inclusion du système dans le champ d'application du règlement, fût-ce au titre de la soumission à un code de conduite, emporte **l'inclusion dans le champ du droit de l'Union, ce qui entraîne l'application du droit primaire, notamment de la Charte des droits fondamentaux de l'Union européenne**, dont on sait à quel point, telle que l'interprète la Cour de justice de l'Union européenne, elle peut être exigeante à l'égard des autorités publiques, en particulier pour celles qui sont en charge de missions touchant à la sécurité. Cet effet doit être sérieusement pris en compte par les pouvoirs publics pour appréhender la portée réelle du texte, au-delà des prescriptions précises qu'il fixe, et apprécier s'il y a lieu de circonscrire le champ d'application du règlement aux SIA des trois catégories pour lesquelles des obligations précises sont fixées.

c/ En troisième lieu, les Etats membres sont susceptibles de conserver des **marges d'adaptation** ménagées par le règlement lui-même et, dans le respect des grands principes des traités, le droit de compléter les exigences résultant du règlement. On notera, sur ce point, que la Commission a fait le choix, qui n'avait rien d'évident, de proposer un **unique instrument juridique**, sous la forme d'un règlement d'effet direct, à la différence du parti retenu en matière de protection des données à caractère personnel, marqué par la dualité entre le RGPD et la directive « police-justice »). S'il est vrai que les principes de fonctionnement des SIA ne diffèrent pas selon la finalité poursuivie, de sorte que les obligations prévues par la proposition peuvent, dans leur principe, s'appliquer indifféremment aux SIA répressifs comme à la généralité des systèmes, on ne peut nier la spécificité des finalités de puissance publique et, singulièrement, celles qui touchent aux activités de police et de justice, qui pourraient appeler un encadrement européen plus souple.

d/ En quatrième et dernier lieu, le droit de l'Union ne préjuge pas du **niveau de norme** – loi ou règlement – nécessaire pour asseoir le recours à telle ou telle catégorie de SIA publics. Il s'agit là d'une question délicate qui se posera rapidement

et, potentiellement, de façon fréquente à l'administration et au juge : dans quels cas la conception et l'utilisation d'un SIA public requerront-ils un **fondement législatif dédié** ? Si certains garde-fous posés par le règlement européen constitueront aussi des règles relatives aux garanties fondamentales accordées aux citoyens pour l'exercice des libertés publiques au sens de l'article 34 de la Constitution, on peut penser que les SIA les plus intrusifs ou coercitifs pourront nécessiter un assentiment exprès du législateur et la fixation de garanties spécifiques. La liste des SIA « à haut risque » peut servir de point de repère à cet égard. On peut citer deux illustrations de cette problématique.

La première porte sur **l'analyse automatisée d'images captées dans l'espace public** par des dispositifs fixes ou embarqués permettant la détection de situations anormales, d'infractions ou de menaces, sans même qu'il soit procédé à l'identification des personnes physiques[179]. En dépit de l'absence d'identification et, en particulier, de mise en œuvre de traitements de reconnaissance faciale, de tels SIA sont susceptibles d'avoir des incidences plus ou moins importantes sur les libertés publiques selon l'usage auquel ils sont destinés. Selon la CNIL, ils présentent « *le risque de généraliser un sentiment de surveillance chez les citoyens, de créer un phénomène d'accoutumance et de banalisation de technologies intrusives* ». Il n'est pas certain que de tels traitements de données à caractère personnel puissent, en toute hypothèse, être créés sans base légale spécifique. Le développement rapide de ces outils plaide, par précaution, pour qu'il soit inséré dans la loi de 1978 un régime-cadre suffisamment souple pour englober l'ensemble des SIA recourant à ce procédé technique, de préférence à une autorisation législative au cas par cas. Ce régime devrait utiliser les souplesses offertes par le RGPD, lorsque ce dernier est applicable (ce qui n'est pas le cas des systèmes déployés à des fins de police administrative ou judiciaire) et, en particulier, les dispositions propres aux traitements ne nécessitant pas l'identification (article 11).

La seconde illustration concerne **l'exploitation massive et automatisée de données à caractère personnel mises en ligne par les personnes concernées**, en particulier sur les réseaux sociaux. Cette pratique, connue sous l'appellation de « *data scraping* » ou de « *web scraping* »[180], intéresse tout particulièrement les autorités investies de missions de contrôle et de sanction. Elle a été autorisée à titre expérimental pour les besoins de la détection de certaines infractions fiscales graves par la DGFIP par l'article 154 de la loi n° 2019-1479 du 28 décembre 2019 de finances pour 2020. L'utilisation de SIA basés sur les techniques d'apprentissage machine permet d'identifier dans cette masse d'informations des personnes ou des choses ou de détecter des anomalies permettant de cibler les contrôles ou de justifier la mise en œuvre d'autres pouvoirs d'enquête. Cet outil relevant du champ de la loi de

[179] V. pour ce cas d'usage l'annexe 9, fiche n° 3.

[180] V. pour une illustration récente s'agissant d'un SIA privé, la mise en demeure par la CNIL, en décembre 2021, de « *Clearview AI* », système de reconnaissance faciale dont la base de données repose sur le *scraping* de photographies et de vidéos publiquement accessibles sur Internet. La société a ainsi réussi à s'approprier 10 milliards d'images à travers le monde. La CNIL a identifié deux manquements : l'illicéité du traitement, en l'absence de base légale, et l'absence de prise en compte satisfaisante des droits d'accès et d'effacement.

finances, il ne peut être déduit de la décision du Conseil constitutionnel n° 2019-796 DC du 27 décembre 2019 que l'ensemble des SIA de cette nature requerrait une base législative dédiée. Mais le contrôle très pointilleux auquel s'est livré le juge constitutionnel et la « clause de revoyure » par laquelle il s'est réservé la possibilité, en cas de pérennisation du dispositif à l'issue de la phase d'expérimentation, de procéder à un nouvel examen de conformité à la Constitution et, le cas échéant, de censurer ces dispositions dans le cas où l'utilité du dispositif ne serait pas avérée, ne peuvent qu'inciter à la prudence. L'insécurité juridique actuelle plaide pour la conduite d'une réflexion dédiée et, s'il y a lieu, l'adoption d'un cadre législatif global que de nombreuses autorités de régulation appellent de leurs vœux. L'article 154 de la loi de finances pour 2020, pour la DGFIP, et l'article 36 de la loi n° 2021-1382 du 25 octobre 2021 relative à la régulation et à la protection de l'accès aux œuvres culturelles à l'ère numérique, pour sécuriser cette pratique mise en œuvre par le service à compétence nationale dénommé « Pôle d'expertise de la régulation numérique » (PEReN), pourraient servir de source d'inspiration.

Dans ce domaine du droit qui reste encore largement à défricher, il est dans l'intérêt des administrations de s'appuyer sur les ressources du Conseil d'Etat, notamment à travers les **demandes d'avis aux sections administratives**.

3.2.2. L'apprentissage de la liberté : les lignes directrices de l'IA publique de confiance

1. L'opportunité d'anticiper l'entrée en vigueur du règlement IA

Dans l'attente de l'entrée en application du règlement, il appartient aux pouvoirs publics de déterminer la nature des contraintes qu'ils entendent faire peser sur les administrations. Si aucun calendrier n'a été avancé officiellement, le règlement européen sur les systèmes d'IA en cours de négociation ne sera probablement pas adopté avant deux à trois ans. A l'instar du RGPD, la proposition de la Commission a prévu qu'il n'entrerait en application que deux ans après son entrée en vigueur (à l'exception des dispositions relatives à la gouvernance et aux organismes de certification, qui entreraient en application dans les trois mois), laquelle interviendrait vingt jours après sa publication. Les obligations qu'il fixe pourraient donc n'être effectives que dans quatre à cinq ans.

La question d'une mise en œuvre anticipée doit être posée, à tout le moins pour les SIA publics. Elle présenterait de multiples avantages.

En premier lieu, ainsi qu'il a été dit, elle contribuerait immédiatement au **renforcement de la confiance** dans ces systèmes, de nature à faciliter la mise en place de la stratégie de déploiement de l'IA publique. Comme on l'a vu, le cadre juridique national est loin d'apporter toutes les garanties que prévoit la proposition de la Commission. En particulier, les SIA ne constituant ou n'impliquant pas des traitements de données à caractère personnel sont très peu encadrés, alors même qu'ils peuvent emporter des conséquences majeures sur les droits et la situation des personnes. Au demeurant, même avec l'adoption du règlement, une partie importante des systèmes d'IA des administrations restera soustraite à l'essentiel des règles qu'il pose, réservées aux systèmes dits « à haut risque ».

En deuxième lieu, l'État serait conduit à effectuer, en amont de l'entrée en application du règlement, un certain nombre de **choix qu'il lui appartiendra en tout état de cause de faire sous l'empire de ce texte**, le moment venu, qu'il s'agisse de l'organisation de la fonction de régulation, de contrôle et de certification, du régime des sanctions administratives, de l'identification précise des systèmes d'IA « à haut risque » ou encore des choix de nature plus politique d'encadrer ou d'interdire certains systèmes dans les domaines les plus sensibles. On ne peut se satisfaire de la relative précipitation et du retard dans lesquels le droit national a été adapté au RGPD et la directive « police-justice » transposée en droit interne.

En troisième lieu, une application « brutale » de la nouvelle réglementation entraînerait inévitablement des **surcoûts** et des **risques de non-conformité**. Le RGPD a montré l'exemple d'une réglementation, certes reconnue comme adéquate et devenue un standard mondial, mais qui a entraîné un coût d'adaptation et d'entrée dans le système particulièrement élevé pour les entreprises et les administrations[181]. Il importe en effet de rappeler qu'il ne suffit pas d'être effectivement « aux normes » : la conformité implique de la documenter, ce qui représente, spécialement pour des administrations qui ne sont pas rompues à cet exercice formel, une lourde charge rédactionnelle. Le lissage de la mise en conformité est de nature à limiter le recours en urgence à des prestataires qui seront très sollicités, souvent enclins à servir prioritairement le secteur privé au risque de délaisser les consultations au titre de la commande publique, et d'inégale fiabilité. Il permettrait une montée en compétence plus progressive des agents.

En quatrième et dernier lieu, si la proposition de la Commission ne prévoit pas de soumettre aux nouvelles exigences les **systèmes mis en service avant son entrée en application**, ces derniers en relèveront néanmoins s'ils subissent d'importantes modifications de leur conception ou de leur destination. A ce titre, le respect des exigences par le système d'origine est de nature à faciliter la mise en conformité de la version modifiée.

2. La forme de l'anticipation : les lignes directrices

Le Conseil d'État ne recommande pas que cette mise en œuvre anticipée prenne la forme d'une législation-cadre nouvelle. Sauf à la circonscrire aux SIA publics, celle-ci introduirait des distorsions de concurrence fâcheuses entre les opérateurs privés français et leurs homologues européens. Elle risquerait de s'avérer inadaptée à la diversité des SIA et à leur évolution rapide, et d'en entraver le développement par inadvertance. En cristallisant le droit national, elle fragiliserait les positions de négociation de la France et pourrait emmener l'action publique dans des directions que le règlement abandonnerait finalement. Enfin, en fixant des règles dont la méconnaissance pourrait être sanctionnée par le juge, elle serait source de contentieux, et ce d'autant plus qu'elle comporterait des termes non définis ou des obligations ambiguës.

[181] Pour les années 2018, le Syntec a évalué le coût de mise en conformité des entreprises françaises au RGPD à près d'un milliard d'euros par an.

Ces écueils peuvent être contournés en recourant à des **lignes directrices qui formaliseraient tout à la fois la stratégie, la doctrine d'emploi et la méthodologie pratique de conception, de déploiement et d'utilisation de systèmes d'IA de confiance au sein de la sphère publique**. Ces orientations guideraient l'action des administrations de l'État, sous la supervision de la DINUM, et auraient vocation à inspirer les initiatives locales. Elles constitueraient aussi une préfiguration des « codes de conduite » dont la proposition de règlement encourage l'élaboration pour les systèmes autres que ceux à haut risque.

a/ Ces lignes directrices emprunteraient leurs principaux concepts à la proposition de règlement IA, contribuant ainsi à **l'harmonisation terminologique** que la première partie de la présente étude appelle de ses vœux. Ces concepts pourraient être précisés, notamment à la lumière des travaux de normalisation, et illustrés.

b/ Au titre de la **stratégie** et de la **doctrine d'emploi**, elles pourraient reprendre les grands axes décrits dans la deuxième partie de la présente étude et, en particulier, rappeler que les SIA ne sont qu'un outil dans la boîte dont disposent les administrations, qui peut s'avérer inadapté aux besoins. Elles pourraient définir les systèmes que l'État s'interdit de déployer pour des raisons éthiques et politiques et distinguer ceux qui appellent une vigilance particulière (SIA à haut risque) de la généralité des systèmes. Elles devraient notamment comporter une **doctrine de recours à la prise de décision automatisée** écartant toute prohibition de principe et toute diabolisation de ce mode de prise de décision. Ce dernier pourrait notamment être privilégié pour la prise de décisions d'acceptation insusceptibles de préjudicier aux tiers ni de compromettre gravement un intérêt public, et pour les situations de compétence liée, dans lesquelles l'administration n'a pas d'appréciation à porter mais des règles précises à appliquer sur la base de faits « objectifs ».

c/ Ces lignes directrices rappelleraient ensuite les **règles de droit d'ores et déjà applicables** en la matière et, surtout, éclaireraient les administrations sur **l'articulation des différents corps de règles existants**. Les administrations qui envisagent ou entreprennent la conception d'un SIA se heurtent tout particulièrement à la multiplicité des normes issus de *corpus* juridiques différents, qui résultent de la sédimentation normative, fruit de la facilité de l'empilement et de la difficulté de la synthèse. Il importe de ne pas se limiter à l'ajout d'une couche supplémentaire de contraintes, mais de leur offrir une vision d'ensemble du cadre dans lequel elles doivent s'insérer. Cette préoccupation a été exprimée s'agissant du règlement IA par les délégations nationales au sein du Conseil, ainsi que par le Comité européen de la protection des données pour ce qui concerne spécifiquement l'articulation avec le RGPD. A l'instar du code du travail numérique, le recours à un SIA pour garantir l'accessibilité et l'ergonomie pratique du « droit de l'IA publique » pourrait s'avérer utile.

Une bonne illustration de la multitude des contraintes qui pèsent sur les administrations est donnée par l'utilisation, déjà évoquée, des SIA pour l'exploitation massive des données en ligne, en particulier par les autorités de régulation.

Les principales questions juridiques soulevées par la collecte et l'exploitation massive de données en ligne par les autorités de régulation

- Les autorités de régulation ont-elles besoin d'une base légale spécifique pour être autorisée à recourir à cette technique, ou leur « statut » législatif suffit-il en ce qu'il leur permet, de façon générale, de procéder aux investigations nécessaires à l'application des règles dont elles contrôlent le respect ?

- Le recours à cette pratique nécessite-t-il une disposition législative au titre des règles concernant les garanties fondamentales accordées aux citoyens pour l'exercice des libertés publiques, notamment en raison de l'atteinte portée au droit au respect de la vie privée et à la liberté d'expression ?

- La collecte et l'exploitation subséquente des données portent-elles une atteinte disproportionnée au droit au respect de la vie privée et à la liberté d'expression ? En particulier, y a-t-il lieu de limiter ce dispositif aux manquements dont la commission est rendue possible ou favorisée par l'usage d'internet et à des manquements d'une particulière gravité ou difficilement détectables par d'autres moyens ? Quelles doivent être les limitations apportées au champ des données collectées (exclusion des données sensibles, des données de tiers, de données accessibles *via* une inscription ou la saisie d'un mot de passe, absence de recours à la reconnaissance faciale…) ? De quelles garanties l'exploitation des données doit-elle être assortie (habilitation d'agents, formation de ces derniers, secret professionnel, destruction à très bref délai des données non pertinentes et durée de conservation limitée, absence de décision ou de déclenchement d'un contrôle sur le seul fondement du traitement algorithmique, débat contradictoire sur les données, droits des personnes intéressées…) ?

- Les conditions générales d'utilisation (CGU) du site ou de l'application sur lequel les données sont collectées sont-elles opposables aux autorités de régulation, notamment en ce qu'elles proscriraient toute pratique de « *scraping* » ou l'utilisation d'API par des autorités publiques ? Doivent-elles être levées par une disposition législative, au titre des principes fondamentaux des obligations civiles et commerciales ? Quelles conséquences juridiques s'attacheraient à la méconnaissance de ces CGU par l'autorité, notamment sur la régularité de la procédure subséquente ?

- Comment concilier cette méthode d'investigation avec les droits de propriété intellectuelle dont peuvent être grevées les données et les bases de données ? Au titre du droit *sui generis*, le producteur de la base de données peut-il interdire certaines modalités d'extractions (articles L. 342-1 et L. 342-2 du code de la propriété intellectuelle) ?

- Quelles sont les conséquences juridiques d'une sollicitation excessive du site ou de l'application (surcharge) par l'autorité administrative entraînant son dysfonctionnement ou son interruption ? Quel régime de responsabilité convient-il d'appliquer en pareil cas ?

e/ Enfin – et ce serait là leur objet principal – les lignes directrices compléteraient ce cadre juridique par l'énoncé d'exigences découlant des principaux généraux de l'IA publique et de la méthodologie de leur mise en œuvre, sous la forme d'une **charte du recours à l'IA dans le secteur public**. Ce volet du document devrait impérativement se concevoir de façon souple et évolutive, afin d'incorporer, en plusieurs étapes si nécessaire et en tenant compte d'un éventuel consensus européen, les principales exigences du règlement. Celles-ci apparaissent en effet largement divisibles : il est tout à fait concevable, par exemple, de fixer dans un premier temps les exigences applicables en matière de données d'apprentissage, avant de définir les conditions du contrôle humain. Les exigences posées pourraient utilement s'inspirer de référentiels de conformité existants, comme celui publié par le Laboratoire national de métrologie et d'essais (LNE). Le document devrait encourager les administrations à faire évaluer les systèmes et à obtenir (en tant que fournisseur) ou exiger (en tant qu'utilisateur) une certification volontaire, dans l'attente des obligations européennes.

Plusieurs précisions doivent être apportées s'agissant de la **méthodologie de déploiement des SIA publics de confiance**.

En premier lieu, les principes généraux précédemment exposés et les exigences opérationnelles qui en découlent ne sont **pas des absolus**. Ils peuvent entrer en conflit ou en tension les uns avec les autres, appelant un **arbitrage** entre eux[182]. Le seul fait qu'un SIA ne respecte pas pleinement une exigence donnée ne fait pas en soi obstacle à sa conception et à son utilisation, pourvu qu'il soit compensé par les bénéfices attendus du système et comportent des mesures correctives suffisantes pour limiter l'ampleur et les effets de cette entorse. Il est crucial à cet égard de ne pas verser dans une vision absolutiste de « l'éthique de l'IA » qui aboutirait à une regrettable autocensure, préjudiciable à l'intérêt général. Les conflits ou tensions entre les exigences doivent autant que possible être résolues par des **modes de gouvernance appropriés**, associant au maximum l'ensemble des parties prenantes. L'acceptable ne peut être défini que collectivement ou, à tout le moins, en tenant compte des nombreux points de vue qui peuvent légitimement s'exprimer (associations d'usagers, organisations syndicales, sociétés savantes…). La question de l'architecture idoine du contrôle éthique est abordée au 4.2.4.

En deuxième lieu, la décision de principe de recourir à un SIA pour l'accomplissement d'une activité de service public devrait être systématiquement précédée d'un **bilan coûts/avantages** intégrant l'ensemble des exigences (création de valeurs attendue contre inconvénients avérés et risques prévisibles), en veillant à documenter les arbitrages rendus. Le degré d'approfondissement de l'exercice et la précision

[182] Ex. : un arbitrage peut être nécessaire entre l'amélioration de la performance du modèle, en particulier son exactitude, par sa complexification, d'une part, et le besoin d'explicabilité et la soutenabilité environnementale, qui plaident pour recourir à des modèles plus simples. De même, la correction d'un modèle en vue d'accroître son équité (en supprimant des biais discriminatoires au détriment d'une catégorie de personnes) peut en réduire la performance au regard des objectifs fixés. Un autre exemple est celui du choix d'un opérateur étranger proposant des services plus performants mais dont le recours peut affaiblir l'autonomie stratégique du pays.

formelle dans la restitution de ses résultats doivent évidemment être fonction de la sensibilité de l'activité et des données utilisées, de la maturité de la technologie utilisée, ainsi que de l'ampleur et de la gravité des risques. Les SIA les plus simples n'appellent qu'une analyse très sommaire et le rappel de principes généraux. Les plus complexes et sensibles, en revanche, nécessiteront une déclinaison opérationnelle des principes et des garanties renforcées, y compris sur le plan procédural. Il importe de rappeler que l'analyse peut avoir été largement défrichée par le fournisseur d'un SIA que l'administration se borne à acquérir « sur étagère ».

En troisième et dernier lieu, l'administration doit impérativement veiller à **l'effectivité des garanties** dont la conception et le fonctionnement du SIA sont assorties. On voit, avec le RGPD, comment une garantie communément admise comme protectrice, comme l'est le consentement préalable, peut s'avérer dans certains cas fictive ou illusoire compte tenu de la réalité des pratiques (le consentement au dépôt de cookies peut procéder de la lassitude ou de l'impatience de l'internaute qui souhaite se débarrasser d'un bandeau gênant et accéder sans tarder au contenu recherché, et non d'une réflexion éclairée sur les conséquences de ce geste). Le droit d'information, même encadré pour que les données fournies soient intelligibles, n'offre souvent qu'une transparence de façade. L'exercice des droits d'accès et de rectification, et plus encore celui des nouveaux droits consacrés par le RGPD (limitation, portabilité), demeure une démarche assez résiduelle au regard du nombre de traitements et de personnes concernées[183]. S'y ajoute le manque de moyens de la CNIL, unanimement constaté par les personnes auditionnées, qui fragilise l'effectivité des droits des personnes concernées.

L'intelligence artificielle accroît considérablement l'acuité du problème en raison notamment de la complexité des modèles et de l'inaccessibilité des codes-sources à l'entendement du citoyen non expert comme de la difficulté pour la personne concernée à percevoir l'intérêt que présente pour elle le développement d'un SIA et le profit qu'elle peut en retirer personnellement, à supposer qu'il ne soit pas conçu pour prendre des décisions défavorables à son égard.

f/ L'élaboration de ces lignes directrices, et, ultérieurement, leur évaluation, devrait associer l'ensemble des parties prenantes, y compris le secteur privé, lequel sera nécessairement impacté par ces orientations qui guideront la rédaction des cahiers des charges de la commande publique. Il serait souhaitable qu'elle fasse l'objet d'une concertation étroite avec la **future autorité de contrôle et de régulation**, dont la préfiguration devrait intervenir le plus tôt possible afin de favoriser une montée en puissance progressive (V. 4.2.3. sur cette question).

[183] A titre d'illustration, s'agissant du droit au déréférencement consacré par l'arrêt de la CJUE du 13 mai 2014, *Google Spain*, C-131/12, Google a recensé, s'agissant de la France, 288 000 demandes concernant 963 000 URL, et s'agissant de l'ensemble des États membres, 1 235 000 demandes concernant 4 800 000 URL. En 2020, la CNIL a reçu 382 plaintes à ce sujet, en baisse de 9,5%. Ces résultats ne sont pas anodins, mais l'on aurait pu s'attendre à ce que les citoyens se saisissent davantage de cette possibilité.

Au regard de leur ampleur et de leur portée, ces lignes directrices, préfiguration expérimentale sous évaluation constante devant éclairer les futurs choix législatifs, devraient associer très étroitement le Parlement, à deux stades. D'abord, lors de la réflexion sur leur élaboration, en permettant aux commissions intéressées (ou spécialement créées, le cas échéant), avec le concours de l'Office parlementaire d'évaluation des choix scientifiques et technologiques, de se prononcer en toute indépendance sur les choix stratégiques et de proposer au Gouvernement des inflexions ou des enrichissements, notamment pour ce qui concerne le déploiement des SIA dans les collectivités territoriales. Ensuite, dans le suivi et l'évaluation de la mise en œuvre des lignes directrices, qui devrait donner lieu à la remise d'un rapport régulier. Ces travaux nourriraient la réflexion parlementaire dans la perspective du débat sur l'adoption des dispositions législatives requises pour assurer la pleine application du règlement IA et, en tant que de besoin, le compléter.

3.3. Le recours juridictionnel, facteur-clé de confiance dans l'IA publique

Il ne peut y avoir d'IA publique de confiance sans droit au recours effectif. Eu égard aux incidences que ces systèmes sont susceptibles d'avoir tant sur la vie des individus que sur celle de la Nation, il est impératif que les personnes qui s'estiment lésées par un tel système puissent faire respecter par les administrations les règles de droit qui s'imposent, et s'imposeront, à elles, dans les conditions précédemment mentionnées, et trouver une réponse dans le prétoire.

Ce droit au recours doit porter aussi bien sur l'annulation des actes relatifs aux SIA publics que sur la réparation des préjudices qu'ils ont causés, et, le cas échéant, sur l'engagement de la responsabilité pénale. Les développements qui suivent ne constituent qu'un survol de ces questions complexes et délicates, qui ne seront éclairées que progressivement par la jurisprudence.

3.3.1. Les voies de recours contre le système et ses effets

Une telle action peut s'envisager à deux niveaux.

Peuvent d'abord être contestés, en aval, les **résultats produits par le SIA**, au premier rang desquels figurent les décisions individuelles prises de manière automatisée et dont les requérants sont les destinataires. Dans ce cas, on peut penser que l'essentiel du débat contentieux sera, en général, étranger à l'intervention du système et se focalisera, comme aujourd'hui, sur la conformité de la décision produite à la règle de droit[184]. Il pourrait toutefois être soutenu, sous l'angle de l'incompétence ou du vice de procédure, que la décision contestée ne pouvait être légalement prise par le système, à l'aune des règles qui encadrent le recours à un tel mode d'édiction des actes administratifs, rappelées en annexe 10.

[184] Le Conseil constitutionnel a rappelé dans sa décision n° 2018-765 DC du 12 juin 2018 que la prise de décision automatisée n'autorisait pas l'administration à adopter des décisions sans base légale ou à appliquer d'autres règles que celles du droit en vigueur.

Peut aussi être envisagé un recours en amont contre la **décision de principe de recourir à un SIA**, l'acte qui le crée ou décide de sa mise en service et/ou celui qui en définit les caractéristiques, afin d'empêcher son déploiement ou de faire cesser son utilisation.

D'ores et déjà, le juge administratif peut être amené à connaître des traitements algorithmiques reposant sur des données à caractère personnel, à l'occasion d'un recours pour excès de pouvoir contre l'acte réglementaire qui le crée, notamment lorsqu'un tel acte est exigé par les articles 31 et 32 de la loi du 6 janvier 1978, ou à l'occasion d'un recours de plein contentieux contre une mesure de sanction prise par la formation restreinte de la CNIL en cas de manquement aux règles de protection des données à caractère personnel.

Toutefois, dans les nombreux cas où le traitement ne fait l'objet d'aucun décret ou arrêté de création, il peut s'avérer plus difficile d'en contester la légalité autrement qu'en saisissant la CNIL, si toutefois il implique des données à caractère personnel. Au moins deux voies de droit pourraient être envisagées.

D'une part, dès l'instant que le recours pour excès de pouvoir est ouvert contre tous les documents de portée générale émanant d'autorités publiques, dès lors qu'ils ont des **effets notables** sur les droits ou la situation de tiers à l'administration[185], le juge administratif pourrait éventuellement être saisi d'un recours contre un document définissant les modalités de fonctionnement d'un SIA ayant de tels effets. Le Conseil d'Etat, statuant au contentieux, a déjà accepté de connaître d'une instruction qui définissait les caractéristiques d'un traitement de données à caractère personnel alors qu'aucun texte n'imposait sa création par un acte réglementaire[186]. Se posera également la question de la possibilité de saisir un **acte révélé** par la mise en service concrète d'un système. S'agissant de l'existence d'effets notables, il est vraisemblable que, à tout le moins, les SIA classés à haut risque par la législation européenne en relèveraient. La liste des critères de classement qui figureraient dans le règlement pourrait aussi servir d'indices[187].

D'autre part, la voie du référé-liberté peut être empruntée, en cas d'atteinte grave et manifestement illégale à une liberté fondamentale résultant de l'utilisation d'un SIA, pour demander en urgence l'arrêt ou la modification de ce système, indépendamment de l'existence d'une décision. Cette voie, comme d'ailleurs celle du référé-suspension, n'est toutefois ouverte qu'en cas de dommage réalisé ou imminent. Ainsi, un décret qui se borne à autoriser la conception et le développement technique d'un traitement algorithmique, mais non sa mise en service opérationnelle, ne crée pas une situation d'urgence justifiant la suspension de son exécution[188]. Dès lors que le contrat liant la

[185] CE, Sec., 12 juin 2020, *GISTI*, n° 418142.

[186] CE, 6 novembre 2019, *Fédération des acteurs de la solidarité et autres*, n° 434376-434377.

[187] V. l'article 7 du projet, qui mentionne, en substance, la destination du SIA, son utilisation effective ou projetée, l'existence réelle ou plausible d'atteintes à la santé, à la sécurité ou aux droits fondamentaux et leur gravité, la dépendance au système, la réversibilité des résultats, l'existence de mesures d'atténuation et de réparation efficaces (hors dommages et intérêts).

[188] JRCE, 26 mai 2020, *Sté Gerbi Avocat victimes & préjudices et autres*, n° 440378, à propos du projet « DataJust ».

Plateforme des données de santé, dite « *Health Data Hub* », et Microsoft, ainsi qu'un arrêté ministériel, prévoient l'impossibilité de tout transfert de données en dehors de l'Union européenne, le traitement de ces données par Microsoft ne constitue pas en lui-même une illégalité grave et manifeste justifiant la suspension du traitement des données par la plateforme mais, en raison du risque que les autorités américaines demandent l'accès à ces données dans le cadre de certains programmes de surveillance, le juge des référés peut demander de prendre une série de mesures de précaution pour garantir la protection des données personnelles, dans l'attente du choix d'un nouveau sous-traitant[189].

Trois exemples étrangers d'interventions juridictionnelles en matière d'IA

Des juridictions étrangères ont déjà enjoint à des administrations de cesser d'utiliser un SIA, qu'il s'agisse d'un système expert ou d'un système fondé sur l'apprentissage machine[190].

La Pologne avait mis en place un système automatisé de profilage des chômeurs, permettant de les ranger dans une catégorie (accompagnement renforcé ou non, droit à une allocation ou non…) sur la base d'un entretien et d'un test portant sur 24 critères. Le système avait fait l'objet de multiples critiques liées à son opacité (impossibilité pour les demandeurs d'emploi de connaître leur score et ses modalités de calcul) et au biais d'automatisation (moins de 1% des suggestions de la machine avaient été remises en cause par les agents). En 2019, le Tribunal constitutionnel polonais a déclaré cet outil contraire à la Constitution.

Dans un jugement du 5 février 2020 (C/09/550982/HA ZA 18-388), la Cour de district de La Haye (Pays-Bas) a déclaré contraire à l'article 8 de la convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales le cadre juridique de l'outil SyRi de détection des fraudes aux prestations sociales. Tout en admettant que le système répondait à une finalité d'intérêt général et en relevant que la prise de décision n'était pas automatisée mais simplement assistée par cet outil, la juridiction a estimé qu'il n'était pas assorti de garanties suffisantes pour assurer une conciliation équilibrée entre le but poursuivi et le droit de mener une vie privée normale, faute d'être suffisamment transparent et vérifiable. En particulier, il a été reproché à l'État de ne pas être en capacité d'expliquer quelles données factuelles pouvaient conduire à regarder un bénéficiaire de prestations sociales comme présentant un risque particulier, ni le type d'algorithme utilisé, et de ne pas mettre les personnes intéressées à même de savoir si leurs données personnelles ont fait l'objet d'un traitement régulier et de contester le fait que le système a émis un « rapport de risque » les concernant. La Cour a également émis des doutes quant à l'absence alléguée de biais discriminatoires.

Aux Etats-Unis, le tribunal fédéral du Texas (cas *Houston Federation of Teachers v. Houston Independent School District*, 4 mai 2017) a regardé comme contraire au « *procedural due process of law* » (exigences procédurales encadrant les

[189] JRCE, 13 octobre 2020, *Association Le Conseil national du logiciel libre et autres*, n° 444937.

[190] V. aussi la décision de la Cour d'appel d'Angleterre et du Pays de Galles citée dans la fiche n° 3 de l'annexe 8 en matière de « police prédictive ».

décisions administratives) le système EVAAS d'évaluation des performances des enseignants de la ville de Houston, qui calculait un score en fonction de l'augmentation moyenne des performances des élèves sur des examens standards, comparée à celle d'autres classes de l'État. Constatant qu'il n'était pas possible aux enseignants d'exiger une vérification indépendante de l'exactitude du score qui leur était attribué, alors que le fournisseur du système admettait que des erreurs étaient possibles et que le score pouvait être pris en compte pour licencier des agents, le tribunal a estimé que les droits de la défense n'étaient pas assurés. La décision juge en revanche que le principe du recours à un tel algorithme n'est pas illégal et que son intelligibilité pour les enseignants concernés est suffisante.

Il convient toutefois de rester conscient de deux limites liées à l'intervention du juge administratif.

D'une part, ce dernier n'a pas vocation à se substituer à l'administration. Cela implique, en premier lieu, que l'administration doit rester en première ligne dans le contrôle de ses propres décisions. Si la Constitution exige que toute personne destinataire d'une décision automatisée ait la possibilité de former un recours administratif traité par un agent public, il pourrait être envisagé qu'un tel recours administratif constitue un **préalable obligatoire** à l'introduction d'un recours contentieux. Ce faisant, l'administration corrigerait ses propres erreurs et, le cas échéant, le modèle algorithmique qui en est à l'origine, afin que soient portés devant le juge des litiges résultant de désaccords de fond, et non de dysfonctionnements techniques ou d'une erreur du système. En second lieu, le juge n'a pas à substituer sa propre appréciation aux **choix de pure opportunité** que l'administration peut faire. Sous réserve d'interdictions de principes découlant des textes (comme l'envisage le projet de règlement européen pour certains usages sensibles), le choix de principe de recourir à un SIA, plutôt qu'à une intervention humaine, paraît devoir rester largement soustrait au contrôle juridictionnel.

D'autre part, le contrôle juridictionnel de la légalité d'un SIA ou des décisions prises sur son fondement peut se heurter à d'importantes **difficultés techniques**, même si la palette des pouvoirs d'instruction du juge s'est opportunément élargie dans la période récente (outre l'expertise, il peut notamment formuler une demande d'avis sur une question technique ou faire appel à un *amicus curiae*). À cet égard, il est nécessaire de développer la formation des juges. Il serait aussi opportun de constituer rapidement un vivier d'experts et de techniciens indépendants, susceptibles d'être sollicités dans le cadre d'un contentieux[191].

[191] Il convient de souligner que le recours à des experts pour examiner les aspects techniques d'un dossier particulièrement complexe sur ce point, pourrait venir allonger les délais de règlement du litige.

Il reviendra enfin au juge administratif de **définir son office** lorsqu'il lui sera demandé d'ordonner l'interruption ou l'abandon d'un SIA de l'administration. On ne peut exclure que l'innovation numérique suscite la créativité jurisprudentielle. En particulier, la place croissante qu'occuperont les systèmes d'IA dans le fonctionnement de l'administration, au point parfois de soulever des questions de dépendance et de réversibilité, et donc de continuité du service public[192], pourra créer des situations dans lesquelles le rétablissement immédiat de la légalité pourrait entraîner, du point de vue de l'intérêt général, un mal plus redoutable que le remède. D'ores et déjà, l'atteinte excessive à l'intérêt général peut justifier la modulation dans le temps des effets de l'annulation contentieuse d'un acte administratif unilatéral, le refus d'annuler un contrat illégal, celui d'ordonner l'effacement des données à caractère personnel illégalement traitées ou de démolir un ouvrage public mal planté, ou encore le refus de rétrocéder un bien illégalement préempté. La question d'une transposition ou d'une adaptation de cette jurisprudence à l'utilisation des SIA ne manquera pas de se poser, lorsque la régularisation de la situation n'apparaîtra pas possible à brève échéance.

3.3.2. La mise en jeu de la responsabilité de l'administration devant le juge administratif

La construction d'un régime de responsabilité « du fait des SIA » publics devrait viser à concilier de la façon la plus équitable possible la sécurité juridique des victimes de dommages provoqués par de tels systèmes, qui sont en droit d'obtenir réparation des préjudices en résultant, afin de garantir la confiance dans ces outils, et celle des fournisseurs et des utilisateurs, afin de ne pas les dissuader d'innover.

Les systèmes d'IA soulèvent des difficultés particulières à cet égard, tenant à l'identification du fait générateur du dommage et du responsable correspondant et à la caractérisation du lien de causalité. Celles-ci résultent principalement de la multiplicité des intervenants et des composants des systèmes, de leur imbrication, dans un même dispositif, avec d'autres briques logicielles, de l'opacité de certains systèmes qui complique la traçabilité des erreurs, de la capacité d'apprentissage en continu de certains algorithmes en phase de déploiement et du rôle de « coproducteur » qui peut ainsi être dévolu à l'utilisateur.

Ainsi, le dysfonctionnement d'un système d'IA peut trouver sa source dans de multiples causes, tenant aussi bien à un **défaut de conception ou de développement imputable au fournisseur –** un jeu de données d'entraînement ou de validation inadapté (données de mauvaise qualité en raison d'un choix inapproprié, d'un travail d'annotation défectueux…), à un mauvais choix d'algorithme, à un défaut de programmation, ou encore à la défaillance d'autres composants que la brique reposant sur l'IA… – qu'à un **défaut d'utilisation** – par exemple, l'utilisation d'un SIA dans un environnement pour lequel il n'a pas été conçu, l'absence de ré-apprentissage pourtant prescrit par le producteur ou un ré-apprentissage vicié, l'absence de signalement d'un dysfonctionnement, la poursuite de l'utilisation du système en dépit des défaillances constatées… – ou encore à **l'intervention**

[192] V. 3.1.1 sur le principe de primauté humaine.

malveillante d'un tiers, ce qui justifie alors la mise en jeu de sa responsabilité civile, sans préjudice de fautes commises par le producteur en raison de l'insuffisante sécurisation du produit ou par l'utilisateur dans son usage imprudent ayant rendu possible ou facilité l'attaque du système.

L'identification du fait générateur du dommage peut ainsi requérir des raisonnements contrefactuels complexes, permettant de déterminer si le même dommage aurait résulté d'un système conçu, maintenu ou utilisé différemment.

Il importe toutefois de ne pas surestimer la novation que représente la diffusion des SIA dans l'action administrative, au regard des questions de responsabilité administrative. Les particularités techniques des SIA précédemment décrites ne modifie pas la nature, mais le degré de complexité des questions posées.

A ce titre, il y a lieu d'exclure d'emblée la thèse audacieuse analysant le SIA comme un **sujet de droit** autonome, disposant d'une personnalité juridique et d'un patrimoine propres, dont la responsabilité pourrait être actionnée par la victime. Il en va ainsi quelles que soient les facultés d'apprentissage dont il est doté, et quand bien même le système prendrait la forme d'un robot humanoïde. Un tel saut conceptuel, dont on peine du reste à appréhender la viabilité pratique[193], apparaît tout à la fois inopportun, en ce qu'il alimente le fantasme de la personnification de l'IA, et largement inutile, eu égard aux régimes de responsabilité existants. Les systèmes d'IA ne sont que des outils, au même titre que tout logiciel auquel l'administration peut avoir recours. Et le principe de primauté humaine précédemment exposé doit se traduire par la mise en jeu de la responsabilité des personnes physiques et morales impliquées[194].

Du point de vue des destinataires de l'action administrative, victimes d'un dommage, on peut penser que la responsabilité pour faute constituera le régime principal à leur disposition pour obtenir réparation des préjudices résultant de l'action administrative automatisée ou assistée par l'IA.

Dans de nombreuses hypothèses, et à l'instar de ce qui a été exposée à propos des actions en annulation, la faute pourra être saisie dans le **résultat** auquel l'utilisation du système a abouti ou concouru, sans qu'il y ait d'intérêt à discuter des raisons pour lesquelles ce système a pu dysfonctionner.

D'une part, une **décision administrative illégale**, qu'elle soit prise ou non par ou à l'aide d'un SIA, est toujours fautive et de nature à engager la responsabilité de l'administration dont elle émane. Il n'est pas concevable de faire peser sur les usagers le risque d'erreur inhérent au recours à tout outil numérique, en particulier aux systèmes reposant sur la logique probabiliste de l'apprentissage machine, qui comporte toujours un certain taux d'erreur. C'est l'administration qui doit en assumer

---

[193] V. sur ce sujet le rapport du groupe de travail de la cour d'appel de Paris, *La réforme du droit français de la responsabilité civile et les relations économiques*, avril 2019, pp. 107 et s.

[194] Il ne sera pas ici question de l'engagement de la responsabilité de l'administration à raison de l'absence ou du refus d'utilisation d'un système d'IA (V. sur cette question les développements consacrés à la revendication d'un « droit à l'IA » en 2.3.1.).

la responsabilité à l'égard des citoyens, ce qui présente l'effet vertueux de l'inciter à recourir aux SIA les plus performants et les plus fiables. Ce raisonnement est plus évident encore en cas de prise de décision assistée, personne ne songeant à exonérer l'administration de sa responsabilité ou à créer un régime spécial de responsabilité parce que les ressources dont elle dispose étaient déficientes – par exemple, parce que sa documentation juridique n'était pas à jour et a pu l'induire en erreur.

D'autre part, le juge apprécie si les **agissements matériels de l'administration** ont été fautifs au regard des obligations de résultat pesant sur elle, qu'elle ait eu recours ou non à un SIA pour l'assister voire pour exécuter l'action (ou pour s'abstenir de l'exécuter). En cas de faute dans la prise en charge d'un patient, peu importe que le médecin ait opéré lui-même ou ait « délégué » cette tâche à un robot : du point de vue du patient, seul le résultat compte, et la responsabilité de l'établissement de santé peut être engagée dans les conditions de droit commun en cas de préjudice subi à raison d'un acte de soins automatisé ou assisté, qu'il s'agisse d'un diagnostic erroné ou d'un acte thérapeutique inapproprié ou mal exécuté. De la même façon, la délivrance de renseignements erronés, qu'elle soit le fait d'un agent public au guichet ou d'un robot conversationnel, constitue une faute de l'administration pouvant ouvrir droit à réparation. Ce régime de responsabilité rejoint la responsabilité civile du fait des choses, en ce sens que le **gardien de la chose** (ici, l'administration utilisatrice du SIA) est responsable des dommages qu'elle cause aux tiers par l'utilisation d'un tel logiciel[195].

Il est toutefois concevable, en particulier en l'absence d'obligation de résultat, que la responsabilité de l'administration soit mise en cause à raison d'un manquement **à une obligation pesant spécifiquement sur elle en tant qu'utilisatrice, voire conceptrice, du système.** Comme on l'a dit, ces obligations, limitées en l'état actuel des textes, seraient sensiblement plus étendues avec le règlement IA. En outre, en l'absence même de dispositions expresses, le juge administratif pourrait être amené, au fur et à mesure de la diffusion des SIA dans l'action administrative, et en s'inspirant des solutions dégagées par le juge civil pour ce qui concerne les systèmes déployés dans le secteur privé, voire du droit comparé, à apprécier si l'administration a agi conformément aux diligences qu'on peut raisonnablement attendre du responsable d'un tel système, en l'état de l'art et des connaissances scientifiques. Le principe de primauté humaine et, en particulier, l'exigence de supervision humaine précédemment évoquée, pourrait servir de guide : une obligation de vigilance pèse sur tout utilisateur d'un SIA, dont l'ampleur dépend de l'occurrence de l'erreur (laquelle est fonction, notamment, du caractère déterministe ou probabiliste du modèle) et de la gravité de l'erreur (ce qui renvoie notamment à la qualification de SIA à haut risque au sens de la proposition de règlement européen). Les diligences attendues de l'administration dans le développement et la maintenance d'un système d'IA interagissant directement avec les usagers ne sont pas très éloignées du régime de

[195] V. en ce sens l'analyse du Conseil d'État dans l'étude annuelle 2017, *Puissance publique et plateformes numériques : accompagner « l'ubérisation »*, p. 116-117.

l'entretien normal d'un ouvrage public[196], dans lequel l'administration doit démontrer qu'elle a accompli les diligences requises pour prévenir le dommage subi par la victime.

Toutefois, à supposer qu'un manquement soit caractérisé, une indemnisation ne peut intervenir qu'en présence d'un **lien de causalité direct entre ce manquement et un préjudice établi**. Ce lien pourrait être délicat à établir en l'absence d'interaction directe entre le système et la victime, et toutes les fois où s'interpose le fait humain (sur la mise en œuvre de l'exigence de supervision humaine, V. 3.1.1.).

En l'absence de faute de l'administration, sa responsabilité pourrait encore être engagée sur le fondement de la rupture d'égalité devant les charges publiques, à raison du préjudice grave et spécial subi par des victimes du fait du recours à un SIA. A titre d'exemple, si, en dépit du respect des règles encadrant le choix des données d'entraînement, il s'avère que l'utilisation d'un SIA entraîne structurellement des conséquences gravement défavorables pour une catégorie bien circonscrite de personnes et dans des cas résiduels, sans que, en l'état des connaissances scientifiques, il soit possible d'y remédier, l'intérêt général pourrait justifier tout à la fois que la mise en œuvre de ce SIA ne soit pas interdite par le juge mais que les personnes lésées puissent prétendre à une réparation.

Les conditions restrictives d'engagement de la responsabilité sans faute pourraient amener les pouvoirs publics à s'interroger sur la nécessité de mettre en place des fonds d'indemnisation, par exemple pour un SIA à très forte valeur ajoutée mais présentant des risques opérationnels irréductibles qu'il ne serait pas équitable de faire supporter aux usagers. Sans préjudice de ces initiatives, le juge sera amené à se poser, au cas par cas, la question de l'élaboration de régimes de responsabilité plus accommodants dans certaines domaines, comme l'a fait, en matière médicale, la jurisprudence *Mme ZYX* (CE, 9 juillet 2003, *AP-HP c/ Mme ZYX ?* n° 220437, Rec.) selon laquelle « *sans préjudice d'un éventuel recours en garantie, le service public hospitalier est responsable, même en l'absence de faute de sa part, des conséquences dommageables pour les usagers de la défaillance des produits et appareils de santé qu'il utilise* » – ce qui inclut d'ailleurs d'ores et déjà l'ensemble des SIA embarqués dans des équipements de santé.

Dans cet esprit, dans sa résolution du 20 octobre 2020 contenant des recommandations à la Commission sur un régime de responsabilité civile pour l'intelligence artificielle (2020/2014(INL)), le Parlement européen a proposé de créer un **régime de responsabilité objective (sans faute) à la charge de tout opérateur d'un SIA à haut risque** et, corrélativement, une obligation d'assurance à la charge de « l'opérateur frontal » d'un tel SIA (défini comme toute personne qui exerce un certain contrôle sur un risque associé à l'exploitation et au fonctionnement du système et tire profit de son exploitation) comme de « l'opérateur d'amont » (défini comme toute personne qui, de manière continue, définit les caractéristiques de la technologie et fournit des données ainsi qu'un service de soutien en amont essentiel et exerce donc également un certain contrôle sur le risque liée à l'exploitation et au

[196] Notion qui recouvre, selon la jurisprudence, tant le défaut d'entretien proprement dit que le vice de conception (CE, 8 mars 1991, *Société Usinor*, n° 70216, Rec.).

fonctionnement du système)[197]. En l'état, indépendamment des obligations qu'elle met à la charge des fournisseurs et des utilisateurs, et qui constitueront autant de points d'appui à des actions en responsabilité pour faute, la proposition de règlement IA de la Commission ne traite pas spécifiquement de la question de la responsabilité.

Il appartient tant aux pouvoirs publics qu'aux juridictions de définir le niveau de risque socialement acceptable en contrepartie des bénéfices attendus de la mise en service de ces systèmes. Dans tous les cas, l'indemnisation des préjudices subis par les destinataires des SIA publics pourrait reposer sur deux principes simples, facteurs de confiance.

D'une part, la complexité technique des SIA ne devrait jamais constituer un obstacle à la mise en jeu de la responsabilité de l'administration, laquelle ne saurait se retrancher derrière elle pour s'en exonérer. **Autrement dit, le choix des moyens par l'administration ne doit pas rejaillir sur le droit à réparation des usagers du service public (principe de « neutralité technologique »)**.

D'autre part, **les citoyens ne devraient pas être contraints d'engager la responsabilité d'un autre acteur que l'administration utilisatrice du système et « gardienne de la chose »**. A la différence des SIA privés, les SIA publics sont, par construction, opérés par des acteurs solvables, vers lesquels les victimes devraient se tourner spontanément, ce qui simplifie considérablement la problématique. L'administration doit les « désintéresser », dès lors que les conditions d'engagement de sa responsabilité sont réunies, et, si elle estime avoir elle-même été induite en erreur ou lésée par un SIA conçu par un tiers, se retourner contre ce dernier dans le cadre de l'action subrogatoire dont elle dispose à son encontre ou d'une action récursoire fondée sur le contrat qui les lie[198]. L'exercice d'imputation, corollaire de la multiplicité des intervenants et de la difficulté de déterminer la cause adéquate du dysfonctionnement d'un système complexe et parfois opaque, interviendra en

[197] Pour les SIA qui ne sont pas à haut risque, le Parlement européen a proposé une exonération de responsabilité en cas d'activation du système à son insu en dépit de mesures de protection raisonnables, et lorsque toute la diligence requise a été déployée en sélectionnant un système adapté au regard des tâches à accomplir et des capacité requises, en mettant correctement en service le système, en contrôlant ses activités et en maintenant la fiabilité opérationnelle par l'installation régulière de toutes les mises à jour disponibles.

[198] On peut imaginer qu'une victime entreprenne d'engager la responsabilité directe du producteur. Si le SIA a été conçu par une autre administration, c'est la responsabilité de cette dernière qui sera actionnée devant le juge administratif. Si le SIA a été conçu par un opérateur privé, se pose notamment la question de l'application de la responsabilité du fait des produits défectueux (art. 1245 à 1245-17 du code civil, lequel s'applique à « tout bien meuble » en vertu de l'art. 1245-2), qui suppose que la victime démontre le caractère défectueux du produit (incapacité à offrir la sécurité à laquelle elle pouvait légitimement s'attendre) et le lien de causalité, mais non la faute du producteur. Ce dernier pourra s'exonérer de sa responsabilité à ce titre s'il prouve que le défaut est né postérieurement à la mise en circulation (notamment en cas de ré-entraînement) ou encore si l'état des connaissances scientifiques et techniques, au moment où il a mis le produit en circulation, n'a pas permis de déceler l'existence du défaut (art. 1245-10). Il doit par ailleurs être rappelé qu'un produit ne peut être considéré comme défectueux par le seul fait qu'un autre, plus perfectionné, a été mis postérieurement en circulation (art. 1245-3).

aval, épargnant au citoyen victime du dommage un ou plusieurs procès long et coûteux. Un soin tout particulier devra être apporté par l'administration à la rédaction des clauses contractuelles organisant la mise en jeu de la responsabilité du fournisseur à son égard. Il apparaît judicieux de prévoir des clauses de conciliation et de désignation d'un expert commun pour déterminer la cause du dommage.

3.3.3. La responsabilité pénale du fait des SIA publics

La réparation financière du dommage causé par l'utilisation d'un système d'IA ne peut suffire à inspirer confiance dans ces systèmes. Au-delà des responsables, les victimes chercheront des coupables ; qu'on songe, par exemple, à des accidents mortels occasionnés par un véhicule ou un système d'armes létales « autonomes », ou un robot chirurgical.

Pas plus qu'en matière de responsabilité civile, il n'y a lieu de sanctionner pénalement le système lui-même, en tant que nouvelle personne morale[199]. Si faute pénale il y a, elle est à chercher du côté soit du fournisseur, soit de l'utilisateur. La mise en œuvre des règles de droit commun de la responsabilité délictuelle semble garantir d'ores et déjà un niveau de protection satisfaisant.

En-dehors des systèmes d'IA conçus ou utilisés à des fins délibérément malveillantes ou selon des modalités constitutives d'une infraction pénale (par exemple, au prix d'un délit d'atteinte aux droits de la personne résultant des fichiers ou des traitements informatiques, comme l'utilisation sans consentement et en-dehors des exceptions prévues par la loi de données à caractère personnel sensibles à des fins d'apprentissage), l'élément intentionnel fera certes souvent obstacle à la caractérisation de l'infraction. Le seul fait que des erreurs soient commises du fait des SIA – et on a vu que les systèmes basés sur l'apprentissage automatique étaient fondés sur la minimisation du taux d'erreur, mais non sur la suppression de ce risque – ne signifie pas qu'il y a eu faute. C'est ce qu'illustre, certes en matière civile mais par un raisonnement potentiellement transposable à la matière pénale pour ce qui concerne la recherche de l'élément intentionnel, l'absence de condamnation de Google, poursuivi pour injure ou diffamation à raison de l'association, par la fonctionnalité *Suggest* du moteur de recherche exploité par cet opérateur, du nom d'une personne et de qualificatifs désobligeants. Dès lors que cette fonctionnalité, basée sur l'apprentissage machine, « *est le fruit d'un processus purement automatique dans son fonctionnement et aléatoire dans ses résultats* », *« l'affichage des « mots-clés » qui en résulte est exclusif de toute volonté de l'exploitant du moteur de recherche d'émettre les propos en cause* » (CCass. 1e civ., 19 juin 2013, n° 12-17591, Bull.). Il en irait de même, par exemple, d'une contrefaçon malencontreusement produite par un système d'IA générant du texte ou une image, tout au moins si l'algorithme a été conçu pour prévenir une telle dérive.

[199] V. pour cette solution le procès fictif « Le carambolage du siècle » organisé à la cour d'appel de Paris en 2018, la formation de jugement ayant déclaré l'intelligence artificielle coupable d'homicide involontaire (s'agissant d'un accident mortel causé par un véhicule autonome) et l'ayant condamnée, outre aux dommages et intérêts, à une « *rééducation algorithmique avec mise à l'épreuve en simulateur* ».

Les **infractions non intentionnelles –** c'est-à-dire celles dont l'élément moral est caractérisé par l'imprudence, la négligence de l'auteur ou le manquement à une obligation légale de sécurité ou de prudence – pourront quant à elles trouver application lorsqu'il est établi que l'auteur n'a pas accompli les diligences normales compte tenu de la nature de ses missions ou de ses fonctions, de ses compétences ainsi que du pouvoir et des moyens dont il disposait. Dans ce cas, les personnes physiques qui n'ont pas causé directement le dommage, mais qui ont créé ou contribué à créer la situation qui a permis la réalisation du dommage ou qui n'ont pas pris les mesures permettant de l'éviter, sont responsables pénalement s'il est établi qu'elles ont, soit violé de façon manifestement délibérée une obligation particulière de prudence ou de sécurité prévue par la loi ou le règlement, soit commis une faute caractérisée et qui exposait autrui à un risque d'une particulière gravité qu'elles ne pouvaient ignorer (art. L. 121-3 du code pénal). C'est à ces dispositions que renvoie l'article L. 123-2 du code de la route[200] qui définit les conditions de la responsabilité pénale du constructeur d'un véhicule autonome en cas d'atteinte involontaire à la vie ou à l'intégrité de la personne.

A titre d'exemple, la mise en service, sur des voies ouvertes à la circulation publique, d'un prototype de véhicule autonome dont le concepteur ne pouvait ignorer qu'il était insuffisamment abouti et excessivement dangereux pour les autres usagers de la route, et qui a causé le décès d'un piéton, pourra justifier le prononcé d'une sanction pénale. L'infraction de mise en danger d'autrui (art. 223-1 du code pénal) est également susceptible de jouer lorsque le concepteur ou l'utilisateur du système avait conscience du risque de mort ou de blessures pouvant résulter de sa mise en service, sans intention de les causer.

C'est dans l'appréciation des diligences attendues du concepteur, du développeur et/ou de l'utilisateur, et la fixation du curseur quant au caractère excusable ou non du manquement, que se situera la difficulté. L'élaboration de normes fixant des standards de qualité, notamment d'exactitude du système, et la certification des systèmes, permettront d'objectiver ce travail. Mais l'expertise restera indispensable pour appréhender l'état de l'art et, ainsi, apprécier si un ou plusieurs protagonistes de la chaîne de déploiement du SIA se sont livrés à des agissements inacceptables, car excessivement dangereux. La comparaison des performances respectives de l'humain et de la machine dans l'accomplissement de la tâche ayant causé le dommage peut utilement éclairer cette appréciation.

Enfin, indépendamment même d'un dommage causé à une victime, la question de l'édiction de nouvelles sanctions pénales, propres à la fourniture et à l'utilisation des SIA, se posera à l'occasion de la mise en œuvre du règlement européen, s'il est adopté. En l'état, et par analogie avec le RGPD, le projet se borne à renvoyer la détermination du régime de sanctions aux Etats membres, pourvu qu'elles soient effectives, proportionnées et dissuasives, et à encadrer le dispositif d'amendes administratives.

---

[200] Issu de l'ordonnance n° 2021-443 du 14 avril 2021 relative au régime de responsabilité pénale applicable en cas de circulation d'un véhicule à délégation de conduite et à ses conditions d'utilisation.

La construction d'une IA digne de confiance suppose d'afficher clairement et d'assumer que tout dommage causé en raison de l'utilisation d'un tel système n'appellera pas le prononcé de sanctions pénales. Il n'y a pas lieu, à cet égard, de « sur-pénaliser » cette activité en raison des risques qu'elle présente, au risque de compromettre la nécessaire innovation. Ainsi, les sanctions pénales qui pourraient être prévues pour la mise en œuvre du règlement IA devraient uniquement porter sur des manquements délibérés à des obligations claires et de première importance, comme la mise en service d'un système d'IA interdit[201] ou celle d'un SIA à haut risque non certifié, ou encore l'utilisation de techniques trompeuses (comme les hyper-trucages) sans information des destinataires. Les lourdes sanctions pécuniaires prévues par le projet de règlement devraient, pour le surplus, suffire à discipliner les acteurs du secteur.

[201] A l'instar de ce qui existe pour la prohibition du profilage des professionnels de justice (art. L. 10 du code de justice administrative et art. L. 111-13 du code de l'organisation judiciaire).

4. Doter la France des ressources et de la gouvernance adaptées à l'ambition

4.1. Doter les administrations des ressources nécessaires

Les ressources nécessaires à la conception, à l'adaptation et à l'utilisation des systèmes d'IA publics sont à la fois humaines, techniques et financières, organisationnelles et juridiques. **Le calibrage respectif de ces ressources dépend d'abord de l'arbitrage des administrations entre « faire », c'est-à-dire développer le SIA en interne, et « acheter », en recourant à un prestataire externe.**

4.1.1. Faire ou acheter ?

L'arbitrage entre « faire » et « acheter » (option au sein de laquelle on pourrait distinguer « faire faire » et « acheter tout fait ») ne peut être fait abstraitement, une fois pour toutes. Il doit résulter d'une analyse projet par projet des besoins induits par sa création, son développement et sa maintenance, notamment en termes de compétences. Il peut ne pas y avoir d'intérêt à re-créer un SIA déjà disponible sur le marché et facilement internalisable, tout comme il peut être opportun de développer soi-même un SIA dont on maîtrisera de bout en bout l'apprentissage et l'industrialisation. Ces deux modalités sont, en outre, susceptibles de se compléter pour un même SIA, qui intégrera des briques développées en interne et d'autres acquises « sur étagère », ou encore qui utilisera un modèle pré-entraîné et adapté, grâce à un réglage fin, au contexte métier dans lequel il sera déployé. En revanche, le recours à un prestataire pour un SIA stratégique et complexe, que l'administration n'a pas les moyens (humains, techniques…) de piloter, peut se révéler désastreux – tant en termes d'outil finalement produit (l'histoire des SI de l'État est pavée d'échecs retentissants) qu'en termes budgétaires.

Dans ce cadre, **penser que confier à un prestataire privé le développement d'un SIA induit automatiquement une économie pour les finances publiques est un leurre** – et le lancement d'un SIA ne doit donc pas être arbitré sous ce prisme. Comme le souligne le récent rapport de la mission d'information parlementaire relative aux différentes missions confiées par l'administration de l'État à des prestataires

extérieurs (*outsourcing*)[202], « *l'externalisation génère rarement des économies pour les prestations numériques et intellectuelles, le taux journalier moyen (TJM) externe étant souvent plus élevé que le même taux en interne* ». C'est d'autant plus vrai pour les SIA que leur maintenance est indispensable au maintien de leur performance, notamment par ré-entraînement du modèle, et que la collectivité peut ainsi être engagée dans la durée, bien au-delà de la phase de conception et de primo-déploiement. A ce titre, la plus grande rigueur doit présider à la gestion de la relation contractuelle, et il peut être opportun de prévoir que les coûts de ce réentraînement soient assumés par le prestataire qui aura tiré des revenus du modèle construit pour une administration, dans le cadre d'autres contrats avec le secteur public voire le secteur privé.

Ce risque financier, particulièrement signalé s'agissant de l'achat de SIA par les collectivités territoriales, commande une réponse adaptée, tenant compte de l'effet de série des systèmes à usage général qui ne requièrent qu'une faible adaptation pour leur déploiement dans les collectivités utilisatrices. L'enjeu est d'éviter que des prestataires qui ont entraîné un modèle grâce à la commande publique, démarchent ensuite d'autres administrations (centrales ou territoriales) en se prévalant de cette expérience et en facturant, à un prix excessif au regard des nouveaux développements réellement nécessaires, un SIA déjà calibré dans le secteur public. La réponse tient notamment dans l'investissement au niveau national dans des solutions mutualisables, permettant de négocier des marges plus faibles auprès des prestataires, et dans la connaissance des systèmes déjà déployés dans d'autres administrations. Cela implique notamment que, s'il peut être envisagé, dans certains cas spécifiques (pour des SIA dans des domaines sensibles notamment), que le SIA public développé en externe ne puisse être revendu par le prestataire à un client privé, un tel droit d'opposition ne soit jamais prévu pour la réutilisation d'un SIA par une autre administration.

L'arbitrage doit se faire, en premier lieu, au regard des compétences internes dont dispose l'administration pilote. Le développement de SIA publics en régie suppose que l'administration dispose d'une équipe d'experts de la donnée ayant la taille critique. S'il n'existe aucune norme en la matière, la disponibilité de deux à trois ingénieurs IA, d'un développeur et d'un ingénieur de la donnée est sans doute un minimum pour développer des SIA modestes[203]. En tout état de cause, il ne faut pas perdre de vue que l'externalisation requiert également des compétences internes afin de définir le cahier des charges, de sélectionner et de piloter les prestataires. S'agissant de la sélection, il est indispensable que l'administration comprenne ce qu'elle achète exactement. À cet égard, un **standard terminologique** gagnerait à être élaboré, sur le modèle de la grille réalisée par la Haute Autorité de Santé sur les dispositifs médicaux intégrant une brique d'intelligence artificielle, afin de faire pièce aux fausses innovations et aux discours commerciaux fallacieux. S'agissant du

[202] Assemblée nationale, rapport n° 4928, janvier 2022.

[203] A titre d'exemple, l'équipe du programme « valorisation de la donnée » de la direction générale des douanes et des droits indirects compte six ingénieurs IA (*data scientists*), un développeur généraliste (dit « *full stack* ») et un ingénieur des données, mais elle fait encore appel de façon soutenue, quoique décroissante, à des prestataires externes.

pilotage, à titre de repère, la Cour des comptes indique dans sa récente communication consacrée à la conduite des grands projets numériques de l'État[204], que « *pour conduire un projet de développement spécifique, les ressources de maîtrise d'ouvrage doivent représenter 30% de celles de maîtrise d'œuvre* ». Ainsi, quel que soit le choix fait en matière d'externalisation, le recours à l'IA publique ne peut se faire sans un investissement RH dès le départ, et de façon soutenue.

Pour les collectivités territoriales, la question du « faire ou acheter » appelle malheureusement, au regard des ressources internes, une réponse très largement contrainte. Sauf dans quelques cas exceptionnels (les plus grandes collectivités, comme les régions ou les métropoles), elles n'ont pas les compétences pour faire elles-mêmes : elles se tournent alors vers un prestataire, sans disposer nécessairement des moyens internes permettant de discuter ses propositions et ses réalisations (choix de l'algorithme, des données d'entraînement, etc.) ce qui représente, comme nous venons de le voir, un risque de surcoût budgétaire important. Le risque est grand d'un transfert indu de richesse (et de données) que les pouvoirs publics doivent se donner les moyens d'éviter. A ce titre, il est prioritaire de doter les collectivités d'une capacité d'expertise propre, restaurant l'égalité des armes avec les prestataires privés.

En deuxième lieu, l'arbitrage entre « faire » et « acheter » doit tenir compte de considérations chronologiques. Alors que les techniques évoluent avec une exceptionnelle rapidité en matière d'IA, le marché propose en principe des systèmes à l'état de l'art. Sous réserve que les contrats conclus ménagent à l'administration la possibilité juridique et technique de changer rapidement ou régulièrement de prestataire, l'achat de solutions développées par le secteur privé lui permet ainsi, en principe, de bénéficier des dernières avancées technologiques. En contrepoint, il convient d'intégrer, dans le bilan coûts/avantages du recours à un prestataire, les externalités positives liées au développement interne et, en premier lieu, l'effet d'apprentissage des équipes, et les coûts potentiels de la réinternalisation, à supposer que la réversibilité soit matériellement possible, bien qu'une telle internalisation soit également susceptible de représenter des économies conséquentes[205]. La conception interne constitue aussi un investissement de long terme, même si les problèmes de fidélisation des experts amènent à relativiser le propos (V. *infra*).

En troisième lieu, il y a lieu de prendre en compte, dans cet arbitrage casuistique, les contraintes liées au droit de la commande publique, très largement commandées par le droit européen et qui ne doivent pas être surestimées.

Il est vrai que ces contraintes, dont il n'est pas nécessaire de rappeler l'utilité par ailleurs, sont peu adaptées au marché des SIA innovants relevant de l'apprentissage machine. Les solutions sont souvent proposées par de jeunes entreprises qui, à supposer qu'elles soient informées des consultations lancées par les collectivités

[204] *La conduite des grands projets numériques de l'État*, Communication à la commission des finances du Sénat, juillet 2020.

[205] Juillet 2020, p. 68.

publiques, peinent à trouver les ressources pour déposer une offre conforme et hésitent à en dégager à cette fin, en raison de l'aléa de toute procédure de mise en concurrence. Ce parcours du combattant de la commande publique contraste avec la facilité de la contractualisation de gré à gré avec les clients du secteur privé. En outre, les exigences de capacité financière imposées par des acheteurs publics préoccupés par le risque de défaillance constituent d'importantes barrières à l'entrée pour ces jeunes pousses à l'équilibre économique encore fragile.

Toutefois, outre les assouplissements sectoriels (en particulier dans le domaine de la sécurité et de la défense), les souplesses existent et on ne peut qu'encourager les administrations à s'en emparer davantage qu'elles ne le font aujourd'hui :

- les **marchés de recherche et développement** sont soustraits à toute obligation de publicité et de mise en concurrence. Toutefois, il est difficile aux acheteurs de cerner le contenu de l'exception, qui peut bénéficier à la réalisation de démonstrateurs technologiques – définis comme les dispositifs visant à démontrer les performances d'un nouveau concept ou d'une nouvelle technologie dans un environnement pertinent ou représentatif (mais non réel), mais pas aux « prototypes de préproduction », qui offrent un « prolongement industriel direct » et se prêtent à une démonstration dans un environnement opérationnel (réel)[206]. En outre, ces marchés doivent être exclusifs de tout objectif de rentabilité ou d'amortissement des coûts de recherche et de développement[207] et ils ne sauraient consister à personnaliser et intégrer la solution dans l'environnement propre à l'acheteur. En pratique, cette exception a plutôt vocation à être utilisée par des centres de ressources partagés afin de co-développer des algorithmes nouveaux susceptibles de répondre à des besoins génériques des acteurs du secteur public, et non spécifiques à telle ou telle administration ;

- le **partenariat d'innovation** présente, lui, l'avantage d'intégrer non seulement les prestations de recherche et développement, mais aussi l'acquisition ultérieure des produits et services en résultant[208]. Mais il suppose d'avoir conduit un sourçage rigoureux afin de s'assurer que la solution n'est pas déjà disponible sur le marché, et il requiert, cette fois, l'accomplissement de formalités de publicité et de mise en concurrence ;

- enfin, si certains marchés relatifs aux SIA peuvent bénéficier de la procédure de passation sans publicité ni mise en concurrence expérimenté par le décret n° 2018-1225 du 24 décembre 2018 et pérennisé par le décret n° 2021-1634 du 13 décembre 2021 au profit des **achats innovants**[209], les administrations ne se sont

[206] 2° de l'art. L. 2512-5 du code de la commande publique.

[207] Art. R. 2122-10 du code de la commande publique.

[208] Art. L. 2172-3 du code de la commande publique.

[209] Selon l'art. L. 2172-3 du code de la commande publique, « *sont considérés comme innovants les travaux, fournitures ou services nouveaux ou sensiblement améliorés. Le caractère innovant peut consister dans la mise en œuvre de nouveaux procédés de production ou de construction, d'une nouvelle méthode de commercialisation ou d'une nouvelle méthode*

encore que très partiellement approprié cette facilité qui, en outre, n'est offerte que pour les marchés dont la valeur estimée du besoin est inférieure à 100 000 euros hors taxes. Or précisément, l'innovation coûte cher et on ne peut guère espérer, à ce prix, que le développement d'une preuve de concept d'un SIA assez modeste. De fait, le bilan (non exhaustif) de l'expérimentation du dispositif tend à montrer que sa mise en œuvre n'a que marginalement porté sur des systèmes d'IA[210].

Il faut ajouter que les délais et le formalisme de la consultation peuvent être adaptés à la nature de l'écosystème des prestataires. En outre, il est fréquent que les jeunes entreprises répondent en groupement avec des acteurs installés, en particulier les grands intégrateurs, qui sont davantage rompus à cet exercice. Enfin, les solutions matures ou les offres de réalisation d'un SIA peuvent faire l'objet d'un **référencement par des centrales d'achat** qui organisent la mise en concurrence. Tel est le cas dans le secteur hospitalier (Uniachat, RESAH) et, plus largement dans la sphère publique, avec les marchés multi-éditeurs sur les logiciels de l'Union des groupements d'achats publics (UGAP) et un lot « intelligence de la donnée »[211] recouvrant un ensemble complet de prestations depuis le cadrage du projet jusqu'à l'accompagnement du changement induit par le système, en passant par la réalisation d'une preuve de concept ou la sélection d'un logiciel standard, l'intégration de la solution et sa prise en main. L'UGAP y a incorporé une prestation d'assistance à la qualification du besoin afin d'éviter un recours inapproprié à ce

organisationnelle dans les pratiques, l'organisation du lieu de travail ou les relations extérieures de l'entreprise ». Le guide pratique de l'achat innovant publié par la direction des affaires juridiques du ministère de l'économie et des finances et l'observatoire économique de la commande publique relève la difficulté de cerner les termes de « nouveaux » et « sensiblement améliorés » et propose de mobiliser un large faisceau d'indices à cette fin. A cette aune, si des SIA matures, proposés « sur étagère » par plusieurs opérateurs depuis plus de deux ans, ne seront généralement pas regardés comme innovants, tel pourra être le cas de SIA « sur-mesure » proposés par un ou un nombre réduit d'opérateurs, dont les caractéristiques diffèrent sensiblement de ceux des systèmes déjà commercialisés, par l'originalité de leur conception (développement d'un algorithme propriétaire, non disponible en source ouverte, utilisation de jeux de données rares ou d'une combinaison originale…) et surtout par la supériorité de leurs performances attendues. Le simple ré-entraînement du modèle dans un nouveau domaine de fonctionnement ne suffit probablement pas à caractériser l'innovation. Cette incertitude juridique et le spectre du délit de favoritisme sont susceptibles de constituer des freins au recours à cette souplesse.

[210] Selon un recensement partiel effectué à l'automne 2021 : deux projets portés par des services départementaux d'incendie et de secours (optimisation des moyens déployés et du recrutement, prévention des accidents…), un par une université (création de parcours personnalisés par analyse automatisée des évaluations, identification des lacunes et rapprochement avec les compétences requises afin de formuler des suggestions de parcours ou d'actions), un par un centre hospitalier universitaire (amélioration et personnalisation de la gestion clinique et thérapeutique des receveurs de greffes), un par un département (outil d'évaluation automatisé des routes) et un par une agence de développement.

[211] Ce lot, qui a un périmètre plus large que l'intelligence artificielle, est plutôt utilisé par de grandes entités, qu'il s'agisse de ministères ou de collectivités territoriales. Il s'agit d'un marché modeste, de quelques dizaines de millions d'euros.

marché par les collectivités clientes. On relèvera toutefois que le référencement par un marché mono-attributaire, privilégié par l'UGAP en raison de la qualité de la relation qu'il permet d'instaurer avec le titulaire, réduit aussi l'éventail des choix qui s'offrent aux collectivités publiques.

4.1.2. Piloter une stratégie RH de l'IA publique

Ce n'est pas faire injure à la fonction publique que de constater, comme l'ont fait de nombreuses personnes auditionnées, que **la culture de l'IA est faible, voire inexistante, chez la plupart des agents publics, y compris à haut niveau de responsabilité**. Les concepts restent nébuleux et les confusions fréquentes ; les enjeux, potentialités et modalités de fonctionnement des SIA sont encore mal appréhendés. La stratégie RH, aujourd'hui concentrée sur le nécessaire recrutement d'experts de la donnée, doit impérativement répondre au besoin d'acculturation des agents et de formation des cadres dirigeants de la fonction publique.

1. Développer une culture de l'IA chez les agents publics

Les **actions de sensibilisation au numérique et aux enjeux de l'IA**, que la première partie de l'étude propose de déployer à destination de l'ensemble des citoyens, en particulier à l'école, doivent être déclinées spécifiquement pour les agents publics, par le biais de formations générales, et s'accompagner d'une évolution de l'environnement culturel administratif dans un sens favorable à l'innovation.

a/ De même qu'on imagine mal qu'un étudiant en médecine puisse accomplir l'ensemble de son cursus sans suivre un module de formation sur l'IA en santé, abordant à la fois les questions techniques, éthiques et juridiques, il n'est pas concevable que la généralité des agents publics prennent à l'avenir leur fonction en ignorant tout de ce qu'est un SIA. Aucun agent ne doit sortir d'une école de service public sans avoir suivi un module d'acculturation aux SIA de quelques heures, lui permettant d'en comprendre l'intérêt pour lui et le public qu'il s'apprête à servir, ainsi que les risques auxquels il doit être attentif.

L'enjeu de la **formation continue** est plus crucial encore, compte tenu de l'évolutivité rapide des techniques et de l'importance de les mettre en regard du vécu administratif et du quotidien métier des agents. Déjà, dans son étude annuelle 2017 *Puissance publique et plateformes numériques : accompagner l'ubérisation*, le Conseil d'État proposait d'introduire dans la formation continue des agents publics une obligation de mettre à jour périodiquement leurs connaissances sur l'environnement numérique de leur métier, dans le but que les techniques et le langage informatiques deviennent, pour l'ensemble des agents publics, une compétence transversale. Dans l'immédiat, deux actions apparaissent prioritaires.

D'une part, la structure hiérarchique des services exige une **formation immédiate de tous les cadres dirigeants de la fonction publique**. Parce qu'aucun SIA ne peut être déployé sans qu'une décision soit prise au sommet, il faut que les sommets comprennent les SIA, et qu'ils encouragent leur déploiement. Ce n'est pas le cas aujourd'hui, et l'incompétence technique des cadres, sur ce sujet, est dramatiquement néfaste : refus de déploiement, par incompréhension ; erreur de

développement, par ignorance ; fautes de management… Or, tous les entretiens menés dans le cadre de cette mission l'ont montré : **l'adhésion des cadres dirigeants est un facteur-clé dans l'émergence comme dans le succès des initiatives**. Une action similaire doit être envisagée au bénéfice des élus locaux, en concertation avec leurs associations. Le Centre national de la fonction publique territoriale a vocation à jouer un rôle-clé dans l'acculturation des directeurs généraux des services et responsables de services ; son offre de service « La transition numérique », déjà riche, pourrait être complétée par des stages spécifiques aux systèmes d'IA.

D'autre part, aucun SIA présentant une certaine sensibilité ne doit être déployé sans une **formation spécifique des agents au SIA déployé**. Une telle formation est au nombre des mesures organisationnelles que le projet de règlement européen prévoit d'imposer, de façon générique, au titre de la supervision humaine, pour les SIA à haut risque. On relèvera que la loi de l'État de Washington du 31 mars 2020 sur la reconnaissance faciale en vue de l'identification à distance des personnes par une autorité publique prévoit expressément cette obligation de formation des personnes responsables du contrôle humain[212]. Elle devrait être complétée, dans la limite des ressources disponibles, par une sensibilisation des autres parties prenantes, notamment la chaîne de commandement ou de contrôle, les usagers et leurs représentants, les agents des partenaires et sous-traitants amenés à intervenir sur le système, à interagir avec lui ou à faire l'objet de mesures assistées par lui.

Qu'il s'agisse de sensibilisation ou de formation, ces actions, par l'ampleur du nombre d'agents concernés et par la technicité potentielle de leur contenu, ne peuvent être organisées à moyens constants. En raison, comme il a été dit, de la nécessité qu'elles soient en prise avec la réalité des métiers et parce que leur efficacité peut se trouver limitée par leur logique descendante et générique, ces formations ne sauraient être confiées à des prestataires extérieurs à l'administration. Elles gagneraient donc à être pilotées et organisées en interne, avec l'appui d'un pôle interministériel de sensibilisation et de formation aux enjeux du numérique et de l'IA[213].

b/ Indispensables, les formations ne seront toutefois pas suffisantes pour que la culture de l'IA infuse au sein des services. La confiance et l'optimisme ne s'inculquent pas par des séminaires, tout comme la peur du lendemain ou, simplement, la conviction du déclin ne se conjurent pas à coup de circulaires. L'évolution culturelle ne peut venir que d'un environnement propice à l'émergence de l'innovation, avec l'ambition ultime de faire des agents publics des prescripteurs actifs de SIA.

Il convient à cet égard de **susciter et d'encourager les initiatives**, dans un cadre managérial fondé sur le **principe qu'aucune idée n'est mauvaise *a priori*** et que **l'échec est aussi créateur de valeur**, fût-ce en termes d'expérience acquise. Alors que la contrainte budgétaire des administrations renforce l'aversion à l'échec, il est d'autant plus nécessaire d'œuvrer à une décrispation de l'environnement culturel

[212] V. pour une description complète de cette loi : W. Maxwell, *La régulation des algorithmes aux États-Unis : quelles leçons pour l'Europe ?*

[213] V. 4.2.1.

administratif qui crée aujourd'hui un effet de dissuasion pour celles et ceux qui nourriraient des velléités créatives et un effet repoussoir pour qui envisagerait de rejoindre le secteur public afin de le transformer par le numérique. Le dispositif des « intrapreneurs » mis en place dans le cadre du programme beta.gouv.fr mérite d'être salué dans son principe. Les projets de SIA réussis et passés à l'échelle devraient être labellisés et largement diffusés comme des exemples à suivre, en s'attachant à montrer en quoi l'existence même du projet, la performance du système et les gains qu'en ont retirés les parties prenantes étaient tributaires de la disponibilité de données de qualité.

Les agents devraient s'emparer eux-mêmes des données et des outils, les manier au quotidien, pour mieux en percevoir l'utilité et susciter des critiques et suggestions d'amélioration. Les systèmes de visualisation des données fournissent un bon terrain d'expérimentation en ce qu'ils visent précisément à rendre les données intelligibles par la construction de représentations visuelles, offrant en même temps aux agents la possibilité de valoriser leur travail.

Le développement de ce type d'initiatives nécessite, outre le volontarisme des hiérarchies, un cadre conjuguant une certaine forme de liberté, y compris pour favoriser la mobilité temporaire d'un agent qui souhaite participer au développement d'un SIA en tant qu'expert métier, et une bonne connaissance des métiers comme des techniques de l'IA. Lorsqu'ils existent, les *Lab IA* des administrations, qui se développent au sein des ministères notamment, peuvent jouer ce rôle. A défaut, un service chargé du pilotage de ces initiatives doit être identifié dans chaque administration.

Ces initiatives doivent également permettre de **préparer la transformation des métiers** due à la mise en œuvre de SIA au sein des administrations. En effet, l'entraînement d'un SIA peut impliquer que des agents soient mobilisés pour annoter les données. Ainsi, l'outil d'anonymisation des décisions de justice développé par la Cour de cassation mobilise une vingtaine d'agents annotateurs, encadrés par une directrice des services de greffe. L'appropriation de ces nouvelles missions sera d'autant plus facile et rapide que les agents auront été sensibilisés en amont aux enjeux liés au développement de SIA au sein de leur administration.

2. *La ressource en experts de la donnée : assurer l'attractivité, fidéliser les talents*

A quelques exceptions près, les ressources humaines consacrées aux projets de SIA restent très modestes, lorsqu'elles existent. La priorité est logiquement donnée à la numérisation de l'activité administrative et à la mise en œuvre d'une politique efficace de la donnée, matière première des SIA, mais aussi, plus largement, des connaissances (des publics, de l'impact des politiques publiques, de moyens disponibles…) indispensables à la conduite de l'action publique.

Démographie des experts de la donnée de l'État

Comme le relève le rapport conjoint INSEE/DINUM, *Evaluation des besoins de l'État en compétences et expertises en matière de donnée* (juin 2021), les administrations de l'État, en-dehors des services de renseignement et des opérateurs, emploient plus de 2 000 experts de la donnée, dont environ 1 150 analystes de la donnée, 290 *data scientists* et 220 ingénieurs de la donnée. Près d'un tiers de ces experts exercent leurs fonctions en dehors de l'INSEE et des services statistiques ministériels. Le recrutement public des analystes et des scientifiques a connu un essor sensible à partir de 2016, à l'instar du secteur privé (sur l'ensemble du marché du travail, la population de data scientists serait passée de l'ordre de 200 en 2015 à environ 4000 en 2019).

Près de la moitié des experts de la donnée sont des agents titulaires. Les contractuels sont majoritaires sur le métier de data scientist (où les fonctionnaires représentent moins d'un quart des effectifs) et minoritaires sur celui de data analyst.

Hormis les services statistiques, les experts de la donnée sont plutôt **disséminés** au sein des administrations, et positionnés en général en dehors des directions numériques. Certains ministères comptant peu d'experts de la donnée disposent uniquement de pôles centraux. Les autres ont en sus recruté des experts dans les directions métiers afin de concevoir et déployer des cas d'usage opérationnels, et de contribuer à transformer le métier lui-même (à l'instar des pratiques observées dans les entreprises). Certaines directions opérationnelles comptent plus de vingt experts de la donnée : c'est le cas de la DGFIP, de la DINUM, de la direction de la sécurité sociale, de la direction générale de l'armement, de la direction générale de l'emploi et de la formation professionnelle, ou encore du pôle numérique ministériel du ministère de la transition écologique.

Les effectifs actuels couvriraient environ 90% des besoins de l'État recensés par la mission. Les besoins ont été évalués à 400 ETP supplémentaires sur deux ans, soit entre 2400 et 2500 experts en 2023 (dont 100 *data scientists*).

Il faut d'abord souligner, au préalable, que les difficultés rencontrées dans le domaine de l'IA sont plus largement celles rencontrées dans l'ensemble des métiers des technologies de l'information, et ne sont pas nouvelles. Elles ont donné lieu à de premières réponses, dans le cadre du plan d'actions pour la filière numérique et des systèmes d'information et de communication (NSIC), formalisé au sein de la circulaire DGAFP-DINSIC du 2 mai 2019 et dont les objectifs étaient « *d'attirer, de recruter et de fidéliser les compétences de la filière NSIC* ». Les enjeux en matière de ressources humaines dédiées au développement de SIA sont de multiples ordres, qui ont été développés, s'agissant plus largement des métiers de la donnée, dans le rapport INSEE/DINUM de juin 2021.

a/ Le premier enjeu est celui du recrutement. L'État dispose déjà de filières d'experts de la donnée avec les corps d'administrateurs et d'attachés de l'INSEE et les corps d'ingénieurs, en particulier celui, désormais interministériel[214], des ingénieurs des systèmes d'information et de communication. L'élargissement des recrutements dans les premiers corps devrait s'accompagner d'un encouragement et d'un accompagnement à y développer davantage de compétences d'analyse des données et surtout d'ingénieurs IA, alors que l'essentiel des ressources sont aujourd'hui consacrées à la production de la statistique publique. De même, les besoins en ingénieurs et architectes des données pourraient être en partie pourvus par les seconds corps, pourvu que davantage de profils soient recrutés dans cette optique.

Quels que soient les efforts de mobilité et de reconversion des agents titulaires en poste, le recrutement d'experts de la donnée sous le statut d'agents contractuels est incontournable. Les faiblesses du secteur public en la matière sont identifiées de longue date. Contrairement à une idée reçue, la question de la rémunération n'est pas apparue, lors des entretiens menés, comme le frein principal, sauf pour les profils les plus expérimentés, très convoités sur le marché du travail, et que l'administration a rarement les moyens de s'offrir. En revanche, les efforts réalisés pour ajuster la grille de rémunération par rapport aux salaires proposés par le privé et le déploiement, depuis 2019, du référentiel interministériel de rémunération des agents contractuels sur les métiers en tension de la filière numérique, qui a supprimé la demande de visa préalable du contrôleur budgétaire et comptable ministériel en-dessous de certains seuils, fluidifiant ainsi les recrutements et facilitant les négociations salariales[215], ont levé l'un des obstacles au recrutement des profils d'experts de la donnée dits « juniors ». Si le secteur public ne s'est pas aligné sur les standards de rémunération du secteur privé pour ces profils, ces derniers semblent enclins à renoncer à des opportunités de rémunération légèrement supérieures dans le second pour servir la cause publique dans laquelle ils croient. La circonstance que des CDD soient proposés en priorité à ces jeunes experts ne constitue pas non plus un obstacle de premier ordre à l'embauche, les candidats eux-mêmes s'inscrivant souvent dans une logique de découverte du monde public et d'« essai ».

À cet égard, doit être souligné le succès du programme « Entrepreneurs d'intérêt général » (EIG), piloté par la DINUM qui les recrute et, le temps de la réalisation du projet auquel ils sont affectés, prend en charge, *via* une subvention allouée aux administrations d'accueil, une partie de leur rémunération. Lancé en 2017, il a compté à ce jour cinq promotions (la dernière promotion a accueilli 39 personnes pour plus de 600 candidatures de qualité), intégrés dans les équipes ministérielles pour le développement d'un projet informatique, appelé « défi ». 64% des EIG sont restés dans l'administration à l'issue de leur défi.

---

[214] Décret n° 2015-576 du 27 mai 2015 portant statut particulier du corps des ingénieurs des systèmes d'information et de communication (ISIC).

[215] Déployé sur 15 métiers en 2019, il concerne, depuis une circulaire DGAFP-DINUM-DB du 15 décembre 2021, 56 métiers de la filière NSIC.

L'attractivité des postes dans l'administration pour les jeunes experts de la donnée pourrait encore être accrue en valorisant davantage les offres d'emploi et en soignant les intitulés, alors que les entreprises du secteur privé rivalisent d'imagination en la matière.

Naturellement, la problématique du recrutement s'envisage aussi du point de vue des administrations, souvent très contraintes par leur **plafond d'emplois**. Si certains appels à projets prennent en compte la masse salariale associée, ce n'est pas le cas de tous (France Relance, FTAP…), et tous les SIA ne bénéficient pas de financements de fonds interministériels. La prise en compte des coûts RH par les appels à projets en matière d'IA devrait être systématique. Il conviendrait également de mener une réflexion sur la possibilité pour les administrations de bénéficier de dotations en personnel plus généreuses pour le développement de SIA, *a minima* en tirant parti, par anticipation, des économies générées par leur déploiement.

En ce qui concerne les **outils juridiques**, le contrat de projet créé par l'article 17 de la loi du 6 août 2019 de transformation de la fonction publique offre d'intéressantes perspectives, en ce qu'il permet, dans les trois versants de la fonction publique, de recruter un agent le temps de mener à bien une mission identifiée, pour une durée comprise entre un et six ans, ce qui est une temporalité cohérente avec le développement d'un SIA.

Enfin, les administrations devraient être encouragées à recruter des **contrats CIFRE** (convention industrielle de formation par la recherche), qui permettent d'accueillir des doctorants au sein d'une collectivité territoriale, d'un EPA, d'un EPIC ou d'un GIP, dans le cadre d'une collaboration avec un laboratoire de recherche. Il serait utile d'envisager l'élargissement des employeurs publics autorisés à accueillir des doctorants dans ce cadre, pour ce qui concerne le développement de projets de SIA, aux autorités administratives indépendantes voire aux ministères et à leurs services déconcentrés.

b/ Le second enjeu, plus aigu, est celui de la fidélisation des talents, en particulier de ceux qui sont recrutés par voie contractuelle. Le public se retrouve trop souvent impuissant à conserver des experts qui, pour les raisons les plus diverses, choisissent de valoriser dans le secteur privé les compétences acquises dans la sphère publique. Ce phénomène trouve sa source dans le décrochage rapide des rémunérations publiques en fonction de la « séniorité », dans les faibles perspectives de carrière au sein des administrations, les experts des données n'étant pas spontanément identifiés comme des prétendants aux postes de direction, et dans un environnement culturel administratif encore marqué par trop de pesanteurs et de complexité, organisationnelle et psychologique, et qui ponctionne parfois plus l'énergie qu'elle ne la suscite ou ne l'amplifie.

Elle s'explique aussi par le manque de fluidité des ressources humaines à l'échelle de l'ensemble de la fonction publique, alors que ces agents sont souvent désireux de diversifier leurs expériences et les projets auxquels ils participent pour maintenir un haut niveau de compétence (logique de « défi » à relever). Les experts des données sont une ressource fondamentalement mobile. À cet égard, la fidélisation ne saurait en aucun cas être confondue avec l'absence de mobilité. Elle doit impérativement

être pensée à l'échelle du secteur public dans son ensemble, et non de chaque service, soucieux de conserver ses « pépites ».

Le développement des *Labs IA* au sein des ministères, qui permettent d'assurer une dynamique collective autour de projets de SIA, doit être souligné, mais également pensé en complémentarité avec celui d'équipes interministérielles mobiles susceptibles de venir ponctuellement en appui sur un projet spécifique. Il conviendrait également d'envisager la déconcentration de pôles chargés de développer les projets d'IA dans les métropoles les plus dynamiques, où le marché du travail est moins concurrentiel qu'en Ile-de-France, sur le modèle du Lab IA de Météo France, implanté à Toulouse.

c/ Des actions spécifiques doivent aussi être déployées en direction des collectivités territoriales. Les exemples très encourageants de SIA déployés par les collectivités dont le Conseil d'État a eu connaissance ont été essentiellement produits par les métropoles, qui ont la possibilité de s'appuyer sur un écosystème local (de formation, de recherche, économique) et de dégager des ressources propres pour créer des preuves de concept puis les faire passer à l'échelle, tout cela en interne. Or, comme il a déjà été dit, de nombreuses collectivités n'ont pas les moyens de recruter les experts leur permettant de déployer elles-mêmes des SIA, ou de contrôler leurs prestataires en cas de recours à la commande publique.

La mise en place d'un dispositif de mise à disposition des experts des données sur des missions dans les collectivités territoriales, piloté par l'ANCT qui agirait comme une structure de portage, pourrait être une piste à explorer. Il ne saurait toutefois se substituer à la nécessaire mobilisation des collectivités elles-mêmes, à travers leurs associations, à l'instar du groupe de travail « IA et data » des Interconnectés, association nationale portée par l'Assemblée des communautés de France (AdCF) et France urbaine en vue de promouvoir les usages numériques au service des territoires.

Le choix du bon échelon de déploiement, notamment en milieu rural (le niveau EPCI comme celui du département peut être envisagé), et de coordination, sera crucial pour construire la possibilité de coopération permettant de recruter les personnels dédiés nécessaires

d/ Enfin, si l'enjeu principal réside dans la ressource en experts techniques de la donnée, **la question de la disponibilité de juristes-experts de la donnée ne doit pas être négligée**. Le programme EIG a ainsi permis de recruter ce type de profils. La réussite de projets prometteurs peut être purement et simplement compromise par le manque de compétences juridiques, soit que des contraintes légales dirimantes ou structurelles n'ont pas été identifiées, soit, au contraire, que des obstacles ont été inventés ou exacerbés sur la base de croyances juridiques, comme c'est parfois le cas en matière de protection des données à caractère personnel. Les développements qui suivent entendent précisément contribuer à remettre à leur juste place les enjeux juridiques de la collecte et de l'utilisation de la donnée.

4.1.3. Libérer le potentiel de valeur de la donnée, carburant des systèmes d'IA

Les données sont la matière première des SIA et le nerf de la guerre que se livrent les opérateurs du numérique. Sans données exploitables, pas de SIA ; à tout le moins, aucun SIA basé sur l'apprentissage machine ne peut voir le jour sans un jeu de données d'entraînement adéquates, dont la constitution représente d'ailleurs la partie la plus lourde du travail des experts de la donnée impliqués dans un tel projet.

Comme le relève le rapport conjoint INSEE / DINUM relatif à l'évaluation des besoins de l'État en compétences et expertises en matière de donnée, la difficulté de disposer de données adéquates « *diminue l'efficacité du travail des experts du traitement de données et engendre parfois des frustrations tant de la part de ces profils experts que des administrations qui les recrutent* », au point que « *certains services de l'administration, qui avaient dans un premier temps recruté des data scientist de haut niveau, ont été contraints de changer d'approche, faute de données disponibles* (…) ».

Du point de vue des données, l'État n'est pas un et unique. Il forme un archipel dont les liaisons inter-insulaires sont souvent très insuffisantes. *A fortiori*, l'existence d'une communauté publique de la donnée, incluant les collectivités territoriales, les établissements publics et les autres personnes chargées d'une mission de service public, reste une vue de l'esprit. Les freins sont multiples et on se bornera ici à évoquer brièvement les obstacles culturels, techniques et, surtout, juridiques.

1. Un enjeu culturel

L'existence d'une politique publique de la donnée structurée et pilotée est encore récente. La mission Etalab a été créée à cette fin en 2011, dans le cadre du plan « France Numérique 2012 », avec pour mission première et pionnière de promouvoir la mise en ligne des documents administratifs et des données publiques (dite d'*open data*). Alors que le portail unique interministériel *data.gouv.fr* vient de fêter ses dix ans et que des progrès tout à fait notables ont été accomplis, dont le rang flatteur de la France dans les classements internationaux porte la trace, le potentiel d'amélioration reste considérable. C'est la raison pour laquelle la circulaire du 27 avril 2021 a entrepris de relancer cette politique, par la nomination d'administrateurs ministériels des données, des algorithmes et des codes-sources et par la publication de feuilles de route ministérielles.

Les administrations françaises peinent à basculer entièrement, résolument et irréversiblement dans l'ère de la donnée. La politique de la donnée s'apparente encore à un exercice imposé s'ajoutant à un plan de charge déjà encombré et occasionnant des coûts supplémentaires, sans retour sur investissement tangible. Les obligations découlant des lois, décrets et circulaires dans ce domaine sont souvent vécues comme des injonctions extérieures et l'expression d'un « catéchisme de la donnée » professé par des entités soupçonnées d'être déconnectées des réalités administratives.

Les freins culturels se manifestent autant dans le partage qu'en amont, dès le stade de la production. Les administrations sont encore trop nombreuses à omettre, par inertie, ou à rechigner, par culture du secret, par précaution face à l'irréversibilité d'une communication excessive de données ou par crainte de soumettre leur activité à la critique, à les diffuser largement au sein de la sphère publique, à supposer du reste qu'on les leur demande. Nombre d'entités, y compris par la voix de leurs dirigeants, ne perçoivent pas – encore – l'importance et l'intérêt d'une « mise en données » de leur activité, en-dehors des traditionnelles statistiques d'activité et des indicateurs de pilotage qui alimentent leurs tableaux de bord. Il est vrai que l'environnement juridique les y incite peu alors, d'une part, que la politique d'open data fait peser sur elles de lourdes charges[216], et, d'autre part, que la valeur des données produites est très souvent **captée par les acteurs privés**, à la faveur du principe de réutilisation libre et gratuite des informations publiques. Il ne s'agit pas ici de minimiser l'importance des retombées socio-économiques de la réutilisation la plus large, considération centrale de la législation européenne[217], mais d'en pointer les effets désincitatifs pour des acteurs publics placés dans l'incapacité de valoriser leur investissement.

L'enjeu prioritaire est de mettre en lumière et de cultiver l'intérêt que présente concrètement la production de données exploitables, pour chaque administration. Les SIA offrent à cet égard des **débouchés nouveaux** qui peuvent convaincre les entités publiques de l'utilité d'y investir davantage, pour peu qu'on leur présente des cas d'usage emblématiques dont la réalisation n'a été rendue possible que par la disponibilité de jeux de données d'entraînement issus des investissements consentis par la puissance publique.

2. Un enjeu technique

L'évolution culturelle que la présente étude appelle de ses vœux doit s'accompagner d'un **investissement massif dans l'informatisation de l'activité administrative, là où elle est incomplète, et dans la mise à niveau des systèmes d'information**. Le financement de nouveaux projets de SIA risque de se faire à fonds perdus si les infrastructures de production et de partage des données ne sont pas adaptées ou deviennent obsolètes faute de maintenance.

L'un des problèmes cruciaux tient à **l'interopérabilité des systèmes** et à l'exploitabilité technique des données en raison de leur **format**. Comme le relève le rapport INSEE/DINUM déjà mentionné, les référentiels non partagés, la non-disponibilité de

[216] Elles peinent d'ailleurs à les assumer. Selon le rapport de la mission confiée par le Premier ministre à M. Eric Bothorel, député, *Pour une politique publique de la donnée* (décembre 2020), près de 90% des collectivités (dont deux départements sur cinq et plus de 34 000 communes) ne respectent pas leurs obligations légales de publication. La mise en ligne de nouveaux jeux de données sur *data.gouv.fr* s'est sensiblement ralentie depuis 2017, après un pic consécutif à l'entrée en vigueur de la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique, et plus d'un tiers de celles qui ont été publiées sur cette plateforme n'ont pas été mises à jour depuis plus de quatre ans. L'essoufflement est particulièrement net pour les administrations de l'État, en particulier déconcentrées.

[217] Directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public.

standards de données, l'hétérogénéité qualitative des données et l'absence de documentation sont des freins majeurs au travail des experts de la donnée.

Une attention particulière doit être portée au choix des **logiciels** auxquels recourent les administrations. L'article 16 de la loi pour une République numérique « encourage » l'utilisation des logiciels libres, par opposition aux logiciels dits « propriétaires »[218] par les administrations, et la circulaire du 27 avril 2021 a prévu la création d'une mission dédiée à l'animation et la promotion interministérielles en matière de logiciel libre[219]. Il y a lieu de rappeler d'ailleurs que les collectivités publiques sont en droit de mettre en concurrence les opérateurs pour l'intégration et l'adaptation à leurs besoins d'un logiciel libre prédéterminé[220]. Du point de vue de la fluidité de la circulation de l'information et de son exploitabilité au sein de la sphère publique[221], l'enjeu n'est toutefois pas d'exclure les logiciels « propriétaires » au profit des logiciels libres, mais de **veiller à ce que la solution, quelle qu'elle soit, utilise des standards (ou formats) ouverts** au sens de l'article 4 de la loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique[222]. Le maître d'ouvrage public doit impérativement veiller à l'interopérabilité la plus large, pour ne pas ériger de nouvelles barrières numériques en sus des frontières institutionnelles entre administrations, ainsi qu'à la portabilité des données et à la réversibilité, pour éviter toute captivité informatique, source de coûts et d'inertie.

A l'heure actuelle, l'article L. 300-3 du CRPA se borne à poser une exigence de standard ouvert, aisément réutilisable et exploitable par un système de traitement autorisé pour la mise à la disposition du public, sous forme électronique, des documents administratifs et informations publiques. L'article 16 de la loi pour une

---

[218] Il s'agit de solutions informatiques grevées de droits de propriété intellectuelle, dont le code source n'est pas modifiable par l'utilisateur, et dont les conditions d'utilisation peuvent être restreintes par leur éditeur.

[219] Le rapport d'information déposé par la mission d'information sur le thème « *Bâtir et promouvoir une souveraineté numérique nationale et européenne* » et présenté par M. Jean-Luc Warsmann et M. Philippe Latombe (n° 4299, enregistré à la présidence de l'Assemblée nationale le 29 juin 2021) propose même de faire du recours aux logiciels libres un principe législatif, et de l'utilisation de logiciels propriétaires une exception à justifier au cas par cas.

[220] CE, 30 septembre 2011, *Région Picardie*, n° 350431, Rec.

[221] Le recours aux logiciels libres répond également à d'autres considérations. En particulier, le code source est connu et adaptable aux besoins de l'administration, la licence d'utilisation est en principe gratuite, et sa sécurisation bénéficie le plus souvent de l'investissement de la communauté des utilisateurs – même si un logiciel libre n'est pas nécessairement plus sûr. Le choix doit naturellement être guidé, à titre principal, par la satisfaction optimale des besoins et tenir compte des coûts directs et indirects de chaque solution (notamment celui de la maintenance et de la modification du code-source des logiciels libres, par les équipes internes ou des prestataires externes).

[222] « *On entend par standard ouvert tout protocole de communication, d'interconnexion ou d'échange et tout format de données interopérable et dont les spécifications techniques sont publiques et sans restriction d'accès ni de mise en œuvre*". En vertu de l'article L. 300-4 du code des relations entre le public et l'administration, « *Toute mise à disposition effectuée sous forme électronique en application du présent livre se fait dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé* ».

République numérique se limite quant à lui à « encourager » l'utilisation des formats ouverts lors du développement, de l'achat ou de l'utilisation, de tout ou partie, de leurs systèmes d'information. La loi pourrait poser le principe du recours à de tels formats, le choix de formats fermés ne pouvant être justifié, à titre exceptionnel, qu'en l'absence de toute autre alternative permettant de satisfaire l'essentiel des besoins exprimés par l'administration, et en soumettant préalablement ce choix à la DINUM. En outre, l'État devrait accompagner prioritairement les collectivités territoriales qui s'engagent à ne recourir qu'à des standards ouverts car l'absence d'interopérabilité des données des systèmes d'information locaux et nationaux est susceptible de pénaliser l'ensemble de la communauté publique.

Le développement des interfaces de programmation (*application programming interface* - API) publiques constitue enfin un chantier essentiel, qui doit être priorisé sur les jeux de données les plus valorisables. La mise en place du portail *api.gouv.fr* par la DINUM a marqué à cet égard un réel progrès. Il doit être poursuivi et amplifié en posant le principe que les données de valeur de tout nouveau système d'information public doivent pouvoir être collectées par cette voie.

3. Un enjeu juridique

Les **freins d'ordre juridique**, réels ou perçus, apparaissent tout aussi pénalisants.

Le cadre juridique du partage intra-public de la donnée s'est certes assoupli dans la période récente. Outre les dispositions qui mettent en œuvre la règle de simplification administrative « dites-le nous une fois », circonscrite à l'instruction des demandes et qui ne présente donc guère d'utilité pour la conception de SIA[223], l'article 1er de la loi pour une République numérique a créé un régime de communication et de réutilisation des documents administratifs entre administrations, directement inspiré des règles de droit commun de l'accès des citoyens aux documents administratifs. Toute administration est tenue de communiquer gratuitement à toute autre qui en fait la demande pour les besoins de l'accomplissement de ses missions de service public, les documents administratifs

[223] L'article 16A de la loi dite DCRA du 12 avril 2000, issu de la loi n° 2011-525 du 17 mai 2011 de simplification et d'amélioration de la qualité du droit et désormais codifié aux articles L. 114-8 et suivants du code des relations entre le public et l'administration a posé une obligation d'échange entre administrations de « *toutes les informations ou données strictement nécessaires pour traiter une demande présentée par le public ou une déclaration transmise par celui-ci en application d'un texte législatif ou réglementaire* », sans que le secret professionnel soit opposable dès l'instant que l'administration destinataire est habilitée à connaître des données qui lui sont fournies. Cette obligation est toutefois circonscrite à l'instruction des demandes et déclarations du public. Elle ne porte de surcroît que sur les domaines et procédures énumérés par décret en Conseil d'Etat, lequel fixe la liste limitative des administrations auprès desquelles la demande de communication s'effectue selon la nature des données (art. 114-9-1 et suivants). Le projet de loi relatif à la différenciation, à la décentralisation, à la déconcentration et portant diverses mesures de simplification de l'action publique locale entend faire sauter le premier verrou et, ainsi, généraliser l'obligation d'échange à l'ensemble des procédures administratives. Toutefois, un décret simple continuerait de déterminer, pour chaque type de données, la liste des administrations responsables de leur mise à disposition.

qu'elle détient, sous réserve des secrets protégés par la loi et dans le respect des règles de protection des données à caractère personnel. Ce régime ne porte pas sur les données mais sur les documents, ce qui suppose soit qu'il existe un support écrit (le cas échéant numérique) sur lequel figurent ces données, soit que le document puisse être produit par extraction des bases de données dont l'administration dispose et sans faire peser une charge déraisonnable sur elle[224].

En l'état, et sauf dispositions particulières, les administrations n'ont, étonnamment, pas plus de droits que le grand public dans l'accès aux données des autres administrations. Et elles peuvent même, en pratique, avoir moins de facilités d'accès qu'un prestataire privé travaillant pour une administration, avec lequel cette dernière peut être amenée à partager des informations confidentielles pour les besoins de l'exécution d'un marché public ou d'un contrat de partenariat.

Par ailleurs, l'échange d'informations publiques entre les administrations aux fins de l'exercice de leur mission de service public ne constitue pas une réutilisation à laquelle s'appliquerait le régime de droit commun prévu au titre II du livre III du CRPA (dernier alinéa de l'art. L. 321-2 de ce code, qui reprend sur ce point l'exclusion figurant au point 11 de l'article 2 de la directive 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public). Paradoxalement, les administrations ne peuvent donc revendiquer le bénéfice du principe de libre réutilisation prévu à l'article L. 321-1 de ce code et les conditions de l'utilisation sont librement fixées par l'administration détentrice des données, sous réserve du principe de gratuité au sein de la seule sphère constituée de l'État et de ses établissements publics administratifs figurant au dernier alinéa du I de l'article 1er de la loi pour une République numérique[225].

Plusieurs pistes de réflexion pourraient être explorées afin de fluidifier la circulation des informations au sein de la sphère publique et d'en faciliter la réutilisation, eu égard à l'intérêt général que présente un tel partage. Il convient de partir d'une double idée : d'une part, **l'administration n'est pas propriétaire des données qu'elle détient : elle n'en est que la gardienne**, et son devoir est de les partager pour qu'en soit tirée toujours plus de valeur, dans l'intérêt général ; d'autre part, ce partage ne doit avoir d'autre borne que la nécessaire confidentialité de certaines données et, en surplomb, la nécessité de ne pas fragiliser la confiance dans les institutions et collectivités publiques auxquelles des données sont remises par des tiers.

[224] CE, 13 novembre 2020, *S.*, n° 432832, T.

[225] Alors que ce principe est d'application générale pour les réutilisateurs privés, sous réserve de dérogations très circonscrites, en vertu de l'art. L. 324-1 du CRPA.

a/ Le cadre juridique du partage et de la réutilisation de données entre administrations devrait être assoupli, dans le respect des exigences constitutionnelles[226].

L'article 1er de la loi pour une République numérique devrait être étendu aux données – en d'autres termes, la loi devrait consacrer un droit d'accès des administrations aux données détenues par les autres, et non pas seulement un droit d'accès aux documents administratifs formalisés, dans le respect des principes en vigueur.

En outre, à l'instar du droit européen de l'accès aux documents et du droit d'accès aux informations relatives à l'environnement[227], il pourrait être prévu que les secrets protégés par la loi ne sont opposables à l'administration demanderesse que si les inconvénients de la communication excèdent l'intérêt public que présente cette communication, eu égard à l'usage que cette dernière entend en faire. Certains secrets particulièrement sensibles (secret de la défense nationale notamment) pourraient être exclus de cet assouplissement. Afin de ne pas dissuader la fourniture volontaire de certaines informations par le public (notamment par les entreprises) et de ne pas pénaliser le partage de données privé-public (v. *infra* 3.), des clauses particulières de confidentialité pourraient être envisagées au cas par cas, tenant compte également de droits de propriété intellectuelle de tiers. En outre, il devrait être clairement énoncé dans la loi que les droits de propriété intellectuelle que peuvent détenir les administrations ne peuvent être opposés à d'autres[228].

S'agissant plus particulièrement de la constitution des jeux de données d'entraînement et de validation aux fins de conception et de développement des SIA publics, il pourrait être envisagé, le cas échéant à titre expérimental, de présumer que l'intérêt public justifie la communication entre administrations des données qu'elles détiennent, même couvertes par un secret protégé par la loi, à charge pour l'administration détentrice de démontrer que les exigences de confidentialité doivent l'emporter. Une habilitation spéciale des agents amenés à traiter de telles données devrait être exigée, de même qu'une complète traçabilité des opérations. Ce régime pourrait s'inspirer de celui de la réutilisation de données de l'administration protégées pour des raisons de confidentialité des informations commerciales ou des données statistiques, de protection des droits de propriété intellectuelle de tiers et de protection des données à caractère personnel, figurant dans le projet de règlement européen sur la gouvernance européenne des données (dit « *Data Governance Act* »). Ce dernier prévoit en particulier la possibilité pour l'administration d'exiger du réutilisateur qu'il effectue ses opérations dans un

[226] V. notamment la décision n° 2021-924 QPC du 9 juillet 2021 censurant les dispositions qui permettaient à toute autorité administrative, d'initiative ou sur demande, de communiquer aux services de renseignement toute information utile à l'accomplissement de leurs missions.

[227] Art. L. 124-4 du code de l'environnement

[228] Une ambiguïté subsiste à cet égard dès lors que le III de l'article 1er de la loi pour une République numérique rend applicable le titre Ier du livre III du CRPA, où figure l'art. L. 311-4 qui dispose que les documents administratifs sont communiqués ou publiés sous réserve des droits de propriété littéraire et artistique.

environnement de traitement sécurisé défini comme « *l'environnement physique ou virtuel et les moyens organisationnels donnant la possibilité de réutiliser les données d'une manière qui permette à l'opérateur de l'environnement de traitement sécurisé de déterminer et de surveiller toutes les opérations de traitement de données, notamment d'afficher, de stocker, de télécharger, d'exporter les données et de calculer les données dérivées au moyen d'algorithmes de calcul* ».

Une réflexion pourrait également être ouverte sur la création d'un régime particulier de partage intra-publique facilité de **jeux de données de haute valeur** identifiés par la DINUM, sur le modèle des « données de référence » pour ce qui concerne la réutilisation des informations publiques[229].

Enfin, afin d'accélérer la résolution des litiges entre autorités administratives, le recours administratif préalable obligatoire devant la CADA, qui s'impose à elles en vertu du II de l'article 1er de la loi pour une République numérique, devrait être supprimé. Le volume de ces litiges est en effet très modeste en comparaison de ceux qui sont portés devant la CADA par des personnes privées. La CADA pourrait se voir confier soit un **rôle de médiateur** en cas de divergence de doctrines entre administrations sur la possibilité de partager des données, soit un **pouvoir décisionnel de règlement des différends**.

b/ Il est également nécessaire de **clarifier la portée du cadre juridique de la protection des données à caractère personnel en ce qui concerne l'échange inter-administrations de ces données et leur utilisation subséquente à des fins de conception ou de modification d'un SIA**, afin d'éviter que ces opérations se heurtent aux divergences d'interprétation entre les administrations et/ou entre délégués à la protection des données. À cet égard, le renvoi général à la loi du 6 janvier 1978 auquel procède l'article 1er de la loi pour une République numérique est non seulement imprécis, mais de nature à induire en erreur quant à la réalité des contraintes pesant sur les administrations, en accréditant l'idée fausse de l'incommunicabilité de principe de telles données.

La fourniture de données à caractère personnel par une administration A à une administration B constitue un traitement de telles données, dont la conformité au RGPD dépend, pour l'essentiel, de celle du traitement auquel B entend procéder à l'aide des données ainsi collectées – en l'occurrence, la conception ou la modification d'un SIA, c'est-à-dire l'entraînement du modèle algorithmique (phase d'apprentissage), à l'exclusion de sa mise en œuvre opérationnelle (phase d'inférence).

1/ En premier lieu, **l'exigence de licéité** posée par l'article 5 de la loi de 1978 ne devrait, en principe, pas soulever de difficulté dès lors que le SIA sera **nécessaire à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement** (e du 1. de l'article 6 du RGPD). L'exigence de nécessité ne paraît pas impliquer que l'administration démontre être dans l'impossibilité de parvenir à ses fins par d'autres moyens que la mise en œuvre du traitement. Il n'existe pas, en effet, de principe de subsidiarité ou

[229] V. art. L. 321-4 et suivants du CRPA.

d'épuisement des solutions alternatives qui relèguerait les traitements de données à caractère personnel au rang de solution de dernier recours. Il faut et il suffit que l'administration établisse que le traitement qu'elle met en œuvre répond aux objectifs d'intérêt général qu'elle s'est fixés. En outre, le traitement ne doit pas porter une atteinte excessive aux intérêts des personnes concernées, le 3. de l'article 6 du RGPD exigeant que le droit de l'État membre qui permet ce traitement réponde à un objectif d'intérêt public et soit proportionné à l'objectif légitime poursuivi. La question de la licéité peut toutefois se poser lorsque l'administration entreprend d'exploiter des données sans finalité précise, afin d'en tirer des corrélations ou connaissances nouvelles (en particulier, en recourant à des modèles d'apprentissage non supervisé pour catégoriser les données).

2/ En deuxième lieu, le **principe de minimisation des données** ne devrait pas davantage constituer un réel obstacle. Il ne signifie pas que le responsable de traitement doit traiter le moins de données possible, mais qu'il ne saurait traiter davantage de données que celles qui sont nécessaires pour satisfaire la finalité assignée au traitement. Il s'agit donc moins d'une obligation de fond aboutissant à appauvrir les jeux de données et à obérer les performances du SIA que d'une obligation de justification au cas par cas de la nécessité du traitement de telles ou telles données. En outre, il importe de ne pas perdre de vue que cette exigence de minimisation peut être satisfaite, dans certains cas, par la pseudonymisation des données (sous réserve que celle-ci ne prive pas le traitement de son intérêt), plutôt que par leur suppression pure et simple. Enfin, il y a lieu de rappeler que l'un des axes actuels de la recherche porte sur la « **frugalité en données** » des modèles (« *few-shot learning* »), c'est-à-dire l'optimisation de l'exactitude du modèle par le recours à un minimum de données (cas des modèles géants auto-supervisés).

3/ En troisième lieu – ce point est déterminant – le **principe de finalité** devrait être appliqué de façon souple pour les besoins du développement des SIA. Le paragraphe 1 de l'article 5, sous b), du RGPD interdit de traiter des données à caractère personnel d'une manière « incompatible » avec les finalités pour lesquelles elles ont été collectées (encadrement des « traitements ultérieurs »)[230]. Lorsqu'une administration A est sollicitée par une administration B souhaitant développer un SIA pour obtenir auprès de la première des données à caractère personnel, l'administration A doit s'assurer que la transmission des données à cette fin, qui constitue un traitement de données, n'est pas incompatible avec les finalités ayant justifié la collecte des données par ses soins.

Le RGPD pose lui-même une présomption absolue de compatibilité pour les traitements ultérieurs à des fins de recherche scientifique, laquelle est comprise dans un sens large comme incluant « *le développement et la démonstration de technologies* » et les études dans le domaine de la santé publique. Certains SIA publics relèveront de cette exception (ex : identification des variables explicatives

[230] Le paragraphe 2 de l'article 4 et l'article 9 de la directive « police-justice » fixent des conditions distinctes pour le traitement ultérieur.

d'une pathologie à des fins de prédiction[231]) mais il est douteux que tous les SIA puissent en bénéficier (ex: ré-entraînement dans un contexte donné d'un algorithme publiquement disponible, sans réelle innovation technologique)[232].

S'agissant d'un traitement reposant sur les nécessités du service public (et non sur le consentement), une liste non exhaustive des critères utilisés pour le « test de compatibilité »[233] figure au paragraphe 4 de l'article 6 du RGPD[234]. On y trouve notamment :

- la pseudonymisation : celle-ci semble devoir être quasi-systématique dès lors que les éléments directement identifiants (nom, prénom…) ne sont, la plupart du temps, d'aucune utilité pour la conception d'un SIA ;
- et les « *conséquences possibles du traitement ultérieur envisagé pour les personnes concernées* ». En soi, la conception d'un SIA n'emporte pas de conséquence directe sur les personnes concernées, tant qu'il n'est pas mis en service. En phase de déploiement, des conséquences défavorables, du point de vue des individus, peuvent s'y attacher, selon la destination du système (par exemple, un modèle d'optimisation du contrôle fiscal ou de lutte contre la fraude sociale) et il importe donc d'examiner si les personnes concernées seraient naturellement enclines, si la question leur était posée, à approuver l'usage de leurs données à cette fin. En d'autres termes : en tenant compte de la finalité d'intérêt général poursuivie, peut-on présumer que les intéressés seraient d'accord pour que leurs données soient utilisées en vue de concevoir le système en cause ? Il y a là une expression indirecte des « *droits des personnes de décider et de contrôler les usages qui sont faits des données à caractère personnel les concernant* » mentionnés à l'article 1er de la loi de 1978.

---

[231] V. par ex. la délibération de la CNIL n° 2020-055 du 14 mai 2020 relative à un traitement ayant pour finalité une étude portant sur le développement et la validation d'algorithmes de prédiction des crises de décompensation cardiaque chez les patients porteurs d'implants cardiaques connectés.

[232] Il n'est sans doute pas anodin que le Conseil d'Etat, statuant au contentieux, n'ait pas fait référence aux assouplissements propres à la recherche que comporte le RGPD, invoquées en défense, lorsqu'il s'est penché sur la légalité du décret du 27 mars 2020 créant le traitement de données à caractère personnel DataJust (CE, 30 décembre 2021, *Sté Gerbi Avocat Victimes et Préjudices et autres*, n° 440376 et a.).

[233] Il serait d'ailleurs plus rigoureux d'évoquer un « test de non-incompatibilité », cette double négation utilisée par le RGPD témoignant de la volonté de ses auteurs de laisser plus de latitude aux responsables de traitement, comme l'avait relevé le G29 dans son avis sur le principe de finalité, rendu sous l'empire de la directive 95/46 (avis WP 203 du 2 avril 2013).

[234] Existence éventuelle d'un lien entre les finalités pour lesquelles les données à caractère personnel ont été collectées et les finalités du traitement ultérieur envisagé ; contexte dans lequel les données ont été collectées, en particulier en ce qui concerne la relation entre les personnes concernées et le responsable du traitement (notamment les attentes raisonnables des personnes concernées) ; nature des données ; conséquences possibles du traitement ultérieur envisagé pour les personnes concernées ; existence de garanties appropriées dans le traitement initial comme dans le traitement ultérieur, comme le chiffrement ou la pseudonymisation.

Il y a place pour la construction, par les pouvoirs publics, en lien avec la CNIL, d'une **grille d'analyse de la compatibilité** permettant de situer le projet entre deux extrêmes :

- Un SIA développé dans le même champ d'activité que celui pour lequel les données à caractère personnel ont été collectées (ex. : système d'aide à l'évaluation du préjudice corporel sur la base des données personnelles contenues dans les décisions de justice), répondant à des motifs d'intérêt général éminents, dont la destination est susceptible de bénéficier aux personnes concernées et, en tous les cas, qui n'est pas susceptible de prendre, de recommander ou de participer à des décisions qui leur seraient défavorables (ex. : améliorer le diagnostic d'une maladie), utilisant des données fournies spontanément à l'administration et non de façon contrainte, ne relevant pas des « données sensibles » et préalablement pseudonymisées dans les règles de l'art, devrait bénéficier d'une présomption de compatibilité ;
- à l'inverse, un SIA développé dans un champ totalement étranger à celui qui avait justifié la collecte, pour des motifs d'intérêt général de second ordre, recourant à des données sensibles non pseudonymisées obtenues de manière coercitive ou contrainte, et susceptibles de prendre des décisions défavorables à l'égard des personnes concernées, serait présumé poursuivre une finalité incompatible.

En toute hypothèse, il convient de procéder à une analyse au cas par cas, en opérant une balance entre l'intérêt public que présente le SIA qu'il s'agit de développer et l'atteinte portée aux intérêts des personnes concernées.

En cas de doute sur le résultat de la balance des intérêts, eu égard notamment à la sensibilité particulière des données en cause, le développement du modèle pourrait intervenir dans un environnement de traitement sécurisé contrôlé par la CNIL.

Les articles 53 et 54 de la proposition de règlement européen sur l'IA prévoient précisément la mise en place de « **bacs à sable réglementaires** » dans lesquels des opérateurs sélectionnés pourraient, à des conditions très strictes, utiliser, pour la conception et le développement de SIA[235], des données à caractère personnel collectées à d'autres fins. En l'état de la proposition, seuls les SIA poursuivant un intérêt public important dans les domaines de la prévention et de la détection des infractions pénales et des menaces pour la sécurité publique, de la sécurité publique et de la santé publique, et en matière de protection de l'environnement pourraient recourir à ce dispositif, ce qui apparaît particulièrement restrictif.

Le recours à un bac à sable réglementaire devrait être possible, plus largement, dès lors que son utilité publique est avérée :

- l'administration responsable du projet de SIA doit démontrer qu'elle ne peut pas le concevoir avec un niveau de performance acceptable sans y recourir, compte tenu de la disponibilité d'autres jeux de données d'entraînement et de la

[235] Ainsi que le relève le CEPD dans l'avis n° 5/2021 sur la proposition de règlement, l'articulation entre ces dispositions et celles du RGPD qui encadre les traitements ultérieurs reste à préciser.

possibilité d'utiliser des données ne présentant pas un caractère personnel (notamment par l'anonymisation de données à caractère personnel) ;

- les avantages attendus du SIA doivent l'emporter sur l'atteinte directe et indirecte portée aux intérêts et aux libertés et droits fondamentaux des personnes concernées.

Les modalités de fonctionnement du bac à sable devraient être fixées au cas par cas dans le cadre d'un protocole convenu entre l'administration et la CNIL, portant notamment sur la nature des données nécessaires, leur pseudonymisation, la limitation de la durée de l'expérimentation et l'évaluation à son issue, les conditions d'information ou de dérogation à l'obligation d'information des personnes concernées, les garanties d'étanchéité de l'environnement utilisé, la traçabilité des opérations et l'habilitation des agents participants.

Enfin, dans le cas particulier où la performance du système exigerait impérativement de poursuivre en environnement réel l'apprentissage engagé dans un bac à sable sur la base de données collectées à d'autres fins, il pourrait être envisagé de l'autoriser pour une durée limitée au strict nécessaire et pour un motif important d'intérêt général, sous la surveillance de la CNIL.

Un dernier obstacle doit être signalé dans l'application du principe de finalité, s'agissant des traitements publics autorisés par un **acte réglementaire** en application des articles 31 et 32 de la loi de 1978, dont l'article 35 prévoit qu'il doit préciser la finalité du traitement. La définition réglementaire des finalités est de nature, par elle-même, à mettre en échec l'application des règles de droit commun du RGPD sur les traitements ultérieurs. La loi de 1978 pourrait être modifiée pour prévoir que l'acte de création du traitement de données à caractère personnel, lorsqu'il est exigé, peut autoriser la réutilisation de ces données aux fins de concevoir et de développer un SIA public, dans le respect du paragraphe 4 de l'article 6 du RGPD et dans le « bac à sable » dont le régime serait également prévu par la loi.

4/ En quatrième et dernier lieu, devrait également être précisée la portée des souplesses offertes par le c) du paragraphe 5 de l'article 14 du RGPD en ce qui concerne le **droit d'information des personnes concernées** sur l'utilisation de données qui n'ont pas été collectées auprès d'elles. Ces dispositions suppriment l'obligation d'information lorsque et dans la mesure où celle-ci exigerait des « *efforts disproportionnés* », ce qui vise notamment la recherche scientifique et le cas dans lequel les objectifs du traitement seraient gravement compromis par la mise en œuvre de cette obligation. Surtout, celle-ci peut également être éludée dans le cas où « *l'obtention ou la communication des informations sont expressément prévues par le droit de l'Union ou le droit de l'État membre auquel le responsable du traitement est soumis et qui prévoit des mesures appropriées visant à protéger les intérêts légitimes de la personne concernée* », ce qui permet d'envisager un régime législatif accommodant pour le partage de données au sein de la sphère publique en vue de la conception et le développement des SIA, sous réserve de prévoir, *a minima*, une obligation de transparence par voie de publication en ligne des caractéristiques du système en projet et de la nature des données utilisées pour le concevoir.

Il y a lieu enfin de rappeler que les droits prévus aux articles 15 à 20 du RGPD (droit d'accès, de rectification, d'effacement, de limitation et de portabilité) ne sont pas applicables au traitement dont le responsable est à même de démontrer qu'il n'est pas en mesure d'identifier la personne concernée et que celle-ci n'a pas fourni d'informations complémentaires permettant de l'identifier.

c/ La problématique du traitement des données à caractère personnel, appliquée aux SIA, soulève également la délicate question des **formalités** nécessaires à la création d'un tel traitement.

Il importe certainement de **rompre sans plus attendre avec la culture de la « formalité préalable »** à la mise en œuvre des traitements de données à caractère personnel par l'administration, qui a survécu à la logique de simplification et de responsabilisation sur laquelle repose le RGPD et qui retarde les projets de SIA publics. A ce titre, la doctrine de l'État doit exclure tout recours à des décrets de confort et, plus largement, à des formalités qui ne seraient pas exigées par les textes. Les actes institutifs, dont les mentions ne rendent pas fidèlement compte du fonctionnement réel des traitements, n'offrent qu'une illusion de sécurité aux responsables de traitements, en même temps qu'ils alourdissent la charge de travail des services, retardent la mise en œuvre des projets et suscitent des actions contentieuses dans le cadre desquelles les conditions de mise en œuvre du traitement ne peuvent être utilement discutées, ce qui est de nature à susciter un sentiment de frustration chez les justiciables.

Il y a lieu de rappeler qu'il est admis que les chefs de service puissent, au titre de leur pouvoir réglementaire d'organisation du service[236], définir eux-mêmes les traitements de données destinés à être utilisés par les services placés sous leur autorité[237]. Une telle création devrait définir les caractéristiques essentielles du traitement (finalités, catégories de personnes concernées, catégories de données traitées et catégories de destinataires). Le registre des activités de traitement qu'il appartient au responsable de traitement de tenir en vertu de l'article 30 du RGPD et de l'article 57 de la loi de 1978 pourrait éventuellement en tenir lieu, sous réserve qu'il soit validé par l'autorité compétente et publié[238].

On relèvera toutefois l'existence d'une double incertitude sur la portée de ce pouvoir, qui dépasse la question des SIA publics.

D'une part, le paragraphe 3 de l'article 6 du RGPD exige que le fondement d'un traitement reposant sur la nécessité pour le service public soit défini par « *le droit de*

---

[236] Issu de la décision CE, 7 février 1936, *Jamart* par laquelle le Conseil d'État a reconnu un pouvoir réglementaire minimal aux ministres en qualité de chef de service, dont ils disposent en l'absence de toute habilitation par une loi ou un décret, afin de prendre les mesures nécessaires à l'organisation de leurs services.

[237] CE, 6 novembre 2019, *Fédération des acteurs de la solidarité et autres*, n° 434376-434377, T.

[238] Selon la CADA, un tel registre constitue un document administratif communicable sur le fondement du CRPA, sous réserve de l'occultation d'information dont la divulgation porterait atteinte à un secret protégé par la loi, comme la sécurité des systèmes d'information (avis n° 20194129 du 12 mars 2020).

l'État membre auquel le responsable de traitement est soumis ». Or la portée de cette formulation reste incertaine. Il semble se déduire des considérants 41 et 45 du RGPD que cette expression n'exige ni un acte législatif adopté par le Parlement, ni même une base légale spécifique pour chaque traitement individuel, pourvu que la base soit claire et précise. Un acte réglementaire de type *Jamart* semble pouvoir entrer dans ces prévisions, mais ce point n'a pas été éclairé par la jurisprudence.

D'autre part, se pose la question de savoir si un traitement relevant du RGPD peut être créé par une décision du chef de service, sans avis de la CNIL, alors même qu'il contient des données sensibles. L'ambiguïté rédactionnelle que recèle le III de l'article 6 de la loi de 1978 devrait être levée par le législateur. Dans cette attente, l'Assemblée générale du Conseil d'État a estimé que ces dispositions n'imposaient pas le recours en pareil cas à un décret en Conseil d'État après avis motivé et publié de la CNIL.

Enfin, le mécanisme d'autorisation de la CNIL pour les traitements de données de santé[239], qui résulte d'un choix de la France et non d'une exigence du RGPD (article 66 de la loi de 1978), peut être de nature à compliquer les projets de recherche en matière d'IA. L'autorisation étant délivrée pour une finalité précise, un léger infléchissement ou enrichissement de celle-ci requiert d'obtenir une nouvelle autorisation. Cette éventualité est d'autant plus vraisemblable que la mise en œuvre des techniques d'apprentissage automatique est susceptible de dévoiler de nouvelles pistes prometteuses. Afin d'accorder plus de souplesse aux organismes de recherche, il pourrait être envisagé que, lorsqu'une autorisation a été accordée dans ce cadre, l'organisme pourrait procéder à une modification limitée des caractéristiques du traitement, sans en remettre en cause la nature ni, de façon substantielle, l'incidence sur les droits et libertés des personnes, moyennant une déclaration à la CNIL. Un autre allègement bienvenu consisterait à supprimer l'avis motivé et publié de la CNIL prévue au 6° de l'article 44 de la loi de 1978, pour les traitements de données sensibles nécessaires à la recherche publique et justifiés par des motifs d'intérêt public importants.

Il y a lieu enfin de rappeler que la création par voie réglementaire d'un traitement de données à caractère personnel, dans le cadre de la loi du 6 janvier 1978, ne préjuge pas de la nécessité d'asseoir sur une base législative dédiée un dispositif technique de collecte dont les modalités seraient de nature à affecter les garanties fondamentales accordées aux citoyens pour l'exercice des libertés publiques, en particulier la vie privée, la liberté d'expression ou la liberté de réunion, au-delà de la question de la protection des données. Ainsi, la circonstance qu'une technique de renseignement, comme des écoutes téléphoniques, implique la mise en œuvre d'un traitement de données à caractère personnel, autorisé par décret, ne permet pas de se dispenser d'une base législative comme celle qui figure à cette fin dans le code de la sécurité intérieure. Tel est aussi le cas de la captation d'images à l'aide de caméras aéroportées, notamment par des drones[240]. Se pose également la question, déjà évoquée, de la

[239] Art. 66 de la loi du 6 janvier 1978.

[240] CE, 22 décembre 2020, *Association La Quadrature du Net*, n° 446155, et art. L. 242-1 et suivants du code de la sécurité intérieure.

nécessité d'une base légale spécifique pour la mise en œuvre de SIA démultipliant les capacités d'exploitation et de surveillance par les services.

d/ Une dernière difficulté d'ordre juridique mérite d'être mentionnée : elle a trait au risque de transfert inapproprié de données à caractère personnel par des opérateurs soumis à des législations étrangères et aux limitations qu'impose le RGPD pour le prévenir, en ce qui concerne le transfert de données en-dehors de l'Union européenne. Cet écueil est tout particulièrement susceptible de pénaliser l'indispensable développement des **plateformes de données**, publiques ou accessibles à un public restreint à des administrations, des chercheurs et des partenaires sélectionnés, qui sont aujourd'hui largement tributaires des services proposés par des groupes soumis au droit américain. L'expérience du *Health Data Hub* l'illustre.

Le *Health Data Hub*

La Plateforme des données de santé (PDS), également appelée « *Health Data Hub* » (HDH), a été créée, sous forme de groupement d'intérêt public, par un arrêté du 29 novembre 2019. Il compte à ce jour 56 parties prenantes : l'État, la CNAM, des établissements de santé, des usagers, des professionnels de santé, des agences, ainsi que des industriels. Depuis juin 2021, il est, avec la CNAM, co-responsable du traitement du système national des données de santé (SNDS).

Le HDH a pour objectif de mettre à disposition des acteurs, sous réserve d'une autorisation de la CNIL lorsque les données ne sont pas anonymisées mais simplement pseudonymisées, un catalogue de données de santé, constitué du SNDS « base principale » et de données cliniques, jugées les plus pertinentes pour la recherche et l'innovation.

A la suite de l'arrêt de la CJUE du 16 juillet 2020, dit *Schrems II*, qui a jugé que la protection des données transférées vers les Etats-Unis par le « *Privacy Shield* » (ou « bouclier de protection des données ») était insuffisante au regard du droit européen, la question de la protection des données du HDH, hébergées par une filiale irlandaise de la société américaine Microsoft dans des serveurs néerlandais, a donné lieu à un âpre débat juridique. Par une ordonnance du 13 octobre 2020 (n° 444937), le juge des référés du Conseil d'État a relevé qu'il ne pouvait être totalement exclu que les autorités américaines, dans le cadre de programmes de surveillance et de renseignement, demandent à Microsoft l'accès à certaines données mais a relevé, d'une part, l'existence d'un intérêt public important à permettre la poursuite de l'utilisation des données de santé pour les besoins de l'épidémie de covid-19 grâce aux moyens de la Plateforme et, d'autre part, que les données étaient pseudonymisées avant leur hébergement, pour en conclure qu'il n'y avait pas lieu pour lui, eu égard à l'office du juge du référé-liberté, d'enjoindre à la Plateforme de suspendre ses activités ou de rompre le contrat la liant à Microsoft. Le JRCE a, en outre, demandé au HDH de continuer, sous le contrôle de la CNIL, à travailler avec Microsoft pour renforcer la protection des droits des personnes concernées sur leurs données personnelles, dans l'attente d'une solution permettant d'éliminer tout risque d'accès aux données personnelles par les autorités américaines.

En janvier 2022, le calendrier prévu pour l'autorisation par la CNIL de l'hébergement, au sein de la plateforme, du SNDS et les autres bases de données dont le HDH est responsable de traitement, a été revu, emportant le retrait de la demande d'autorisation formée auprès de la CNIL et un nouveau report du lancement opérationnel de la plateforme.

Cet exemple montre une fois encore l'importance du développement d'une offre de services maîtrisée par des acteurs français et européens.

Les difficultés du HDH ne se réduisent toutefois pas à la problématique de la maîtrise des données. Ainsi, malgré l'intérêt évident de cette initiative, notamment pour l'accélération des projets de recherche et le déploiement de l'IA en santé, celle-ci n'a pas encore emporté l'adhésion d'une majorité des acteurs de la santé, et notamment d'une partie des établissements hospitaliers qui sont nombreux à privilégier un hébergement local des données de santé issues de leurs bases. La perspective d'une captation, par des opérateurs privés, de la valeur ajoutée créée par le secteur public, sans retour sur investissement direct pour ce dernier, est également de nature à retarder ou handicaper son développement, comme la culture encore insuffisante du partage de la donnée.

Cette problématique du partage des données devra être abordée de manière spécifique aux **collectivités territoriales**. Détentrices de données d'une richesse exceptionnelle, elles sont les plus vulnérables au risque de pillage par des prestataires ou des tiers utilisant sans vergogne – et sans retour pour les collectivités qui les auront élaborées et traitées – cette masse considérable de données. De même, il est aujourd'hui difficile pour une collectivité d'envisager une politique autonome de valorisation d'usage ou d'échange de ses données, qui demande une expertise que même l'État commence seulement à acquérir : pourtant, laisser en jachère cette ressource handicaperait significativement l'essor de l'usage des SIA publics. On ne peut ici qu'esquisser les bases d'un questionnement, qui, en pleine collaboration avec les élus et leurs associations, devra réfléchir à l'intérêt d'un régime spécifique, à des formes nouvelles de coopération, à la question du juste retour et de la bonne échelle de décision, aux besoins propres de chaque type de collectivité…

4. L'accès des administrations aux données du secteur privé et leur exploitation

La politique d'*open data* des administrations publiques et le régime juridique de la réutilisation des informations publiques ont mis l'accent sur la fourniture par le secteur public des données utiles au développement de services par le secteur privé. Ils ne doivent toutefois pas occulter, en sens inverse, l'importance de l'accès du premier aux données produites par le second (flux dit « *Business to Government* » – B2G).

L'immense gisement de données générées par les activités privées, notamment commerciales, et l'extraordinaire potentiel de connaissances qu'il renferme doivent être davantage exploités pour accroître l'efficacité des politiques publiques et la performance des services publics, y compris par le développement de SIA les utilisant comme données d'entraînement. L'administration doit parfois consentir des efforts

importants pour produire elle-même des données qu'elle pourrait se procurer beaucoup plus aisément auprès d'acteurs privés. Ces données peuvent être brutes ou résulter d'un traitement par l'opérateur privé, lequel peut ne pas même les exploiter pour ses propres besoins. Elles peuvent présenter une certaine sensibilité (données à caractère personnel, secret des affaires…) ou aucune (certaines données statistiques par exemple).

A l'heure actuelle, la fourniture de ces données peut résulter d'une obligation légale (obligation de déclaration, dépôt d'une demande, données recueillies lors d'un contrôle, obligation de rendre compte de l'exécution d'une délégation de service public…), donner lieu à un partenariat contractuel entre l'administration et un acteur privé, faire l'objet d'un marché public d'achat de données, ou encore s'inscrire dans une démarche citoyenne (exemple du « *Self Data territorial* » porté par l'association La Fing et déployée dans plusieurs villes dont Nantes, Lyon et La Rochelle)[241]. Il s'agit souvent de transferts de données ponctuels, au coup par coup, et pour des finalités précises, dont les autres administrations ne bénéficient pas le plus souvent pour les raisons déjà indiquées. Elles sont le plus souvent étrangères à la conception d'un SIA. On peut ici mentionner, à titre d'illustration, l'indice des prix réalisé par l'INSEE au moyen des données de la grande distribution, sur le fondement de l'article 3 de la loi du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques.

Les obstacles à l'intensification de ce flux d'informations privé-public sont nombreux : le manque de sollicitations de la part d'autorités administratives qui connaissent plus mal encore que les leurs les données dont disposent les acteurs privés ; le coût du transfert de données (mise en place d'une infrastructure sécurisée, coûts de transaction pour la conclusion de partenariats…) ; le manque d'intérêt direct et d'incitation pour le fournisseur privé de données, la crainte d'un effet-boomerang dans le cas où les données seraient utilisées pour réguler davantage son secteur d'activité, le risque de mise en cause de sa responsabilité à raison de données de mauvaise qualité ou obsolètes utilisées pour prendre des décisions publiques dommageables, ou encore l'absence de cadre juridique clair, notamment en droit de la propriété intellectuelle et en droit de la concurrence[242].

Il est impératif d'encourager la conclusion de partenariats avec les acteurs privés sur la base d'un recensement des données présentant un intérêt particulier pour l'action publique[243]. La circulaire du Premier ministre du 27 avril 2021 a annoncé, sur ce sujet, la mise en place d'une mission de préfiguration de la fonction de médiateur de la donnée d'intérêt général qui devait rendre ses conclusions le 1er décembre 2021 mais qui n'a pas encore été lancée. La mission des AMDAC devrait inclure

[241] M. Molins et autres (La Fing), *Kit Self Data territorial*, octobre 2019.

[242] V. sur l'ensemble de ces points le rapport du groupe d'experts de haut niveau sur le partage de données privé-public, *Towards a European strategy on business-to-government data sharing for the public interest*, rapport final, 2020.

[243] On notera que le projet de règlement sur la gouvernance des données (« *Data Governance Act* ») entend promouvoir « l'altruisme en matière de données », qui désigne la démarche de mise à disposition volontaire de données par des particuliers, des associations ou des entreprises, pour le bien commun.

explicitement cette dimension, en diffusant au sein de leur périmètre le **réflexe du partage de données privé-public**, notamment lorsqu'un besoin identifié fait naître l'idée de développer un SIA nécessitant certaines données non disponibles au sein de la sphère publique.

Il est toutefois à craindre que les démarches entreprises par les autorités publiques ne suffisent pas à surmonter les réticences des acteurs privés. Il conviendrait de réfléchir à des incitations financières (dédommagement financier – comme c'est le cas pour les réquisitions judiciaires, inclusion du partage de la donnée dans les conditions d'attribution des marchés publics ou comme critère de choix des offres…) ou symboliques (à l'instar du projet de registre des « organisations altruistes en matière de données » prévu dans la proposition de règlement européen sur la gouvernance des données), mais aussi à des **procédures plus coercitives de réquisition de données** comme celle qui est envisagée par la proposition de règlement sur les données (dit « *Data Act* »).

Cette proposition prévoit la possibilité, en-dehors du champ police-justice, d'une **mise à disposition contrainte des données du secteur privé au profit des administrations publiques en cas de « besoin exceptionnel »**, c'est-à-dire soit pour répondre à ou prévenir une situation d'urgence publique, soit lorsque l'absence de données disponibles fait obstacle à ce qu'une administration s'acquitte d'une tâche d'intérêt général spécifique qui lui est confiée par la loi. Dans ce dernier cas, la procédure peut être mise en œuvre soit lorsque l'administration n'a pu se procurer ces données par d'autres moyens (y compris l'achat sur le marché, le recours à des obligations existantes ou l'adoption d'une nouvelle mesure législative en temps utile), soit lorsque ce mécanisme permet de réduire substantiellement la charge administrative pour les détenteurs de données ou d'autres entreprises. Cette prérogative doit respecter les intérêts légitimes des détenteurs de données, y compris au regard du secret des affaires et des investissements consentis pour se procurer lesdites données. Dans la mesure du possible, les données requises ne doivent pas avoir un caractère personnel. Les différends seraient réglés par une autorité nationale compétente à désigner.

La conception d'un SIA public devrait pouvoir justifier la mise en œuvre d'une procédure de réquisition de données de ce type. Une référence à cette hypothèse dans le *Data Act*, par renvoi au règlement IA le cas échéant, serait opportune. A défaut d'être couverte par ce texte[244], il conviendrait de veiller à ce que ce dernier ne fasse pas obstacle à la création, en droit national, d'une telle prérogative, qui pourrait s'inspirer de la logique, connexe[245], de l'expropriation pour cause d'utilité publique : une administration pourrait requérir la transmission de données en vue d'une finalité

[244] Il n'est pas certain que la réalisation d'un SIA relève en toutes circonstances du cas dans lequel les données sont nécessaires pour que l'administration « *s'acquitte d'une tâche d'intérêt général spécifique qui lui est confiée par la loi* ». La plupart du temps, l'administration peut se dispenser du SIA, fût-ce au prix d'une dégradation de la qualité du service.

[245] S'agissant de données, dont les opérateurs requis ne sont pas privés, l'atteinte portée à leur droit de propriété, au titre du droit de décider de leur usage, se limite à l'obligation de les partager.

d'intérêt général, tel que le développement d'un système d'IA, si cette réquisition est nécessaire (ce qui n'est notamment pas le cas si ce SIA peut être développé avec un niveau de performance satisfaisant en utilisant d'autres jeux de données accessibles) et si le bilan entre les avantages attendus du SIA en projet et les inconvénients de toute nature que comporte un tel recueil de données coercitif (atteinte à des secrets protégés par la loi, atteinte portée aux intérêts des personnes concernées par les données à caractère personnel…) est positif. L'exercice d'une prérogative aussi intrusive devrait impérativement être entouré de garde-fous procéduraux adaptés, y compris, le cas échéant, une autorisation juridictionnelle ou délivrée par une autorité administrative indépendante, qui pourrait être l'autorité de régulation des SIA[246].

4.1.4. Se doter des ressources techniques et financières adaptées

1. Les besoins en infrastructures

A titre liminaire, il est bon de rappeler que **trop d'administrations ne sont pas encore au niveau requis de déploiement des outils numériques de base** (ordinateurs, logiciels performants, réseaux privés virtuels permettant le travail à distance…) et que l'on ne peut, sans susciter rejet, ironie ou amertume, leur parler efficacement de SIA. Partout, durant les entretiens qui ont permis de réaliser cette étude, est revenue cette antienne : pas de SIA si les besoins élémentaires (en matériel et en logiciels) ne sont pas couverts.

Naturellement, il ne serait pas raisonnable d'attendre la mise à niveau numérique de l'ensemble des administrations pour se doter des ressources techniques nécessaires au déploiement de SIA publics[247]. Les besoins spécifiques des projets d'IA doivent toutefois être intégrés dès le départ dans les politiques de mise à niveau informatique, comme le fait la stratégie nationale « *Cloud* au centre » – même si, en matière de *cloud*, le niveau technique requis par l'entraînement de certains SIA ne sont pas encore pleinement couverts.

Il a déjà été question de la disponibilité des infrastructures et, en particulier, de la **puissance de calcul**. On se bornera ici à rappeler que la France possède plusieurs supercalculateurs publics, dont ceux du Commissariat à l'énergie atomique (dont le Joliot-Curie, dédié à la recherche), de Météo France, ou du CNRS (le Jean Zay)[248]. Toutefois, ces supercalculateurs peinent à se maintenir dans la course mondiale et européenne à l'augmentation des capacités de calcul. L'horizon de la classe exascale

[246] V. 4.2.3. sur cette autorité chargée de la régulation des SIA.

[247] Pour une évaluation – un peu ancienne – de ces besoins, V. notamment le rapport Allistene, *Infrastructure de recherche pour l'intelligence artificielle*, janvier 2018.

[248] Le supercalculateur Jean Zay, installé en 2019 à l'Institut du développement et des ressources en informatique scientifique (IDRIS) du CNRS sur le plateau de Saclay à la suite du rapport Villani, est l'un des plus puissants d'Europe. Ses capacités ont été augmentées en 2020, lui permettant d'atteindre les 28 pétaflops (soit 28 millions de milliards d'opérations à la seconde) et, à cette époque, le 10e rang mondial. Fin 2021, une nouvelle augmentation de ses ressources a été annoncée, ses nouveaux GPU (*Graphics Processing Unit*, ou processeur graphique) devant permettre de doubler sa capacité de calcul dédiée à l'IA. Toutefois, le rang mondial de ce supercalculateur ne cesse de reculer.

(un milliard de milliards d'opérations par seconde), déjà réalité dans plusieurs pays, initialement envisagée à l'horizon 2022, n'est finalement pas prévue en France avant 2024 ou 2025. En parallèle, la France ne s'est pas associée au projet de supercalculateur européen pré-exaflopique, appelé Lumi, implanté en Finlande et l'Union européenne fait figure de retardataire au regard des investissements américains, chinois et japonais.

Ainsi qu'il a déjà été dit à propos de l'exigence d'autonomie stratégique, la compétition internationale appelle en la matière une mutualisation des investissements, tant entre Etats membres qu'entre les sphères publique et privée, assortie de la fixation claire des droits de tirage respectifs sur les infrastructures.

Le déploiement de ces infrastructures implique par ailleurs de penser, en parallèle, leur neutralité carbone. Ainsi, l'Établissement public d'aménagement Paris-Saclay a signé une convention avec le CNRS pour récupérer la chaleur produite par le supercalculateur Jean Zay, permettant de récupérer l'équivalent de la consommation en chaleur de 1 000 logements. Il convient également de noter que les derniers supercalculateurs peuvent être refroidis avec de l'eau tiède, ce qui améliore sensiblement leur performance énergétique. Grâce à ces mesures, la croissance de la consommation énergétique des centres données n'aurait augmenté que de 6% entre 2010 et 2018, alors que leur puissance de calcul a, sur la même période, augmenté de 55%[249].

2. Le développement et le partage des algorithmes

Indépendamment des infrastructures, il est essentiel que l'accès aux ressources techniques soit pensé dans le cadre d'une **mutualisation renforcée des briques logicielles entre administrations**.

Il est nécessaire de s'appuyer sur un socle de développements réussis afin de promouvoir une politique de dissémination plus méthodique, fondée sur des briques standardisées et interopérables (jeux d'apprentissage, prise en compte des variations langagières, modèles pré-entraînés) et des briques métiers plus spécifiques. L'objectif est d'éviter les doublons et l'achat fragmenté de données (alors que les besoins peuvent se recouper, par exemple ceux de l'IGN, de la DGFIP et de l'ASP pour les images satellite et aériennes) ou de solutions comparables aisément redéployables ou ré-entraînables (comme les agents conversationnels). Pour ce faire, les enjeux en termes de propriété intellectuelle[250], de commande publique et de concurrence doivent être anticipés. Une réflexion sur la création d'une structure juridique dédiée à un tel partage mériterait d'être lancée.

[249] E. Masanet, A. Shehabi, N. Lei, S. Smith, J. Koomey, *Recalibrating global data center energy-use estimates*, Science, 367. 984-986, 2020.

[250] Une réflexion spécifique est indispensable sur la question des droits de propriété intellectuelle grevant les systèmes d'IA développés pour l'administration par des prestataires privés et les données qui en sont issues. Le CCAG TIC 2021 mériterait d'être complété d'un volet dédié aux SIA, eu égard à leurs spécificités.

D'ores et déjà, le site code.gouv.fr référence l'ensemble des codes source utilisés ou développés par les administrations publiques. Le projet de l'association France Urbaine de créer une « Bibliothèque nationale de France » de l'IA territorialisée, qui serait à la fois un centre de ressources permettant la mise en commun de codes et de projets et un lieu d'échanges entre acteurs territoriaux, illustre l'intérêt d'une telle démarche de mutualisation. Celle-ci ne doit pas se déployer en reproduisant la logique en silos, bien connue mais souvent injustifiée, entre administrations de l'État et collectivités, mais de façon transversale et coordonnée.

A titre d'exemple, certains pays comme la Suède, le Danemark ou l'Espagne ont développé des programmes de déploiement transversal des technologies numériques du langage dans la sphère publique. **Or le champ du traitement du langage naturel mérite un investissement particulier** pour deux raisons : d'une part, l'administration (française) fonctionne très largement sur la base de l'écrit et de la production documentaire ; d'autre part, la performance des modèles algorithmiques peut varier considérablement d'une langue à l'autre et la France a tout intérêt à développer des modèles propres à la langue française, qui pourront d'ailleurs être partagés avec l'ensemble de la communauté francophone. Le projet PIAF piloté par la DINUM illustre l'intérêt d'une telle démarche.

Le projet PIAF

L'objectif de ce projet porté par Etalab est de développer des algorithmes de traitement du langage naturel en français, plus particulièrement pour les questions-réponses (qui ? quand ? où ? pourquoi ? que signifie ? comment faire ? quelle différence entre … ? oui/non ?…), permettant de concevoir et d'améliorer des outils de type robots conversationnels et des moteurs de recherche pour les administrations publiques. A l'heure actuelle, les données d'entraînement disponibles sont issues d'un travail d'annotation réalisé sur des textes rédigés en anglais, et traduits par un logiciel de traduction automatique. Or il apparaît que l'utilisation de données d'entraînement issues de l'annotation de textes rédigés en langue française permet d'améliorer significativement la performance des modèles.

Pour ce faire, le lab IA a mis en place une plateforme participative permettant à toute personne intéressée d'annoter des extraits d'articles de l'encyclopédie en ligne Wikipédia, en français. La plateforme, hébergée sur le site *GitHub*, est toutefois en langue anglaise (comme c'est l'usage pour les logiciels) et d'une utilisation réservée, de fait, à des personnes relativement averties et disposant des droits d'administrateur sur leur poste informatique, dans la mesure où elle suppose d'installer une application. La campagne a permis de collecter 9 500 couples de questions-réponses grâce à l'implication de 700 contributeurs. Sur cette base, le Lab IA a développé un modèle pré-entraîné de questions-réponses disponible sur la plateforme *HuggingFace*.

Ce SIA a été expérimenté sur trois cas d'usage : le site d'information *servicepublic.fr* avec la mise en place d'une interface permettant à toute personne de se renseigner sur les règles applicables à un projet qu'elle envisage, les dispositifs auxquels elle est éligible ou toute situation formulée en langage courant ; la foire aux questions de la CNIL ; et la foire aux questions sur le droit de la donnée d'Etalab.

3. Les ressources financières

Le coût de conception, de développement et de déploiement d'un SIA peut être extrêmement variable. Dans un contexte fortement contraint pour les budgets publics et, notamment, pour les budgets informatiques, y compris dans les collectivités territoriales, et alors que le maintien en conditions opérationnelles ou la remise à niveau des outils métiers « classiques » constitue un chantier prioritaire, le financement d'un SIA innovant peut s'avérer problématique pour une administration.

Les dispositifs de subventions existants permettent de répondre en partie à cette difficulté.

Les principales sources de financement des SIA publics

Outre les fonds propres des administrations et d'éventuels financements sectoriels, doivent être cités :

Le Fonds d'innovation et transformation numérique (FITN), créé en 2020 dans le cadre du plan France Relance, et doté de 292 millions d'euros, est copiloté par la DINUM et la DITP. Il peut bénéficier aux administrations de l'État ainsi qu'aux collectivités, et financer des projets d'automatisation de tâches ou d'amélioration du cycle de la donnée.

Le Fonds pour la transformation de l'action publique (FTAP), créé en 2017 et doté de 780 millions d'euros sur 5 ans, est piloté par la DITP. Il finance de nombreux projets IA, dont, à titre d'illustration, la Plateforme de données scientifiques pour la recherche archéologique, porté par l'INRAP, ou le projet Sécurité économique augmentée de la DGE et du SISSE, visant à mieux identifier et traiter les menaces pour les intérêts économiques du pays.

La DINUM et la DITP ont, en outre, lancé, dans le cadre du Programme d'investissements d'avenir (PIA), deux appels à projets dédiés au déploiement de SIA dans les administrations publiques (AMI-IA), dont le second a été doté de 4 millions d'euros. Ils ont financé 21 projets, dont deux abandonnés en raison de l'indisponibilité des données nécessaires.

S'agissant des financements destinés aux collectivités territoriales, outre le FITN, une enveloppe de 30 millions d'euros a été dégagée dans le cadre du Plan France Relance, pilotée par l'Agence nationale de la cohésion des territoires (ANCT). Il convient enfin de mentionner les « Challenges IA » de la Banque publique d'investissement (BPI) qui, au titre du PIA, financent, dans le cadre de partenariats publics-privés (financement d'une entreprise « sponsorisée » par une administration), des projets d'amélioration des services publics.

Il pourrait être envisagé d'**élargir les « Grands défis »** sélectionnés par le Conseil de l'innovation à l'utilisation des SIA dans l'action publique, en incluant la problématique de l'acceptabilité sociale et de l'évaluation de la confiance.

Il convient de mentionner deux limites susceptibles d'affecter les programmes de subvention, qui appellent une attention particulière des pouvoirs publics :

- d'une part, ceux-ci permettent souvent d'accompagner la réalisation des « preuves de concept », mais pas nécessairement le **passage à l'échelle**, alors que cette phase d'industrialisation est à la fois critique et coûteuse. Les fonds investis dans la phase de conception et de développement peuvent ainsi l'être en pure perte. Plus largement, le soutien financier à l'investissement informatique ne peut éluder la problématique des coûts de fonctionnement subséquents ;
- d'autre part, le financement de **dépenses de personnel** (titre 2) par ces subventions est souvent exclu ou, lorsqu'il est prévu (comme c'est le cas du programme Entrepreneurs d'intérêt général), il ne s'accompagne d'aucune souplesse sur le plafond d'emplois. Cette contrainte peut amener les porteurs de projet à recourir à des prestataires externes, souvent à grands frais, plutôt que d'embaucher les experts nécessaires (ingénieurs IA, analystes de la donnée…), fût-ce temporairement. Or, ainsi qu'il a été dit, l'intégration de ces experts au sein des équipes du responsable de projet présente de nombreuses vertus. Le choix entre les deux formules ne devrait pas être contraint par les règles d'octroi des aides.

4.1.5. Diffuser des repères méthodologiques communs

En raison de la complexité technique, organisationnelle et managériale d'un projet de SIA, et de la difficulté à appréhender par avance l'ensemble des potentialités que renferme les technologies d'intelligence artificielle, la conception et le déploiement d'un SIA appellent une méthodologie particulièrement rigoureuse et une agilité accrue dans sa mise en œuvre au cas par cas.

Cette étude n'a pas pour objet d'établir un guide méthodologique complet, directement utilisable par les administrations publiques. Toutefois, **quelques points de repères**, qui pourraient utilement alimenter un tel guide pratique, à réaliser, sur un mode participatif, par la communauté de l'IA publique sous l'égide de la DINUM, doivent être soulignés. Ils se rapportent, d'une part, au principe du recours à un SIA et, d'autre part, aux modalités de ce système.

1. Recourir à un système d'IA ou pas ?

La connaissance préalable des potentialités des systèmes d'IA et des cas d'usage déployés par d'autres administrations, que la présente étude contribue à diffuser, peut faire germer l'idée de recourir à l'IA pour faire évoluer voire transformer une politique publique, une organisation administrative, une action ou un processus interne ou la relation aux citoyens. Dans tous les cas, il importe de conserver à l'esprit que l'IA n'est qu'un outil parmi bien d'autres à la disposition de l'administration et avoir conscience des exigences qui conditionnent la réussite d'un projet de SIA et des ressources disponibles.

Il est crucial, au préalable, que l'administration fixe clairement l'objectif qu'elle s'assigne, ce qui suppose de **définir avec précision le problème qu'elle cherche à résoudre, les besoins qu'elle entend satisfaire, et le résultat auquel elle souhaite parvenir** (notamment la performance minimale acceptable du système), qui pourra

s'enrichir ou être adaptée au fil du projet. Les bénéfices attendus doivent être identifiés, pour l'ensemble des parties prenantes, en premier lieu l'usager du service et les agents qui en ont la charge.

À cet égard, l'administration doit notamment se poser la question de savoir s'il est préférable d'automatiser entièrement la tâche ou simplement de fournir une assistance à l'agent dans l'exécution de cette tâche. Plusieurs déterminants du choix peuvent être cités, sans que cette liste ne prétende à l'exhaustivité : la fréquence de réalisation de la tâche, son caractère fastidieux voire dévalorisant ou source de risques psycho-sociaux ou, à l'inverse, l'appétence des agents pour sa réalisation, son caractère formateur, la performance comparée entre l'agent et le système, les conséquences d'une erreur et la possibilité de garantir un niveau minimal d'absence de biais, la possibilité de garantir la supervision humaine en cas d'automatisation complète, le besoin du service de réorienter ses ressources vers d'autres priorités, etc.

La question de l'**acceptabilité sociale du recours à l'IA** doit irriguer ces premières réflexions. Une erreur « classique » à éviter, et qui a conduit à l'arrêt du développement de plusieurs SIA publics pourtant prometteurs, est l'appréhension du projet dans sa seule dimension technique, au détriment de facteurs de réussite d'égale importance : l'association des futurs utilisateurs et la transparence, à l'égard des agents comme des citoyens. Cela implique de s'interroger sur la réelle plus-value que la transformation envisagée doit apporter par rapport à l'existant, le degré de maturité numérique du service et la capacité pour les futurs utilisateurs du système de se l'approprier, en tenant compte des potentielles évolutions des règles et des métiers (en amont ou en parallèle du déploiement futur du SIA) et de l'engagement des cadres de haut niveau à porter et expliquer le projet auprès des agents.

Il convient également d'identifier quels **moyens humains, financiers et en termes de données** sont d'ores et déjà disponibles et pourront éventuellement être dégagés pour mener à bien le projet, et si ces ressources sont en adéquation avec son ambition. Le calibrage des ressources sera notamment fonction de la complexité du projet et du recours à la commande publique pour tout ou partie du système mais, en tout état de cause, l'administration doit pouvoir mettre en place une équipe projet, associant au minimum une personne « du métier » (représentant du service utilisateur, connaissant l'environnement, apte à exprimer les besoins et à (ré)orienter le projet), un data scientist, un développeur et un représentant de la direction du numérique compétente. Pour évaluer les besoins en termes de ressources, il peut être opportun de se rapprocher d'autres administrations (centrales ou territoriales) ayant déployé un SIA similaire.

Enfin, la **réflexion juridique et éthique** doit être menée dès cette première étape. Le SIA doit paraître conforme, dans sa finalité et dans son mode de fonctionnement général, aux principes exposés ci-dessus (V. troisième partie de cette étude). L'administration doit également se poser la question de savoir si une règle de droit s'oppose à la conception du SIA, dans son principe-même, et si ce dernier implique de disposer d'une base juridique *ad hoc*, et, le cas échéant, de quel niveau. Le cas échéant, les démarches CNIL qui devront être réalisées doivent être anticipées dès cette première phase.

2. Quelle approche d'IA choisir et selon quelles modalités ?

Une fois prise la décision de développer un SIA, une avalanche d'arbitrages doivent être faits.

Le premier choix à faire est celui du type d'IA : **système déterministe ou apprentissage machine** ? La réponse dépend de nombreux paramètres – en partant du présupposé que le système-expert requis ne serait pas lui-même d'une complexité excessive et en rappelant qu'un même processus administratif peut parfaitement recourir à l'un et l'autre, à des stades différents. On se bornera ici à énumérer quelques-uns de ces paramètres :

- **La nature de la tâche** : un système-expert est plus souvent adapté lorsque la tâche est bien normée (format des données d'entrée, nature des opérations de traitement à effectuer, règles à suivre, type de sorties désirées…), alors que l'apprentissage machine est en général à privilégier lorsqu'il est difficile de définir *a priori* des règles et instructions permettant d'obtenir un résultat déterminé à partir d'une situation de départ, notamment lorsqu'il s'agit de « comprendre » du texte ou de la voix, de catégoriser des images ou autres contenus, de prédire un résultat ou un évènement susceptible d'être influencé par de nombreuses variables (dont certaines peuvent ne pas être connues au préalable), de personnaliser une prestation en fonction de nombreux paramètres ;
- **Les attentes et exigences de l'écosystème concerné** : un système-expert devra être privilégié si une erreur est susceptible d'avoir des conséquences disproportionnées au regard des bénéfices attendus de l'automatisation, s'il est nécessaire de garantir une explicabilité parfaite du résultat issu de l'outil ou encore si la défiance des parties prenantes dans l'IA est importante ;
- **Les ressources disponibles** : un système-expert mobilisera davantage les « gens de métier » (qui doivent formaliser l'ensemble des opérations devant conduire au résultat) et moins les experts de la donnée (*data scientist* en particulier) ;
- **La disponibilité des données exploitables** : sans données en quantité et en qualité suffisantes, la construction d'un nouveau système basé sur l'apprentissage sera le plus souvent compromise et seule l'utilisation d'un système pré-entraîné sera envisageable. L'administration doit donc faire le point sur les jeux de données dont elle dispose déjà et ceux qui pourraient être constitués, et anticiper les ressources nécessaires pour rendre ces jeux suffisamment pertinents, complets, fiables et non-biaisés.

Dès lors que l'administration fait le choix du recours à un SIA basé sur l'apprentissage machine, deux axes de travail doivent l'occuper, d'égale importance : d'abord, l'identification, la collecte, le nettoyage, la sécurisation du stockage et l'annotation des données ; ensuite, la construction et la validation du modèle, qui comprend notamment le choix de l'algorithme, son codage, son entraînement puis sa validation et, le cas échéant, sa certification.

Quel que soit le type d'IA choisie, il convient en outre de prêter attention aux points suivants : d'abord, la licéité et le caractère « éthique dès la conception » du SIA, par l'actualisation des éléments établis dans la première phase de réflexion sur le

recours à l'IA elle-même et l'engagement des démarches réglementaires et/ou administratives requises ; ensuite, les besoins spécifiques de la phase d'industrialisation du système (par exemple, un nouvel entraînement sur la base de nouveaux jeux de données dans le cas d'un SIA basé sur l'apprentissage machine, ou l'éventuelle actualisation des règles en cas de recours à un système-expert) ; enfin, la surveillance et la maintenance du SIA après sa mise en service, qui suppose des contrôles techniques réguliers et l'organisation de retours d'expérience, avec tant les agents que les usagers.

Enfin, ainsi qu'il a été développé au point 4.1.1, l'administration doit décider du degré d'externalisation de la conception du système, en tenant compte de l'existence de produits disponibles sur étagère, en opérant une « **balance des coûts** » incluant ceux de long terme, et, le cas échéant, en veillant au respect des règles de la commande publique et en soignant la rédaction des clauses du contrat, en particulier en ce qui concerne la propriété intellectuelle du système, la maîtrise des données et la responsabilité en cas de dommage causé aux tiers.

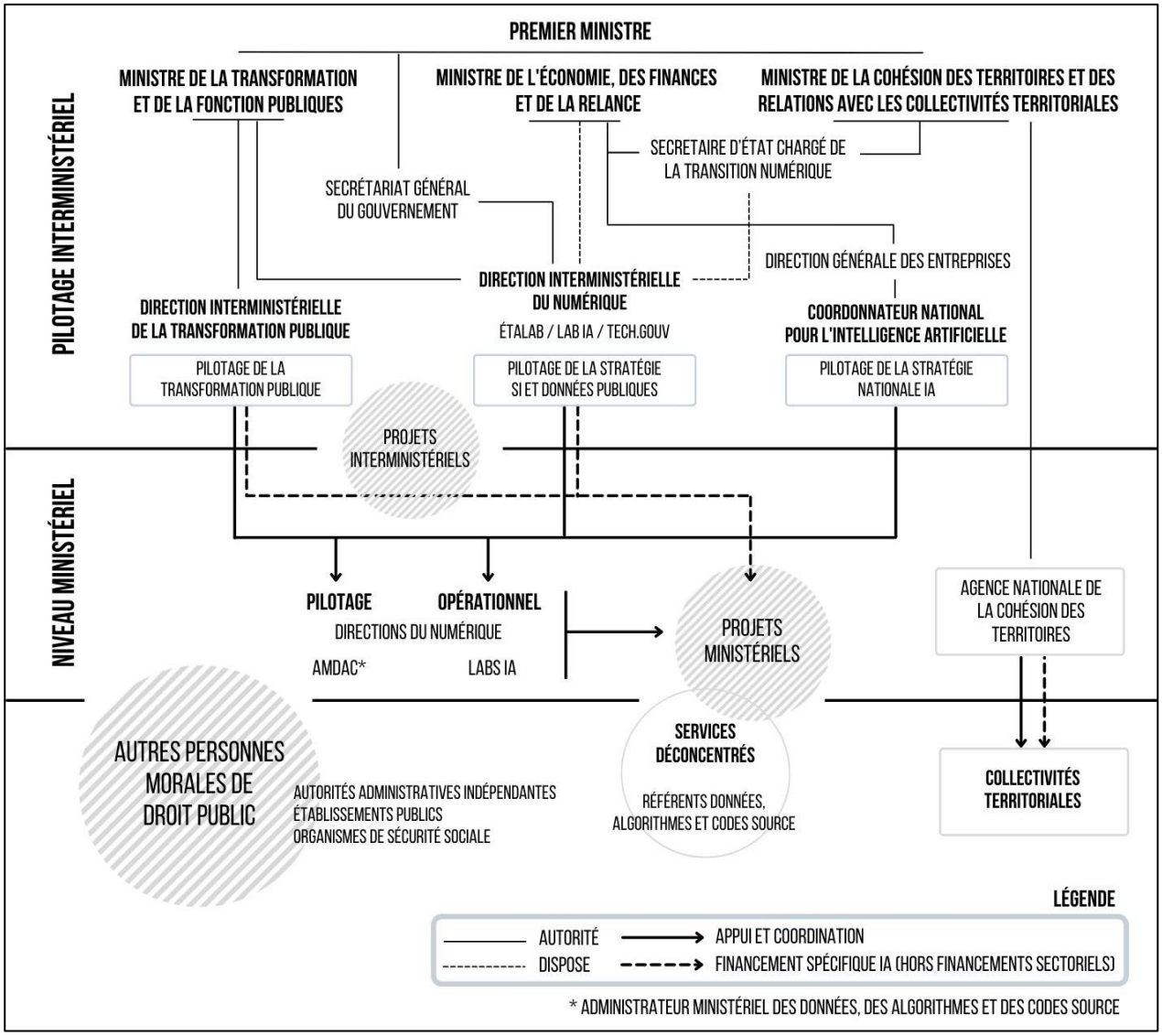
4.2. Construire une gouvernance tournée vers l'innovation et la confiance

La gouvernance de l'IA publique a vocation à assurer plusieurs fonctions : la définition et le pilotage de la stratégie ; la mise en œuvre opérationnelle par les administrations, incluant la conduite des projets et la conformité interne ; la régulation externe des systèmes, dans un cadre en voie d'harmonisation à l'échelle européenne ; et, enfin, la réflexion sur l'usage éthique de l'IA et l'implication des citoyens.

4.2.1. Assurer le pilotage de la stratégie de l'IA publique

Il est nécessaire de disposer, dans l'administration de l'État, d'une structure de pilotage chargé de concevoir et d'animer la mise en œuvre de la stratégie des SIA publics, de mettre à la disposition des administrations les ressources de toute nature dont elles ont besoin pour concevoir et déployer des systèmes d'IA, d'incarner et de porter les actions de communication et de sensibilisation du grand public, de représenter la France dans les instances européennes et internationales et d'y déployer une stratégie d'influence et, enfin, d'assurer une fonction de veille juridique, technique et d'« intelligence administrative » (identification des bonnes pratiques mises en œuvre à l'étranger).

Ces fonctions sont aujourd'hui assurées, à des degrés divers, par une multiplicité d'acteurs :



Trois entités sont ainsi principalement en charge du pilotage stratégique des SIA publics (V. annexe 8) :

- la **direction interministérielle du numérique** (DINUM), créée en 2019 et rattachée au secrétaire général du gouvernement, pilote la transformation numérique des politiques publiques. En son sein, le département Etalab, qui coordonne la politique de l'*open data* de l'État, assure, *via* son « Lab IA », la conduite et l'accompagnement des projets de SIA publics. Sa contribution au déploiement d'outils d'IA dans les administrations de l'État, et l'appui qu'il peut

fournir dans la construction d'outils mutualisables (par exemple le projet PIAF, présenté *supra*) doit être soulignée ;

- la **direction interministérielle de la transformation publique** (DITP), placée sous l'autorité du ministre chargé de la réforme de l'État, est principalement chargée de la mise à disposition de ressources financières (programmes de subventions) et humaines (intervention d'experts au sein des administrations ayant des projets de transformation) ;
- enfin, le **coordonnateur national pour l'intelligence artificielle**, placé auprès du directeur général des entreprises mais directement missionné par le Premier ministre, est en charge de piloter la stratégie nationale pour l'IA déjà évoquée dans la deuxième partie de l'étude, laquelle inclut un volet relatif à la transformation publique par les SIA. Il joue un double rôle de stimulation des initiatives et de fluidification du partage d'informations.

Il faut reconnaître à cette organisation le mérite d'éviter trois écueils.

Le premier serait celui d'une **sanctuarisation complète de la politique d'intelligence artificielle publique dans une entité dédiée**. Comme on l'a dit, les systèmes d'IA s'inscrivent dans un **continuum de solutions techniques**. En dépit du potentiel unique de transformation qu'ils renferment, ils ne doivent pas être sacralisés dans un ministère ou une direction d'administration centrale qui leur serait propre, mais, au contraire, articulés avec l'ensemble de la politique numérique et de transformation publique de l'État, comme elle l'est tant à la DINUM qu'à la DITP.

Le deuxième écueil serait celui d'une **rupture franche entre la stratégie de l'IA publique et celle de soutien à l'IA privée**. Les acteurs et les problématiques de l'IA dans l'action publique entretiennent des **liens étroits avec ceux de la sphère privée** : de nombreux cas d'usage sont analogues ; les administrations ont besoin de s'appuyer sur des prestataires privés spécialisés, y compris d'ailleurs dans des domaines propres à la sphère publique, en particulier dans le domaine régalien (armement, maintien de l'ordre, renseignement…) ; les données des uns et des autres sont potentiellement utiles à tous. Ces considérations ne justifient pas en soi de concevoir une stratégie unique confiée à une seule entité, alors que, du point de vue des politiques publiques, les objectifs diffèrent (transformation de l'action publique, d'un côté ; politique économique et sociale, de l'autre). Mais il est essentiel que cette double dimension soit prise en compte dans la définition de la gouvernance de l'IA publique. À cet égard, si le positionnement du **coordonnateur national** à la direction générale des entreprises du ministère de l'économie, des finances et de la relance l'oriente davantage vers la problématique de l'innovation privée que vers celle de la transformation publique, il a, en pratique et en raison notamment de son parcours professionnel, activement investi ce second terrain.

Le troisième et dernier écueil est celui de l'**hypertrophie**, au risque de mobiliser des ressources rares qui seraient mieux utilisées dans les administrations en charge du déploiement des SIA, et d'accroître le risque de déconnexion avec le terrain. Dans ce domaine où les priorités opérationnelles ne peuvent être définies de façon centralisée et où la créativité et l'agilité sont de rigueur, la subsidiarité doit jouer à plein. Il est d'ailleurs souvent préférable de confier la responsabilité d'un projet,

d'une fonction, ou l'animation d'un réseau à une entité (un ministère, un établissement public…) cheffe de file, qui en est elle-même responsable dans son propre champ de compétence, plutôt qu'à une structure centralisée. En l'état, cependant, **les ressources dont disposent les trois structures de pilotage mentionnées sont très – et sans doute trop – modestes** : le coordonnateur national est assisté d'un adjoint ; Etalab, qui compte 28 collaborateurs, en consacrent 4 à l'IA.

Telle est, précisément, l'une des faiblesses de la gouvernance actuelle. Il est manifeste qu'aucune de ces entités, notamment pas la DINUM, ne dispose de la **taille critique** lui permettant d'exercer un véritable effet de levier, à grande échelle, sur la politique de développement des SIA publics, et d'être identifié par les administrations comme un centre de services partagé. En outre, si ces structures sont amenées à échanger entre elles, soit dans le cadre du comité de pilotage mensuel animé par le coordonnateur national, soit sur une base informelle, **l'articulation de leur action n'est pas formalisée**. Enfin, le positionnement, la modestie et la multiplicité des structures font naître un **problème de visibilité et d'identification**. Nombre des acteurs rencontrés par la mission ne savaient pas spontanément vers qui se tourner dans l'éventualité du lancement d'un SIA.

La dispersion institutionnelle est d'ailleurs aussi à l'origine d'un manque de visibilité de **la stratégie française en matière d'IA publique au niveau européen**. Ainsi, l'étude de référence de la Commission européenne relative à l'IA dans le secteur public, qui compare les stratégies de déploiement de l'IA dans le secteur public de l'ensemble des Etats membres de l'Union européenne, conclut que la stratégie française ne comporte que 4 des 17 actions formant une stratégie complète aux yeux des auteurs (campagnes d'information, formation générale des agents, projets-pilote et élaboration d'un cadre éthique), contre 14 pour l'Estonie et 9 pour les Pays-Bas ou le Portugal. Or à l'évidence, un certain nombre d'actions menées par notre pays n'ont pas été prises en compte, ce qui semble traduire un manque de valorisation de la stratégie elle-même ou des activités conduites par les institutions chargées de sa mise en œuvre.

Sans doute s'épuiserait-on à rechercher une organisation idéale de la fonction de pilotage de la stratégie de l'IA publique qui n'existe pas, en raison de la multiplicité des considérations, parfois contradictoires, qui doivent entrer en ligne de compte. Toutefois, afin d'accompagner le changement de rythme et d'échelle que l'étude appelle de ses vœux, deux évolutions pourraient être envisagées.

Sur le plan opérationnel, le **renforcement du département Etalab** de la DINUM apparaît nécessaire si l'on veut en faire un véritable centre de ressources partagées pour l'ensemble des collectivités publiques, y compris, le cas échéant, par le biais d'un portage de ressources humaines. Une telle fonction, si elle devait être consacrée dans les textes et servie par des moyens conséquents susceptibles de bénéficier aussi bien aux administrations centrales qu'aux services déconcentrés de l'État et aux collectivités territoriales, pourrait justifier la transformation de ce service d'administration centrale en **service à compétence nationale**. Les ingénieurs IA pourraient d'ailleurs partager leur temps entre le déploiement dans les administrations pour le développement des cas d'usage et l'activité au sein d'Etalab,

sur le modèle du *Health Data Hub* (où la proportion respective de ces activités est de l'ordre de 80 % / 20 %).

Sur le plan institutionnel, le **pilotage politique** de la stratégie d'IA, publique comme privée, pourrait être confié à un membre du gouvernement en charge du numérique (comme l'est actuellement le secrétaire d'État au numérique) rattaché à la fois au ministre en charge de de l'économie et au ministre chargé de la réforme de l'État ou de la transformation publique, pour marquer clairement la dualité des objectifs que le numérique doit poursuivre. Sous l'autorité directe de ce membre du gouvernement (ou, en son absence, de ces deux derniers ministres), le **coordonnateur national pour l'IA devrait se voir plus nettement reconnaître un rôle-pivot dans la stratégie de l'IA publique**. Nommé en conseil des ministres dans le cadre de la procédure prévue au cinquième alinéa de l'article 13 de la Constitution (en raison de l'importance de la fonction tant pour la garantie des droits et libertés que pour la vie économique et sociale de la Nation), et toujours missionné par le Premier ministre, le coordonnateur national devrait disposer en propre d'une équipe resserrée (d'une dizaine de personnes au maximum) et s'appuyer, s'agissant des SIA publics, sur les agents qui s'y consacrent à la DINUM, à la DITP et à l'ANCT, dont il coordonnerait l'action et les différents dispositifs dans son domaine de compétence. Outre l'organisation du comité de pilotage interministériel qu'il assure déjà, le coordonnateur national devrait animer une **instance de partage avec les collectivités territoriales**, associant étroitement l'ANCT, à laquelle participeraient notamment les associations de collectivités et la Caisse des dépôts et consignations.

L'efficacité d'une telle organisation reposerait sur deux conditions : d'une part, une capacité des services, par-delà la diversité de leurs rattachements hiérarchiques, à fonctionner de façon décloisonnée, sans territorialité ; d'autre part, la capacité de l'équipe du coordonnateur national à analyser et qualifier les besoins qui pourraient lui être adressées par les collectivités publiques et à les orienter vers d'autres interlocuteurs dès l'instant que ces besoins peuvent être satisfaits sans recourir à un système d'IA innovant.

Outre le pilotage de la stratégie et la coordination de l'appui opérationnel aux administrations, le coordonnateur national serait explicitement chargé, en lien avec l'Ambassadeur pour le numérique, qui s'y emploie déjà, de déployer une **stratégie d'influence** à l'échelle européenne et internationale, non seulement dans les instances de normalisation, comme indiqué dans la première partie de l'étude, mais aussi dans les organisations internationales. Il serait en effet fâcheux de laisser à d'autres la promotion de règles universelles en matière d'IA, qu'il s'agisse de puissances cherchant à légitimer des pratiques contraires aux valeurs nationales et européennes ou d'acteurs économiques revendiquant une liberté toujours plus ample sans nécessairement accroître celle des usagers qui contribuent à leur profitabilité. La réflexion internationale a d'ores et déjà accouché d'un certain nombre d'initiatives de nature principielle – Recommandation sur l'éthique de l'intelligence artificielle de l'UNESCO[251], Principes de l'OCDE sur l'intelligence artificielle, Recommandation du Conseil de l'OCDE sur l'intelligence artificielle… En

[251] Adoptée le 24 novembre 2021 par la 41e conférence.

outre, 25 pays dont la France (représentée par le coordonnateur national), participent au Partenariat mondial sur l'intelligence artificielle (PMIA) qui réunit des experts et représentants des gouvernements et de la société civile afin de « *guider le développement et l'utilisation responsables de l'IA, dans le respect des droits de la personne, des libertés fondamentales et de nos valeurs démocratiques communes, conformément à la Recommandation de l'OCDE sur l'IA* ». Se posera à terme, lorsque les régulations régionales (en particulier européennes et américaines) auront acquis une certaine maturité, la question de leur convergence puis celle de la négociation, au sein du système onusien, d'un accord international et la création d'une organisation mondiale consacrée au numérique, de la même façon que certaines organisations internationales ont pu, historiquement, mettre fin définitivement à (OACI[252]) ou réduire significativement (OMPI[253]) l'anarchie dans des secteurs où elle était source d'insécurité et de pertes économiques.

4.2.2. Favoriser la mise en œuvre opérationnelle des SIA

S'agissant de la conduite des projets au niveau ministériel, **aucune organisation-type ne se dégage des observations effectuées ;** et l'on notera d'emblée que la problématique ne se pose pas très différemment au sein des collectivités territoriales.

Assez logiquement, les directions numériques (DNUM) sont les plus mobilisées. Outre leurs missions classiques de pilotage de l'équipement informatique de leur ministère et de suivi des projets relatifs aux applicatifs (dont relèvent les SIA), nombre de DNUM assument également la fonction d'administrateur ministériel des données, des algorithmes et des codes sources (AMDAC) et ont la charge, à ce titre, de déployer la politique publique de l'*open data* et de l'exploitation des données publiques. Si, à la suite du rapport Bothorel, les AMDAC – qui, à l'origine, étaient seulement des AMD dont la mission se limitait à la politique de la donnée – ont vu leurs fonctions élargies à l'ouverture des algorithmes et des codes sources, force est de constater que la nouvelle dynamique recherchée peine à être trouvée et le niveau d'engagement des AMDAC, s'agissant de l'IA elle-même, est très disparate. Plusieurs facteurs peuvent l'expliquer : peu d'AMDAC assument uniquement cette fonction, qu'ils cumulent en général, comme il a été dit, avec celle de DNUM voire, pour une partie d'entre eux, avec des fonctions de directeur d'administration centrale « métier » ; si les AMDAC ont eu à formaliser leur feuille de route (au total, ces feuilles de route ont identifié 75 actions relatives à l'exploitation des données – *dataviz*, IA, hubs…), ils ne disposent d'aucun moyen supplémentaire pour leur mise en œuvre.

Il convient enfin de souligner que certains ministères ont mis en place un « **Lab IA** » (ministère de l'Intérieur, ministère de la transition écologique avec Ecolab au sein du Commissariat général du développement durable), une « Fabrique numérique » (ministères sociaux) ou, de façon plus ambitieuse encore, un service à compétence nationale servant l'ensemble du ministère (cas de l'Agence de l'innovation de

[252] Organisation de l'aviation civile internationale.

[253] Organisation mondiale de la propriété intellectuelle.

défense au ministère des armées, qui a un objet beaucoup plus large que les données). Ils permettent de rassembler les compétences « métier » et techniques autour d'une dynamique projet, et semblent être un modèle intéressant, qui permet d'éviter le risque du saupoudrage entre une multitude de structures n'ayant pas la taille critique pour exercer l'effet de levier nécessaire à un changement d'échelle. Une certaine centralisation du budget dédié à l'innovation numérique apparaît pertinente pour mettre en compétition les projets et soutenir plus massivement et plus durablement les plus prometteurs.

La fonction de pilotage ministériel du déploiement de l'IA gagnerait donc à faire l'objet d'une structuration plus affirmée, sans que soit imposé un modèle standardisé qui risquerait de déstabiliser des organisations rodées ou des dynamiques naissantes.

Quelle que soit l'organisation choisie *in fine*, les administrations gagneraient à **structurer certaines fonctions**.

La première est la **fonction de conformité**. Les directions et services des affaires juridiques, lorsqu'ils existent, ont vocation à jouer leur rôle de gardien du respect des règles posées par le règlement IA et complétées par le droit national. Cette fonction inclut le respect de l'obligation de **certification** des SIA identifiés à haut risque par le règlement IA : si l'administration se borne à acquérir un tel système, elle devra seulement s'assurer, en tant qu'utilisatrice, de ce que ce dernier a fait l'objet de la certification par un « organisme notifié » désigné par l'autorité nationale de contrôle (cf. 4.2.3. ci-dessous) ; si elle conçoit elle-même un SIA à haut risque, elle devra veiller à l'évaluer dans le cadre de la procédure de contrôle interne prévue par le projet de règlement (article 43 et annexe VI)[254]. L'administration devra donc se doter des moyens d'auto-certifier ses systèmes. Il serait toutefois souhaitable que le règlement ouvre expressément la possibilité pour l'administration de faire certifier un système qu'elle met en service par un organisme notifié, choix qui sera en général plus onéreux, mais aussi plus sécurisant, y compris dans la perspective d'un recours contentieux, et gage de confiance pour le public. D'ores et déjà, certaines administrations de l'État, notamment dans la sphère régalienne et des collectivités territoriales, font appel à l'expertise du LNE en matière d'évaluation[255].

[254] Par exception, le projet de règlement prévoit que les SIA d'identification biométrique sans consentement font obligatoirement l'objet d'une procédure externe d'évaluation de la conformité (décrite à l'annexe VII) si le fournisseur n'a pas appliqué les normes harmonisées. L'administration peut aussi décider volontairement d'y recourir.

[255] La France dispose à cet égard d'un avantage dans la mesure où le LNE est le seul organisme européen d'évaluation de la conformité à s'être d'ores et déjà doté d'une équipe dédiée à l'évaluation des SIA, à l'origine du premier référentiel de certification au monde des processus et du cycle de vie du système. Le NIST américain et le National Institute japonais ont également fait le choix d'un service dédié.

Le délégué à la protection des données, d'ores et déjà amené à intervenir pour tout SIA recourant à des données à caractère personnel, pourrait voir sa mission étendue à l'ensemble de ces systèmes. En l'état, le règlement IA ne prévoit certes pas d'obligation de désigner un délégué dédié. Mais la relation établie avec la CNIL, qui a vocation à jouer un rôle important dans la régulation des SIA, *a fortiori* si elle est désignée comme autorité nationale de contrôle (V. 4.2.3. ci-après), pourrait être mise à profit pour faciliter le dialogue de conformité entre les services et l'autorité.

L'intégration du recours à l'IA dans la **fonction de tutelle** des établissements et organismes constitue un deuxième enjeu. Elle n'est pas encore systématique alors que les SIA peuvent constituer un levier de transformation de ces opérateurs. A titre d'illustration, s'agissant des organismes de sécurité sociale, le développement de SIA n'est pas explicitement évoqué alors que la CNAV, la CNAF ou l'URSSAF Caisse nationale ont déjà une stratégie volontariste en la matière.

Enfin, les administrations centrales de l'État gagneraient à structurer une fonction d'**appui aux services déconcentrés** de l'État dans lesquels le déploiement de SIA est rare et dépendant d'initiatives individuelles. Un premier pas a été franchi avec la désignation de référents « données, algorithmes et codes sources », placés auprès des préfets de région, mais des cadres informels de dialogue pourraient être institués, dans une double logique ascendante de remontée des « irritants » susceptibles de déboucher sur un projet de SIA et descendante de pédagogie et de test des preuves de concept avant passage à l'échelle.

4.2.3. Confier le pilotage de la régulation des SIA à une CNIL transformée

La proposition de règlement IA dessine une architecture relativement complexe de la régulation des SIA.

Le principe de base est simple : chaque État membre désigne une autorité nationale de contrôle, chargée de la mise en œuvre et de l'application du règlement, de la coordination des activités de l'État membre, du rôle de point de contact unique pour la Commission et de la représentation de l'État membre au sein du comité européen de l'IA. Cette mission générique semble notamment inclure le contrôle des systèmes et la sanction. Cette entité joue normalement le rôle d'autorité de surveillance du marché des SIA, et celui d'« autorité notifiante », c'est-à-dire d'organisme responsable de l'accréditation et de l'évaluation des « organismes notifiés » qui seront chargés de certifier la conformité des SIA à haut risque. Le règlement autorise cependant les Etats membres à dissocier ces fonctions, s'ils justifient du bien-fondé de ce choix.

La proposition de règlement implique d'emblée une telle dissociation pour les SIA relevant des 70 législations d'harmonisation sectorielle, qui prévoient déjà l'intervention d'une autorité de surveillance du marché. De façon très opportune, l'article 43 de la proposition intègre, dans une logique de guichet unique, l'évaluation de la conformité des SIA relevant de ces secteurs dans celle des produits auxquels ils s'intègrent ou qu'ils constituent, et qui est déjà prévue par cette législation. A titre d'exemple, un dispositif médical embarquant un SIA serait certifié par un organisme

notifié relevant du règlement 2017/745 du 5 avril 2017 relatif aux dispositifs médicaux (en France, le GMED, filiale du LNE), sous la surveillance de la Haute autorité de santé.

Pour le reste, les Etats membres conserveraient une large marge d'appréciation dans la désignation de l'autorité de contrôle compétente. Celle-ci pourrait d'ailleurs être un service d'administration centrale puisque, en l'état, le texte exige seulement qu'elle soit organisée de manière à garantir l'objectivité et l'impartialité de ses activités.

Il serait parfaitement envisageable de créer une autorité ou un service dédié. Ce choix ne paraît toutefois pas optimal. Il en résulterait non seulement des surcoûts liés, notamment, aux fonctions support qui sont nécessaires à son fonctionnement, mais aussi une complexification institutionnelle supplémentaire, alors que l'autorité de contrôle doit s'appuyer sur le « réseau » des autorités de surveillance de marché existantes. Un très grand nombre de SIA impliquant aussi des traitements de données à caractère personnel, les responsables se trouveraient contraints de conduire un dialogue de régulation avec cette autorité dédiée et la CNIL, avec le risque de divergences doctrinales, accru par l'incertaine articulation entre le RGPD et le projet de règlement IA. Si les deux corps de règles sont distincts et poursuivent des objectifs qui ne se superposent pas complètement, de sorte que les manquements à l'un et l'autre devraient, juridiquement, pouvoir être sanctionnés cumulativement, un dédoublement répressif doit, dans la mesure du possible, être évité.

Au total, **la très forte adhérence entre la régulation des SIA et celle des données, en particulier des données à caractère personnel, et l'intérêt d'une internalisation institutionnelle de l'articulation des deux régimes juridiques, plaident assez naturellement pour que la CNIL se voie confier les deux fonctions**. S'agissant des SIA relevant des institutions de l'Union, la proposition de règlement confie d'ailleurs ce rôle au Contrôleur européen de la protection des données, et on trouvera aussi intérêt, sur un plan opérationnel, à ce que les membres du comité européen de la protection des données (CEPD) et ceux du comité européen de l'IA envisagé par la proposition de règlement soient, le plus possible, les mêmes, même s'ils pourront être représentés, le cas échéant, par des personnes différentes.

En raison de la compétence des autorités sectorielles de surveillance des marchés, et, plus largement, parce que la régulation des SIA ne saurait se concevoir indépendamment de celle des activités qu'ils servent, le rôle de la CNIL comme autorité nationale de contrôle du règlement IA devrait être conçu comme celui d'une **autorité de coordination, de supervision, de tête de réseau**. En effet, une large part des tâches à accomplir suppose autant une bonne connaissance des SIA que des secteurs dans lesquels ils doivent être déployés. On peut à ce titre imaginer que, notamment, les grands régulateurs ou des structures ordinales puissent participer à la régulation des SIA relevant de leur champ de compétence, sous la supervision de l'autorité de contrôle. Cela pourrait notamment concerner les régulateurs des industries de réseau (ARCEP, CRE, ART…), ceux à vocation horizontale (ADLC, AMF, ACPR) ou ceux propres à certains secteurs (Autorité de régulation des jeux en ligne, ARCOM…). Pourrait ainsi être envisagé un mécanisme de **délégation** permettant à des autorités sectorielles d'accréditer ou de participer à l'accréditation des

organismes notifiés, voire de proposer des sanctions, le cas échéant en tant qu'autorité de poursuite. Cette mutualisation des forces présenterait aussi un intérêt budgétaire, en évitant la création de doublons au niveau de la CNIL.

Un tel choix suppose toutefois une transformation profonde de la CNIL, à deux égards.

D'une part, comme l'exige d'ailleurs expressément le projet de règlement IA, il y a lieu de changer d'échelle en ce qui concerne les moyens, en particulier humains, accordés à cette autorité. La crédibilité des pouvoirs publics dans le développement des SIA au regard de leur acceptabilité par l'opinion suppose en effet un **investissement immédiat massif et déterminé pour augmenter les capacités du régulateur**. Or, alors que la CNIL, forte de son ancienneté et de son expérience, a réussi par son dynamisme à devenir l'une des toutes premières instances de régulation dans le concert européen, elle demeure l'un des plus modestes, qu'on raisonne en valeur absolue ou en rapportant ses moyens à la taille du pays. Et alors que l'entrée en vigueur du RGPD a par exemple conduit la Grande-Bretagne, qui s'y savait pourtant soustraite, à doubler les effectifs de son autorité de contrôle nationale (ICO), l'effort budgétaire réel en faveur de la CNIL est resté très en-deçà des besoins (la CNIL n'a bénéficié que de 25 postes supplémentaires en 2022).

D'autre part, le positionnement de la CNIL comme son image doivent évoluer pour qu'elle ne soit plus seulement (perçue comme) une autorité de protection des individus contre une menace, mais un véritable **régulateur**, soucieux de conjuguer le développement de notre pays par le soutien à l'innovation et le respect des droits fondamentaux des personnes concernées, et un facilitateur de **l'innovation technologique au service de la société**, soucieux de créer un environnement favorable à l'innovation responsable. Il ne s'agit pas ici de faire passer au second plan sa mission de protection des données personnelles, qui fait de la CNIL une référence au niveau européen, mais bien de s'appuyer sur cette expérience pour cultiver la nécessaire complémentarité entre protection et innovation. Les auditions ont montré que nombre d'administrations appelaient de leurs vœux cette évolution. La logique « produits » du règlement européen, s'il est adopté, y conduit également.

Parmi les modifications à apporter à cette fin, trois d'entre elles apparaissent prioritaires, indépendamment d'un **changement d'appellation** qui marquerait symboliquement cette transformation.

En premier lieu, la **structure du collège** devrait être profondément modifiée, non seulement pour y accueillir des spécialistes des systèmes d'IA, mais aussi, plus largement, pour que les acteurs de l'innovation y soient davantage représentés. Ce faisant, l'inévitable tension qui existe entre cet enjeu d'innovation, de performance administrative et de compétitivité économique, d'une part, et la protection des droits et libertés des individus, d'autre part, serait internalisée au sein d'une même structure dont les arbitrages seraient rendus au terme d'un débat équilibré entre les sensibilités les plus diverses.

En deuxième lieu, la capacité d'adaptation de la CNIL, notamment dans la construction de sa doctrine sur des sujets particulièrement évolutifs, suppose une certaine **diversité des profils**, entre juniors et seniors, et entre agents titulaires et contractuels. Or la CNIL, à l'instar d'autres administrations, rencontre aujourd'hui, en raison des conditions salariales proposées, des difficultés de recrutement de profils seniors, ce qui pose la question des moyens budgétaires qui lui sont alloués pour assurer de tels recrutements.

En troisième et dernier lieu, il est impératif que les moyens supplémentaires qui lui seraient alloués ne soient pas exclusivement, ni même majoritairement, affectés à la fonction d'autorisation, de contrôle et de sanction, mais qu'un effort particulier soit consenti pour renforcer significativement la fonction d'**accompagnement et d'appui à la conformité**. De nombreuses administrations ont exprimé le besoin de disposer, le plus en amont possible, d'un interlocuteur éclairé pour les conseiller, orienter leurs réflexions et faciliter d'éventuelles démarches et formalités préalables, et ont pointé le risque de retard voire d'abandon de projets lié à une intervention trop tardive et, parfois, trop rigide. Les administrations sont également désireuses de pouvoir disposer d'outils méthodologiques, de recommandations et de lignes directrices leur donnant de la visibilité sur l'interprétation des règles applicables. Les contentieux éventuels sur ces actes de droit souple, portés directement devant le Conseil d'Etat, permettent d'en obtenir rapidement – en un an environ, sous réserve d'éventuelles questions préjudicielles à la Cour de justice de l'Union européenne – la confirmation ou l'infirmation.

L'ampleur et la difficulté d'une telle transformation ne doivent pas être sous-estimées. Elles requerront un portage politique au plus haut niveau et une vigilance particulière dans les nominations.

4.2.4. Structurer la réflexion sur l'éthique et l'implication civique

1. La « fonction éthique »

La prise de conscience de l'importance des enjeux d'éthique est contemporaine des balbutiements des systèmes d'IA, et les craintes en la matière sont, pour ce qui concerne les services publics, beaucoup plus développées que les systèmes eux-mêmes. La réflexion prospective et l'anticipation des risques n'en restent pas moins cruciales pour garantir la confiance, au-delà du respect de la règle de droit. **Une structuration de la « fonction éthique »**, comprise comme la réflexion sur les incidences des SIA sur la société, les institutions et les individus et sur les valeurs qui doivent guider leur développement, **est donc nécessaire.** Cette réflexion est indispensable pour asseoir une prise de décision publique responsable et acceptable.

Au plan national, la nouvelle autorité de contrôle des SIA, devrait naturellement développer une telle réflexion, comme la loi en confie la compétence à la CNIL en matière de traitement des données à caractère personnel[256]. Il serait artificiel d'imaginer un système où l'éthique serait une voie parallèle et étanche à la normativité juridique. Dans la mesure toutefois où son intervention serait d'abord

[256] V. le e) du 4° du I de l'art. 8 de la loi du 6 janvier 1978.

opérationnelle et où ses ressources ne lui permettraient probablement pas d'animer efficacement un débat éthique permanent sur l'IA, il est essentiel qu'elle s'appuie sur les travaux du Comité consultatif national d'éthique (CCNE) qui a créé un comité national pilote d'éthique du numérique à l'origine de plusieurs avis extrêmement riches (sur les agents conversationnels, le numérique en santé, la reconnaissance faciale, les « véhicules autonomes »…). Ces avis pourraient donner lieu à un débat au sein d'une structure dédiée de l'autorité nationale de contrôle, indépendante (y compris du collège), composée de personnalités qualifiées de toutes disciplines, du président ou d'un membre du comité d'éthique du numérique et de représentants d'associations, de syndicats, d'opérateurs et d'usagers, qui assurerait l'interface entre la réflexion sur l'éthique de l'IA et la doctrine d'application des règles de droit encadrant l'utilisation des SIA, en particulier publics.

Cette réflexion éthique généraliste doit inspirer la réflexion éthique opérationnelle, fonction qu'il convient de structurer en proximité. La difficulté consiste ici à concilier, d'une part, la nécessaire articulation de cette fonction avec les équipes en charge de la conception et du déploiement des SIA, afin d'éviter une déconnexion qui aboutirait à priver d'effectivité les recommandations éthiques, et, d'autre part, la nécessaire indépendance de la réflexion éthique, à défaut de laquelle il est illusoire d'espérer inspirer confiance.

Il n'y a certainement pas lieu d'assortir tout déploiement de SIA d'un comité d'éthique indépendant dédié. Une telle tentation se heurterait rapidement à la pénurie de personnalités qualifiées pour les composer. La multiplication des structures serait en outre dommageable en termes de cohérence et de coût, si elle entraîne des conflits de normes (avec la réglementation générale), des orientations contradictoires (entre des comités multiples non coordonnés) ou une impression d'inefficience (si l'éthique n'est qu'une recommandation sans contrainte ni sanction). La constitution d'un tel comité ne peut se justifier que pour des SIA à haut risque, et, en leur sein, pour ceux qui présentent la sensibilité la plus élevée. Pour les autres systèmes, les réflexes déontologiques des participants au projet, l'intervention des administrateurs des données, des algorithmes et des codes-sources et le rôle de correspondant éthique qui pourrait être dévolu au délégué à la protection des données, constituent le premier niveau de contrôle. Le second, en surplomb, pourrait revenir, au sein de chaque administration (à la maille ministérielle et dans les grandes collectivités territoriales), à une **instance collégiale**, associant la société civile, adressant des recommandations sur les usages qu'il conviendrait de s'interdire en l'absence d'acceptabilité suffisante et sur la prise en compte des risques identifiés dans les choix de conception des SIA en projet.

2. L'implication civique

On a déjà souligné à quel point l'approche en termes de droits individuels pouvait paraître insuffisante. D'une part, pour les systèmes les plus sensibles (sécurité et défense), elle demeure assez distante (contrôle indirect, sur des traitements parfois classifiés, ou non publiés) ; d'autre part, elle n'aboutit qu'à rectifier des situations individuelles et repose sur la mobilisation de chaque personne concernée, dont on sait la très grande faiblesse en raison d'une asymétrie structurelle d'information, de

moyens et de compétences ; enfin, et surtout, elle ne porte que sur des données personnelles. Or, un SIA pouvant décider du sort de questions cruciales peut parfaitement ne traiter aucune donnée personnelle, et ainsi se prêter assez mal, pour le cœur de son contrôle, à une approche individuelle : la justice ou l'équité du système sont des questions globales, reposant autant sur la nature des données traitées que sur les choix faits pour orienter le fonctionnement du système, ou les variables qui vont déterminer son évolution. Une approche collective est alors nécessaire.

Installer l'usager, le bénéficiaire (ou la victime) du système, et plus largement le citoyen, dans la régulation des SIA publics, est une ambition qui soulève des questions dépassant de beaucoup la question de l'intelligence artificielle : la participation des citoyens, autrement que par la voie de consultations formelles et autres panels de citoyens, appelle une grande variété de réflexion, sur les évolutions des statuts associatifs, l'accès à l'expertise, la création de droits nouveaux, dépassant d'évidence le cadre de la présente réflexion.

Il reste néanmoins possible d'affirmer que **seule la présence effective de citoyens directement ou par l'intermédiaire de partenaires sociaux ou d'associations autour de la table de la conception, du déploiement, de la critique et de la correction des SIA est de nature à prévenir une partie de la défiance et de la critique**.

Il ne s'agit évidemment pas de tenter de détourner une critique indépendante, dont l'administration ne peut que profiter, non seulement quand elle est formulée devant le juge, qui ne devrait être que l'arbitre ultime des différends, mais avant tout auprès d'elle et sur la place publique, en instaurant un dialogue nourri par la transparence des déploiements et par le souci partagé du bien commun au regard des objectifs visés. Il ne s'agit pas non plus d'évincer les structures internes de la concertation, mais de faire en sorte que cette concertation épouse l'agilité des projets eux-mêmes, sans être prisonnier des processus consultatifs traditionnels : à ce titre, l'association des organisations syndicales ne peut se borner à une présentation ponctuelle, projet par projet, en réunion du comité social, mais donner lieu à une véritable concertation stratégique, au niveau interministériel et ministériel. Elle doit en outre être envisagée dans le cadre d'une consultation plus large associant l'ensemble des parties prenantes, sans que ne soient artificiellement opposés les intérêts des usagers et ceux des agents.

Aller plus loin est donc nécessaire, recourant à des instances de validation, d'écoute et de suivi, dont le droit positif donne des modèles dans d'autres secteurs qui pourraient avantageusement être expérimentés dans le champ des SIA : les commissions locales d'information, dans le domaine environnemental[257], les comités de suivi des plans de sauvegarde et de mise en valeur dans le domaine du patrimoine[258] et, naturellement, les commissions des usagers des établissements de

[257] Art. L. 125-17 du code de l'environnement.

[258] Art. L. 313-1 et suivants du code de l'urbanisme.

santé[259] et les comités de protection des personnes[260], dans le domaine sanitaire, sont autant d'exemples d'une implication opérationnelle de personnes intéressées dont il est possible de s'inspirer.

Il reviendra au législateur et à l'administration de déterminer les modalités de cette participation civique. D'ores et déjà, il apparaît souhaitable d'expérimenter, au fur et à mesure des déploiements de SIA, des modes d'implication civique aptes à instaurer la confiance, combinant implication des associations sectorielles et des citoyens volontaires, en s'interrogeant sur leur nécessaire formation et sur les moyens mis à leur disposition pour permettre leur implication concrète, et à définir par chaque porteur de projet en s'appuyant, le cas échéant, sur des instances déjà existantes. A ce titre, plusieurs pistes pourraient être explorées : rendre obligatoire la prise en compte des attentes du public (dès le lancement du projet et en continu, pour rendre possibles des évolutions du système répondant à ces attentes), permettre le lancement de contrôles suggérés par les usagers, instaurer des cadres de dialogue direct avec les usagers du système (dans le cadre d'une forme de médiation permanente), etc. Le but n'est pas de neutraliser la critique, mais de rechercher les conditions d'une coopération respectueuse des intérêts des uns (le service et sa mission) et des autres (le public et ses droits) pour bâtir une confiance propice à un bon usage de l'IA.

[259] Art. L. 1112-3 du code de la santé publique.

[260] Art. L. 1123-7 du code de la santé publique.

Conclusion

Les avancées technologiques récentes ont ouvert une nouvelle ère dans l'histoire de l'intelligence artificielle. Le secteur public ne peut se contenter d'en être le spectateur passif ou le régulateur distant. Pour écarter le risque d'être dépassé par des techniques qu'il doit pourtant contrôler, pour le bien commun, et de voir contestée sa capacité à fournir aux citoyens une qualité de service au niveau de ce qu'offrent les acteurs privés les plus avancés, les acteurs publics doivent s'engager résolument dans la voie du développement de systèmes d'intelligence artificielle, et emprunter le chemin de déploiements raisonnés, éthiques, soutenables et responsables. C'est la condition d'une performance que les citoyens réclament. C'est aussi une façon, pour la France, de continuer à jouer, comme elle le fait depuis 1978, le rôle d'inspiratrice des bons usages des nouvelles technologies numériques. Le tissu économique prometteur qui est le substrat sur lequel se déploient les systèmes d'intelligence artificielle a besoin d'une vision, d'une stratégie, d'un élan de la sphère publique qui lui donnent les moyens de se développer. Ce développement doit se faire d'emblée sur des bases respectueuses de l'éthique publique et protectrices. Tel est le principal appel qui conclut ce rapport.

La réussite de cette ambition reposera largement sur la capacité des pouvoirs publics et des administrations publiques à inspirer confiance dans cette démarche et dans les outils qui en sont les produits. A l'heure de la réforme de l'État permanente, l'intelligence artificielle, si elle n'est pas perçue comme une fin en soi mais comme un outil au service d'une meilleure administration, si elle s'accompagne de l'assurance de la loyauté et de l'effectivité des garanties données, peut constituer la nouvelle étape de la transformation de la relation au service public. L'adhésion à ce postulat ne ressort pas de l'évidence. C'est donc avant tout – avant de concevoir lois, règlements, structures et budgets – sur l'accompagnement de ces changements administratifs que doit se concentrer l'effort, en mobilisant, en premier lieu, les agents publics dans leur ensemble, ainsi replacés au centre des réflexions sur les évolutions souhaitables de leurs métiers.

La conduite de ce chantier devra permettre d'anticiper l'évolution du cadre normatif en gestation au niveau européen. Si les marges de manœuvre laissées, à ce stade, aux Etats membres sont réelles, l'envergure des transformations que le projet de règlement européen porte en germe invite, à tout le moins, à l'anticipation. La rédaction de lignes directrices d'ici à son entrée en vigueur doit, à ce titre, être perçue tant comme un levier de sécurisation des acteurs que comme un outil d'acculturation aux enjeux et aux principes garants d'une intelligence artificielle « de confiance ».

Si cette étude a pu évoquer, dans leur diversité, la plupart des problématiques s'attachant à la mise en œuvre d'une stratégie de déploiement de systèmes d'intelligence artificielle au sein des administrations publiques, son ambition n'était ni d'épuiser la variété des questions posées, ni de proposer un examen approfondi de chacune d'elles. D'autres études devront les éclairer, qu'il s'agisse, par exemple, de la circulation des données au sein de la sphère publique et de la valorisation de celles qui sont produites par les systèmes d'IA publics, de la compatibilité du déploiement massif de ces systèmes avec les contraintes environnementales, ou encore de ses conséquences sociales, économiques et budgétaires.

L'ampleur de la tâche est immense. Le Conseil d'Etat, avec l'ensemble de la juridiction administrative, entend y prendre toute sa part pour ce qui le concerne, en y associant l'ensemble des agents et des acteurs du droit.

Annexes

Annexe 1 : Lettre de mission

Annexe 2 : Composition du groupe de travail

Annexe 3 : Liste des auditions conduites par la mission

Annexe 4 : Glossaire

Annexe 5 : Quelques dates-clés de l'IA

Annexe 6 : Synthèse du projet de règlement européen sur l'intelligence artificielle (« règlement IA »)

Annexe 7 : Méthodologie et bilan du questionnaire adressé aux administrations

Annexe n° 8 : Présentation des acteurs de l'IA publique en France

Annexe 9 : « Cartographie » des cas d'usage des systèmes d'IA dans l'action publique

Annexe 10 : Note sur le cadre juridique national des systèmes d'IA publics

Annexe 1 : Lettre de mission

Le Premier Ministre

0742/21/SG

Paris, le 24 JUIN 2021

Monsieur le Vice-Président,

Le recours à des procédés d'intelligence artificielle, c'est-à-dire reposant sur des programmes informatiques dits auto-apprenants, est devenu, en quelques années seulement, un phénomène massif et structurant pour les sociétés entrées dans la transition numérique.

Il paraît aujourd'hui acquis que la sphère publique ne pourra pas rester en dehors de ce mouvement, quelles que soient par ailleurs les inquiétudes légitimes qu'il suscite. Dans le domaine de l'action publique, ces interrogations d'ordre général, relatives notamment à un risque de perte de maîtrise de l'outil informatique, se doublent d'une vigilance particulière quant au respect des principes du service public dans la prise de décisions individuelles. En outre, l'État occupe une place à part dans ces évolutions, en sa double qualité d'utilisateur et bénéficiaire potentiel de ces technologies, mais aussi de garant, dans l'ordre interne et par son action internationale, d'un développement qui ne mette pas en cause des équilibres éthiques ou démocratiques.

Il convient donc de poser un regard lucide sur les transformations de l'action publique induites par les progrès de l'intelligence artificielle, sur les gains qui peuvent en être attendus en termes d'efficacité et d'amélioration de la prise de décision et sur les conditions qui doivent l'entourer pour se prémunir d'un certain nombre de risques tels qu'une perte de souveraineté de l'Etat, une utilisation des données à caractère personnel irrespectueuse du droit au respect de la vie privée ou un examen insuffisamment circonstancié des situations individuelles.

C'est la raison pour laquelle je souhaite que le Conseil d'État conduise une étude sur ce sujet complexe, aux enjeux encore mal appréhendés et en évolution constante.

Sur la base d'une clarification des concepts, cette étude pourra d'abord établir une cartographie des outils existants, dans les services de l'État et des collectivités publiques en général, d'ores et déjà susceptibles de se rattacher à la notion d'intelligence artificielle ou de faire l'objet, à bref délai, de tels traitements.

L'étude s'attachera ensuite, au regard de l'évolution du droit européen, à évaluer l'impact potentiel, en termes de qualité de l'action publique, de l'introduction ou du développement de l'intelligence artificielle pour certaines missions telles que,

notamment, la santé, la justice, l'éducation, l'emploi, la sécurité intérieure ainsi qu'au sein de services disposant de pouvoirs d'enquête (fiscalité, concurrence, douanes). En ce qui concerne la sécurité intérieure, la complexité et la sensibilité du sujet nécessiteront sans doute des études spécifiques complémentaires.

Enfin l'étude aura pour objet de définir les conditions d'un bon usage de l'intelligence artificielle dans l'action publique, au regard des risques exposés ci-dessus, et de déterminer si son développement appelle des garanties juridiques particulières.

Je souhaite que cette étude puisse être rendue avant la fin de l'année 2021.

Je vous prie de croire, Monsieur le Vice-Président, en l'assurance de ma haute considération.

Annexe 2 : Composition du groupe de travail

Président : Thierry Tuot, *conseiller d'État, président-adjoint de la section de l'intérieur*

Rapporteurs :

Alexandre Lallet, *conseiller d'Etat, assesseur à la 10e chambre de la section du contentieux*

Thalia Breton, *auditrice de 1ère classe, rapporteure à la 4e chambre de la section du contentieux*

Membres du Conseil d'État

Laurent Cytermann, *rapporteur public à la 3e chambre de la section du contentieux*

Jean Lessi, *rapporteur à la section sociale*

Timothée Paris, *rapporteur à la section de l'intérieur*

Personnalités extérieures

Nadi Bou Hanna, *Directeur interministériel du numérique (DINUM*), représenté par **Paul-Antoine Chevalier,** *responsable du Lab IA et du pôle exploitation de données d'Etalab, direction interministérielle du numérique (DINUM)* et **Périca Sucevic**, chef du pôle Droit et société d'Etalab

Thierry Lambert, *délégué interministériel à la transformation publique*, représenté par **Grégoire Tirot**, chef du département « pilotage du plan de la transformation publique », **Sarah Allix**, chef de projet, et **Cécile le Guen**, responsable du pôle outils de pilotage par la donnée

Antoine Louvaris, *professeur agrégé des facultés de droit (droit public), ancien élève de l'ENA, président du comité de pilotage de l'institut Droit Dauphine et codirecteur du Master 2 droit des affaires*

Renaud Vedel, *coordonnateur national pour l'intelligence artificielle*

*

Le secrétariat de la section du rapport et des études a apporté à la mission un appui précieux pour l'organisation des réunions et des entretiens, et la mise au point de l'étude.

Charleene Eymard et Elisabeth Buisson, stagiaires à la section du rapport et des études, ont contribué à la réflexion et à la rédaction de l'étude. Roxane Abou et Charles Morin, en la même qualité, ont également participé aux travaux.

Annexe 3 : Liste des auditions conduites par la mission

(Par ordre alphabétique. Les fonctions mentionnées sont celles occupées à la date de l'audition)

Isabelle Adenot, membre du collège de la Haute Autorité de santé (HAS) et présidente de la Commission nationale d'évaluation des dispositifs médicaux et des technologies de santé (CNEDiMTS) de la HAS, accompagnée de Corinne Collignon, adjointe au chef de service de l'évaluation des dispositifs et de Pierre-Alain Jachiet, responsable de la stratégie des données

Jamal Atif, professeur à l'université Paris Dauphine-PSL, chargé de mission "science des données et intelligence artificielle" à l'Institut des sciences de l'information et de leurs interactions (INS2I) du Centre national de la recherche scientifique (CNRS).

Guillaume Avrin, responsable du département évaluation de l'intelligence artificielle du Laboratoire national de métrologie et d'essais (LNE)

Francis Bach, directeur de recherche à l'Institut national de recherche en sciences et technologies du numérique (INRIA)

Emmanuel Bacry, directeur de recherche au CNRS, directeur scientifique du Health Data Hub

Nathalie Bakhache, directrice adjointe du cabinet de la ministre de l'action et de la fonction publiques, et M. Antoine Michon, conseiller chargé de la transformation numérique de l'État, des affaires européennes et internationales

Stéphane Barritault, secrétaire général de l'Institut de cardiométabolisme et nutrition (IHU ICAN)

Adrien Basdevant, avocat et membre du Conseil national du numérique (CNNUM)

Laure Bédier, directrice des affaires juridiques au ministère de l'économie, des finances et de la relance, accompagnée de Serge Doumain, chef du bureau « Economie, statistiques et techniques de l'achat public » (Observatoire économique de la commande publique)

Alexandra Bensamoun, professeure de droit privé à l'université Paris-Saclay

Alain Bensoussan, avocat, accompagné de M. Éric Bonnet, directeur de la communication juridique du cabinet Lexing Alain Bensoussan Avocats

Benoît Bergeret, directeur exécutif du Metalab, accompagné de Julia Fenart, Head of European Affairs, France Digitale et de Louis Fleuret, directeur adjoint de La French Tech

Patrick Bezombes, président de la commission de normalisation Intelligence artificielle de l'Association française de normalisation (Afnor), accompagné de Isabelle Blanc**,** Chief Data Officer, ministère de l'enseignement supérieur et de la recherche

Gérard Biau, professeur, Sorbonne Université, directeur du Sorbonne Center for Artificial Intelligence (SCAI) accompagné de Raja Chatila, professeur émérite en robotique et en intelligence artificielle

Isabelle Blanc, Chief Data Officer, ministère de l'enseignement supérieur et de la recherche, AMDAC

Annabelle Bouchet, représentante du syndicat Snepap-FSU

Hélène Brisset, directrice du numérique aux ministères chargés des affaires sociales, AMDAC

Michel Cadot, Délégué interministériel aux Jeux olympiques et paralympiques, accompagné de Christophe Delaye, conseiller en charge de la sécurité

Anne-Florence Canton, cheffe du service du numérique au ministère de la justice (AMDAC), accompagnée de Fabien Antoine, directeur de projet Stratégie data, de Marine Kettani, chargée de mission près de la cheffe du service de l'expertise et de la modernisation et de Camille le Douaron, chargée de mission data au service de l'expertise et de la modernisation

Jean-Yves Capul, administrateur des données, ministère de l'éducation nationale, de la jeunesse et des sports (AMDAC)

Emmanuel Chiva, directeur de l'Agence de l'innovation de défense, ministère des armées, accompagné de Michaël Krajecki, directeur de projet IA

Julien Chiaroni, directeur du Grand Défi en Intelligence Artificielle au Secrétariat général pour l'investissement

Olivier Colliot, directeur de recherche, Centre national de la recherche scientifique (CNRS)

Stéphanie Combes, directrice du Health Data Hub (plateforme nationale des données de santé)

Stéphane Commans, responsable Portfolio Projets scientifiques et alliances à l'Institut de cardiométabolisme et nutrition (IHU ICAN)

Michel Cottura, directeur général adjoint chargé du pilotage des programmes et de la maîtrise d'ouvrage, Pôle Emploi

Bertrand Decaix, directeur de cabinet de l'Agence centrale des organismes de sécurité sociale (ACOSS), accompagné de Jean-Baptiste Courouble, directeur des systèmes d'information, de Carole Leclerc, directrice de l'innovation et de Xavier Bonnet, directeur de l'audit du pilotage de la performance et de la stratégie

Amaury Decludt, chef de la délégation à la stratégie de la direction générale des douanes et droits indirects

Maud Decraene, responsable du pôle Juridique & Valorisation à l'Institut de cardiométabolisme et nutrition (IHU ICAN), déléguée à la protection des données (DPO) ICAN

Nicolas Deffieux, directeur du pôle d'expertise de la régulation numérique (PEReN), ministère de l'économie, des finances et de la relance, accompagné de Florent Laboy, directeur adjoint, et de Lucas Verney, expert technique

Romain Delassus, chef du service du numérique au ministère de la culture, accompagné de Christine Debray, cheffe du département stratégie et pilotage du numérique, ainsi que de Romain Joron et Aurélien Cornaux

Marie-Laure Denis, présidente de la Commission nationale de l'informatique et des libertés (CNIL), accompagnée de Louis Dutheillet de Lamothe, secrétaire général, Bertrand Pailhes, directeur des technologies et de l'innovation et Thomas Dautieu, directeur de la conformité

Stéphane Donne, directeur du département statistiques, système d'information et big data de la Caisse nationale des allocations familiales (CNAF), accompagné de Agnès-Laurence Nal, attachée de direction à la direction du réseau au ministère du budget, des comptes publics, de la fonction publique et de la réforme de l'État

Stéphane Duhieu, délégué de recherche à l'Institut de la vision (IHU Foresight)

Nathalie Demont, secrétaire fédérale de la Fédération générale des fonctionnaires force ouvrière (FGFFO)

Thomas Dumortier, conseiller juridique à la Commission nationale consultative des droits de l'homme (CNCDH), accompagné de Célia Zolynski, personnalité qualifiée et de Lucien Castex, membre

Olivier Esper, gestionnaire principal des politiques publiques chez Google, accompagné de Ludovic Peran, chef de produit pour la recherche IA, Inès Kouraïchi, responsable commercial secteur public France, Italie, Espagne et Portugal, et Léa Manenti, responsable collectivités territoriales

Luc Farré, secrétaire général de l'Union nationale des syndicats autonomes (UNSA) Fonction publique

Gabriel Ferriol, directeur du service de vigilance et de protection contre les ingérences numériques étrangères (Viginum) au Secrétariat général de la défense et de la sécurité nationale (SGDSN)

Fabien Fieschi, directeur du numérique au ministère de l'Europe et des affaires étrangères (AMDAC)

Xavier Fischer, Chief Executive Officer, DatakaLab

Corinne Fortin, secrétaire générale de l'Institut du cerveau

Nicolas Goniak, conseiller pour les affaires intérieures à la représentation permanente de la France auprès de l'Union européenne, Benoît Blary, conseiller en charge des télécommunications, du numérique et des postes, Pauline Dubarry, conseillère Justice, Jonathan Cole, conseiller en charge des relations avec le Parlement européen

Etienne Grass, Directeur général « secteur public », Capgemini Invent

Emmanuel Grégoire, premier adjoint de la maire de Paris, accompagné de Pierre Musseau, conseiller ville intelligente et durable

David Gruson, directeur du programme Santé, Jouve

Stéphane Hatem, directeur de l'unité mixte de recherche 1166 (maladies cardiovasculaires et métaboliques, faculté de médecine Sorbonne Université

Samuel Heuzé, chef de la mission d'organisation des services du Premier ministre (AMDAC)

Sylvain Humbert, secrétaire général adjoint du Conseil d'État, chargé des juridictions administratives

Mylène Jacquot, secrétaire générale de l'Union des fédérations de fonctionnaires et assimilés (Uffa-CFDT)

Dominique Jamme, directeur général des services de la Commission de régulation de l'énergie, accompagné de Didier Lafaille, chef du service de la prospective et de l'innovation

Edward Jossa, président de l'Union des groupements d'achats publics (UGAP), accompagné de Frédéric Trinquecoste, directeur des achats informatiques, de Lionel Ferraris, directeur en charge des politiques publiques et d'Emilia Soeiro-Terme, cheffe de département Prestations intellectuelles informatiques

Nicolas Kanhonou, directeur de la promotion de l'égalité et de l'accès aux droits, Défenseur des droits accompagné de Sarah Benichou, adjointe au directeur et de Gaëtan Goldberg, chargé de mission numérique, droits et libertés

Pascal Kessler, président de la Fédération autonome de la fonction publique (FA-FP)

Thierry Kirat, directeur de recherche au Centre national de la recherche scientifique (CNRS), directeur de l'école doctorale « Sciences de la décision, des organisations, de la société et de l'échange » (SDOSE), université Paris Dauphine-PSL

Claude Kirchner, directeur de recherche émérite de l'Institut national français de recherche en sciences et technologies du numérique (INRIA)

Jérôme Lang, directeur de recherche au Centre national de la recherche scientifique (CNRS), Senior Researcher, Lamsade, université Paris Dauphine-PSL

Ivan Laptev, directeur de recherche à l'Institut national français de recherche en sciences et technologies du numérique (INRIA)

Benoit Le Blanc, directeur de l'École nationale supérieure de cognitique, ENSC Bordeaux, président de l'Association française pour l'intelligence artificielle (AFIA)

Marc Le Floch, directeur adjoint du réseau de la Caisse nationale des allocations familiales (CNAF), accompagné de Yasmine Leroueil, adjointe au directeur collaboratif et système d'information des fonctions supports

Pascal Le Luong, secrétaire général de la Cour de cassation, accompagné d'Estelle Jond-Necand, conseillère référendaire et directrice du projet open data et de l'équipe-projet « pseudonymisation des décisions de justice »

Georges-François Leclerc, préfet de la région Hauts-de-France, accompagné d'Amélie Puccinelli, secrétaire générale adjointe de la préfecture du Nord, d'Olivier Rovère, directeur territorial adjoint Nord de l'Agence régionale de santé des Hauts-de-France, Jean-Yves Bessol, inspecteur d'académie, directeur académique des services de l'éducation nationale du Nord, et Jean-François Papineau, directeur zonal Nord de la sécurité publique.

Philippine Lefèvre-Rottmann, déléguée aux affaires publiques de Numeum, Jawaher Allala, CEO de Systnaps et Katya Lainé, CEO de Kwalys, administratrices, Valentin Hueber, délégué Industrie du futur, innovation et technologies, Lucile Lecomte, déléguée aux usages numériques

Fabrice Lenglart, directeur de la recherche, des études, de l'évaluation et des statistiques au ministère des solidarités et de la santé (AMDAC)

Thomas Lesueur, commissaire général au développement durable, ministère de la transition écologique (AMDAC), accompagné de Thomas Cottinet, directeur d'Ecolab et Marc Léobet, directeur de projet IA & Transition écologique auprès du directeur d'Ecolab

Jérôme Letier, directeur du numérique (DNUM) du ministère de l'intérieur (AMDAC), Jean-Martin Jaspers, délégué ministériel à l'intelligence artificielle et Christophe Marquaille, chef du bureau Laboratoire valorisation des données à la DNUM, datalab

Gaëlle Martinez, déléguée générale fonction publique de l'Union syndicale Solidaires

Nicolas Mayer-Rossignol, président de la métropole Rouen-Normandie

Françoise Mercadal-Delasalles, co-présidente du Conseil national du numérique (CNNum), accompagnée de Jean Cattan, secrétaire général, de Justine Cassell, directrice de recherche INRIA, Gilles Dowek, chercheur INRIA et de Philippine Régniez, rapporteure

Rémi Meunier, Directeur « secteur public » de Dataiku etRomain Doutriaux, vice-président marketing et communication Europe

Louise Meyfroit, chargée d'opérations scientifiques à l'Institut de cardiométabolisme et nutrition (IHU ICAN)

Sarah Michot, Junior Advocacy & Campaign Manager et Anne Mollen, Policy & Advocacy Managerin, AlgorithmWatch

Nicolas Monsarrat, directeur général Accenture Health, accompagné de Gabriel Bellenger, Health & Public Service Consulting lead

Laurent Nunez, coordonnateur national du renseignement et de la lutte contre le terrorisme

Cédric O, secrétaire d'État chargé de la transition numérique et des communications électroniques

Akim Oural, adjoint au maire et délégué ville numérique, Ville de Lille

Benoît Parizet, directeur de la transformation numérique de la Caisse des dépôts et consignations et de la stratégie digitale de la Banque des territoires, accompagné de Matthieu Blanc, responsable du Pôle Data

Lior Perez, responsable du département des développements à la direction des systèmes d'information Météo France et Christophe Morel, directeur de la stratégie de Météo-France

Manon Perrière, directrice adjointe de Tracfin, accompagnée de Mélanie Gourié, cheffe du département des systèmes d'information et Delphine Lé, cheffe du pôle Data Science

Edouard Philippe, maire de la Ville du Havre

Lionel Ploquin, administrateur général des données de la DGFIP, ministère de l'économie, des finances et de la relance, accompagné de Su Yang, responsable du pôle donné de la délégation à la transformation numérique et de François Terrier, directeur de recherches au CEA, directeur du programme Intelligence Artificielle de CEA Tech et de l'inflexion IA de confiance du CEA List

Fabrice Popineau, professeur à l'Ecole supérieure d'électricité (Supélec)

Guillaume Poupard, directeur général de l'Agence nationale de la sécurité des systèmes d'information (ANSSI)

Annie Prévot, directrice de l'Agence du numérique en santé, accompagnée de Marc Loutrel, directeur expertise

Laurence Prevost, directrice de la division Consulting Secteur public de Sopra Steria, accompagnée de Nicolas Conso, directeur conseil Secteur public

Simon Raout, directeur de la performance au centre hospitalier de Valenciennes

Pierre-Louis Rolle, directeur des programmes *Société Numérique* et *Nouveaux Lieux Nouveaux Liens & Mission incubateur de services numériques* et AMDAC de l'Agence nationale de la cohésion des territoires (ANCT)

Isabelle Ryl, directrice du centre de recherche de Paris de l'Institut national de recherche en sciences et technologies du numérique (INRIA)

Benoît Sagot, directeur de recherche à l'Institut national français de recherche en sciences et technologies du numérique (INRIA)

Philippe Schall, chef du bureau Programmation des contrôles et analyse des données à la direction générale des finances publiques (DGFIP)

Mehdi Siaghy, directeur de la recherche et de l'innovation, CHRU de Nancy

Sébastien Soriano, directeur général de l'Institut national de l'information géographique et forestière (IGN)

Bruno Sportisse, directeur général de l'Institut national de recherche en sciences et technologies du numérique (INRIA) accompagné d'Isabelle Herlin, coordinatrice de l'équipe recherche IA de l'INRIA

Périca Sucevic, chef du pôle Droit et société d'Etalab (DINUM) et Paul-Antoine Chevalier, responsable du Lab IA et du pôle exploitation de données d'Etalab

Jérôme Teillard, chef de projet Réforme de l'accès à l'enseignement supérieur du ministère de l'enseignement supérieur, de la recherche et de l'innovation (MESRI)

Stéphane Trainel, AMDAC des ministères économiques et financiers

Francky Trichet, vice-président de Nantes Métropole en charge de l'innovation, du numérique et de l'international et Claire Sacheaud, administratrice générale de la donnée, en charge de la stratégie data de la collectivité

Mohammed-Adnène Trojette, conseiller action publique et numérique du Président de la République et conseiller technique numérique du Premier ministre

Olivier Vallet, Président directeur général de Docaposte, accompagné de Pierre-Etienne Bardin, chief data officer du groupe La Poste

Renaud Vedel, Coordinateur national pour l'intelligence artificielle

Henri Verdier, ambassadeur pour le numérique

Julien Vignon, directeur de projets IA, service de l'économie numérique à la direction générale des entreprises, ministère de l'économie, des finances et de la relance

Cédric Villani, député, président de l'office parlementaire d'évaluation des choix scientifiques et technologiques

Renaud Villard, directeur général de la Caisse nationale d'assurance vieillesse (CNAV), accompagné de Véronique Puche, directrice des systèmes d'Information de la CNAV

Vincent Vuiblet, néphrologue au CHU de Reims, maître de conférences des universités, directeur de l'Institut d'intelligence artificielle en santé université de Reims Champagnes-Ardenne

*

Conseil de l'Europe

Muriel Décot, secrétaire de la Commission européenne pour l'efficacité de la justice (CEPEJ)

Yannick Meneceur, conseiller spécial auprès du secrétariat de la Commission européenne pour l'efficacité de la justice (CEPEJ)

Parlement européen

Iban Garcia Del Banco, député européen

Axel Voss, député européen

Dragos Tudorache, député européen

Commission européenne

Cabinets

Lucrezia Busa, membre du cabinet du commissaire européen en charge de la justice (Didier Reynders)

Werner Stengg, membre du cabinet de la vice-présidente de la commission européenne Margrethe Vestager

Nuria Subirats-Rebull, assistance chargée des politiques au cabinet du commissaire européen pour le marché intérieur (Thierry Breton)

DG CONNECT

Kilian Gross, chef de l'unité A2 développement et coordination des politiques en matière d'intelligence artificielle

DG JUST

Eike Gräf, gestionnaire des politiques pour les droits fondamentaux

DG HOME

Zsuzsana Felkai Janssen, coordinatrice IA

Dan Rotenberg, adjoint au chef d'unité D4 Sécurité dans un monde numérique

Gilles Robine, END

Agences

Darek Saunders, directeur de l'Observatoire de la sécurité des frontières et chercheur dans l'unité Recherche et Innovation de Frontex

Carlos Torrecilla Salinas, Ignacio Sanchez, Luca Tangi et Emilia Gomez, Joint Research Centre (JRC)

Annexe 4 : Glossaire

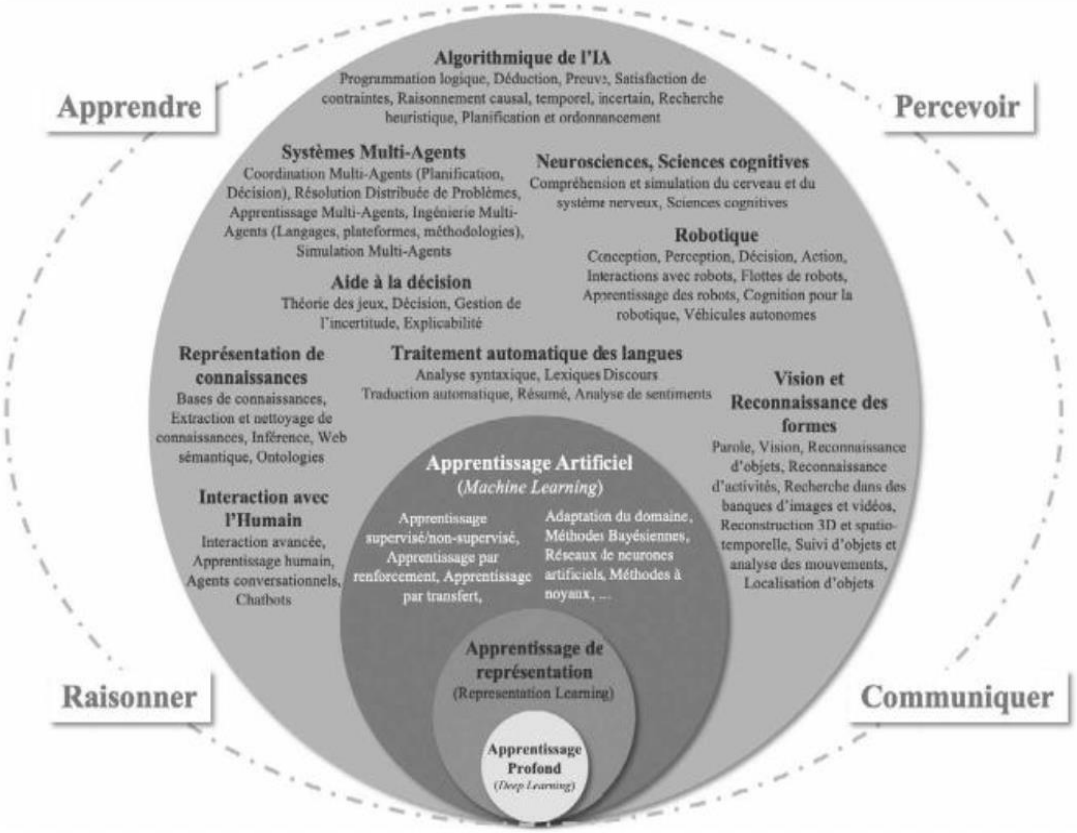


Schéma synthétique du vocabulaire de base de l'intelligence artificielle[261]

[261] J. Guilhem, *Sciences et sens de l'intelligence artificielle*, *Dalloz*, 2020, p. 16

Agent intelligent

Système capable de percevoir son environnement à l'aide de capteurs et, sur la base d'une analyse pouvant inclure le raisonnement, la modélisation et l'apprentissage, d'interagir avec lui à l'aide d'effecteurs afin de réaliser une tâche prédéfinie.

Une typologie classique de ces agents, résultant des travaux de référence de S. Russel et P. Norvig[262], distingue :

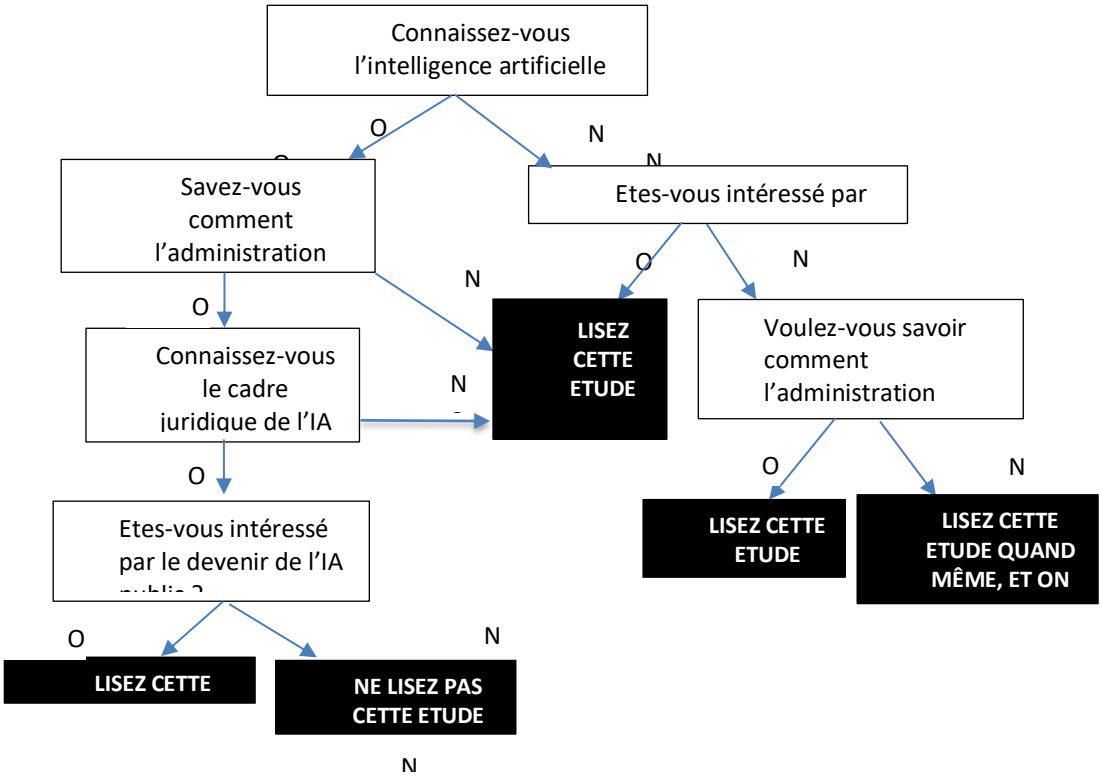
- les agents qui se bornent à prendre acte de l'environnement perçu pour déterminer l'action à effectuer (*simple reflex agents*) : un lampadaire public équipé d'un détecteur de mouvement s'allumera quelques secondes s'il fait noir et que quelqu'un passe, et restera éteint dans le cas contraire ;
- ceux qui utilisent des modèles pour analyser les données passées et tenir compte de la façon dont cet environnement évolue (spontanément) et des conséquences des actions du système sur l'environnement (*model based reflex agents*) : le lampadaire public capable de détecter la pluie pourra neutraliser cette perturbation météorologique, lorsqu'elle se produit, pour décider de s'allumer ou non. Ces agents interviennent en présence d'un environnement qui n'est pas « entièrement observable », au sens où une décision rationnelle ne peut être prise qu'en complétant les informations issues de l'observation directe par les capteurs par d'autres informations « externes » ;
- ceux à qui est assigné un objectif et qui sélectionnent une façon de l'atteindre, en comparant les conséquences de chaque action possible (*goal-based agents*) : un système d'assistance à la lutte contre les feux de forêts doit comparer les conséquences des différentes options (intervenir à tel ou tel endroit…) pour arrêter une stratégie permettant d'éteindre le feu avant qu'un bois ait entièrement brûlé ;
- ceux qui, en sus, évaluent la performance respective de chaque action et retiennent celle qui maximise le bénéfice, c'est-à-dire la meilleure façon d'atteindre l'objectif (*utility-based agents*) : un robot de nettoyage urbain à qui est assigné la tâche de nettoyer les trottoirs d'une rue en un minimum de temps devra s'interroger sur les conséquences des différentes options possibles (commencer par le trottoir de droite ou de gauche ? remonter ou descendre la rue ?) en tenant compte de différents paramètres (obstacles, circulation…), pour optimiser son parcours ;
- et, enfin, ceux qui intègrent un apprentissage, en mémorisant les résultats passés de leur action et en s'adaptant en conséquence (*learning agents*).

[262] S. Russel et P. Norvig, *Intelligence artificielle – Une approche moderne*, Pearson, 4e édition.

Algorithme

Ensemble de règles opératoires dont l'application permet de résoudre un problème énoncé au moyen d'un nombre fini d'opérations (Larousse) ; description d'une suite d'étapes permettant d'obtenir un résultat à partir d'éléments fournis en entrée (CNIL).

Exemple d'algorithme simplifié (et biaisé) de recommandation : dois-je lire ou non cette étude du Conseil d'État ?



- ***Algorithme basé sur les règles*** : algorithme qui reproduit un raisonnement logique de type déductif à base de règles programmées préalablement.
- ***Algorithme heuristique*** **:** algorithme dont le but n'est pas de proposer une solution optimale mais une solution satisfaisante en un minimum de temps. Une heuristique s'impose quand les algorithmes de résolution exacte sont d'une complexité excessive[263].
- ***Algorithme frugal*** **:** algorithme conçu pour utiliser un minimum de données (frugalité en données) ou de puissance de calcul (frugalité environnementale).

[263] J. Sohier, D. Sohier, *Logistique*, Vuibert, 2017, V. « Glossaire », p. 183-190.

Annotation / Étiquetage (*« Data Labeling »*)

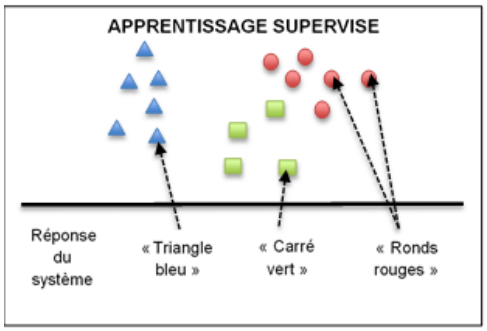
Tâche consistant à assigner une valeur ou un nom à des données (ex : « chat » sur une photo de chat). Cela permet ensuite à l'IA d'apprendre à reconnaître les différentes catégories de données et à les distinguer. C'est une étape indispensable de l'apprentissage automatique supervisé (V. ci-dessous)[264].

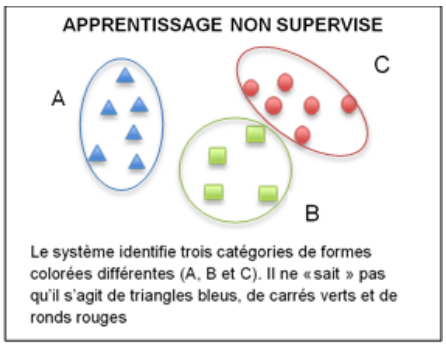
API (*Application Programming Interface* – Interface de programmation d'application)

Solution informatique permettant à des applications de communiquer entre elles et de s'échanger des données.

Apprentissage

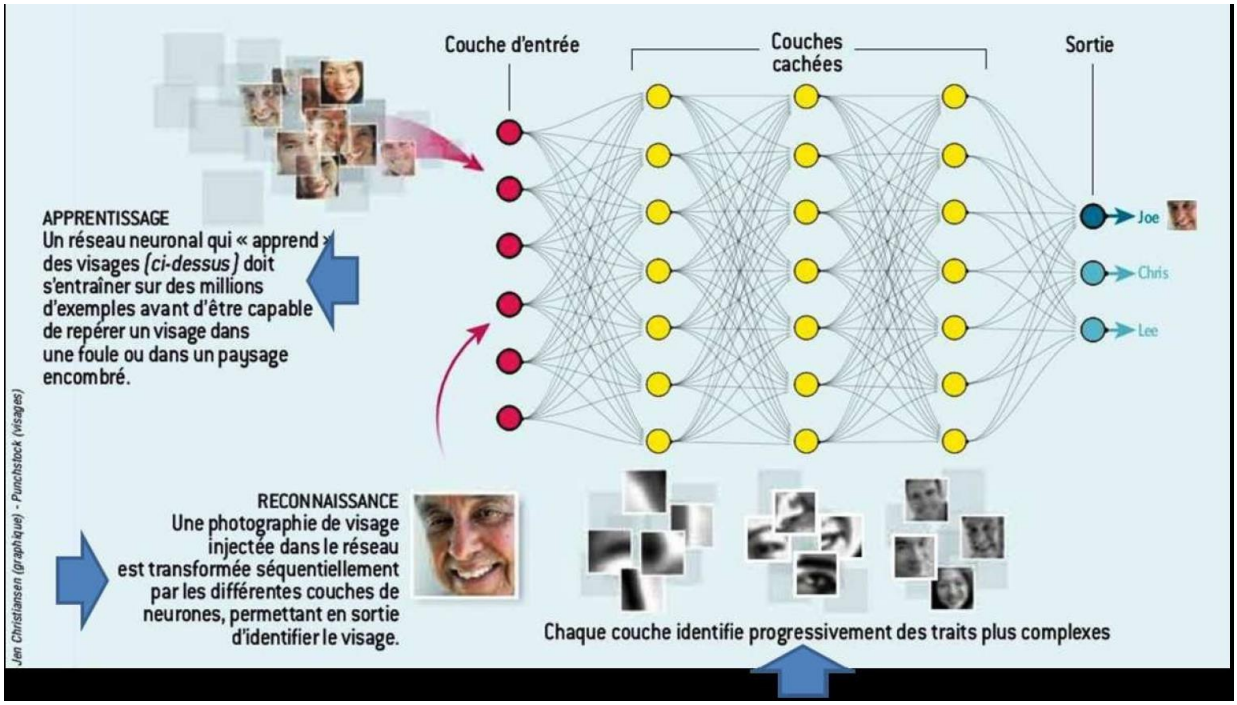
- ***Apprentissage auto-supervisé*** *:* méthode permettant d'utiliser des données non étiquetées afin de générer des étiquettes de façon automatique en se fondant sur la structure ou sur les caractéristiques latentes des données et en exploitant leurs corrélations. Exemple : modèle ALBERT développé par Google pour le traitement du langage naturel et la reconnaissance vocale dans les requêtes de recherche.
- ***Apprentissage automatique*** : approche d'intelligence artificielle qui consiste à construire un modèle algorithmique capable de modifier lui-même ses paramètres entraînables afin de maximiser l'exactitude des résultats qu'il produit à partir des données d'entrée. Ainsi, si le résultat produit n'est pas satisfaisant, le modèle, « nourri » par de nouvelles données, s'ajustera afin de produire un meilleur résultat. Exemple : un chatbot.
- ***Apprentissage supervisé*** : méthode d'apprentissage dans laquelle on indique au SIA les résultats attendus à partir de certaines données d'entrée, en recourant à des données annotées / étiquetées, afin qu'il adapte en conséquence ses paramètres entraînables et se rapproche le plus possible de la « bonne réponse ».
- ***Apprentissage non supervisé :*** technique d'apprentissage automatique ne recourant pas à des données annotées/étiquetées et ayant pour objet d'identifier la structure des données (caractéristiques communes, récurrences…) afin de les classifier (catégoriser).





[264] Pour un exemple concret d'utilité de l'étiquetage en matière de reconnaissance linguistique, V. par ex. S. Amri et L. Zenkouar, « Traitement automatique de la langue amazighe : algorithmes d'apprentissage pour l'étiquetage morphosyntaxique », *Études et Documents Berbères*, vol. 44, n° 2, 2020, p. 197-208.

- ***Apprentissage par renforcement*** **:** méthode d'apprentissage fondée sur l'obtention d'une récompense (en cas de succès) et/ou d'une punition (en cas d'échec) qui conduit le modèle à modifier ses paramètres entraînables afin de maximiser ses succès. Exemple : AlphaGo, premier programme à battre un joueur de go, puis le champion du monde de ce jeu, ou encore le programme *Deep Blue* en matière de jeu d'échecs.
- ***Apprentissage par transfert*** : méthode qui consiste à s'appuyer sur un modèle (source) pré-entraîné pour développer un autre modèle (cible). Ce type d'apprentissage est souvent utilisé lorsqu'il n'y a pas suffisamment de données pour entraîner le modèle cible. Exemple : La reconnaissance de voitures par un algorithme peut être transférée à un autre modèle destiné à reconnaître des camions.
- ***Apprentissage profond*** : méthode d'apprentissage reposant sur des réseaux de neurones artificiels, composés de plusieurs couches dont le nombre est déterminé par l'humain, qui est entraîné de telle sorte que la machine modifie la pondération (le poids) des neurones afin d'améliorer l'exactitude du modèle (c'est-à-dire de minimiser l'erreur)[265]. Exemple : Le système Horus de reconnaissance visuelle destiné aux aveugles.



[265] V. notamment sur les enjeux du *deep learning* : J. Guilhem*, Sciences et sens de l'intelligence artificielle*, *Dalloz*, 2020, p. 16, citant notamment Y. Le Cun, Y. Bengio et G. Hinton, « Deep learning », Nature, n° 521, p. 436–444 ; également à ce sujet : F. G'sell, *Le Big data et le droit*, *Dalloz*, 2020, p. 208

Bac à sable

Environnement technique isolé et sécurisé utilisé pour concevoir et tester un système d'IA.

Biais

Terme générique désignant un écart à la réalité. Dans la terminologie de l'IA, cette notion est utilisée pour désigner de nombreux phénomènes différents :

- ***Biais d'automatisation*** : tendance de l'humain à privilégier les recommandations d'un système d'IA sur ses propres analyses ou celles d'autres humains et, par extension, à avaliser de manière systématique ces recommandations ;

- ***Biais de renforcement*** **:** ce biais peut se produire lorsque le résultat produit par un système aboutit à des actions qui vont amplifier le phénomène mesuré (exemple d'un SIA qui recommande à la police d'intervenir plus fréquemment dans un quartier en raison d'une délinquance évaluée comme supérieure : si la police suit la recommandation, le durcissement des contrôles qui en résultera a toutes les chances d'aboutir à un surcroît d'interpellations, confirmant statistiquement la prévalence de la délinquance, alors même que la réalité objective peut être très différente).

- ***Biais de sélection (d'échantillonnage)*** **:** survient lorsque le jeu de données d'entraînement n'est pas représentatif en raison de la sur- ou de la sous-représentation de certaines catégories d'individus. Ce biais peut intervenir aussi bien au stade de la collecte qu'à celui de la préparation des données. Il peut résulter des préjugés des humains qui traitent les données, ou simplement refléter l'état historique du monde réel (ex. : un SIA entraîné au recrutement de cadres dirigeants sur la base de la population actuelle de ces cadres privilégiera le recrutement d'hommes, qui sont historiquement sur-représentés).

- ***Biais statistique*** : écart du résultat produit par le système par rapport au résultat attendu (V. l'entrée « compromis biais-variance »).

« *Big data* » (ou « mégadonnées »)

Désigne un très grand volume de données informatiques issues de sources hétérogènes (moteurs de recherche, réseaux sociaux etc.). Il recouvre les données collectées, stockées, traitées et analysées dans de courts délais, que ces données soient structurées ou non structurées[266]. Par extension, le « big data » désigne l'essor de l'utilisation à grande échelle d'une quantité massive de données, notamment par le secteur privé.

[266] V. l'article de F. G'sell, précité.

Classification automatique

Catégorisation algorithmique d'objets, qui consiste à attribuer une classe ou catégorie à chaque objet (ou individu) à classer, en se fondant sur des données statistiques.

***Chatbot* (agent conversationnel, dialogueur)**

Logiciel spécialisé dans le dialogue en langage naturel avec un humain, qui est capable notamment de répondre à des questions ou de déclencher l'exécution de tâches (commission d'enrichissement de la langue française).

***Cloud computing* ou « informatique en nuage »**

Mode de traitement des données d'un client, dont l'exploitation s'effectue par l'internet, sous la forme de services fournis par un prestataire[267]. Au lieu de s'appuyer sur des ressources (infrastructures informatiques, logiciels…) qu'elle acquiert ou contrôle, l'entité utilisatrice utilise des serveurs distants détenus ou contrôlés par un prestataire qui lui offre un ensemble de services pouvant inclure le stockage de données, la mise à disposition de puissance de calcul et la mise en réseau des terminaux (« *infrastructure as a service* »), ainsi que l'exécution d'applications (« *software as a service* »).

Code-source

Ensemble de fichiers texte qui présente les instructions composant un programme sous une forme lisible. Il est une sorte de traduction informatique du modèle algorithmique[268].

Extrait du code-source (public) de l'application « Calculatrice » de Windows

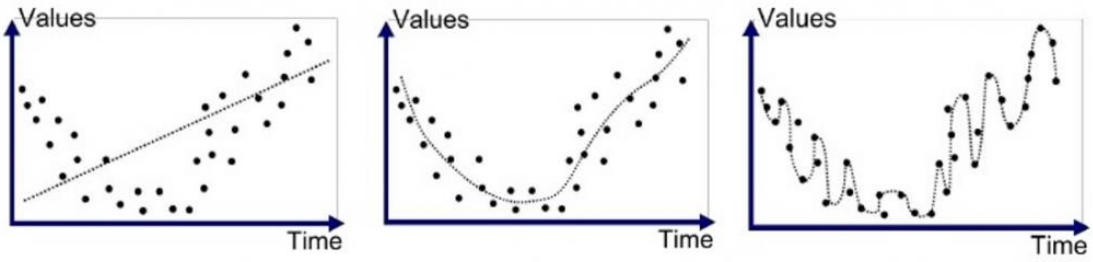
```
void TraceLogger::LogInvalidInputPasted(....)
{
  if (!GetTraceLoggingProviderEnabled()) return;
  LoggingFields fields{};
  fields.AddString(L"Mode", NavCategory::GetFriendlyName(mode)-
>Data());
  fields.AddString(L"Reason", reason);
  fields.AddString(L"PastedExpression", pastedExpression);
  fields.AddString(L"ProgrammerNumberBase",
GetProgrammerType(...).c_str());
  fields.AddString(L"BitLengthType",
GetProgrammerType(bitLengthType).c_str());
  LogTelemetryEvent(EVENT_NAME_INVALID_INPUT_PASTED, fields);
```

[267] Vocabulaire de l'informatique et de l'internet, JORF n° 0129 du 6 juin 2010.

[268] Rapport de la mission Bothorel, *Pour une politique publique de la donnée*, décembre 2020, p. 19.

Compromis (ou équilibre) biais/variance

Le biais permet de calculer l'erreur d'un modèle et, la variance, sa résilience, c'est-à-dire sa capacité à généraliser à des données autres que celles ayant servi à son entraînement. L'équilibre biais/variance vise ainsi à obtenir d'un modèle un bon compromis entre un taux d'erreur acceptable et une bonne capacité de généralisation. Un modèle qui présente un taux d'erreur très faible sur le jeu de données d'entraînement mais une forte variance est sur-ajusté (surentraîné).



Modèle sous-ajusté
La relation entre les deux variables est modélisée par une fonction très simple (linéaire), peu sensible à des données aberrantes, mais qui donne un résultat insatisfaisant (écart excessif entre le résultat attendu – les points – et le résultat proposé par le modèle – la ligne). Le modèle est trop fruste.

Modèle robuste
L'entraînement permet de dégager une fonction relativement simple qui rend compte de façon satisfaisante de la relation entre les deux variables.

Modèle sur-ajusté
La prévision est excellente sur les données d'entraînement (la fonction passe par tous les points ou presque), mais l'application à de nouvelles données donnera de mauvais résultats car le modèle, qui cherche à épouser la distribution des données sans en appréhender la structure sous-jacente, sera sensible au moindre « bruit » ou donnée anormale.

Destination (du système d'IA)

Usage auquel un système est destiné, incluant la finalité, le contexte spécifique et les conditions d'utilisation. Exemple : système permettant de traduire en anglais des textes littéraires écrits en français du XIXème siècle.

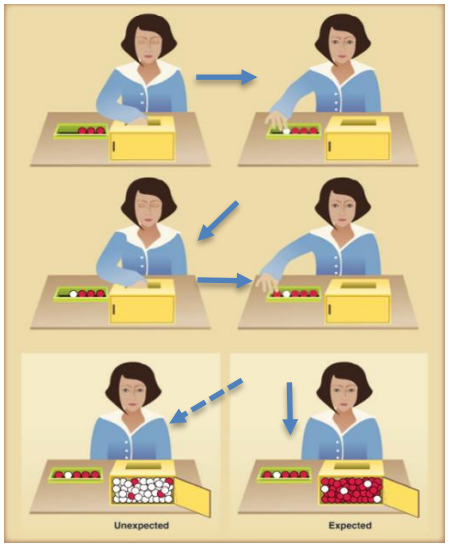
Données

- ***Donnée structurée*** **:** donnée formatée selon un référentiel prédéfini, à laquelle sont associées des métadonnées, et qui peut être aisément trouvée et traitée (organisée dans un entrepôt de données, avec des champs normés) : Exemple : numéros de téléphones et adresses dans un fichier de personnel ; références de produits dans une base de données d'un fabricant ; montant des transactions sur un compte bancaire ; référence du dossier contentieux sur une décision de justice...

- ***Donnée non structurée* :** donnée stockée dans un format quelconque (format d'origine en général) et dont l'exploitation requiert des outils complémentaires. Exemple : lac de données, contenu d'un courrier électronique, motifs d'une décision de justice, image satellite, film…
- ***Données d'entrée :*** données fournies à un système d'IA ou obtenues directement par lui et sur la base desquelles il produit un résultat (projet de règlement IA)
- ***Données de sortie (résultats)*** : données produites par un système d'IA en appliquant un traitement algorithmique à des données d'entrée.
- ***Données d'entraînement*** : données d'apprentissage qui permettent au système d'IA d'apprendre à effectuer la tâche qui lui est assignée.
- ***Données de validation*** : données utilisées pour fournir une évaluation d'un système d'IA entraîné et ajuster ses paramètres non entraînables et son processus d'apprentissage, notamment pour éviter le surajustement (projet de règlement IA). Il s'agit de données entièrement ou partiellement différentes des données d'entraînement.
- ***Données de test*** : données utilisées pour fournir une évaluation indépendante d'un système d'IA entraîné et validé afin de confirmer sa performance attendue avant de le mettre sur le marché ou en service (projet de règlement IA). Ces données sont utilisées par les organismes de certification indépendants pour délivrer une certification.

Estimation bayésienne (ou inférence bayésienne)

Technique statistique consistant à estimer la probabilité qu'une règle soit vraie (ou qu'un phénomène se produise) en fonction de l'observation de données de résultats (alors que la statistique classique, dite fréquentiste, consiste, à l'inverse, à évaluer la probabilité d'un résultat en fonction d'une règle donnée).



En fonction de la couleur des boules tirées, on peut évaluer la probabilité que le contenu de la boîte corresponde à celui de gauche (très faible probabilité) ou à celui de droite (probabilité forte). Il s'agit d'une opération d'inférence (compte tenu de l'effet, quelle est la probabilité de la cause ?). On parle de probabilité « conditionnelle ».

On peut s'en servir, par exemple, pour évaluer la probabilité qu'une personne soit atteinte de la covid-19 sachant que son test PCR est positif, compte tenu du taux moyen de faux positif.

L'estimation bayésienne peut être utilisée pour l'analyse « prédictive », en appliquant les probabilités calculées à des situations nouvelles.

Visuel issu de https://scienceetonnante.com/2012/10/15/linference-bayesienne-bayes-level-2/

Exactitude

Principale mesure de la performance des systèmes d'IA. L'exactitude peut elle-même être mesurée par des indicateurs différents : justesse (*accuracy*), précision (*précision*), sensibilité ou rappel (*recall*), moyenne harmonique (*F1 score*)…

A titre d'exemple, l'exactitude d'un SIA de détection des tumeurs sur une radio des poumons pourra être mesurée comme suit :

Vérité / Réponse système	Le patient présente une tumeur	Le patient ne présente pas de tumeur	TOTAL
Tumeur détectée	490	50	540
Pas de tumeur détectée	10	4450	4460
TOTAL	500	4500	5000

La justesse du système est de 98,8% ([490+4450]/5000) : dans 98,8% des cas, le système a vu juste.

La précision du système est de 90,7% (490/540) : le système a eu raison de signaler la présence d'une tumeur dans 90,7% des cas.

La sensibilité du système est de 98% (490/500) : 98% des tumeurs présentes sur les images ont été détectées par la machine.

La moyenne harmonique est de 94%.

Experts des données

Ensemble des métiers spécialisés dans l'analyse, l'exploitation et le traitement des données, et comprenant notamment les métiers suivants :

- ***Analyste des données*** : métier consistant à tirer des données exploitées des informations et des connaissances organisées et restituées de manière didactique et sous des formes favorisant la prise de décision (statistiques, tableaux de bord, graphiques, indicateurs…).
- ***Ingénieur IA*** **(data scientist) :** métier consistant à concevoir les solutions complexes, en particulier les modèles algorithmiques constitutifs des systèmes d'IA, permettant de tirer le maximum de valeur des données pour l'organisation.
- ***Architecte des données*** **(data architect*) :** métier consistant à concevoir la stratégie, les principes et les standards de collecte, de stockage et d'exploitation (accès, utilisation…) des données d'une organisation, et à préconiser les solutions techniques à mettre en œuvre à cette fin.
- ***Ingénieur des données*** **(data engineer) :** métier consistant à développer et assurer la maintenance de l'infrastructure de données afin d'en assurer la performance et le respect des règles encadrant leur utilisation, sur la base des travaux de l'architecte des données. L'ingénieur de la donnée est aussi chargé, comme développeur informatique, de coder les modèles algorithmiques conçus par l'ingénieur IA.

- ***Intendant des données*** **(data steward) :** métier, d'apparition récente et aux contours mal définis, consistant à s'assurer de la qualité et de l'exploitabilité des données ainsi que du bon fonctionnement des flux de données et de leur conformité au droit et aux règles internes de l'organisation.

  Le ***développeur*** n'est pas considéré comme un « expert de la donnée » mais il est indispensable à la conception et au développement d'un SIA. On parle souvent de « développeur *full stack* » pour désigner un développeur polyvalent, capable à la fois de coder le modèle algorithmique sous forme de programme informatique (« *back-end* ») et de gérer l'interface utilisateur, qu'il s'agisse d'une application ou d'un site web (« *front-end* »).

Généralisation

Capacité du système d'IA à produire des résultats corrects à partir de données autres que celles qui ont été utilisées pour son entraînement. On parle également d'inférence pour désigner la phase de mise en œuvre du système sur de nouvelles données d'entrée.

GPU (*graphics processing unit* – processeur graphique)

Composant électronique prenant la forme d'un circuit intégré souvent implanté sur une carte graphique, dont la puissance de calcul est utilisée pour la conception des modèles algorithmiques complexes. Sont aussi utilisés les unités de traitement de tenseur (*Tensor Processing Units* – TPU).

Intelligence artificielle (IA)

On se reportera, sur cette question, à la première partie de l'étude. Parmi les innombrables définitions proposées de l'IA, on peut citer les suivantes :

- Champ interdisciplinaire théorique et pratique qui a pour objet la compréhension de mécanismes de la cognition et de la réflexion, et leur imitation par un dispositif matériel et logiciel, à des fins d'assistance ou de substitution à des activités humaines (Commission de terminologie de la langue française) ;
- Un système d'IA est un système fondé sur une machine qui peut, pour un ensemble donné d'objectifs définis par l'humain, fait des prévisions, des recommandations ou prendre des décisions influençant les environnements réel et virtuel. Ils sont conçus pour fonctionner à des degrés d'autonomie divers (OCDE, 2019).
- Systèmes (et potentiellement infrastructures) conçus par l'humain qui, étant donné un but complexe, agit dans la dimension physique ou numérique en percevant son environnement à travers l'acquisition de données, en interprétant les donnée structurées et non structurées collectées, en raisonnant à partir des connaissances ou en traitant l'information dérivée de ces données, et en décidant des meilleures actions pour atteindre un but donné (Groupe d'expert de haut niveau de l'Union européenne, 2019).

- Un ensemble de technologies connexes utilisées pour résoudre des problèmes de façon autonome et accomplir des tâches afin d'atteindre des objectifs définis sans pilotage humain (Cadre éthique, Australie, 2019).
- Un système d'IA est un système artificiel qui accomplit des tâches dans des circonstances variées et non prévisibles sans supervision humaine significative, ou qui peut apprendre de l'expérience et améliorer ses performances en utilisant des jeux de données (US National Defense Authorization Act, 2018)
- Ensemble de sciences, théories et techniques dont le but est la conception et le développement de machines capables de reproduire des capacités cognitives d'un être humain (Comité ad hoc sur l'intelligence artificielle du Conseil de l'Europe).

Logiciel

Ensemble des programmes, procédés et règles, et éventuellement de la documentation, relatifs au fonctionnement d'un traitement de données[269]. Le cœur du logiciel est constitué du ou des programmes informatiques, qui consistent en des instructions écrites dans un langage de programmation ayant vocation à être exécutées par un processeur.

- ***Logiciel libre*** : logiciel que toute personne est libre, techniquement et juridiquement, d'utiliser, de copier, d'étudier et de modifier, ainsi que d'en rediffuser des versions modifiées[270]. Exemple : Firefox, OpenOffice, Thunderbird et PDFCreator sont des logiciels libres.
- ***Logiciel propriétaire*** : logiciel ne présentant pas tout ou partie des quatre libertés caractéristiques du logiciel libre. Exemple : Microsoft Word, Adobe Acrobat Reader et Adbe Photoshop sont des logiciels propriétaire.
- ***Logiciel en tant que service*** (*software as a service* – SaaS) : logiciel qui n'est pas installé sur les machines de l'utilisateur mais sur un serveur externe et auquel le client recourt dans le cadre d'un contrat de prestation de service (et non d'acquisition d'un logiciel) – V. aussi l'entrée « *Cloud computing* »

[269] Arrêté du 22 décembre 1981, « enrichissement du vocabulaire de l'informatique ».

[270] V. sur la notion de logiciel libre l'article de J. Crémer et A. Gaudeul,« Quelques éléments d'économie du logiciel libre », *Réseaux*, vol 124, n° 2, 2004, p. 111-139

Matrice de confusion

Outil utilisé en apprentissage supervisé pour synthétiser les performances du modèle algorithmique du point de vue de l'exactitude.

		RESULTAT ATTENDU (VERITE)	
		POSITIF (Ex. : beau temps)	**NEGATIF** (Ex. : mauvais temps)
RESULTAT PRODUIT PAR LA MACHINE	**POSITIF** (Ex. : Il va faire beau demain)	VRAI POSITIF Le système ne s'est pas trompé en annonçant qu'il allait faire beau	FAUX POSITIF Le système a prédit du beau temps, alors qu'il pleut
	NEGATIF (Ex. : Il ne va pas faire beau demain)	FAUX NEGATIF Le système a prédit qu'il ne ferait pas beau, alors que le soleil brille	VRAI NEGATIF Le système ne s'est pas trompé en annonçant qu'il ne ferait pas beau

Méthodes ensemblistes

Ensemble de techniques consistant à entraîner plusieurs modèles à partir du même type d'algorithme et à confronter et « fusionner » leurs résultats afin d'améliorer la stabilité et la précision du modèle, en réduisant le « bruit » et les biais. A titre de comparaison, il est plus sûr de demander à cent météorologistes, plus qu'à un seul, quel temps il fera demain, et de retenir la position majoritaire. Parmi ces méthodes, on trouve notamment le « *bagging* » et le « *boosting* ».

Méthodes d'exploration et d'optimisation (*search and optimization methods*)

Technique consistant à déterminer la meilleure façon d'atteindre un objectif en explorant le champ des possibles, soit de façon complète, soit en utilisant des informations externes permettant d'explorer prioritairement les voies les plus prometteuses (V. aussi ci-dessus « algorithme heuristique »)

Partitionnement de données

Méthode consistant à diviser un ensemble de données en différents paquets homogènes, partagent des caractéristiques communes, qui correspondent le plus souvent à des critères de proximité (dite similarité informatique) que l'on définit en introduisant des mesures et classes de distance entre objets.

Prise de décision assistée

Décision prise par un humain avec l'aide d'un système d'IA

Prise de décision automatisée

Décision prise par un système d'IA

Profilage

Toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique (article 4 du RGPD).

Réseaux de neurones

Modèle algorithmique inspiré du fonctionnement des neurones biologiques. Composé *a minima* d'une couche d'entrée accueillant des données brutes, reliée à une couche cachée qui traite ces données, elle-même reliée à une couche de sortie qui produit un résultat. Si le réseau comporte plus d'une seule couche cachée, il s'agit d'un apprentissage profond. Les réseaux de neurones qui interagissent entre eux travaillent sur plusieurs tâches et chaque couche se verra assigner une tâche. Exemple : Dans le cas de la reconnaissance faciale, chaque couche réalise une tâche d'abstraction (l'identification des formes, l'identification des tailles etc.)

RPA (*Robotic process automation* – Automatisation robotisée des processus)

Ensemble de technologies ayant pour but de remplacer l'exécution humaine de tâches normées, souvent répétitives, par son exécution par un logiciel (« (ro)bot »).

Surentraînement (ou surajustement) :

V. ci-dessus « Compromis biais-variance »

Système d'IA

Selon la définition du projet de règlement européen en cours de discussion, un système d'IA est un « *système qui : / (i) reçoit des données et des entrées provenant de machines et/ou de l'homme, / (ii) déduit comment atteindre un ensemble donné d'objectifs définis par l'homme en utilisant l'apprentissage, le raisonnement ou la modélisation mis en œuvre avec les techniques et les approches énumérées à l'annexe I, et / (iii) génère des sorties sous forme de contenus (systèmes d'IA générative), prédictions, recommandations ou décisions, qui influencent les environnements avec lesquels il interagit* ».

Système d'IA public

Système d'IA conçu et/ou utilisé par une personne publique ou privée dans le cadre de la mission de service public dont elle est chargée.

Système-experts

Systèmes d'IA capables de simuler le savoir-faire d'un expert humain dans un domaine précis, en exploitant des bases de données qui sont spécifiques à ce domaine et en recourant à un moteur d'inférence pour simuler différents raisonnements déductifs. Ces systèmes experts sont utilisés par exemple dans les logiciels de gestion de contrat traditionnels[271] ou encore pour des activités de *credit scoring*.

Traitement automatique des langues / du langage naturel

Fonctionnalité de SIA qui analyse et traite le texte, et d'en générer. Le traitement automatique des langues est à la base des outils de traduction. Exemple : logiciels de traduction (Google Traduction, DeepL, applications d'apprentissage des langues, logiciels de reconnaissance vocale etc.)

Vision par ordinateur

Fonctionnalité de SIA qui analyse et traite les images et vidéos, et permet d'en générer. Du point de vue de l'ingénierie, elle cherche à comprendre et à automatiser les tâches que le système visuel humain peut effectuer.

Visualisation de données

Technique consistant à représenter des données sous une forme (graphique, image, infographie, tableau de bord…) permettant de mieux les comprendre, les interpréter et les exploiter.

[271] Florence G'sell, *op. cit.*, p. 26

Annexe 5 : Quelques dates-clés sur l'intelligence artificielle

VIII^e siècle avant l'ère chrétienne : dans la mythologie grecque, certains récits de création de machines ayant des capacités humaines relèvent des automates, comme les servantes d'or d'Héphaïstos, et d'autres s'approchent de la recherche d'une création mécanique ayant les capacités de l'intelligence humaine – comme le gardien de la Crète, Talos, forgé en bronze par le même Héphaïstos, dieu du feu et forgeron céleste. La création d'une intelligence humaine n'est pas envisagée sans son apparence humaine, l'archétype en étant sans doute Pandore.

Vers 330 avant l'ère chrétienne : Aristote dans sa Politique, évoque l'utilisation d'esclaves artificiels.

Fin du III^e siècle avant l'ère chrétienne : premiers automates humanoïdes, par Philon de Byzance.

860 : le calife al-Ma'mûn demande aux frères Banou Moussa un ouvrage, publié à Bagdad à la maison de la sagesse, intitulé le Livre des mécanismes ingénieux, qui décrit notamment un automate programmable.

1206 : Ismail Al-Jazari publie, après 25 ans de travaux, le Livre de la connaissance des procédés mécaniques, qui décrit de nombreux automatismes mais aussi des robots humanoïdes programmables.

XII^e siècle : Raymond Lulle conçoit des « machines à penser », qui visent à automatiser des raisonnements logiques.

Après quelques automates au Moyen-Âge (dont une célèbre horloge offerte à Charlemagne par le calife Haroun Al Raschid), il faut attendre la Renaissance pour que les ancêtres des systèmes d'IA naissent : le Clos Lucé conserve la maquette, conçue d'après des plans de Léonard de Vinci, d'un chariot dont on a récemment émis l'hypothèse qu'il était moins une automobile qu'un véhicule au trajet programmable. On voit également au musée de Berlin un automate chevalier capable de mouvements élémentaires.

Fin du XVI^e siècle : On prête au rabbin Juda Loew Ben Bezalel, dit le Maharal, l'invention du Golem (initialement, dans le Talmud, créature d'argile avant que son âme ne lui soit donnée) créé (selon certaines versions) pour défendre la communauté juive des pogroms ; ses actions sont programmées en lui introduisant un message dans la bouche.

1642-1645 : Blaise Pascal invente la première machine à calculer, la Pascaline, ou « machine d'arithmétique ».

1725 : Les premières machines programmables, par cartes perforées, sont inventées (métier à tisser automatisé de Bouchon et de Falcon).

Le XVIII^e siècle, tout animé de l'esprit scientiste des Lumières, voit fleurir les créations d'automates (le fameux canard de Jacques de Vaucanson, exposé en 1744 au Palais Royal) qui pour certains sont des androïdes – Vaucanson crée également

un joueur de tambourin, de galoubet, de flûte. A Vienne, Friedrich von Knauss crée un écrivain automate, mécanisme perfectionné par Pierre Jacquet Droz, qui crée en 1770 l'écrivain (répertoire de phrases programmables de 40 caractères), puis le dessinateur, et enfin la musicienne (5 morceaux, avec simulations de mouvement de tête et de la respiration.

1770 : Le hongrois Wolfgang von Kempelen invente le Turc mécanique, joueur d'échecs animé par des rouages, qui joue, notamment contre Napoléon et Benjamin Franklin, et gagne toutes ses parties. En réalité, un homme est caché derrière les engrenages et anime la marionnette.

1818 : Marie Shelley publie *Frankenstein, ou le Prométhée moderne*.

1836 : Charles Babbage conçoit, à la suite de la « machine à différences », la première machine analytique, ancêtre de l'ordinateur programmable.

1843 : Ada Lovelace élabore le premier algorithme pour la machine analytique de Babbage, pour calculer les nombres de Bernoulli. Il est considéré comme le premier programme informatique, faisant de Lovelace la première « développeuse ».

1910 : Premier ordinateur prédictif (mécanique), la machine (surnommée *old brass brains*) à prédire la hauteur et l'horaire des marées en fonction du lieu, mise en œuvre par le service côtier et géodésique américain. Un ordinateur électronique la remplace en 1965.

1915 : Hammond et Miesner créent le premier robot avec capteurs (un « chien », en réalité une boîte à roulettes sensible aux sources lumineuses). En 1928, le Français Henri Piraux le perfectionne. A l'exposition universelle de 1939 à New York, sont présentés Elektro, un robot humanoïde de deux mètres qui a un vocabulaire de 700 mots, et son chien Sparko, qui obéit à quelques ordres donnés par le premier.

1920 : Le Tchécoslovaque Karel Capek fait créer sa pièce de théâtre R.U.R. (*Rossumovi Universalni Roboti*) dans laquelle apparaît, pour la première fois, le mot « robot » (travail, ou servage, en tchèque).

1941 : « La Bombe », ou Enigma, qui désigne l'ensemble du système de « bombes », premier ordinateurs (conçus notamment par Turing) destinés à décrypter le code secret allemand utilisant des machines à rotor à positionnement variable.

1944 : Premier ordinateur à cartes perforées IBM, le Harvard Mark, directement inspiré des métiers à tisser de Bouchon et Falcon. Il est programmé par un petit groupe de trois personnes, dont Grace Hopper, considérée comme l'une des premières femmes programmeuses.

1948 : Norbert Wiener publie Cybernetics or Control and Communication in the Animal and the Machine, ouvrage dans lequel il définit la cybernétique.

1950 : Publication dans la revue « Mind » (LIX, 236, pp 433-460) de l'article fondateur d'Alan Turing « *I, computing machines and intelligence* », où il pose la question : « les machines peuvent-elles penser ? ». Il y propose le test de Turing (comment évaluer une intelligence artificielle, dans un dialogue entre un humain d'une part, et un système d'IA et un humain, tous deux non visibles du premier, d'autre part),

l'incapacité à discerner si les réponses viennent du système ou de l'humain établissant que le système a atteint le niveau humain). Il prédit qu'en l'an 2000, le perfectionnement des machines sera tel que plus de 30% des experts seront trompés après un dialogue de 5 minutes avec une machine.

1954 : Norbert Wiener publie The human use of Human beings, ouvrage fondateur de la critique du recours aux systèmes d'IA et analysant leurs aspects éthiques et sociologiques.

1955 : Isaac Asimov publie Franchise, nouvelle où apparaît une intelligence artificielle générale, Multivac.

1955 : Allen Newell et Herbet Simon conçoivent le « théoricien logique », qui démontre plusieurs théorèmes mathématiques. Il est considéré comme le premier programme d'intelligence artificielle.

1956 : Conférence d'été de Dartmouth, séminaire de recherche de six semaines dirigé par le mathématicien John McCarthy, qui définit un programme de recherche définissant l'intelligence artificielle – première utilisation du concept – comme le champ tendant à simuler les performances cognitives et les fonctions du cerveau humain.

1958 : Première démonstration par Frank Rosenblatt, au laboratoire de recherche navale (Etats Unis), du Perceptron, premier système d'apprentissage machine supervisé.

1961 : Premier robot industriel (manutention et soudure).

1967 : Le laboratoire d'intelligence artificielle de Stanford crée Shakey, robot mobile contrôlé par ordinateur (capable de transporter seul un objet d'une pièce à l'autre, après une heure de calculs).

1969 : Marvin Minsky et Symour Papert publient Perceptrons, *An Introduction to Computational Geometry*. Le premier « hiver de l'IA » commence.

Années 1970 : Les systèmes experts connaissent un succès grandissant.

1972 : Première chaîne de production entièrement automatisée (Nissan ; Renault, 1976).

1973 : L'université de Waseda au Japon crée Wabot, robot humanoïde qui marche et soutient une conversation élémentaire.

1982 : Le réseau de neurones de John Hopfield participe à la renaissance de l'intérêt pour l'IA connexionniste.

Un nouvel « hiver de l'IA » commence à la fin des années 1980. Les programmes de la *Defense Advanced Research Projects Agency* (DARPA) et du *National Research Council* américains sont arrêtés, en raison du manque de résultats rapides et à la hauteur des promesses affichées. En 1991, le gouvernement japonais arrête son programme, lancé en 1981, pour des ordinateurs de 5e génération, dont l'objectif était de pouvoir mener une conversation.

1997 : Deep Blue, crée en 1990 par IBM et capable de calculer 200 millions de coups par seconde, est le premier système informatique à battre Garry Kasparov, champion du monde d'échecs.

2005 : Un véhicule « autonome », élaboré par des chercheurs de Stanford, parcourt plus de 210 kilomètres lors du *DARPA Grand Challenge*.

2010 : Google commence à développer la Google car, véhicule autonome (2017, service expérimental de taxis à Phoenix, Arizona).

2011 : Lancement de *Google Brain*, projet de recherche d'apprentissage profond.

2011 : La machine Watson d'IBM triomphe des champions télévisuels du jeu Jeopardy (il s'agit à nouveau seulement d'un calculateur).

2011 : Apple Siri est le premier assistant personnel intelligent installé sur un téléphone. Tout au long des années 2010, l'IA permet aux téléphones de devenir des « *smartphones* ».

2012 : Lors de la conférence *European conference on computer vision*, réunissant les meilleurs chercheurs en vision par ordinateur, le modèle de reconnaissance d'images, développé par Geoffrey Hinton, Alex Krizhevsky et Ilya Sutskever à l'aide de réseaux de neurones, affiche un taux d'erreur de l'ordre de 17%, très loin devant les autres modèles (environ 27%).

2014 : Deux professeurs de l'université de Reading prétendent avoir trompé 10 des 30 interrogateurs ayant parlé 5 minutes avec un jeune ukrainien de 13 ans, Eugene Goortmans – qui est en réalité un système d'IA.

2016 : Premier accident mortel d'une voiture autonome (le pilote de test ne reprend pas le contrôle et le véhicule percute un camion qui lui coupe la route).

2016 : AlphaGo, de l'entreprise DeepMind, (qui repose sur l'apprentissage automatique, notamment en jouant des parties contre lui-même) bat le champion du monde du jeu de go, Lee Sedol, après trois jours d'apprentissage.

2016 : Microsoft lance *Tay*, un chatbot auquel on peut s'adresser librement sur Twitter. Le chatbot est arrêté après 16 heures de fonctionnement où il n'a cessé de déverser des injures racistes, et ne sera pas remis en service.

L'année 2016 marque également un tournant dans l'intérêt que portent les responsables politiques au développement de l'IA (Barack Obama y consacre un long entretien au magazine *Wired*), ainsi que son irruption dans les campagnes électorales. Lors des élections étasuniennes de 2016, les *deepfakes*, les faux comptes créés automatiquement sur les réseaux sociaux (dits *bots*) ou encore le scandale *Cambridge Analytica* révèlent l'influence que peuvent avoir les SIA sur les processus démocratiques.

2018 : Le premier tableau peint par un système d'IA est vendu aux enchères pour 432 000 dollars.

2019 : AlphaGo Zero, machine fondée sur l'apprentissage par renforcement, bat AlphaGo après quatre heures d'apprentissage.

Annexe 6 : Synthèse du projet de règlement européen sur l'IA

La Commission européenne a présenté le 21 avril 2021 une proposition législative présentée comme le premier cadre juridique sur l'intelligence artificielle dans le monde[272]. La présidence slovène a soumis au Conseil un « texte de compromis » en décembre 2021, dont le contenu est synthétisé ci-après.

I - Objectifs principaux et logique générale

Le règlement a pour but d'assurer le bon fonctionnement du marché intérieur des systèmes d'IA et de garantir des standards élevés en matière d'éthique et de respect des droits fondamentaux. Il s'agit, plus précisément :

- d'assurer la sûreté des systèmes d'IA mis sur le marché de l'Union européenne et le respect des droits fondamentaux ;
- de garantir la sécurité juridique pour faciliter l'innovation ;
- de faciliter le développement d'un marché unique pour des applications d'IA légales, sûres et dignes de confiance.

La proposition repose sur :

- **un champ d'application large**, en raison d'une définition englobante des systèmes d'intelligence artificielle, de l'inclusion de l'ensemble des secteurs d'activité, publics et privés, et de l'étroitesse des exclusions ;
- **une approche produit**, encadrant la fabrication et la mise sur le marché (ou la mise en service, lorsque le produit n'est pas destiné à être commercialisé) des systèmes ;
- **une approche par les risques**, qui conduit à proscrire les SIA présentant un risque inacceptable, à imposer des contraintes exigeantes sur les « systèmes d'IA à haut risque » (risque élevé) et à ne soumettre les autres systèmes (qui ne présentent aucun risque ou un risque faible) qu'à des obligations d'information en incitant à l'adoption de codes de conduite.

II - Champ d'application

Champ d'application matériel

Le système d'IA est défini comme un système :

- qui reçoit de la part d'opérateurs humains ou de machines des données et des entrées ;
- qui déduit (infère) la façon d'atteindre un ensemble d'objectif définis par l'humain en utilisant l'apprentissage, le raisonnement et la modélisation mis en œuvre par les techniques et approches énumérées à l'annexe I (apprentissage

[272] Proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union, COM(2021) 206 final, 2021/0106 (COD).

automatique, y compris apprentissage profond ; approches fondées sur la logique et les connaissances comme la représentation des connaissances, la programmation inductive, les bases de connaissances, les moteurs d'inférence et de déduction, le raisonnement (symbolique) et les systèmes experts ; approches statistiques, estimation bayésienne, méthodes de recherche et d'optimisation…) ;

- et de produire des résultats sous la forme de contenus (systèmes d'IA génératifs), des prédictions, des recommandations ou des décisions, qui influencent les environnements avec lesquels il interagit ;

Les exclusions à raison de l'objet portent :

- en premier lieu sur les systèmes d'IA développés ou utilisés exclusivement à des fins militaires ou de sécurité nationale ;
- en second lieu, sur les systèmes d'IA, y compris leurs résultats, qui sont spécifiquement développés et mis en service pour les seuls besoins de la recherche scientifique et du développement. Il est en outre précisé que le règlement ne doit pas affecter les activités de recherche et développement dans la mesure où celles-ci ne conduisent ni n'impliquent de mettre sur le marché ou en service un système d'IA.

En outre, le règlement ne s'appliquerait pas aux systèmes d'IA à usage général ou « à des fins générales », c'est-à-dire des SIA qui sont capables d'assurer des fonctions générales (reconnaissance d'images ou de paroles, création de son ou de vidéo, détection de formes, fourniture de réponses à des questions, traduction…), sans destination précise, et qui seront utilisés pour concevoir des SIA ayant une destination spécifique et, comme tels, soumis en principe au règlement.

Champ d'application organique

Le règlement a vocation à s'appliquer :

- aux **fournisseurs** (publics et privés) qui mettent sur le marché ou en service des SIA : le fournisseur est l'opérateur qui développe ou fait développer le SIA et qui assume la responsabilité de cette mise sur le marché ou en service ;
- et aux **utilisateurs** (publics et privés) de ces systèmes.

Champ d'application géographique

Le règlement s'appliquerait dès lors que le SIA est mis sur le marché ou en service dans l'Union, quel que soit le lieu d'établissement du fournisseur, ou utilisé par un utilisateur présent ou établi dans l'Union.

En outre, les fournisseurs et utilisateurs établis en dehors de l'Union y seraient également soumis lorsque les résultats générés par le système sont utilisés dans l'Union, ce qui imposera une forme de traçabilité des données produites par un SIA, d'autant plus exigeante que la définition du SIA est large.

III - Systèmes d'IA interdits

La proposition prévoit d'interdire quatre pratiques particulièrement sensibles, qu'elles soient ou non intentionnelles :

- manipulation mentale (techniques subliminales d'influence au-dessous du seuil de conscience d'une personne ayant pour objet ou pour effet d'altérer substantiellement son comportement, avec un risque raisonnable de préjudice physique ou psychologique) ;
- abus de faiblesse : exploitation des vulnérabilités liées à l'âge, au handicap, à la situation économique ou sociale, avec un risque raisonnable de préjudice physique ou psychologique ;
- crédit social : classement de personnes physiques en fonction de leur comportement social ou de caractéristiques personnelles avec un traitement préjudiciable ou défavorable, lorsque le contexte social d'utilisation est dissocié du contexte d'origine ou qu'il y a disproportion ou absence de justification ;
- identification biométrique en temps réel dans des lieux publics à des fins répressives (par les forces de l'ordre), sauf, sur autorisation judiciaire ou administrative préalable, pour certains usages (recherche ciblée de victimes potentielles spécifiques ; prévention d'une menace spécifique et substantielle pour la sécurité physique, la santé, la vie de personnes ou des infrastructures critiques, ou d'une attaque terroriste ; détection, localisation, identification et poursuite d'une infraction pénale entrant dans le champ du mandat d'arrêt européen (passible d'une peine d'au moins 3 ans de prison et inscrite sur la liste fixée par la décision-cadre du 13 juin 2002). Une exception ne peut jouer qu'après évaluation du rapport bénéfices/risques (nécessité, proportionnalité…).

IV - Régime juridique des SIA à haut risque

Identification

Ces SIA à haut risque relèvent de deux grandes catégories :

1° les systèmes utilisés dans des secteurs industriels énumérés (dispositifs médicaux, jouets, ascenseurs, équipements radioélectriques, équipements de protection individuelle, aviation civile, transports…) et soumis à une évaluation de la conformité par un tiers ;

2° les systèmes identifiés comme étant par nature à haut risque, énumérés à l'annexe III de la proposition.

DOMAINE	SYSTÈMES VISÉS
Systèmes biométriques	Identification biométrique des personnes physiques à distance (en temps réel et *a posteriori*) sans leur accord.
Infrastructures critiques et protection de l'environnement	Composants de sécurité dans la gestion et l'exploitation du trafic routier et dans la fourniture d'eau, de gaz, de chauffage et d'électricité Composants de sécurité ou de contrôle d'infrastructures digitales Systèmes de contrôle des émissions et de la pollution
Education et formation professionnelle	Accès, admission et affectation dans les établissements d'enseignement ou de formation à tous niveaux Evaluation des résultats des personnes physiques et pilotage des programmes d'enseignement et de formation dans les établissements à tous niveaux
Emploi, gestion de la main d'œuvre et accès à l'emploi indépendant	Recrutement et sélection de personnes physiques (diffusion des offres d'emploi, pré-sélection, filtrage) Promotion, licenciement, attribution de tâches basée sur le comportements ou les caractéristiques personnelles de chacun, suivi et évaluation des performances et comportement des travailleurs sous contrat
Accès et droit aux services privés, aux services publics et aux prestations sociales	Gestion des prestations et services d'aide sociale (éligibilité, octroi, retrait, récupération…) Evaluation de la solvabilité des personnes physiques ou établissement d'une note de crédit (hors « petits fournisseurs » pour leurs besoins propres) Gestion des appels d'urgence (priorisation des interventions des services de secours, pompiers…) Fixation des primes d'assurance, gestion des souscriptions et réclamations

Systèmes d'aide aux autorités répressives, ou à une autre autorité agissant en leur nom	Evaluation de la probabilité de commission d'une infraction ou de récidive Prédiction de la survenance ou de la réitération d'infraction sur la base du profilage individuel ou de l'évaluation des traits de personnalité, des caractéristiques et des antécédents judiciaires Evaluation du risque encouru par les victimes potentielles d'infractions pénales Détection des mensonges, des émotions et des hypertrucages (deepfakes) Evaluation de la fiabilité des preuves
Gestion de la migration, de l'asile et des contrôles aux frontières, par les autorités compétentes ou en leur nom	Détection de mensonges et d'émotions ; vérification de l'authenticité des documents de voyage et pièces justificatives Evaluation du risque sécuritaire, sanitaire ou migratoire que représente un étranger Vérification d'éligibilité des demandeurs (d'asile, de visa, de titre de séjour) à un statut
Administration de la justice et processus démocratiques	Utilisation par le juge (ou pour son compte) pour l'interprétation des faits et de la loi, application de la loi à un ensemble concret de faits

Au terme d'un réexamen annuel, la Commission serait autorisée à compléter cette liste, dans les domaines limitativement énumérés par cette annexe, pour des usages présentant des risques équivalents ou supérieurs pour la santé, la sécurité, ou l'atteinte aux droits fondamentaux.

Régime des SIA à haut risque et obligations des fournisseurs

La proposition de règlement prévoit de subordonner la mise sur le marché ou en service des SIA à haut risque au respect d'un ensemble d'obligations destinées à garantir leur fiabilité et leur innocuité. La portée de ces obligations sera précisée, le cas échéant, par des normes harmonisées à définir (ou, à défaut, des spécifications communes adoptées par la Commission) :

- **Mise en place d'un système de gestion des risques** : identification et analyse des risques connus et prévisibles ; estimation et évaluation des risques potentiels en conditions normales et anormales d'utilisation ; adoption de mesures appropriées de gestion des risques, sur la base de tests ;
- **Utilisation de données d'apprentissage, de validation et de test respectant des standards de qualité** portant notamment sur les choix de conception, la collecte, les traitements de préparation des données, la formulation d'hypothèses pertinentes, l'évaluation préalable de la disponibilité, de la quantité et de l'adéquation des jeux de données nécessaires, le repérage des biais et la détection et la résolution des lacunes ou déficiences dans les données. Les jeux

de données doivent être pertinents, représentatifs, exempts d'erreurs et complets, et ils doivent tenir compte du contexte (géographique, comportemental, fonctionnel…) d'utilisation du SIA (cette dernière exigence étant présumée respectée si les données d'entraînement et de validation relèvent elles-mêmes de ce contexte) ;

- Etablissement d'une **documentation technique** préalablement à la mise en service et tenue à jour ;
- **Enregistrement automatique des évènements** (journaux) permettant d'assurer la traçabilité du fonctionnement (en particulier pour l'identification biométrique) et d'en assurer la surveillance après commercialisation ;
- **Transparence suffisante des SIA à haut risque** à l'égard des administrations qui les utilisent, leur permettant de connaître la destination du système (incluant le domaine géographique, comportemental ou fonctionnel dans lequel il a vocation à être utilisé), d'interpréter les résultats et de les utiliser de manière appropriée (notice d'utilisation accessible…) ;
- **Contrôle humain :** au stade de la conception et/ou de l'utilisation, des personnes physiques doivent exercer un contrôle effectif permettant d'appréhender totalement les capacités et limites du système, de détecter et traiter les anomalies, d'interpréter correctement les résultats, de prévenir le biais d'automatisation (tendance à se fier systématiquement ou excessivement aux résultats produits par le système), d'ignorer ou d'inverser le résultat issu du système, et d'arrêter rapidement le système (« bouton stop »). Pour l'identification biométrique, la validation par au moins deux personnes physiques est requise ;
- Respect d'exigences d'**exactitude** (évaluée selon des critères précisés dans la notice), de **robustesse** et de **cybersécurité** (résilience en cas d'erreurs, de défaillances ou d'incohérences, mise en place de plans de sauvegarde et autres mesures de sécurité, protection contre les intrusions, l'empoisonnement des données, les exemples adverses et l'exploitation des défauts du modèle). Lorsque les SIA continuent leur apprentissage après la mise en service, les biais liés aux boucles de rétroaction (utilisation des données de sortie comme données d'entrée) doivent faire l'objet d'un « *traitement adéquat* » par des « *mesures d'atténuation appropriées* ».

Ces obligations pèsent sur le fournisseur du système. Ce dernier doit mettre en place un **système de gestion de la qualité,** comprenant des politiques, des procédures et des instructions écrites, afin de garantir la conformité du système au règlement, incluant notamment la description de la politique de la donnée, le système de gestion des risques, la procédure de notification d'incidents, les processus d'échanges d'informations avec les autorité de contrôle et « *un cadre de responsabilisation définissant les responsabilités de l'encadrement et des autres membres du personnel* ». **Ce système doit être proportionné à la taille de l'organisation du fournisseur**.

Avant la mise en service du SIA à haut risque, le fournisseur doit se soumettre à une **procédure d'évaluation de la conformité** du système aux obligations prévues par le

règlement. Pour les SIA à haut risque relevant de l'annexe III (ce qui représente la quasi-totalité des SIA à haut risque mis en œuvre par l'administration), l'évaluation de la conformité est réalisée dans le cadre d'une **procédure de contrôle interne**, sans intervention d'un organisme notifié. Le fournisseur vérifie que le système de gestion de la qualité est conforme aux exigences de l'article 17, il examine les informations contenues dans la documentation technique pour apprécier la conformité, et il s'assure que le processus de conception et de développement du SIA et son système de surveillance après commercialisation sont cohérents avec cette documentation. Pour les SIA aux fins d'identification biométrique, les fournisseurs peuvent, alternativement à la procédure de contrôle interne, opter pour une procédure particulière d'évaluation de la conformité, prévue à l'annexe VII, associant un organisme notifié (ce rôle étant joué par l'autorité de surveillance du marché en matière répressive et d'immigration). L'autorité de surveillance du marché peut autoriser des dérogations à la procédure d'évaluation de la conformité en cas d'urgence sécuritaire, sanitaire, environnementale…

Obligations de l'utilisateur

L'**utilisateur à titre professionnel** d'un SIA à haut risque est tenu d'utiliser le système et d'en surveiller le fonctionnement conformément à la notice d'utilisation établie par le fournisseur. Il doit veiller à la pertinence des données d'entrée et à la tenue des journaux d'opérations, s'il est prévu qu'il exerce un contrôle sur ces points. Les informations contenues dans la notice d'utilisation sont prises en compte dans l'analyse d'impact sur la protection des données à caractère personnel que le responsable de traitement doit réaliser en vertu du RGPD ou de la directive police-justice, le cas échéant.

V. – Obligations de transparence applicables à l'ensemble des SIA

Qu'ils soient ou non classés « à haut risque », les SIA sont soumis à des obligations de transparence dans certaines circonstances :

- Lorsqu'un SIA interagit avec une personne physique, celle-ci doit être en mesure de le savoir (sauf en matière répressive) ;
- Lorsqu'une personne physique est exposée à un SIA de reconnaissance des émotions ou de catégorisation biométrique (hors domaine répressif), elle doit être informée de son fonctionnement ;
- Les utilisateurs d'un SIA d'hypertrucage (*deepfake*) doivent préciser que les contenus produits ont été générés ou manipulés artificiellement, en-dehors du domaine répressif, du droit de création artistique, de la liberté scientifique…

VI - Codes de conduite

Sous réserve des obligations de transparence qui viennent d'être mentionnées, les systèmes d'IA qui ne sont pas classés à haut risque ne sont soumis à aucune obligation au titre du règlement. Ce dernier se borne à encourager l'élaboration de codes de conduite en vue de l'application volontaire des règles propres aux SIA à haut risque.

En outre, les codes de conduite ont vocation à prendre en compte d'autres exigences : la viabilité environnementale, l'accessibilité aux personnes handicapées, la participation des parties prenantes à la conception et au développement des SIA, la diversité des équipes de développement…

VII - Articulation avec les règles relatives au traitement des données à caractère personnel

Le principe de base, explicité dans l'exposé des motifs, est que le règlement IA s'appliquerait sans préjudice du RGPD et de la directive police-justice, c'est-à-dire que les deux corps de règles s'appliqueraient cumulativement aux fournisseurs et utilisateurs dès lors que les SIA peuvent être qualifiés de traitements de données à caractère personnel. Le considérant 41 précise que la conformité au règlement IA n'emportera pas conformité au RGPD.

Deux dispositions méritent toutefois d'être mentionnées.

D'une part, il est prévu que les opérations de « débiaisement » des SIA peuvent justifier le traitement de données sensibles, sous réserve d'apporter des garanties appropriées (limitations techniques à la réutilisation, pseudonymisation et cryptage…).

D'autre part, la Commission propose de créer un cadre sécurisé pour le développement, la mise à l'essai et la validation de SIA innovants avant mise en service, dénommé « **bacs à sable réglementaires** », qui permettraient de traiter à cette fin des données à caractère personnel collectées à d'autres fins. La proposition semble ainsi pouvoir s'interpréter comme autorisant le principe d'un tel traitement ultérieur au titre du RGPD dès lors que les conditions posées par le règlement IA sont remplies, et alors même que celles qui sont prévues par le RGDP (article 6, paragraphe 4) ne le seraient pas.

L'éligibilité du recours à ce dispositif est subordonnée à deux conditions :

- d'une part, ne seraient éligibles à ces bacs à sable que les systèmes d'IA poursuivant un intérêt public important en matière de prévention et détection des infractions pénales et des menaces pour la sécurité publique, de sécurité publique et de santé publique, et de protection de l'environnement ;
- d'autre part, il doit s'agir de satisfaire aux exigences posées pour les systèmes d'IA à haut risque, sans qu'il soit possible d'anonymiser les données ou d'utiliser des données synthétiques ou à caractère non personnel (logique de subsidiarité).

La proposition de règlement encadre les modalités d'utilisation des bacs à sable réglementaires :

- en premier lieu, doivent être mis en place un suivi des risques pour les droits fondamentaux, un mécanisme d'atténuation ou d'arrêt, une séparation logique, un isolement et une protection de l'environnement, avec des accès réservés et contrôlés (et interdiction de transférer les données à un tiers ou de le laisser consulter les données) ;

- en deuxième lieu, les traitements réalisés dans ce cadre ne doivent pas déboucher sur des décisions affectant les personnes concernées ;
- en troisième lieu, les données doivent être supprimées au plus tard à l'issue de l'utilisation et les journaux conservés jusqu'à un an après la fin de la participation, de même que la description complète et détaillée du processus et de la justification de l'entraînement, des tests et de la validation du système d'IA ;
- en quatrième lieu, un résumé succinct du projet d'IA développé doit être rendu public ;
- en cinquième et dernier lieu, le fonctionnement du bac à sable doit associer les autorités nationales compétentes, lesquelles conservent leurs pouvoirs de sanction, et les modalités de fonctionnement (critères d'admissibilité, procédure de demande, sélection, participation et sortie, droits et obligations des participants…) seront fixées par des actes d'exécution.

VIII – Sanctions

Sans préjudice des sanctions (notamment pénales) mises en place par chaque État membre, le règlement impose la mise en place d'un régime d'amendes administratives :

Manquement	Plafond de la sanction
Conception ou utilisation d'un SIA interdit (cf. III)	30 M€ ou, si ce plafond est plus élevé, 6% du chiffre d'affaires mondial de l'entreprise
Non-conformité aux exigences relatives aux données s'imposant aux SIA à haut risque (utilisation de jeux de données biaisés, erronés, incomplets…)	
Non-conformité d'un SIA à haut risque aux autres exigences	20 M€ ou, si ce plafond est plus élevé, 4% du chiffre d'affaires mondial de l'entreprise
Fourniture d'informations inexactes, incomplètes ou trompeuses aux organismes notifiés ou aux autorités nationales compétentes (entrave…)	10 M€ ou, si ce plafond est plus élevé, 2% du chiffre d'affaires mondial de l'entreprise

La proposition prévoit la création d'un **comité européen de l'intelligence artificielle** réunissant les autorités de contrôle nationales et le Contrôleur européen de la protection des données.

Il comporte en outre des règles concernant la désignation des **autorités nationales compétentes** pour assurer la mise en œuvre du règlement. La notion d'« autorité nationale compétente » renvoie à trois autorités :

- **l'autorité de contrôle nationale** : elle est définie comme l'autorité chargée de la mise en œuvre et de l'application du règlement, de la coordination des activités de l'État membre, du rôle de point de contact unique pour la Commission (elle doit notamment assurer la synthèse et la transmission à la Commission des résultats des activités de surveillance du marché) et de la représentation de l'État membre au sein du comité européen de l'IA ;
- **l'autorité notifiante** : il s'agit de l'autorité qui accrédite et évalue les organismes notifiés, chargés de procéder à l'évaluation de la conformité des SIA à haut risque avec le règlement ;
- **l'autorité de surveillance du marché** : il s'agit de l'autorité chargée du bon fonctionnement du marché des SIA et qui exerce les compétences dévolues aux autorités de surveillance du marché par le règlement (UE) 2019/1020. Le règlement lui confie notamment le soin d'accorder les dérogations à la procédure d'évaluation de la conformité.

S'agissant des SAI industriels et financiers, la proposition renvoie aux autorités sectorielles de surveillance du marché lorsqu'elles existent[273]. S'agissant des SIA utilisés à des fins répressives ou d'immigration (y compris l'asile), l'autorité de surveillance du marché est soit l'autorité de contrôle en matière de protection des données à caractère personnel (CNIL), soit les autorités nationales compétentes pour surveiller ces activités.

S'agissant des autres SIA, le règlement prévoit seulement l'obligation de désigner (au moins) une autorité de surveillance du marché, en rendant applicable le règlement (UE) 2019/1020 aux SIA en général. Il appartient ainsi aux Etats membres de désigner une telle autorité (ou plusieurs) pour les SIA qui ne se rattachent pas déjà à l'une des 70 législations d'harmonisation énumérées en annexe à ce règlement 2019/1020.

La proposition pose le principe selon lequel la même autorité assume ces trois rôles (contrôle, notification et surveillance du marché). Mais elle permet aux Etats de désigner des autorités différentes pour des raisons organisationnelles et administratives portées à la connaissance de la Commission – ce qui est inévitable

[273] En France, la principale autorité de surveillance du marché est la DGCCRF (et la DGDDI pour les importations). Il existe des autorités de surveillance sectorielles : ANSM pour les dispositifs médicaux et les produits cosmétiques, DGT pour les équipements de protection individuelle et les machines, DGPR du ministère de la transition écologique pour les équipements à pression ou les feux d'artifice, DGALN du même ministère pour les ascenseurs, DGAC pour les drones, ANFR pour les équipements radio infrarouges…

dès l'instant que la proposition elle-même prévoit la désignation de plusieurs autorités de surveillance du marché selon la destination des SIA. La proposition ne désigne pas l'autorité compétente pour infliger les sanctions qu'elle prévoit. Pour les SIA mis en œuvre par l'Union, ce rôle est dévolu au Contrôleur européen de la protection des données.

La désignation des autorités devrait intervenir dans un délai de trois mois à compter de l'entrée en vigueur du règlement.

En l'état, le règlement ne prévoit pas la désignation, par les responsables des systèmes, de délégués de type « délégué à la protection des données ».

X - Dispositions transitoires

En l'état de la proposition, les obligations découlant du règlement s'appliqueraient :

- aux SIA mis en service à compter de l'entrée en application du règlement, soit deux ans après son entrée en vigueur (laquelle interviendrait elle-même le 20ème jour suivant sa publication) ;
- aux SIA mis en service avant cette date mais qui, à compter de celle-ci, subissent d'importantes modifications de leur conception ou de leur destination.

Annexe 7 : Méthodologie et bilan du questionnaire adressé aux administrations

Une enquête par questionnaire a été menée auprès des administrations publiques, à laquelle près d'une cinquantaine d'entre elles ont répondu (le contenu du questionnaire et la liste des répondantes sont indiqués à la fin de cette annexe). Elle visait à, d'une part, interroger le rapport des administrations (centrales, autorités indépendantes, locales, sanitaires et sociales) à l'intelligence artificielle et, d'autre part, à dresser un panorama des projets de SIA envisagés, développés, achetés, ou abandonnés par celles-ci, et à mieux connaître les opportunités et difficultés rencontrées dans ces déploiements.

Réalisée sur la base du volontariat, l'enquête rend impossible toute velléité généralisatrice. En effet, il fait peu de doute que les administrations ayant accepté de répondre au questionnaire sont celles où une culture, une réflexion et une adhésion à l'IA sont déjà présentes, ne serait-ce qu'à un stade embryonnaire. Ce biais explique sans doute le premier résultat du questionnaire selon lequel plus de 80% des administrations auraient pour projet de recourir à brève échéance à un SIA pour accomplir leurs missions, ainsi que l'adhésion à plus de 90% à l'idée que l'IA peut « accroître sensiblement la performance » de l'entité.

Le compte rendu de ce questionnaire n'a donc de prétention ni à l'exhaustivité, ni à la représentativité. Toutefois, les réponses reçues, complémentaires des entretiens menés, permettent de dresser un panorama des pratiques en matière d'IA de certaines administrations, qui peuvent être considérées comme signifiantes, tant en ce qui concerne les opportunités et les obstacles du recours à des SIA publics qu'en matière d'attentes d'une partie de l'administration vis-à-vis de l'État et du rôle qu'il peut jouer pour appuyer le développement d'une IA d'utilité publique.

1. Le recours à l'IA fait l'objet d'un véritable élan au sein des administrations répondantes, qui en reconnaissent la pertinence pour améliorer la qualité et l'efficacité du service

1.1. Une pluralité d'initiatives centrées sur la mise en œuvre des politiques publiques dont les administrations ont la charge

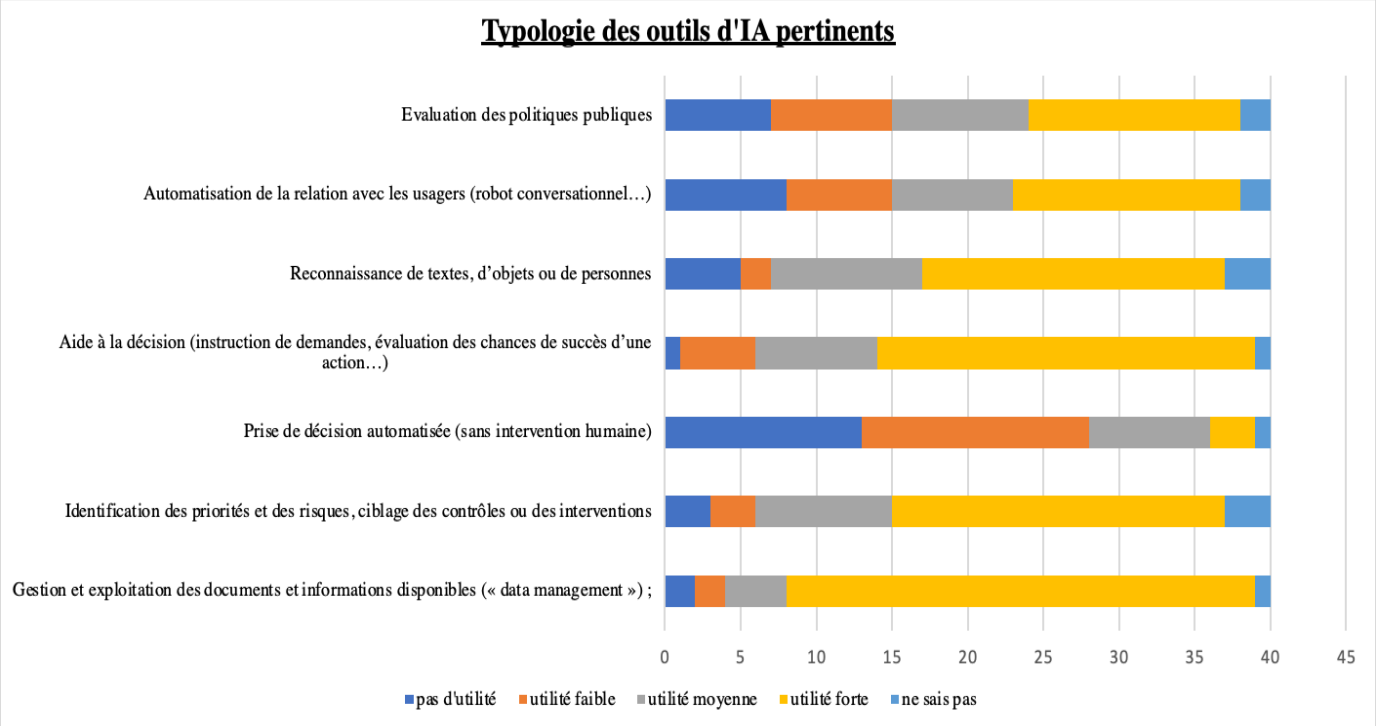
Les différentes administrations répondantes ont très largement témoigné de leur créativité et de leur forte inspiration à mettre à profit la technologie de l'IA pour réinventer une partie de leurs missions. Il ressort en effet des réponses, sans surprise, que **les administrations déploient des SIA pour améliorer l'efficacité des politiques publiques** dont elles ont la charge :

- Les métropoles et les régions développent particulièrement des projets d'IA en lien avec les problématiques de gestion des territoires. C'est le cas par exemple en Guadeloupe où l'IA a d'ores et déjà été mise au service des enjeux d'élagage, aux Antilles où des dispositifs contrôlent l'affluence des sargasses, ou encore pour la métropole de Toulouse travaillant à la cartographie des îlots de chaleur urbains.

- Les administrations centrales quant à elles, abordent l'intégration de l'IA dans leurs outils de mission avec leurs propres problématiques, mais il convient de souligner la grande attention qu'elles portent à l'utilisation de la masse des données qu'elles collectent. A titre d'illustration, le HCERES a mis en avant le fait que l'IA pouvait être un moyen de valoriser les nombreuses données conservées, à titre plus ou moins expérimental.
- Les autorités administratives indépendantes mettent sur pied des projets d'intelligence artificielle en lien avec leurs missions spécifiques. Par exemple, l'Agence française de lutte contre le dopage complète le passeport biologique des athlètes et détecte de nouveaux moyens de dopage grâce à des outils d'IA. Le dénombrement des femmes et hommes à l'écran, réalisé par le CSA/ARCOM, est rendu possible grâce à un outil IA d'identification du genre.

1.2. Une vision partagée des opportunités que représente un recours accru à l'IA

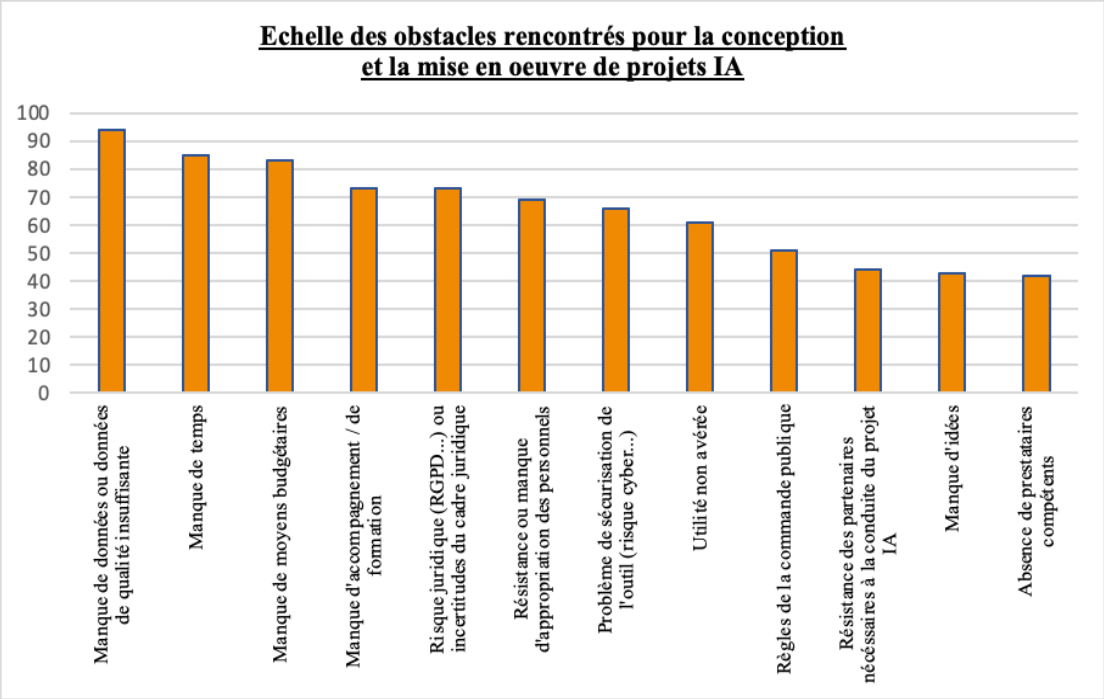
Principalement, l'objectif d'une **amélioration de la qualité et de l'efficacité du service public** – et accessoirement de son image auprès des usagers -, s'associe à l'ambition d'une revalorisation des métiers de la fonction publique pour nourrir un élan d'innovation. Toutefois, l'utilité des outils qui permettent d'automatiser la relation avec les usagers n'est pas plébiscitée.



S'agissant des usages à plus forte valeur ajoutée, les systèmes de gestion, d'exploitation des documents, de classement des priorités ou encore d'aide à la décision sont considérés par la majorité des répondantes comme d'une utilité très forte. Par contraste, si l'importance de ne pas accumuler de retard, et donc

d'anticiper les révolutions technologiques présentes et à venir, est également un des moteurs de l'utilisation progressive de l'IA, surtout auprès des administrations centrales, **le recours à un SIA de prise de décision automatisée sans assistance humaine fait l'objet d'un rejet fort et partagé**, les administrations soulignant à cet égard le risque éthique lié à l'usage de ce type de SIA.

2. Les obstacles à un déploiement accru des SIA publics tiennent à la fois à un déficit de ressources et à une crainte d'une perte de contrôle



2.1. Le manque de ressources en données, en temps ou en moyens financiers, le principal obstacle opérationnel au déploiement de SIA publics

Le premier frein, mentionné par la quasi-totalité des administrations répondantes, est **le manque de données, ou la qualité insuffisante des données mises à disposition**. En effet, plusieurs administrations remarquent qu'en raison de la piètre qualité des données collectées à l'état brut, le temps de nettoyage et de montée en qualité des jeux de données nécessaire est souvent peu compatible avec les exigences administratives de résultat.

Pourtant, ce temps pourrait aisément être mis à profit compte tenu de la nécessité d'associer les différents corps de métiers impactés à cette transformation technologique. En effet, la nécessité d'adopter une culture de l'IA et d'adapter les projets de SIA aux réalités du terrain se fait sentir lorsque les administrations rapportent soit une résistance des personnels à embrasser ce nouvel outil, soit un **défaut d'appropriation des projets par les métiers concernés**. Ainsi les ministères

comme les régions remarquent de concert que le passage de l'expérimentation des projets à une industrialisation de l'usage des outils d'IA paraît encore incertain.

Les règles de la commande publique et, dans une moindre mesure, la **résistance des partenaires potentiels** susceptibles de mettre en place les outils d'IA, sont en effet des obstacles fréquemment relevés par les répondantes. A cette difficulté le souhait, exposé dans la troisième partie de cette annexe, de faciliter la mutualisation des ressources de données et la conclusion de partenariats, notamment avec des acteurs privés.

Il est alors à ce stade important de remarquer que, si le **manque de budget** n'est pas unanimement cité par les administrations comme le principal frein à l'innovation, il tient néanmoins une place de premier plan parmi les obstacles au déploiement de SIA. Si la nécessité d'un budget dédié à la transition technologique est globalement partagée, **plusieurs administrations exposent les stratégies adoptées pour limiter le coût de la transition technologique et même assurer la réalisation d'économies.** Par exemple, la métropole de Toulouse, dont le projet consiste à mettre en place une plateforme de big data en régie (IADATA), supporte ainsi uniquement le coût de l'emploi de six nouvelles personnes affectées à ce projet. Les économies peuvent également être réalisées par un gain d'efficience des services, comme en Guadeloupe grâce à la récupération d'indemnités d'occupation du domaine public par exemple.

Enfin, le manque d'idées n'a pas fait partie des obstacles principaux à la mise en place des outils d'IA au sein des administrations. Mais le **besoin d'un plus grand partage et d'une meilleure communication interne sur les initiatives et les outils mis en place** a néanmoins été largement souligné, nombre d'administrations estimant leurs systèmes d'IA reproductibles par d'autres.

2.2. Les obstacles liés aux risques du recours à l'IA tiennent essentiellement au cadre juridique et à l'explicabilité des algorithmes

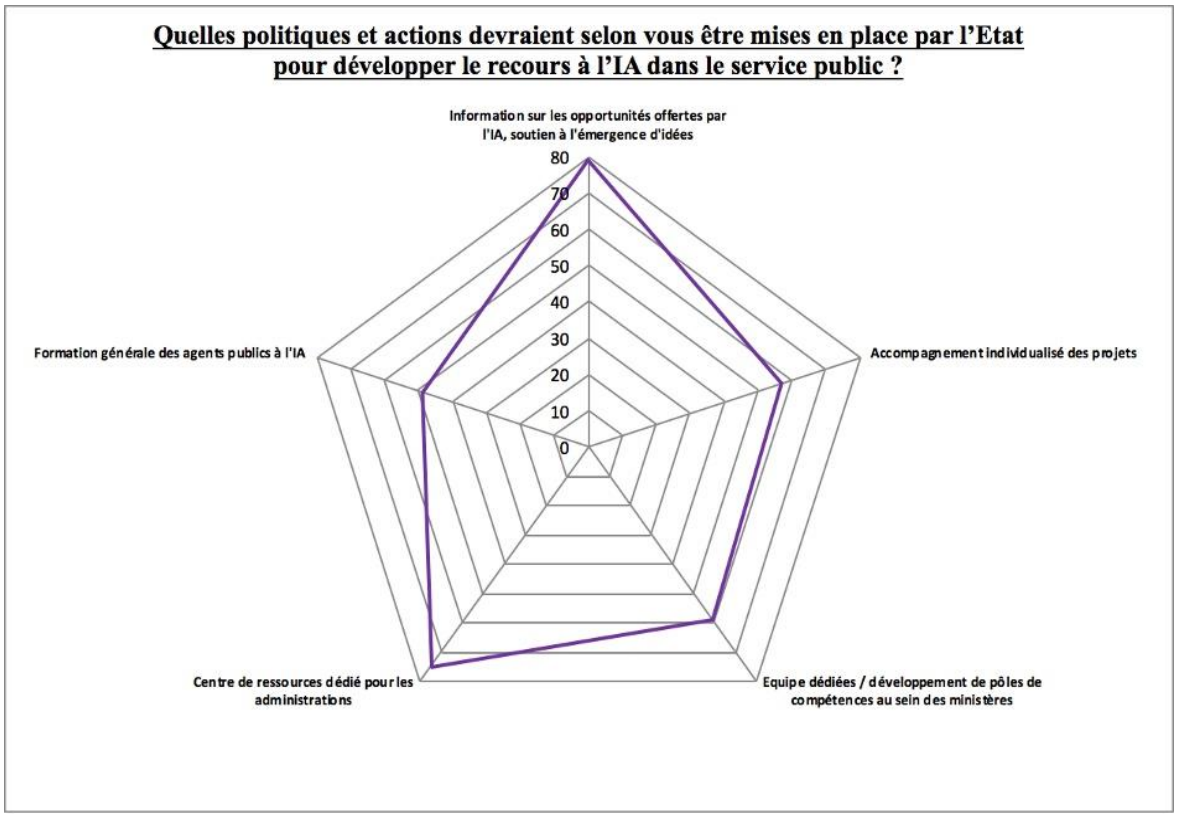
Les risques mis en avant par les administrations comme freins potentiels sont de trois natures, toutes reliées les unes aux autres. Les risques liés à la sécurité juridique relèvent tout d'abord du **souci de se conformer aux règles du RGPD**. Mais, elles sont renforcées notamment en raison du second risque identifié, qui est celui de la **cyber sécurité**. En effet, la création de bases de données nettoyées et utilisables revient en quelque sorte à créer une nouvelle richesse nécessitant une forte protection. Combinant ces deux facteurs, et les dépassant, l'ultime risque identifié est de type **éthique**. De manière attendue, les thèmes de la responsabilité, plus précisément de l'imputabilité, et de **l'explicabilité des décisions prises sur la suggestion d'un algorithme** sont les plus souvent cités.

Toutefois, il est important de noter que quelques administrations considèrent n'identifier pour l'instant aucun risque de ce type, ce qui peut s'expliquer dès lors que la prise de décision automatisée est écartée par la majorité des administrations. En réalité, la nécessité d'être capable pour les administrations de retracer le chemin emprunté par l'algorithme d'IA pour aboutir à la suggestion de telle ou telle décision laissée à la discrétion de l'opérateur humain demeure un véritable enjeu. Il rejoint la

nécessité soulevée par la plupart des répondants de former en leur sein des équipes dédiées à la gestion des outils d'IA.

3. Les attentes vis-à-vis de l'État pour lever les freins au développement de l'IA relèvent essentiellement de l'accompagnement méthodologique et de la mutualisation des initiatives

Les objectifs de la mission et l'orientation du questionnaire aidant, les administrations répondantes ont mis en avant de fortes attentes vis-à-vis de l'État en matière de déploiement de SIA.



L'attente qui apparaît le plus clairement dans les contributions écrites est la volonté d'**avoir accès à un centre de ressources dédiées à l'intelligence artificielle** auprès duquel les administrations pourraient obtenir des informations sur ce qui a déjà été réalisé en matière de SIA, ce qui a réussi, échoué, et grâce à quels financements et avec quels partenaires. En somme, l'accès à des informations simples et transparentes en matière d'IA est ce qui est le plus exprimé par les répondants. Cette attente est particulièrement manifeste au sein des collectivités territoriales, notamment chez les métropoles qui ont répondu au questionnaire. Une nouvelle fois, ce n'est pas le manque d'idées qui empêche pour les répondants le déploiement complet de SIA mais bien le manque d'informations quant aux voies de concrétisation de celles-ci.

Si ce n'est pas l'enjeu premier mis en avant par les répondants, **la question de la formation des personnels** paraît importante. Celle-ci s'exprime aussi différemment selon le type d'administration répondante. Naturellement, certains ministères et autorités indépendantes ont exprimé leur souhait de voir intégrer à la formation des cadres de l'État des modules intégrant les principaux enjeux de l'IA. La préfiguration du nouvel Institut national du service public (INSP), qui a depuis remplacé l'ancienne Ecole nationale d'administration, apparaît pour certaines administrations comme l'occasion d'intégrer de tels modules aux formations. Au-delà des cadres dirigeants, le souhait a été formulé par la majorité des administrations de voir **l'ensemble des personnels sensibilisés au fonctionnement des SIA**, ne serait-ce que pour faire reculer les craintes évoquées précédemment et augmenter l'acceptabilité de tels outils en interne.

En creux de ces attentes se distingue aussi la volonté des administrations, aux différentes échelles, de **développer des logiques partenariales dès les premières phases de développement des SIA**. Premièrement, le développement de partenariats entre administrations en matière d'IA est déjà une réalité prometteuse, notamment au sein des administrations centrales. Par exemple, la délégation à l'intelligence artificielle (DMIA) du ministère de l'Intérieur met en commun le développement de certains projets expérimentaux avec le ministère des Armées et celui de la Justice.

Deuxièmement, outre les partenariats interministériels, certaines administrations ont décidé de renforcer les partenariats avec des entités externes. Ainsi, les Hospices de Lyon ont particulièrement mis en avant ces logiques coopératives dans le déploiement de leur stratégie IA. Pour cette institution, les partenariats sont noués à la fois avec les acteurs de la recherche locaux (dispositifs Cifre par exemple, appui sur les grandes écoles lyonnaises) et avec certains acteurs du privé. Les partenariats du CHU de Lyon reposent sur une véritable logique d'établissement pilotée directement par les instances dirigeantes du CHU, appuyée par une commission d'experts dédiée à l'IA. Enfin, les logiques partenariales peuvent aussi se décliner à l'échelle internationale. Le succès du groupe d'experts de la donnée franco-canadien « GPAI » pour *Global Partnership on Artificial Intelligence*, portée par la France et le Canada, a notamment été souligné par le ministère de l'Intérieur. Au-delà de ce groupe d'experts, l'échelle européenne a été notée par certaines administrations comme un levier intéressant pour de tels partenariats.

Si la volonté de développer de tels partenariats est belle et bien présente chez les répondants, il ressort également que l'État pourrait mieux les accompagner. Parmi les attentes des répondants, est par exemple exprimée **l'opportunité de faire évoluer et simplifier les règles actuelles de la commande publique**. Selon certaines métropoles, une telle évolution permettrait de créer un réel « écosystème de la donnée » à l'échelle métropolitaine. Ces consortiums locaux pourraient être l'occasion pour les entreprises de trouver dans les administrations des « partenaires » plus que de simples « clients ».

*

Questionnaire transmis aux administrations publiques

Développement de projets recourant à l'IA

Avez-vous développé ou financé – ou projetez-vous de le faire à brève échéance – des projets recourant à l'intelligence artificielle – IA (ou dont vous estimez qu'elles utilisent l'intelligence artificielle) pour l'accomplissement des missions qui vous sont confiées (ou aux missions qui sont confiées aux opérateurs/établissements dont vous assurez la tutelle) ?

Si vous avez répondu "oui" à la première question, veuillez préciser les objectifs du projet, la méthode employée (association des personnels…), les techniques et algorithmes utilisés (apprentissage machine – supervisé, non supervisé, par renforcement… – apprentissage profond…), les moyens associés, le coût de conception et de maintenance de l'outil, le ou les partenaires et prestataires publics ou privés sollicités (tâches internalisées/externalisées), l'état d'avancement (en projet, en cours de conception, déployé, abandonné...) et les résultats obtenus.

Si vous avez répondu "oui" à la première question, quels sont les facteurs clés de succès et d'échec que vous avez identifiés ?

Si vous avez répondu "oui" à la première question, est-ce que ce ou ces projets s'inscrivent dans une stratégie numérique ou IA plus vaste ? Si oui, quels sont les piliers de cette stratégie ? Veuillez nous orienter vers la documentation présentant cette stratégie.

Si vous avez répondu "oui" à la première question, ce ou ces projets vous paraissent-ils reproductibles dans d'autres administrations ? Si oui, à quelles conditions ?

Opportunités et risques du recours à l'IA

Quels obstacles avez-vous identifiés dans le déploiement de projets recourant à l'IA ? Items proposés : utilité non avérée, manque d'idées, manque de temps, manque de données et données de qualité insuffisante, manque de moyens budgétaires, manque d'accompagnement et de formation, absence de prestataires compétentes, résistance ou manque d'appropriation des personnels, résistance des partenaires nécessaires à la conduite du projet IA, risque juridique ou incertitudes du cadre juridique, règles de la commande publique, problème de sécurisation de l'outil (risque cyber). Pour chaque item, le répondant pouvait indiquer qu'il ne constituait aucun obstacle, un obstacle très faible, un obstacle faible à moyen, un obstacle moyen à fort et un obstacle fort.

Veuillez expliciter vos réponses à la question précédente.

Si vous avez identifié d'autres risques que ceux mentionnés ci-dessus, veuillez les indiquer.

Pensez-vous que l'IA puisse accroître sensiblement la performance de votre entité ?

Si vous avez répondu "oui" à la question précédente, à quel(s) titre(s) ? (plusieurs réponses possibles) Items proposés : amélioration de la qualité / accessibilité

continuité du service public, élargissement du périmètre de l'action publique, économies budgétaires, revalorisation des fonctions des agents publics, image du service public, fiabilité des données,

Quels types d'outils recourant à l'IA vous paraît-il utile de concevoir pour les besoins de votre entité ? Items proposés : gestion et exploitation des documents et informations disponibles (« data management »), identification des priorités et des risques, ciblage des contrôles ou des interventions, prise de décision automatisée (sans intervention humaine), aide à la décision (instruction de demandes, évaluation des chances de succès d'une action…), reconnaissance de textes, d'objets ou de personnes, automatisation de la relation avec les usagers (robot conversationnel…), évaluation des politiques publiques. Pour chaque item, le répondant pouvait indiquer qu'il le considérait comme pas utile, d'une utilité faible, moyenne, forte ou « je ne sais pas ».

Si vous avez identifié d'autres outils que ceux mentionnés ci-dessus, veuillez les indiquer.

Quelles politiques et actions devraient selon vous être mises en place par l'État pour développer le recours à l'IA dans le service public ? Items proposés : information sur les opportunités offertes par l'IA, soutien à l'émergence d'idées, accompagnement individualisé des projets, équipes dédiées / développement de pôles de compétences au sein des ministères, centre de ressources dédié pour les administrations, formation générale des agents publics à l'IA, formation d'experts IA au sein de l'administration, renforcement des liens avec la recherche, financements, faciliter le recours à des prestataires privés. Pour chaque item, le répondant pouvait indiquer qu'il le considérait comme pas une priorité, une priorité faible, moyenne, forte.

Si vous avez identifié d'autres politiques et actions que l'Etat, ou une autre administration, pourrait mettre en œuvre pour promouvoir le déploiement de l'IA dans le secteur public, veuillez les indiquer.

Quels risques juridiques associez-vous à l'utilisation de l'IA pour l'exercice des missions qui vous sont confiées ?

Quels risques éthiques associez-vous à l'utilisation de l'IA pour l'exercice des missions qui vous sont confiées ?

Administrations répondantes

Ministères

- Ministère de l'Agriculture et de l'alimentation
- Ministère de la Culture
- Ministère de l'Intérieur
- Ministère de l'Éducation nationale, de la jeunesse et des sports
- Ministère de l'Économie, des Finances et de la Relance
- Ministère des solidarités et de la santé
- Ministère du Travail, de l'emploi et de l'insertion
- Ministère de la justice
- Ministère des Armées

Autorités indépendantes

- Commission nationale de l'informatique et des libertés (CNIL)
- Haut Conseil du Commissariat aux Comptes (H3C)
- Autorité de régulation des communications électroniques, des postes et de la distribution de la presse (ARCEP)
- Autorité de contrôle des nuisances aéroportuaires (ACNUSA)
- Commission nationale de contrôle des techniques de renseignement (CNCTR)
- Conseil supérieur de l'audiovisuel (CSA)
- Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES)
- Autorité de régulation des transports (ART)
- Agence française de lutte contre le dopage (AFLD)
- Haute Autorité pour la transparence de la vie publique (HATVP)
- Autorité des marchés financiers (AMF)
- Autorité de sûreté nucléaire (ASN)
- Médiateur national de l'énergie
- Haute Autorité pour la diffusion des œuvres et la protection des droits sur internet (HADOPI)

Collectivités territoriales

- Ville de Nîmes
- Ville de Lyon
- Ville de Marseille
- Métropole Rouen Normandie
- Métropole Nice Côte d'Azur
- Métropole de Lyon
- Bordeaux Métropole
- Aix Marseille Provence Métropole
- Montpellier Méditerranée Métropole
- Strasbourg Métropole

- Grenoble-Alpes Métropole
- Orléans Métropole
- Toulouse Métropole
- Région Hauts de France
- Région Occitanie
- Conseil Régional de Guadeloupe
- Région Grand Est
- Régions de France
- Nantes Métropole et Ville de Nantes

Secteur sanitaire et social

- Caisse nationale des allocations familiales (Cnaf)
- Caisse nationale de solidarité pour l'autonomie (Cnsa)
- Hospices civils de Lyon
- CHU de Nice
- Pôle Emploi
- Caisse des dépôts et consignations

Annexe 8 : Présentation des acteurs de l'écosystème de l'IA publique

Les administrations centrales dédiées

Bien que son décret d'attribution n'en fasse pas expressément mention, le secrétaire d'État chargé de la transition numérique et des communications électroniques joue un rôle de sponsor et d'impulsion dans la mise en œuvre de la stratégie nationale pour l'IA. Il est compétent, plus largement, sur l'ensemble des questions relatives à la promotion et à la diffusion du numérique.

La ministre de la transformation et de la fonction publiques est plus directement chargée du déploiement de SIA publics, dès lors que son décret d'attribution prévoit qu'elle est responsable de la transformation numérique de l'État et, à ce titre, du développement et de l'amélioration des usagers et services numériques ainsi que de la politique d'ouverture et de circulation des données.

Le **coordinateur national pour l'intelligence artificielle** et son adjoint, placés auprès du directeur général des entreprises du ministère de l'économie, des finances et de la relance, ont pour mission de piloter la Stratégie nationale pour l'intelligence artificielle. Le coordinateur national, qui ne dispose pas d'équipes en propre, anime le réseau des correspondants ministériels à travers un comité de pilotage mensuel auquel participe également le Lab IA de la direction interministérielle du numérique (V. *infra*). Il assure en outre la représentation de la France au sein du Partenariat mondial pour l'intelligence artificielle.

Sur le plan opérationnel, deux administrations centrales concourent à la mise en œuvre de la stratégie au sein de l'État, en sus des actions menées par le coordonnateur national.

1/ La **direction interministérielle du numérique** (DINUM), créée par le décret n° 2019-1088 du 25 octobre 2019, est rattachée au secrétaire général du gouvernement, et placée sous l'autorité du ministre de l'action et des comptes publics et à la disposition du ministre de l'économie et des finances et du secrétaire d'État chargé du numérique. Elle a notamment pour mission de contribuer à la transformation numérique des politiques publiques, de soutenir le développement des compétences de l'État dans le domaine du numérique, d'élaborer et mettre à disposition des ressources numériques partagées ou encore d'étudier l'opportunité de « *recourir à des technologies en voie de maturation issues du monde de la recherche* ». Son directeur, qui joue, selon le décret, le rôle d'administrateur général des données (la circulaire du 27 avril 2021 le qualifiant d'administrateur général des données, des algorithmes et des codes sources), est systématiquement informé des projets informatiques d'un montant d'au moins 5 millions d'euros, et il peut s'opposer à ceux qui excèdent 9 millions d'euros.

En son sein, le **département Etalab** coordonne la conception et la mise en œuvre de la stratégie de l'État dans le domaine de la donnée, incluant son ouverture (politique

d'» *open data* » avec, notamment, la plateforme interministérielle de référence *data.gouv.fr*), sa circulation entre administrations (qui s'appuie notamment sur une offre d'interfaces de programmation *via* le portail *api.gouv.fr*) et son exploitation, notamment par la promotion des sciences des données et de l'intelligence artificielle dans la sphère publique.

Cette dernière mission est principalement assurée par le **pôle « exploitation des données (Lab IA et data sciences) »** d'Etalab, composé de quatre experts de la donnée. Il pilote depuis la fin 2018 le « Lab IA » dont l'objectif est de conduire et d'accompagner des projets en science de la donnée et de co-construire des outils recourant à l'intelligence artificielle, de faciliter l'expérimentation et le partage de bonnes pratiques et de ressources, d'acculturer les agents publics à l'IA et d'animer des communautés d'experts et d'amateurs. Outre les événements qu'il anime (sept « *datadrinks* » en 2020, réunissant plus de 350 participants) et sa lettre d'actualité, le Lab IA a conçu et mis à en ligne plusieurs guides et webinaires à destination des administrations (guide des algorithmes publics, guide méthodologique pour la conduite de campagnes d'annotation dans le cadre des projets d'apprentissage supervisé, guide pour l'utilisation de l'IA dans la pseudonymisation de documents). Il pilote en outre la démarche collaborative PIAF (« Pour des IA francophones »).

Etalab pilote en outre le **programme « Entrepreneurs d'intérêt général » (EIG)** par lequel des experts du numérique, en immersion pendant dix mois au sein d'administrations publiques, relèvent des « défis » proposés par ces dernières, consistant à concevoir des outils améliorant le service public par le numérique. Une dizaine de défis, sur environ 90 au total (dont 39 en data science), ont porté spécifiquement sur la conception de SIA ou d'outils recourant à des SIA (en sus des nombreux projets visant à rassembler, structurer, transférer et mettre à disposition des données), notamment DataJust et le code du travail numérique. Ce programme reçoit un accueil très favorable des administrations, dont certaines ont pu pérenniser l'embauche d'EIG.

Enfin, la DINUM pilote pour l'essentiel le fonds d'innovation et transformation numérique (FITN) dont plusieurs volets sont susceptibles de soutenir financièrement la conception et le déploiement de SIA par l'administration. Il a par exemple bénéficié au projet SIMARA mené par les archives nationales et recourant à un algorithme de reconnaissance de caractères manuscrits afin de numériser les inventaires des services d'archives. Un dispositif comparable est ouvert aux collectivités territoriales *via* le fonds de transformation numérique des collectivités territoriales.

2/ La **direction interministérielle de la transformation publique** (DITP), créée par le décret n° 2017-1584 du 20 novembre 2017, assure le suivi des réformes prioritaires de transformation des acteurs publics et leur territorialisation, par la promotion d'organisations centrées sur l'usager et la responsabilisation des agents. Elle dispose, à travers une équipe de plus de 80 consultants et experts, d'une capacité d'intervention au soutien des administrations qui développent des projets de transformation ambitieux et cohérents avec la stratégie nationale, pour des missions de quelques jours à six mois. Elle gère le fonds de transformation de l'action publique

(FTAP), doté de 780 millions d'euros de 2017 à 2022, qui a permis de financer de nombreux projets incluant un SIA (*Health Data Hub*, Intelligence Emploi, LabIA de Météo France, Projet « foncier innovant » de la DGFIP, projet « Améliorer la sécurité de la navigation maritime grâce à l'intelligence artificielle » de la direction des affaires maritimes, l'Observatoire de l'artificialisation des sols de l'IGN…).

La DINUM et la DITP ont lancé conjointement **deux appels à manifestation d'intérêt** (AMI-IA) pour l'accompagnement technique et métier de projets incluant des SIA dans le secteur public. 21 projets ont été sélectionnés. Deux ont été abandonnés, dont l'un en raison de l'inaccessibilité des données. Si la faisabilité et la maturité des initiatives sont inégales, la démarche apparaît pertinente et efficace pour faire émerger des idées et favoriser le développement de preuves de concept. Une phase d'évaluation des outils doit être menée afin d'apprécier l'opportunité d'une industrialisation, d'une adaptation ou d'une extension, et la publication des codes sources.

En outre, **l'écosystème de la recherche en intelligence artificielle**, fréquemment salué pour son excellence, peut être mobilisé pour soutenir les efforts d'innovation des administrations. Outre le CNRS, l'INRIA et le CEA sont les deux principaux organismes de recherche investis dans l'IA et en lien direct avec les administrations publiques. Trois projets sélectionnés dans le cadre de l'AMI-IA bénéficient ainsi d'un accompagnement scientifique par l'INRIA, grâce au partenariat conclu avec la DINUM, comme l'illustre l'accompagnement du projet de système de détection des divergences de jurisprudence de la Cour de cassation. La collaboration entre l'équipe Datashape de l'INRIA, le Service hydrographique et océanographique de la marine (SHOM) et la DINUM a permis de développer un algorithme d'estimation de la profondeur des fonds sous-marins à partir d'images LIDAR qui a vocation à être intégré dans l'outil métier des agents du SHOM. Cinq nouveaux projets sont accompagnés en 2022 *via* ce partenariat. Afin d'accroître leur impact sur le déploiement des SIA dans l'action publique, il est impératif, d'une part, de soutenir davantage l'effort de recherche, d'autre part, d'intensifier les liens avec les administrations, en veillant à ne pas mobiliser la recherche sur des activités opérationnelles qui ne lui incombent pas et qui doivent être prises en charge par les services, et, enfin, de prioriser les programmes de recherche en fonction, notamment, de leur incidence opérationnelle sur les services publics, dans le cadre d'une stratégie de l'IA publique.

Les autres administrations centrales

La conception et le déploiement des SIA dans les administrations centrales ne répondent pas à une logique organisationnelle unique. Cette mission est tantôt dévolue à la direction du numérique du ministère, tantôt à un service dédié de type « Lab ». Dans les administrations les plus investies, les ressources peuvent être réparties entre les services centraux et les services déconcentrés (cas de la DGFIP par exemple).

La circulaire n° 6264/SG du 27 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources entend donner une nouvelle impulsion à cette politique, conformément aux recommandations du rapport remis par la mission

Bothorel. En effet, si la France s'est rapidement dotée d'un cadre juridique propice à l'*open data* et si de nombreux progrès ont été réalisés dans la période récente, notamment la disponibilité accrue des interfaces de programmation - API), les pratiques de diffusion publique, de partage entre services et de structuration des données en vue d'en permettre et d'en faciliter la réutilisation n'ont pas évolué au rythme initialement escompté. Ce retard pénalise l'ensemble de l'écosystème de l'IA, public et privé confondus.

Conformément à cette circulaire, chaque ministère a désigné un **administrateur ministériel des données, des algorithmes et des codes sources** (AMDAC) chargé de coordonner la politique du ministère et d'animer un réseau de correspondants au sein des directions métier, et s'est doté d'une feuille de route décrivant la stratégie du ministère dans ce domaine et les actions programmées ou envisagées, notamment en matière de formation. Cet effort bienvenu de formalisation ne s'accompagne toutefois d'aucun moyen supplémentaire – l'AMDAC étant parfois le directeur de l'administration centrale compétente, qui ne peut matériellement consacrer à cette fonction le temps qu'elle mérite.

Les administrations déconcentrées

A ce stade, la politique de la donnée comme la conception et le déploiement des SIA sont encore largement tributaires d'initiatives « personnelles » d'agents formés ou sensibilisés à ces questions et désireux d'insuffler dans les services qui les emploient une dynamique d'innovation. La circulaire du 27 avril 2021 a prévu la désignation de référents « données, algorithmes et codes sources » auprès des préfets de région afin d'animer localement cette politique.

Les collectivités territoriales

L**'Agence nationale de la cohésion des territoires** (ANCT), chargée d'accompagner et de soutenir leur transformation numérique, a mis en place un « Incubateur des territoires » qui appuie le développement de services numériques pour les usagers, y compris des projets innovants, à l'aide de l'enveloppe de 30 millions d'euros prévue dans le cadre du plan France Relance. L'objectif consiste à créer un cadre général pour le développement de SIA robustes et répondant à des exigences éthiques élevées, ce qui correspond à une demande forte des élus locaux, soucieux de garantir l'acceptabilité sociale de ces outils. Le lancement de cette activité d'accompagnement est récent et les ambitions pour l'heure modestes. L'action de l'ANCT reste prioritairement orientée vers la lutte contre la fracture numérique, avec la formation de 4000 conseillers numériques « France Services ».

Les établissements publics et les organismes de sécurité sociale

Ils bénéficient d'une très grande liberté organisationnelle pour mettre en œuvre leur politique de la donnée et de conception de SIA, avec des degrés de maturité très variables d'une entité à l'autre.

Les conventions d'objectifs et de gestion qui lient l'État aux caisses nationales de sécurité sociale prévoient prioritairement une modernisation des outils informatiques et leur interopérabilité entre régimes et avec les systèmes

d'information de l'État. Elles n'évoquent que marginalement ou implicitement le développement de SIA, qui sont pourtant déjà nombreux dans certains réseaux (essentiellement dans les branches vieillesse et famille et au sein de l'URSSAF Caisse nationale, voir infra). Le schéma stratégique des systèmes d'information de la sécurité sociale 2018-2022 prévoit, au titre de l'action « *Partager les initiatives Big data* », la réalisation d'un rapport d'étude portant notamment sur l'échange de données et la mutualisation des compétences spécifiques au datamining et à l'intelligence artificielle et le bilan des résultats des preuves de concept réalisées, ainsi que la création d'un grand prix de l'innovation numérique. Au titre de « l'automatisation des processus métiers », identifié comme l'axe n° 1, les efforts portent sur la conception d'une architecture favorisant les échanges de données à travers des référentiels et des dispositifs techniques communs. Lors de leur signature en 2018, ces conventions n'identifiaient donc pas les SIA comme un levier de transformation des organisations et des process, mais plutôt comme un facteur d'amélioration incrémentale du service rendu.

Annexe 9 : Cartographie des cas d'usage des systèmes d'IA dans l'action publique

Fiche n° 1 : Gestion des territoires

Fiche n° 2 : Défense

Fiche n° 3 : Sécurité, activités d'enquête, de contrôle et de sanction

Fiche n° 4 : Justice

Fiche n° 5 : Travail et emploi

Fiche n° 6 : Education

Fiche n° 7 : Protection sociale

Fiche n° 8 : Santé

*

Fiche n° 1 : Gestion des territoires

La gestion et l'aménagement du territoire et de l'espace (voiries et réseaux, urbanisme…) constituent un terrain d'application privilégié pour le déploiement des SIA. Cette famille de cas d'usage a été largement popularisée par le concept fourre-tout, aussi nébuleux que celui d'intelligence artificielle, et porteur d'un fâcheux *a contrario*, de « ville intelligente » (*smart city*) ou de « territoire intelligent », de plus en plus utilisé afin d'englober les espaces périurbains et ruraux[274]. Cette notion, qui prolonge celle de « ville connectée », renvoie, à grands traits, à un espace dans lequel les données collectées par l'ensemble des canaux disponibles (données produites par les services, données fournies par les citoyens, données issues de capteurs connectés disséminés sur des bâtiments ou équipements publics[275]…) procurent une connaissance fine, transverse et actualisée (voire en temps réel) des ressources, des besoins des habitants et usagers et des situations, et permettent aux collectivités publiques d'adapter les services qu'elles proposent à ces besoins ainsi qu'aux exigences du développement durable, en tenant compte des initiatives privées. En résumé, il s'agit d'une ville optimisée par le numérique – cet optimum s'appréciant au regard des choix politiques décidés par la collectivité, en concertation ou en lien avec les acteurs de la ville (citoyens et usagers, Etat, entreprises publiques et privées…).

Cet objectif – voire cet idéal – peut s'inscrire dans une stratégie globale planifiée (Dijon à partir de 2015, Angers Loire Métropole de 2020) ou relever d'une démarche plus incrémentale et casuistique (Issy-les-Moulineaux). Il peut s'appuyer sur des outils innovants qui procurent une connaissance de plus en plus fine des caractéristiques du territoire et une maîtrise plus étroite des leviers d'intervention publique. On peut citer à cet égard les plateformes de données urbaines (*urban data platforms*), les « hyperviseurs », qui peuvent prendre la forme d'un centre physique de pilotage des décisions et des équipements (exemple du « poste de commandement » de Dijon, dans le cadre du projet de ville intelligente « OnDijon »), ou encore les « jumeaux numériques ».

Le jumeau numérique d'un territoire est sa réplication virtuelle en trois dimensions augmentée d'informations qualitatives et quantitatives et de la représentation graphique de métriques, qui permet tout à la fois de mieux comprendre l'état des lieux et le fonctionnement de l'espace, et de simuler l'impact de décisions de politique publique (par exemple, le réaménagement d'une place ou d'un quartier, la piétonnisation de voies ouvertes à la circulation automobile…) ou d'évènements externes en vue d'adopter la meilleure stratégie de planification et de gestion.

---

[274] V. sur cette notion le rapport de la direction générale des entreprises, *De la Smart city à la réalité des territoires connectées – L'émergence d'un modèle français ?*, octobre 2021.

[275] Comme par ex. les 3 000 bancs connectés déployés par la ville de Paris, destinés à mesurer le trafic piéton, l'utilisation de l'espace et de ces équipements, la température et la pression atmosphérique, la satisfaction des usagers…

Certains villes, comme Rotterdam[276] et Herrenberg[277], expérimentent cet outil complexe, coûteux mais prometteur. Se rattachent à cette démarche l'ambitieux projet Lidar HD de cartographie en trois dimensions du territoire national piloté par l'Institut national de l'information géographique et forestière, le projet Road-AI porté par le Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement (CEREMA) et l'INRIA qui vise à construire un jumeau numérique dynamique de la route afin d'éclairer les gestionnaires d'infrastructures, ou encore le projet exploratoire du Bureau de recherches géologiques et minières en partenariat avec l'Université d'Orléans dans la plaine de Beauce.

Les SIA constituent l'un des moyens, parmi bien d'autres, de concevoir et de mettre en œuvre un projet de « territoire intelligent », tant pour la définition des politiques publiques que pour leur évaluation. On perçoit aisément, par exemple, l'utilité des techniques d'apprentissage machine pour identifier des corrélations et facteurs explicatifs « occultes » d'un phénomène (ex : sous-fréquentation d'un équipement, saturation d'un axe routier, apparition de nuisibles…), qu'une analyse qualitative classique, une enquête de terrain ou un questionnaire ne permettraient pas d'identifier. Le progrès dans la connaissance de la cité présente également l'avantage d'objectiver les enjeux et les déterminants de la prise de décision, et concourt ainsi à améliorer aussi bien l'impact des procédures de participation du public que l'acceptabilité des choix de politiques publiques.

Selon l'étude « Territoire intelligent et donnée publique » de la direction générale des entreprises d'octobre 2021[1], « *l'IA en est encore à ses balbutiements dans les services déployés par les collectivités* ». De nombreux cas d'usage peuvent néanmoins être recensés, tous ne mobilisant pas nécessairement les techniques d'apprentissage machine. Ne seront pas traités ici, bien qu'ils puissent contribuer à une meilleure gestion du territoire, les usages répressifs (comme la détection des infractions routières et du stationnement, la police de l'urbanisme…), qui relèvent d'une autre logique et soulèvent des questionnements éthiques et juridiques distincts (V. sur les usages en matière de contrôle et de sanction la fiche n° 3).

1. La mobilité

En matière de **transport et de mobilité**, les SIA sont déjà largement déployés commercialement, notamment dans les applications de calcul d'itinéraires et de guidage routier accessibles sur les téléphones mobiles. Du point de vue des autorités publiques, il s'agit d'un champ d'expérimentation prioritaire en raison des enjeux liés à la congestion urbaine (sécurité routière, qualité de l'air, nuisances sonores, facilitation des activités économiques et des échanges…). Ces systèmes peuvent notamment :

- aider à la gestion des flux et à l'optimisation du plan de circulation en prévoyant le trafic routier rue par rue et quartier par quartier en fonction d'un très grand nombre

[276] Université Erasmus de Rotterdam, *Governance, Trust and Smart City Business Models : the Path to Maturity for Urban Data Platforms*, 11 novembre 2020, H2020-SCC-2016.

[277] F. Dembski et a., *Urban Digital Twins for Smart Cities and Citizens : The Case Study of Henrrenberg, Germany*, Sustainability, 16 mars 2020, 12, 2307.

de paramètres explicatifs278 et de capteurs fixes ou mobiles (transports publics, véhicules dédiés, mais aussi l'ensemble des usagers de la route acceptant de partager des données…), en ajustant les leviers de circulation (ouverture/fermeture de voies, extinction/allumage et changement de rythme des feux tricolores279…) en temps réel en fonction du trafic constaté ou des besoins des services d'urgence (pompiers, ambulances). La capitale estonienne, Tallinn, expérimente depuis 2018 un SIA recourant à la vision par ordinateur, qui utilise des caméras situées à certains carrefours pour compter des véhicules de différents types (et, à l'avenir, les piétons, cyclistes et autres usagers de la route), afin de prévoir la congestion dans certaines zones : le taux d'exactitude, affecté notamment par les conditions météorologiques et l'encrassement des capteurs, a été évalué à 70 % ;

- faciliter le stationnement par l'identification des véhicules par lecture des plaques d'immatriculation (cas d'usage désormais courant dans les parkings collectifs) ou en informant en temps réel les automobilistes des places disponibles et qui devraient le rester lorsqu'ils arriveront sur place (solution « ParkAm » en Israël, utilisant des caméras sur la voie publique et un algorithme de vision par ordinateur) ;
- accroître l'efficacité des services de mobilité partagée, notamment par l'appariement performant entre l'offre et la demande (covoiturage, vélo…) ;
- automatiser des transports collectifs ou en commun de surface, notamment les taxis (projet de Waymo à San Francisco, tests de taxis autonomes en Chine) et les bus (premier essai commercial à Singapour en 2021, expérimentation à Malaga en Espagne…) ou les petites navettes autonomes (Lyon, Rouen Normandie, Vernon-Giverny…). A terme, des transports aériens autonomes de proximité (« taxis volants ») sont également concevables (projet « VoloCity » à Rome par exemple) ;
- orienter les décisions de réfection et d'amélioration de la voirie routière à partir des données relevées par des capteurs embarqués (le cas échéant un simple smartphone fixé au pare-brise d'un véhicule administratif), en utilisant un algorithme d'apprentissage (vision par ordinateur) permettant d'identifier et de catégoriser les détériorations de la chaussée (fissures, nids de poules…).

L'**entretien des infrastructures de transport** peut aussi se prêter au déploiement de véhicules autonomes, en particulier en cas d'intempéries (expérimentation en mars 2018 de chasse-neige autonomes sur un aéroport norvégien, capables de déneiger plus de 300 000 m² en une heure - projet Yéti).

Les SIA peuvent aussi directement contribuer à la **mesure et à la gestion de la qualité de l'air**, en prévoyant plusieurs heures à l'avance la concentration de particules fines à tel ou tel endroit280. Il est aussi possible de mieux détecter et gérer les **nuisances sonores**.

---

278 V. par ex. O.I Olayode, L.K Tartibu et M.O Okwu, *Application of Artificial Intelligence in Traffic Control System of Non-autonomous Vehicles at Signalized Road Intersection*, Procedia CIRP, Vol 91, 2020. Il s'agit d'un champ de recherche ancien (V. John F. Gilmore et Khalid J. Elibiary, *AI In Advanced Traffic Management Systems*, AAAI Technical Report WS-93-04, 1993).

279 V. par ex. le projet allemand KI4LSA en test dans la ville de Lemgo.

280 D. Iskandaryan et a., *Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data : A Review*, Applied Sciences, 10(7):2401, avril 2020.

2. La gestion des services publics locaux

En matière de **gestion des déchets**, le déploiement de SIA s'appuie en général sur la collecte de données issues de poubelles connectées. A Rotterdam, des capteurs ont été posés sur 6540 poubelles afin de mesurer le taux de remplissage, d'adapter les tournées en conséquence et d'améliorer la maintenance des containers. L'objectif de ce système déployé depuis 2018 est de passer de 203 trajets fixes à 165 trajets dynamiques, réduisant de 20% le nombre de kilomètres parcourus et, en conséquence, les émissions de CO_2. Un SIA peut être utilisé, par exemple :

- au niveau des poubelles individuelles, pour calculer le montant de la redevance d'enlèvement des ordures ménagères en fonction du poids des déchets (cas de la communauté de communes Pays Haut Val d'Alzette) ;
- au niveau de containers situés sur la voie publique, pour identifier, classer et trier les déchets dès leur dépôt à l'aide d'un algorithme de reconnaissance d'objet (exemple de la « Bin-e » créée en Pologne) ou pour éclairer l'usager sur le compartiment dans lequel il doit jeter un détritus analysé par la machine (système Oscar au Canada)[281]. Le cas échéant, ces containers peuvent être reliés directement au terminal de collecte par un réseau de tubes souterrains qui « aspirent » les sacs lorsque la borne est pleine (cette collecte pneumatique, qui existe depuis plusieurs dizaines d'années dans certaines villes, se développe en région parisienne à partir de l'expérimentation conduite à Romainville) ;
- au niveau de bennes à ordures situées en déchetteries, comme c'est le cas de plus de 200 bennes connectées déployées dans l'aire urbaine toulousaine (syndicat Décoset) : le SIA peut identifier le type de déchets déposés et le nombre de fois où la benne a été tassée, calculer le taux de remplissage et optimiser les flux (diminution d'environ 7% des rotations de camions).
- au stade du centre de tri, pour classer et trier les déchets en tenant compte de leur type, de leur composition et de leur éventuelle contamination par des produits chimiques, en utilisant notamment la vision par ordinateur, le traitement algorithmique de données issues de capteurs divers (infrarouges, détecteurs de métaux, scanners 3D, radio-identification par les « tags RFID »…) et la robotique pour le geste de tri proprement dit. Un SIA est susceptible de permettre de procéder à un tri plus rapide, plus fiable et surtout plus fin, et peut contribuer à la diversification des filières de tri et à la performance de l'économie circulaire. En l'état des technologies, le tri des déchets de construction semble largement automatisable, alors que celui des déchets ménagers s'avère plus

[281] Certains districts en Chine ont quant à eux opté pour un algorithme de reconnaissance faciale qui identifie la personne se présentant devant une benne à ordures et ouvre le couvercle s'il s'agit d'un résident local. La poubelle contrôle la correspondance entre le sac poubelle jeté et le bac dans lequel il l'est et des points, des crédits ou des bons d'achat, correspondant le cas échéant à la valeur de marché des déchets jetés, peuvent être accordés aux citoyens exemplaires dans le recyclage de leurs déchets.

délicat en raison de la diversité des matériaux et des matières[282]. Il fait l'objet de nombreuses expérimentations (tri optique sur une chaîne de tri à Amiens ; système de tri des plastiques souples à Sydney…) ;

- au niveau global, pour améliorer la qualité du service à l'usager, la qualité de vie au travail et l'allocation optimale des ressources : les SIA permettent de prévoir les conditions et les proportions dans lesquelles les déchets sont jetés et adapter les décisions publiques en conséquence (format des poubelles, fréquence des ramassages…), d'optimiser les organisations de travail (détermination des tournées…) et de favoriser l'appariement entre l'« offre » et la « demande » (économie circulaire, problématique du gaspillage alimentaire…).

En matière de **gestion de l'eau**, les infrastructures de production et de traitement de l'eau sont déjà équipées d'un très grand nombre de capteurs connectés, dont les données sont principalement utilisées dans la mesure de la qualité de l'eau et la recherche des causes d'incidents qui se sont produits. Les SIA ouvrent la voie à :

- une meilleure anticipation et prévention des fuites et de la dégradation de la qualité de l'eau, le cas échéant en combinaison avec des données d'information générale – météorologiques par exemple (algorithme d'apprentissage machine développé par le groupe de production d'eau Ramboll en Finlande), le paramétrage des pompes en vue d'ajuster finement l'offre et la demande (projet CHAIN à Aarhus, au Danemark, basé sur l'apprentissage machine) ;
- la réduction de la consommation d'énergie nécessaire au fonctionnement du système de pompage (expérimentation d'un système de coordination des mouvements de pompage en fonction de la quantité d'eau traitée nécessaire pour une journée donnée, basé sur un algorithme à Melbourne en Australie) ;
- au niveau des consommateurs, à la détection précoce des fuites en raison d'une surconsommation manifeste, ce que doit permettre, par exemple, la généralisation des compteurs d'eau « intelligents » à Paris, pilotée par Eau de Paris ; à la mesure du taux d'humidité des sols des parcs qui permet, en combinaison avec les prévisions météorologiques, d'évaluer la nécessité d'arroser (projet développé à Montpellier).

3. L'espace public et l'aménagement du territoire

L**'éclairage public** est un secteur fréquemment et prioritairement investi par les collectivités dans le cadre des démarches de type « territoire intelligent ». Outre l'installation de capteurs sur les lampadaires d'éclairage public pour la collecte de données utiles à des finalités tierces (trafic routier, flux de piétons, mesure de la qualité de l'air, évaluation des nuisances sonores et détection de bruits anormaux…), le système d'éclairage peut être connecté et paramétré afin de moduler son utilisation en fonction de la fréquentation des sites et de réduire tout à la fois la pollution lumineuse nocturne et l'empreinte écologique de ces équipements. Le classique détecteur de mouvements peut être augmenté par un modèle de vision

[282] V. Wilts, H., Garcia, B.R., Garlito, R.G., Gómez, L.S., Prieto, E.G. *Artificial Intelligence in the Sorting of Municipal Waste as an Enabler of the Circular Economy*. Resources 2021, 10, 28, qui présente une expérimentation de tri automatisé de déchets ménagers près de Barcelone.

par ordinateur qui distingue une voiture ou un humain (ce qui justifie l'allumage du lampadaire à son passage) et un chat ou un chien (ce qui ne le justifie pas).

En ce qui concerne la **propreté**, l'utilisation de capteurs de données (sans collecte nécessaire de données à caractère personnel) couplée avec des systèmes de reconnaissance d'objets (vision par ordinateur), le cas échéant conjuguée avec des dispositifs numériques de signalement de dépôts sauvages par les habitants, peut permettre d'évaluer le taux de propreté (ou de saleté) d'une rue et de déclencher une opération de nettoyage ciblée (dans les cas les plus critiques) ou d'adapter les tournées. Le nettoyage pourra à terme être effectué par des balayeuses autonomes circulant la nuit et capables de moduler leur vitesse ainsi que la puissance d'aspiration et de nettoyage (expérimentation « Trombia Free » à Helsinki : le véhicule de nettoyage est capable à ce stade de couvrir entre 5000 et 15 000 m² de voirie par heure[283]).

Les politiques de **gestion et d'affectation des espaces urbains** peuvent également être assistées par des SIA, en particulier de vision par ordinateur des images satellites. A titre d'exemple, Rotterdam a entrepris, à l'aide d'un tel outil, de cartographier les 14,5 km² de toitures-terrasses de la ville afin d'identifier leur usage actuel et de déterminer s'ils pourraient être utilisés pour des réservoirs d'eau, des espaces verts ou la production d'énergies renouvelables. La connaissance acquise par les différentes sources de collecte de données devrait être restituée régulièrement voire en temps réel par des interfaces accessibles au public, prolongeant les efforts entrepris dans la réalisation des systèmes d'information géographique. Pourrait ainsi être envisagée la réalisation d'un géoportail amélioré utilisant le traitement du langage naturel et permettant, pour une parcelle ou un immeuble donné, d'interroger l'interface en langage courant et d'obtenir en réponse des données à caractère non personnel issues de l'ensemble des bases de données publiques et privées disponibles (par exemple, dans la perspective d'un achat immobilier, la chronique récente de la concentration en polluants, le niveau et la nature des nuisances sonores existantes et prévisibles, l'existence de parcelles constructibles et de projets de construction à proximité – voire une évaluation de la probabilité qu'un tel projet voie le jour dans le voisinage, le temps de trajet pour rejoindre le supermarché ou l'hôpital le plus proche, le délai prévisionnel dans lequel un ravalement de façade sera nécessaire…).

La **délivrance des autorisations d'urbanisme** apparaît également comme un domaine prometteur pour l'automatisation par les SIA, à la faveur de la dématérialisation des dossiers de demande en vigueur à compter du 1er janvier 2022 : sous réserve des appréciations « subjectives » qu'appelle l'instruction de ces dossiers (atteinte portée au caractère ou à l'intérêt des lieux…), qui sont plus difficiles à traiter par un modèle algorithmique, l'essentiel des points de contrôle des projets d'urbanisme peuvent être réalisés ou assistés par la machine, qu'il s'agisse de règles purement objectives (hauteur de la construction, distance aux limites séparatives et aux voies publiques, emprise au sol, gabarit-enveloppe…) ou de dispositions appelant une analyse d'impact prospective qu'un modèle algorithmique est mieux à même de réaliser qu'un humain

[283] Les rues de Paris (hors trottoirs et voies piétonnes, qui soulèvent des difficultés particulières de circulation pour un véhicule autonome) représentent environ 15 millions de m².

(incidence du projet sur la commodité et la sécurité de la circulation publique ou sur l'accessibilité pour les engins de secours, impact environnemental…). Le raccourcissement drastique des délais d'instruction qu'entraînerait le déploiement de SIA, y compris d'ailleurs pour la saisine de services associés (architecte des bâtiments de France…) constituerait un puissant facteur d'incitation au dépôt de demandes dématérialisées. Les effectifs des services d'urbanisme pourraient être redéployés vers l'examen des dossiers les plus délicats, l'amélioration de la qualité de la réglementation d'urbanisme locale et la police de l'urbanisme.

4. La relation aux usagers et la vie démocratique locale

Comme toutes les autres administrations en relation directe avec le public, les collectivités territoriales et leurs groupements peuvent aussi exploiter le potentiel des SIA pour améliorer la **relation avec les citoyens et usagers des services publics**, notamment par des agents conversationnels fournissant des renseignements sur la base de demandes formulées en langage courant (exemple de la métropole d'Aix-Marseille), des outils d'identification de l'interlocuteur compétent au sein de la sphère publique, des systèmes de traitement automatique des demandes... Dès lors qu'ils ne se substituent pas à l'interlocuteur humain ou qu'une telle substitution s'accompagne d'une politique volontariste d'accessibilité numérique, et qu'ils permettent une amélioration sensible du service rendu, notamment par leur disponibilité permanente et la réduction des délais de traitement des demandes, le développement de ces services de « guichet » numérique peut constituer un puissant levier d'acceptation des SIA par les usagers et citoyens. A l'instar de services commerciaux basés sur l'IA quotidiennement utilisés par des millions de personnes en France, ils concrétisent la promesse de création de valeur des SIA au bénéfice direct des usagers du service public, alors que les retombées positives de systèmes déployés en « *back-office* » sont souvent moins perceptibles.

Grâce aux outils de traitement automatique de texte et à l'analyse sémantique, les SIA peuvent contribuer à une **meilleure compréhension des besoins des citoyens et une meilleure connaissance de leur opinion**, par l'exploitation de contenus qu'ils publient en ligne (V. le système *The Dublin Beat* qui recourt à un modèle entraîné à comprendre les opinions, réactions et sensibilités des habitants de Dublin pour exploiter chaque mois les tweets publics et restituer un « état de l'opinion » sous forme de tableaux de bord). Ils peuvent aussi inciter les collectivités à recourir plus fréquemment à des **consultations publiques**, les contributions pouvant être structurées et exploitées à l'aide d'un système adapté (exemple du système Pierino en Italie, qui a permis de traiter en moins d'une semaine 270 000 contributions en réponse à la consultation publique de l'Education nationale « La bonne école »).

Ces exemples de cas d'usage montrent l'exceptionnel potentiel des SIA pour la gestion de leur territoire par les collectivités territoriales. En raison des ressources et de la vision nationale dont il peut disposer, l'État pourrait, plus encore qu'aujourd'hui, se positionner comme fournisseur d'outils puissants pour les collectivités territoriales, dont la libre administration, pour être pleinement effective, doit être éclairée. Les SIA peuvent ainsi contribuer à parachever, dans les faits, la décentralisation des compétences et l'autonomie locale.

Fiche n° 2 : Défense

Le mythe du robot tueur par erreur a beaucoup fait pour discréditer le recours à des SIA, en particulier dans le monde de la défense. La réalité des faits est plus prosaïque, même si le secteur des armées pose de manière exacerbée les questions de la dépendance (comment conserver une défense efficace en dépit de la perte éventuelle de tout ou partie des capacités que lui donne l'IA), de la vulnérabilité (comment se défendre contre les ravages accrus d'attaques s'en prenant aux SIA), de la compétitivité (comment rester au niveau des adversaires potentiels), et de l'éthique (comme pour toute fonction de souveraineté, les questions de contrôle humain deviennent encore plus importantes que pour tout autre SIA).

Forte d'une recherche qui demeure compétitive dans le domaine de la défense, en dépit de la disparité des moyens comparés à ceux mobilisés par les Etats Unis[284], ou, peut-on supposer, par la Chine, la France suit depuis 2019 une feuille de route rendue publique par la ministre des armées le 5 avril 2019, qui a conduit à créer de nouvelles structures au sein du ministère pour coordonner les efforts et les développements.

La défense est sans doute le secteur dans lequel le déploiement de SIA est le plus ancien et le plus étendu, dans la mesure où les systèmes d'armes les plus sophistiqués incorporent tous un ou plusieurs SIA et/ou dépendent pour leur fonctionnement du recours à des SIA. C'est aussi l'un des domaines dans lesquels l'usage des SIA s'opère à toutes les échelles, depuis la gestion collective jusqu'à l'assistance individuelle.

L'administration des forces, loin des images futuristes véhiculées par Hollywood, est le principal usage actuel, et l'un des plus prometteurs, des SIA. Peu différent des usages constatés dans d'autres ministères, il peut s'agir de gestion du personnel (notamment en recourant à des *chatbots* pour les questions statutaires, ou le calcul des rémunérations, particulièrement complexe au regard des différentes positions et affectations), la gendarmerie ayant à cet égard pris une avance certaine en procédant à une action de sensibilisation de l'ensemble du personnel et déployé des SIA d'aide à la gestion ou à la relation avec certaines catégories d'agents. Il s'agira aussi le plus souvent de logistique et de maintenance, notamment pour les systèmes les plus complexes (aéronefs) pour lesquels la masse des informations à traiter à l'issue de chaque utilisation est très au-delà des capacités de traitement et d'interprétation des agents humains. La maintenance prédictive, la bonne gestion des stocks et du parc, les relations avec l'ensemble des partenaires industriels peuvent en être améliorées de manière très significative : contenir le coût de possession des systèmes d'armes (en termes d'effectifs associés et de coûts) et en améliorer la disponibilité sont des objectifs majeurs qu'on peut assigner aux SIA).

[284] Le rapport en crédits publics d'investissements directement affectés à l'IA est de l'ordre de 1 à 30 entre la France et les États-Unis.

Les premiers usages plus opérationnels sont liés aux capteurs et au traitement des données qui en sont issues. C'est tout particulièrement dans le domaine de **l'exploitation de l'imagerie** que des SIA s'avèrent performants (images satellites, images aériennes, notamment collectées par des drones), le rôle d'un SIA pouvant s'étendre de l'interprétation à la collecte, en interprétant en temps réel les images reçues par le système d'information d'un drone pour réorienter la programmation de son vol afin de se concentrer sur des zones dans lesquelles des objets jugés dignes d'intérêt méritent une observation accrue ou plus fréquente. Plus généralement, les SIA contribuent à donner une image complète d'un théâtre, d'un champ ou d'un terrain d'opérations en collationnant l'ensemble des données, issues des capteurs ou recueillies en milieu ouvert et en proposant leur interprétation – des avantages décisifs pouvant en résulter. L'information recueillie peut être exploitée depuis les états-majors jusqu'au combattant sur le terrain.

Leur capacité de traitement massifié de données complexes permet aux SIA de **contribuer à la conduite des opérations** en assistant celle-ci, d'une manière qui n'est pas très différente de l'assistance à la gestion de n'importe quel système complexe, qu'il s'agisse d'une centrale nucléaire, d'un avion, ou d'une formation tactique – en permettant d'avoir en temps réel une image consolidée, facilement accessible et partagée de paramètres variés, qui peuvent intégrer le dialogue entre des systèmes d'armes, pour combiner leur appréhension champ de bataille, ou les informations qu'ils donnent quant aux ressources encore disponibles ou aux capacités utilisables. Elles contribuent à l'accélération du cycle de décision à tous les niveaux, gage, aujourd'hui plus que jamais, de la conservation de l'initiative. En libérant tant les opérateurs que les états-majors de tâches automatiques ou routinières, les SIA leur permettent de se concentrer sur l'appréciation de situation, l'analyse et la conduite des opérations.

En s'approchant encore plus du combat, la France a, à ce stade, refusé tout déploiement de systèmes d'armes létaux autonomes (SALA), au sein desquels le déclenchement d'une action létale est entièrement laissé au fonctionnement d'un SIA à base de capteurs. Toutefois **un degré élevé de recours à des ouvertures du feu automatiques,** pouvant de ce fait avoir des conséquences létales, est d'ores et déjà observé dans le déploiement des matériels en usage aujourd'hui. Il en va ainsi par exemple des systèmes de défense antimissile (notamment ceux embarqués sur les navires), dont le délai de réaction, pour pouvoir être utilement employés, exclut le recours à une décision humaine (autre que de principe en amont, en activant le système) dont le délai d'intervention les priverait de toute efficacité. Certes, il ne s'agit ici pas de prendre la décision d'ouvrir le feu sur un être humain, mais la nuance est mince. De très nombreux autres automatismes de réaction, notamment pour la défense rapprochée de matériel sont d'ores et déjà à l'œuvre (avions, navires, véhicules blindés…). Les missiles de toute nature, même quand la décision de les tirer est encore humaine, n'atteignent et ne « choisissent » leur cible qu'au terme du fonctionnement du SIA qui en régit le vol et l'impact. De même, à bord des avions de combat moderne, la décision de principe demeure certes humaine, mais avant que le système d'armes lui-même ne définisse en toute autonomie l'ensemble des conditions qui amèneront à un tir efficace pour le déclencher de manière opportune.

La réflexion, la recherche et l'expérimentation portent aujourd'hui sur **l'autonomie relative de nombreux vecteurs associés** soit à un combattant au sol, le plus souvent embarqué dans un véhicule d'où il assure le « dialogue » avec les vecteurs associés, soit à un aéronef piloté, soit à un navire. La possibilité de déployer un « essaim » de tels drones terrestres, aériens ou maritimes (ou à changement de milieu) agissant à la demande, et sous contrôle humain, pour des tâches qui peuvent être de reconnaissance, d'observation, de défense rapprochée, qui peuvent permettre d'agir dans plusieurs dimensions, et de pénétrer sans risque pour la vie humaine en milieu hostile ou dangereux, constitue à ce stade l'une des pistes les plus riches de déploiement des SIA. C'est tout particulièrement dans ce cas que la spécificité des enjeux éthiques du déploiement de SIA militaire s'avère marquée : là comme ailleurs, pour être utile, la réflexion ne doit pas se concentrer sur les apparences (comme la « décision d'ouverture du feu » – dont l'opportunité et les conditions ne peuvent être appréciées que dans un contexte) mais intégrer l'ensemble des déterminants d'un processus et proposer des grilles de lecture d'une situation permettant à chaque fois, tant dans la définition de la doctrine au moment de sa mise en œuvre le respect des principes éthiques arrêtés.

La technologie déployée pour le développement de SIA militaires n'est pas substantiellement différente de celle à la base des SIA civils. La **dualité de ces technologies crée une vulnérabilité** particulière. L'un des domaines privilégiés de l'essor des SIA militaires portera notamment sur la sécurité de leur fonctionnement et la détection des intrusions comme la lutte contre les agressions par d'autres SIA hostiles. Dans ce registre, les SIA militaires doivent mettre un accent tout particulier sur l'intégrité et la sécurité de leurs données, notamment d'apprentissage, en raison de leur spécificité (les SIA liés à un matériel dépendent par exemple des données de leur fonctionnement et de leurs performances qui, par nature, sont des informations hautement confidentielles). Une autre particularité vient de l'exigence fréquente de frugalité des SIA en raison de la rareté de certaines données, notamment celles concernant les adversaires et leur matériel. Enfin, l'entraînement doit aussi apprendre à s'en passer, pour pouvoir continuer à agir en « mode dégradé ».

Le **nouveau champ de bataille immatériel qu'est l'espace cyber, et le nouveau champ de la lutte informationnelle**, à la frontière entre le renseignement et les opérations de défense, sont naturellement les lieux privilégiés du déploiement de SIA à vocation militaire. Le recours à des SIA à des fins défensives, leur utilisation pour parvenir à trouver les sources des agressions, mais aussi le cas échéant, une fois la décision politique prise, pour conduire des actions contre les forces hostiles, du type de celles dont la presse se fait régulièrement l'écho et qui sont invariablement prêtées, selon la nature des victimes, soit aux Russes et aux Chinois, soit aux Israéliens, font partie des usages possibles des SIA au service de la Défense nationale. Comme toutes les opérations de ce type, elles appellent un contrôle politique adéquat respectant le secret essentiel à leur succès.

Pour les armées, la « révolution du SIA » est déjà engagée. Elle se déroule très loin des fantasmes de robots tueurs, alors même que l'amélioration de la gestion (au sens économique comme au sens opérationnel), par exemple du groupe aérien du porte-avion, amène des économies et une efficacité incommensurables. Le changement

que ces techniques introduisent est sans doute aussi structurant que le fut l'entrée dans l'âge électronique à partir des années 1940, ou dans celui du moteur à explosion dans les années 1910. Comme toujours, la puissance ne résulte cependant pas de la technique – si son défaut assure la défaite, sa détention n'est jamais à elle seule la source de la victoire – mais dans la manière dont elle est employée, ce qui demande une révolution intellectuelle autant que technologique. À cet égard, les enjeux de formation et de recherche, plus encore peut-être que dans le domaine civil, sont essentiels pour que la France demeure dans le peloton de tête de la compétition internationale, auquel elle appartient, malgré la disproportion de ses forces.

Fiche n° 3 : Sécurité et activités d'enquête, de contrôle et de sanction

Ce sont sans doute les domaines de la sécurité, du renseignement et de l'ordre public et, plus largement, les activités d'enquête, de contrôle et de sanction qui suscitent, au regard de l'emploi des systèmes d'IA, le plus de craintes et de fantasmes sur les outils déployés par les pouvoirs publics. Loin des mythes portés par le cinéma d'espionnage et d'action, l'usage des SIA les plus innovants et, plus encore, leur efficacité opérationnelle s'avèrent encore modestes en la matière, surtout quand on les compare aux initiatives des opérateurs privés qui font l'objet d'une moindre vigilance citoyenne, alors qu'elles ne sont parfois guère moins intrusives.

Les SIA offrent d'importantes potentialités pour la détection des infractions, des manquements et des menaces pour les populations et les biens, la conduite des investigations (notamment judiciaires), ainsi que l'organisation de la réponse préventive et répressive. Ils peuvent être déployés dans de très nombreux contextes opérationnels (renseignement, police judiciaire, police administrative, contrôles administratifs, sécurité civile…) qui seront regroupés dans cette fiche, pour les besoins de la présentation des usages, en dépit d'une grande variabilité des enjeux politiques, opérationnels et juridiques qui s'y attachent[285].

1. L'acquisition de connaissances aux fins d'optimisation des contrôles et des décisions

Dans cette première famille de cas d'usage, les SIA constituent un moyen d'acquérir des connaissances sur les schémas infractionnels, les déterminants des comportements déviants ou la survenance de risques de sécurité par l'exploitation des données historiques. La mise en évidence de variables susceptibles d'influer sur le risque de commission d'un manquement ou sur l'existence d'une menace permet de mieux calibrer et orienter les moyens de contrôle déployés sur le terrain ou les décisions. On se bornera ici à citer quelques-uns des nombreux exemples d'utilisation de SIA[286] à cette fin.

Le ciblage du contrôle fiscal

La DGFIP a entrepris à partir de 2014 d'utiliser l'IA pour optimiser la programmation des contrôles fiscaux des entreprises, en recourant aux techniques d'apprentissage machine. L'outil développé dans le cadre du projet dit « Ciblage de la Fraude et

[285] Sur les usages en matière de sécurité, on pourra utilement se reporter au rapport au Premier ministre du député Jean-Michel Mis, *Pour un usage responsable et acceptable par la société des technologies de sécurité*, 2021.

[286] On peut aussi mentionner l'outil MonitorFish du centre national de surveillance des pêches qui définit un facteur de risque par navire de pêche afin d'aider au ciblage des contrôles, sur la base de l'historique de contrôle, la catégorie de pêcherie, la constatation en temps réel ou différé d'infractions, le projet CibNav prévoyant la construction d'un modèle statistique à la disposition des agents de la direction des affaires maritimes pour cibler les navires présentant un profil de risque particulier en termes de sécurité ou d'infraction aux règles de la navigation maritime…

Valorisation des Requêtes » (CFVR) offre un intéressant exemple de complémentarité homme/machine. Sur environ 180 facteurs de risques identifiés, une trentaine est issue de l'apprentissage supervisé à partir des données des contrôles passés et des données externes (notamment celles qui résultent de l'échange automatique d'information entre Etats), les autres résultant de la connaissance du métier, c'est-à-dire de l'expérience des agents chargés du contrôle, qui ont eux-mêmes et de longue date identifié des profils « à risque de fraude » et des paramètres-clés. De façon périodique, les entreprises sont passées au crible de ces risques et la cotation du niveau de risque permet d'effectuer une forme de hiérarchisation des dossiers, plus rapide que le travail de recoupement manuel sur la base de données issues de sources cloisonnées. Les informations sont mises à la disposition du directeur local, qui décide de sa stratégie de contrôle en en tenant compte. L'outil ne procède jamais de lui-même au déclenchement d'un contrôle et, a fortiori, à la notification d'un redressement. Ce ciblage a été étendu aux particuliers, qui font l'objet d'une cotation sur la base d'une cinquantaine de risques.

La DGFIP a également développé des modèles permettant la segmentation de populations homogènes du point de vue du risque de manquement à la loi fiscale, et des modèles non supervisés pour la détection d'anomalies – comme, par exemple, un écart significatif entre le montant d'une transaction immobilière et les prix habituellement pratiqués dans une situation comparable. Certaines tentatives de fraude au fonds de solidarité mis en place pour soutenir les entreprises frappées par la crise sanitaire ont été également été déjouées par le recours à un système automatisé simple, décelant les demandes redondantes et la discordance des numéros SIREN[287].

L'un des indicateurs de performance renseignés dans les documents budgétaires porte précisément sur la part des contrôles programmés à l'aide de l'intelligence artificielle et du « data mining ». Ce taux est passé de moins de 14% en 2018 à plus de 32% en 2020, avec une cible fixée à 50% en 2022, ce qui témoigne de l'impressionnante dynamique de cette forme d'optimisation du contrôle. Le montant des droits et pénalités mis à la charge des contribuables en 2020 à l'aide de ces contrôles avoisine 800 millions d'euros, ce qui est proche du taux observé pour la programmation purement humaine des contrôles, pour un coût significativement moindre puisqu'il est envisagé de redéployer plusieurs centaines d'agents sur les quelques 2000 équivalents temps plein affectés à l'analyse de risques et à la programmation des opérations de contrôle fiscal.

Le ciblage des contrôles douaniers

De façon comparable à la DGFIP et en s'appuyant sur l'expérience acquise par celle-ci, la direction générale des douanes et des droits indirects (DGDDI) développe des SIA destinés à déceler des atypies dans les flux de marchandises à l'import et à l'export ou les valeurs déclarées et à orienter les contrôles et le déploiement routier des brigades,

[287] Peut aussi être mentionnée le projet, accompagné par le dispositif EIG, d'identification et de contrôle des résidents fiscaux français percevant des revenus de l'étranger, par croisement de données nationales et internationales.

en exploitant les données massives issues des déclarations et des données d'autres administrations (analyse de la délinquance, données de l'aviation d'affaire…).

La détection des infractions financières par Tracfin

Tracfin, qui est notamment destinataire de déclarations de soupçons émanant en particulier des banques et comportant des milliers de données qui ne peuvent être exploitées humainement avec la même performance que la machine, a développé en interne un outil, qui comprend entre autres une brique basée sur l'apprentissage supervisé, permettant d'identifier des récurrences (schémas de fraude), de coter le risque de fraude et de construire une cartographie visuelle des situations qui méritent plus particulièrement des investigations. Un modèle algorithmique de traitement du langage naturel, réentraîné pour les besoins spécifiques du service, est utilisé pour la détection des entités nommées.

Il est intéressant d'observer que la performance de Tracfin est largement tributaire de la qualité de la détection des soupçons par les déclarants, en particulier les établissements bancaires, et qu'il est donc légitime que l'autorité publique compétente s'assure de la fiabilité et de la robustesse des traitements algorithmiques mis en œuvre par ces derniers à cette fin.

L'analyse dite prédictive en matière pénale

L'un des usages des SIA les plus anciens[288] et médiatisés dans le domaine de la sécurité porte sur l'anticipation du risque qu'une infraction pénale soit commise, par une personne donnée ou dans une situation donnée.

L'usage le plus controversé concerne ce qu'on peut appeler « l'analyse prédictive personnelle », c'est-à-dire **l'évaluation du risque de commission d'infraction et, en particulier, de récidive** en fonction des caractéristiques d'une personne, afin, le cas échéant, de fixer le quantum de la peine ou de prendre des décisions d'aménagement de peine.

Le système COMPAS, utilisé dans certains États des États-Unis, utilise 137 données d'entrée pour fournir un score sur deux ans concernant le risque de récidive d'un prévenu ou d'un condamné, le risque de comportement violent et le risque de non-comparution en l'absence de détention provisoire. Dans un arrêt très commenté (*State v. Loomis*), la Cour suprême du Wisconsin a jugé que l'utilisation d'un tel SIA ne méconnaissait pas en soi le droit du prévenu à un procès équitable, en dépit des études (notamment celle de l'organisation Propublica) qui tendent à montrer une surreprésentation des personnes d'origine afro-américaine dans les situations de « faux positif » (cas dans lesquels l'individu se voit affecter un score élevé de récidive mais ne récidive pas dans les faits) par rapport aux populations blanches. Cette distorsion vient notamment de l'utilisation du lieu de résidence, qui, compte tenu de la composition ethnique des quartiers, fournit une probabilité d'appartenance à telle

[288] V. notamment le rapport – critique – du ministère de la justice américain de novembre 2014, *Predictive analytics in law enforcement : a report by the Department of Justice*.

ou telle origine ethnique, alors même que cette dernière donnée n'est pas renseignée dans COMPAS.

Le système LSI-R (*Level of Service Inventory-Revised*), également en service aux Etats-Unis, calcule de la même façon un score de risque de récidive sur la base de 54 données d'entrée, porté à la connaissance de l'administration pénitentiaire et de la justice afin d'examiner les demandes de libération conditionnelle et, plus particulièrement, d'allouer davantage de ressources d'accompagnement et de suivi aux détenus à haut risque. La transparence du système a été mise en cause dans l'affaire *Kansas c/ Walls* de 2017, dans laquelle la cour d'appel du Kansas a jugé que le détenu devait pouvoir accéder, au-delà de la fiche de résultats, au rapport complet du système pour pouvoir en contester efficacement la fiabilité.

La police de Durham, au Royaume-Uni, a quant à elle développé le système HARM (*Harm Assessment Risk Tool*) qui attribue un score en fonction du risque faible, moyen ou élevé de récidive ou, plus largement, de commission d'une nouvelle infraction dans les deux années suivant la condamnation. A cette fin, les données personnelles de plus de 100 000 délinquants ont été compilées (âge, sexe, origine géographique, parcours judiciaire…) qui permettent d'identifier et de pondérer les variables supposément explicatives de la délinquance.

En Espagne, l'outil VioGen évalue le risque qu'une personne soit victime d'une agression de la part d'un prévenu ou d'un condamné pour violences conjugales ou sexuelles, orientant les efforts de protection. La police espagnole estime que ce système a significativement contribué à réduire l'occurrence des agressions.

L'autre catégorie d'usages porte sur **l'évaluation de la probabilité qu'une infraction soit commise à un endroit et à un moment donnés**. C'est « l'analyse prédictive situationnelle »[289]. En complément de l'expertise des agents fondée sur leurs observations de terrain, l'exploitation automatisée des données de la criminalité et de la délinquance passées permet d'envisager la construction de modèles plus ou moins sophistiqués calculant le risque sécuritaire en un lieu et à un moment donné, en fonction de variables diverses qui ne se réduisent pas aux infractions déjà commises à tel ou tel endroit, mais qui intègrent des variables explicatives plus « profondes », comme les caractéristiques démographiques ou socio-économiques, l'heure du jour ou de la nuit, la température, la desserte et la proximité de certaines installations… L'ambition sous-jacente est de dégager, à partir des données historiques, des « lois d'évolution de la délinquance » qui échappent aujourd'hui largement aux experts.

Largement répandus aux Etats-Unis, ces outils connaissent également un certain succès en Europe, notamment aux Pays-Bas avec le « système d'anticipation des infractions »[290] et le « *Burglary Predictor* » qui fournit une probabilité qu'un ou

[289] V. sur ce sujet, notamment, le rapport de l'Institut d'aménagement et d'urbanisme d'Ile-de-France, *La police prédictive - Enjeux soulevés par l'usage des algorithmes prédictifs en matière de sécurité publique*, avril 2019.

[290] V. pour une analyse critique : S. Oosterloo et G. van Schie, *The Politics and Biases of the « Crime Anticipation System » of the Dutch Police*, http://ceur-ws.org/Vol-2103/paper_6.pdf

plusieurs cambriolages soient commis dans un quartier de la ville d'Utrecht au cours d'une semaine, en tenant compte des conditions météo, des vacances et évènements particuliers, des caractéristiques socio-démographiques du quartier,de l'éclairage public, des caméras de vidéoprotection, des arbres, places de stationnement, bennes à ordures… La police de Zürich utilise également un logiciel de « prédiction » des cambriolages.

La France n'est pas en reste. L'ancien Observatoire national de la délinquance et des réponses pénales utilisait à cette fin un algorithme dénommé RTM (*Risk Terrain Modeling*) basé sur des facteurs économiques et sociaux identifiés comme des variables environnementales de la délinquance. De façon plus opérationnelle, la gendarmerie nationale expérimente depuis près de dix ans des outils inspirés des démarches de police prédictive, fondés à la fois sur la commission des actes passés (historique des infractions et dynamique récente) et sur 600 variables d'ordre socio-économique ou territorial constituant autant de facteurs d'influence. Dans un esprit proche, certaines communes se dotent d'outils d'analyse statistique et cartographique de la délinquance leur permettant d'orienter l'action de la police municipale. La ville de Marseille conduit, avec l'aide de financements européens, un projet « Big data de la tranquillité publique » dont l'ambition est de prévoir et prévenir la survenance de certains évènements par l'exploitation de données issues des services publics opérant dans la ville.

Les outils relevant de l'analyse prédictive situationnelle en matière de sécurité font l'objet de plusieurs critiques : d'une part, il est soutenu que leur efficacité n'est pas clairement démontrée sur le plan opérationnel, en comparaison des connaissances dont disposent déjà les agents de terrain, ce qui a conduit certaines collectivités à en abandonner l'usage (cas du logiciel PredPol[291] délaissé par la police de Santa Cruz en Californie ou celle du comté du Kent au Royaume-Uni) ; d'autre part, est dénoncé le phénomène de « prophétie auto-réalisatrice » qui résulte de ce que le renforcement des contrôles dans un secteur géographique donné augmente la probabilité que des infractions y soient détectées, donc celle que le modèle recourant à ces nouvelles données d'entrée recommande un nouveau durcissement du dispositif répressif ; enfin, et en lien avec la critique précédente, les biais discriminatoires dont les modèles peuvent être affectés sont dénoncés en raison de la corrélation historique entre l'occurrence des contrôles dans un quartier et la proportion de la population étrangère ou immigrée.

Le déploiement des moyens d'intervention de la sécurité civile

La compréhension des déterminants des catastrophes et des sinistres et la capacité à anticiper leur survenance constituent un enjeu majeur pour les services en charge de la sécurité civile, tant pour les choix d'implantation des moyens que pour leur dimensionnement au jour le jour. Les SIA offrent des perspectives prometteuses en la matière.

[291] Cet outil propose une carte de « points chauds » sur la base d'une analyse cinétique de la criminalité inspirée de la propagation sismique.

La **prévention des risques naturels** se prête tout particulièrement au déploiement de SIA afin de mieux les prévoir et de déclencher le plus tôt possible des alertes, qu'il s'agisse des inondations (chaire Hydr.IA à Alès, lancement de projets expérimentaux à Louvain et Rhode-Saint-Genèse en Belgique, déploiement d'un système d'alerte en Inde et au Bangladesh…), des feux de forêts (cartographie des zones de végétation à débroussailler pour la limiter et protéger les habitations et les réseaux en Australie, outil Radfire du Laboratoire national Nord-ouest Pacifique aux Etats-Unis pour détecter les départs de feux déclenchés par la foudre à partir des images satellite et contrôler l'utilisation des produits ignifuges, prévision de la propagation des incidences en Bolivie…), des tremblements de terre (identification et cartographie des failles sismiques grâce à la vision par ordinateur appliquée aux images satellite[292], compréhension du phénomène par l'exploitation des statistiques à l'aide de l'apprentissage profond[293], prévision précoce des sinistres[294], anticipation des répliques[295]) ou des avalanches.

Sur un plan opérationnel, le service départemental d'incendie et de secours du Doubs a développé à partir de 2018, en partenariat avec un laboratoire de recherche (Femto-ST, IUT de Belfort-Montbéliard), et mis en service en février 2021 un outil baptisé PrédictOps, qui utilise 1200 variables (données météorologiques, éphéméride, données de trafic routier, qualité de l'air, niveau des cours d'eau, données épidémiologiques…) et un modèle entraîné sur les données issues des 40 000 interventions effectuées par les sapeurs-pompiers de ce département chaque année, afin de prévoir aussi bien les périodes de crue que les pics d'accidents de la circulation ou de suicide, ou la survenance d'arrêts cardiaques, et, en conséquence, de déterminer le niveau de sollicitation du SDIS à une échéance comprise entre une heure et 72 heures, selon le type de sinistre. Les moyens humains et matériels peuvent être adaptés en conséquence, en mobilisant les personnes sous astreinte, les pompiers volontaires voire les équipes du département le plus proche. Les résultats obtenus en milieu urbain semblent probants.

Un système comparable, limité aux accidents autoroutiers, a été développé aux Pays-Bas par le ministère des infrastructures et de l'environnement (« Incident Management »). L'exploitation des données d'accidentologie disponibles pour les années 2012 à 2016 a permis d'identifier et de pondérer les variables explicatives, de fournir la probabilité d'un accident à tel endroit à brève échéance, et d'organiser en conséquence le déploiement des 260 inspecteurs routiers et des 160 gestionnaires de trafic pour intervenir le plus rapidement possible et éviter la congestion du trafic.

---

[292] L. Mattéo et a., *Automatic Fault Mapping in Remote Optical Images and Topographic Data With Deep Learning, Journal of Geophysical Research*, Solid Earth, 2021.

[293] Q. Kong, D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, P. Gerstoft, *Machine Learning in Seismology: Turning Data into Insights*. Seismol. Res. Lett. 90, 3-14 (2019).

[294] S. Mostafa Mousavi et a., *Earthquake transformer – an attentive deep-learning model for simultaneous earthquake detection and phase picking*, Nature Communications, 11, n° 3952, 2020.

[295] A. Mignan, *Forecasting aftershocks : Back to square one after a Deep Learning anticlimax*, Temblor, 2019, critiquant le recours à l'apprentissage profond à cette fin et plaidant pour des modèles simples d'apprentissage machine et de régression.

2. La détection des infractions et menaces par l'exploitation massive des données en ligne

Noyées dans un gigantesque flot permanent de données, les données révélant des actes de cybercriminalité comme celles qui sont utiles à la prévention et à la répression d'infractions commises dans le monde physique deviennent de fait invisibles à l'œil humain. Les repérer, les qualifier et en tirer les conséquences opérationnelles devient une priorité de l'action des services compétents, qu'ils ne peuvent assumer qu'à l'aide de systèmes d'IA. Aucun service n'est en effet en mesure de réunir des milliers de collaborateurs pour veiller constamment sur la toile, dans toutes les langues. La capacité de traitement surhumaine des SIA offre au contraire la possibilité d'une exploitation massive des flux de données générées, directement ou *via* d'autres systèmes d'IA, par l'activité humaine – pratique parfois connue sous le terme anglo-saxon de « *web scrapping* » ou « *data scrapping* ».

Concrètement, les SIA sont à même de filtrer et trier les données de toute nature (texte, voix, image…) circulant sur internet, qu'il s'agisse de données de contenus (texte d'un courriel, vidéos mises en ligne sur un site internet, messages échangés *via* une application de messagerie instantanée…) ou de « méta-données » (comme les données relatives à l'origine, à la destination et à la durée des échanges, les données de localisation des matériels utilisés…). Ce filtrage peut être effectué selon différents procédés : il est possible de recourir à des critères déterminés par les agents sur la base de leurs connaissances métier, acquises par l'expérience ou par le renseignement humain préalable (repérage d'un suspect préalablement identifié par exemple) ; d'utiliser, ainsi qu'il a été dit au point 1, des critères révélés par une analyse automatisée des infractions et menaces passées, qui ont permis de mettre en évidence des schémas récurrents, des corrélations et, après un travail de qualification, des variables causales ; ou encore, sans même utiliser de critères de ciblage pré-déterminés et en recourant le cas échéant à des techniques d'apprentissage non supervisé, d'identifier des singularités dans un flux de données suffisamment normé, susceptibles de révéler des manquements, qui ne sont rien d'autres que des écarts à la norme, c'est-à-dire des anomalies (statistiques).

Les systèmes ne sont pas limités au simple paramétrage de mots-clés (« drogue », « bombe »…) permettant de faire remonter les courriels ou les sites qui les contiennent, mais recourent à l'analyse sémantique pour comprendre le thème, le sens général et certaines caractéristiques d'une conversation, en utilisant s'il y a lieu des outils de traduction automatisée afin de s'affranchir de la barrière de la langue.

Ces outils ne permettent donc pas seulement d'effectuer une recherche ou une surveillance ciblée dans une masse de données considérable (on sait ce qu'on cherche ou qui on cherche et le système permet de démultiplier les moyens d'investigation pour le retrouver dans un vaste espace), mais aussi d'identifier les infractions ou les menaces elles-mêmes, avant de procéder à des investigations ciblées (on sait qu'on cherche des infractions d'une nature donnée et le système extrait du flux de données celles qui lui paraissent mériter de plus amples vérifications). Il y a lieu d'observer qu'il n'y a là aucun changement de nature dans l'activité de contrôle elle-même : la police peut tout aussi bien sillonner un quartier

à la recherche d'un cambrioleur en fuite qui lui a été signalé ou y patrouiller de façon plus « passive » dans l'éventualité qu'une infraction y soit commise, en repérant les comportements suspects. Ce qui fait la spécificité des systèmes, et des risques qui les accompagnent, c'est la démultiplication de la surveillance et l'ampleur des données auxquelles elle s'applique, qui peut révéler une partie importante de la vie privée des utilisateurs d'internet. Ces dispositifs traduisent ainsi une ingérence dans le droit au respect de la vie privée, plus ou moins forte selon qu'elle porte ou non sur des données que les personnes ont délibérément et manifestement choisi de rendre publiques elles-mêmes, mais aussi dans la liberté d'expression, les citoyens étant d'autant plus susceptibles de s'autolimiter ou de s'autocensurer dans l'expression qu'ils savent que ce qu'ils diront et feront « pourra être retenu contre eux » (cet effet de dissuasion est désigné par le terme anglo-saxon « *chilling effect* »).

Là encore, les cas d'usage existants ou en développement sont nombreux et on se bornera à n'en citer que quelques-uns.

La **technique dite de « l'algorithme »**, expérimentée à partir de 2015 puis pérennisée par la loi n° 2021-998 du 30 juillet 2021[296], est utilisée par les services spécialisés de renseignement aux fins de détecter des menaces terroristes révélées par les données de connexion. Les opérateurs de communications électroniques sont tenus d'appliquer aux flux de données qu'ils gèrent des traitements dont les caractéristiques sont définies par les services de renseignement sur la base de signaux (faibles ou forts) susceptibles de caractériser la menace terroriste. Ces traitements déclenchent des « alertes » qu'il appartient aux services de qualifier à brève échéance sur la base de données anonymisées : soit il s'agit, *a priori*, d'une fausse alerte et les données doivent être supprimées ; soit il existe une suspicion suffisamment sérieuse de menace terroriste méritant des investigations plus approfondies, et l'anonymat des personnes concernées par les données peut être levé sur autorisation du Premier ministre, après avis de la Commission nationale de contrôle des techniques de renseignement.

La **cybersécurité** est également un champ de déploiement crucial des SIA. Sur la base de l'apprentissage du comportement standard d'une infrastructure, il est possible d'identifier des phénomènes anormaux pouvant en compromettre la sécurité. Plus largement, l'exploitation automatisée des flux de données peut permettre de mettre au jour des pratiques suspectes susceptibles de révéler des menaces, attaques et tentatives d'intrusion diverses (V. le projet VDI de l'Autorité de sécurité nationale norvégienne qui entend déployer un modèle d'apprentissage automatique pour l'identification de virus). Les SIA ont également la capacité technique de traiter automatiquement une grande quantité d'informations disponibles sur le web conventionnel ou le *dark web* afin d'identifier les attaquants actifs, les modes opératoires, les vulnérabilités les plus exploitées… ce qui permet d'orienter la pose des « marqueurs techniques » pour la détection des évènements pouvant affecter la sécurité des systèmes d'information (V. sur ce dernier point l'article L. 2123-2-1 du code de la défense et l'article L. 33-14 du code des postes et des communications électroniques).

[296] Art. L. 851-3 du code de la sécurité intérieure.

Le **service Viginum**, logé au sein du Secrétariat général de la défense et de la sécurité nationale, opère un traitement de données à caractère personnel[297] dans le but d'identifier les ingérences numériques étrangères, notamment en période électorale. La détection de tentatives de manipulation de l'information (« infox ») est effectuée par application de critères pré-déterminés par l'expérience des agents aux contenus accessibles au public. Il s'agit, à ce stade, d'un modèle algorithmique basé sur les règles, sans recours à l'apprentissage automatique. Le résultat de la collecte conduit ensuite à des analyses plus approfondies par les experts des services sur l'existence de tentative de manipulation et leur origine, qui sont alors portées à la connaissance du gouvernement – lui seul décide des suites à donner et de leur nature. Doté à terme d'une soixantaine d'agents, le service Viginum serait dans l'incapacité matérielle de traiter par des moyens humains ne serait-ce qu'une petite fraction des données que le système déployé analyse.

La DGFIP a été autorisée, par l'article 154 de la loi n° 2019-1479 du 28 décembre 2019 de finances pour 2020, à expérimenter un **outil de collecte et d'exploitation de données publiquement disponibles en ligne, sur les sites internet des opérateurs de plateforme en ligne, à des fins de contrôle de certaines infractions fiscales**. Le décret n° 2021-148 du 11 février 2021, qui en détaille les caractéristiques, prévoit une phase d'apprentissage au cours de laquelle seront conduites « *la modélisation et l'identification des caractéristiques des comportements susceptibles de révéler la commission des infractions et manquements* » entrant dans le champ du dispositif et l'identification d'indicateurs et de critères de pertinence, et une phase d'exploitation au cours de laquelle le modèle sera appliqué aux données recueillies sur les plateformes.

Le **pôle d'expertise de la régulation numérique (PEReN)**, service à compétence nationale créé par le décret n° 2020-1102 du 31 août 2020 afin d'appuyer les services de l'État dans leur mission de régulation des plateformes numériques, recourt également à la collecte et au traitement massifs de données sur ces plateformes, dans des conditions précisées par l'article 36 de la loi n° 2021-1382 du 25 octobre 2021. A l'aide de services *cloud* externes et, désormais, de serveurs internes utilisant des GPU, il développe des SIA permettant, par exemple, de s'assurer que le modèle algorithmique d'un comparateur proposant des recommandations (de restaurants, d'hôtels, de voyages...) fonctionne conformément aux allégations de son gestionnaire, sans biais de nature à induire en erreur les consommateurs. L'un des axes de travail, en partenariat avec l'INRIA, est l'audit des algorithmes en « transparence faible », qui permet, par l'analyse statistique de traces, d'inférer certains principes de fonctionnement du modèle algorithmique à défaut de disposer d'un accès direct au code source. Il concourt également à la détection de contenus viraux sur les réseaux sociaux (évaluation d'outils utilisant les métadonnées des messages et les graphes de diffusion plutôt que leur contenu) et les comportements suspects sur les plateformes, notamment l'utilisation de robots susceptibles de nuire à leur bon fonctionnement par la réalisation d'opérations multiples et artificielles (par exemple, pour la manipulation de cours de bourse). Le PEReN a également

[297] Décret n° 2021-1587 du 7 décembre 2021.

contribué au développement d'un outil permettant de suivre les évolutions, le cas échéant subreptices, des conditions générales d'utilisation des plateformes.

Les **autorités de concurrence** envisagent également de recourir à des SIA pour détecter et qualifier des ententes par collusion algorithmique, qui peuvent prendre les formes les plus diverses, d'un contrôle automatisé des prix pratiqués au sein d'un système organisé à la fourniture par un même prestataire du même algorithme (ou d'algorithmes coordonnés) à des concurrents permettant un alignement tarifaire[298]. La caractérisation de l'intention collusive peut d'ailleurs s'avérer extrêmement délicate en raison de l'effet de « boîte noire » des modèles reposant sur les réseaux de neurones.

A l'étranger, le Land de Rhénanie-du-Nord Westphalie a déployé un outil d'analyse automatisée des images circulant sur Internet afin de détecter des **contenus pédopornographiques**. Le taux d'exactitude du modèle est évalué à plus de 90%.

Au Royaume-Uni, le ministère de la justice a entraîné un réseau de neurones en matière de traitement de langage naturel, à partir de 500 rapports d'inspection des prisons (représentant 250 000 phrases) et l'a ré-entraîné grâce aux nouveaux rapports rédigés, afin d'identifier les situations à risque et les facteurs explicatifs d'incidents et d'orienter les décisions et les inspections en conséquence.

Dans un tout autre domaine, le ministère des transports britannique a utilisé l'apprentissage non supervisé pour regrouper les garages effectuant le contrôle technique des véhicules en fonction de la façon dont ils effectuent les contrôles. Croisé avec les données historiques sur les procédures d'infractions engagées à l'encontre des garagistes ne respectant pas les standards de qualité, l'outil a permis d'affecter un score de risque à chaque professionnel et de cibler les contrôles en conséquence. Le temps consacré à la préparation des contrôles inopinés effectués par les examinateurs a baissé de 50 % selon le gouvernement.

Le projet expérimental « *Test Balloon* » en Suède vise à détecter des **publicités** déguisées et/ou trompeuses dans les publications en ligne et, en particulier dans les post diffusés sur l'application Instagram par des influenceurs, grâce à un outil issu du marché et adapté aux besoins de l'agence de protection des consommateurs (*Konsumentverket*), combinant la fonctionnalité de reconnaissance d'images et celle de traitement du langage naturel.

3. La détection d'infractions ou de risques dans le monde physique

Les SIA peuvent être adossés à des infrastructures physiques de collecte de données de terrain afin de détecter des infractions ou des menaces. Cette famille de cas d'usage recourt essentiellement à la vision par ordinateur[299].

---

[298] V. sur ce sujet l'étude conjointe de l'Autorité de la concurrence et du Bundeskartellamt, *Algorithmes et concurrence*, novembre 2019.

[299] Peuvent également être utilisés des SIA de traitement automatique du son ou de traitement des données issues de capteurs thermiques (cas de l'expérimentation menée par la ville de Metz pour la détection des piétons aux carrefours afin de moduler le temps du feu

Un des usages les plus avancés est la **reconnaissance faciale à des fins d'identification des auteurs d'infraction**. Le traitement des antécédents judiciaires (TAJ), utilisé par la police et la gendarmerie nationales, dispose depuis 2013 d'un module permettant de faire des rapprochements entre photographies anthropométriques, mais aussi avec des images de vidéosurveillance ou des réseaux sociaux, sur requête d'un agent. En 2019, 375 000 demandes avaient été effectuées[300]. Interpol dispose également d'un système automatisé de reconnaissance faciale, alimenté par les États membres, sans, toutefois, s'agissant de la France, de transmission automatique des photographies du TAJ.

Les dispositifs de contrôle automatisé de la vitesse (**radars automatiques**) utilisent cette brique logicielle : un SIA lit les caractères de la plaque d'immatriculation du véhicule « flashé » telle qu'elle apparaît sur la photo. Le système d'immatriculation des véhicules permet alors d'identifier le titulaire du certificat d'immatriculation, auquel il convient d'adresser l'avis de contravention. Désormais, le contrôle porte également sur les défauts d'assurance, par croisement avec le fichier des véhicules assurés. Il y a lieu de préciser que la verbalisation n'est pas intégralement automatisée, puisqu'un officier de police judiciaire doit constater l'infraction. Plus récemment, dans le cadre du projet IA Flash qui s'inscrit dans le programme « Entrepreneurs d'intérêt général », la vision par ordinateur est utilisée pour vérifier la cohérence entre le numéro de plaque, d'une part, et la marque et le modèle du véhicule, d'autre part, et, ainsi, limiter les erreurs de verbalisation. Des investigations peuvent être lancées pour identifier l'origine de l'incohérence, qui peut aussi bien être une erreur dans la confection de la plaque qu'une erreur de lecture ou une usurpation.

La vision par ordinateur permet aussi de détecter des automobilistes tenant un téléphone portable en main – tel est le cas en Australie, en Belgique (expérimentation de l'institut Vias), aux Pays-Bas, et, désormais, en France, avec les « radars-tourelle » en cours de déploiement – ou de caractériser des infractions à des carrefours (outil CANARD en Pologne).

Il peut être recouru à la lecture automatique des plaques d'immatriculation (LAPI) pour de nombreux autres usages. Des véhicules de police municipale équipés de caméras sont ainsi utilisés pour le contrôle du **stationnement** payant et l'envoi des forfaits post-stationnement. Certaines communes y ont eu recours pour la verbalisation du stationnement gênant et ont fait l'objet d'une mise en demeure de la CNIL en août 2020, en l'absence de texte autorisant leur utilisation aux fins de la recherche d'infractions pénales.

vert pour fluidifier et sécuriser la traversée – cet outil ne recourant toutefois pas à l'apprentissage automatique).

[300] Assemblée nationale, Avis au nom de la commission des lois sur le projet de loi de finances pour 2021, Tome VII, Sécurités.

Les **images aériennes et satellites** constituent également un gisement de données considérable que les services en charge de missions de contrôle et d'enquête peuvent exploiter afin de démultiplier les investigations. Deux domaines sont particulièrement emblématiques des potentialités des SIA :

- La **police de l'urbanisme** : la direction départementale des territoires et de la mer (DDTM) de l'Hérault conduit à cette fin un projet basé sur l'apprentissage supervisé, sélectionné dans le cadre de l'appel à manifestation d'intérêts piloté par la DINUM, visant à détecter les occupations et utilisations irrégulières du sol, en rapprochant les constructions et aménagements identifiées à l'aide d'un SIA de vision par ordinateur des autorisations d'urbanisme délivrées par les autorités compétentes. La DGFIP a de son côté lancé le projet dit « Foncier innovant » qui s'appuie sur les données de l'IGN et un modèle de reconnaissance de formes et d'objets basé sur l'apprentissage machine afin d'identifier des bâtiments nouveaux, des extensions de bâtiments existants ou des aménagements (comme des piscines) réalisés sans autorisation ou en contravention avec une autorisation d'urbanisme (en ce qui concerne la surface, le positionnement…). En cas de discordance entre les informations ainsi collectées et les éléments déclarés au titre des impôts directs locaux, un agent de l'administration fiscale est averti, vérifie l'existence ou non d'un manquement et met en demeure le contribuable de régulariser sa situation s'il y a lieu. Le projet est expérimenté dans neuf départements et centré à ce stade sur la détection des piscines. Sa généralisation est prévue en 2022 ;

- La **politique agricole** : plusieurs pays recourent à des SIA exploitant les images satellite et aériennes pour contrôler l'usage des sols et vérifier si les conditions d'octroi des subventions dans le cadre de la Politique agricole commune sont remplies. En Wallonie, les moyens humains permettaient, par les visites de terrain, de contrôler environ 5% des terres agricoles ; le taux est de 100% grâce à un SIA. En Estonie, le système SATIKAS est capable de détecter si une prairie a été fauchée ou non grâce à un réseau de neurones appliqué aux données satellitaires de Copernic ; l'objectif du gouvernement estonien est d'être en capacité d'identifier la nature des cultures et les espèces d'arbres.

Les fonctionnalités de vision par ordinateur sont également susceptibles d'être exploitées à des fins de **sécurisation de lieux ouverts au public**, selon deux modalités distinctes.

La moins intrusive consiste à **détecter des anomalies ou situations à risque, sans procéder à l'identification d'une personne physique**. Ces caméras dites intelligentes (ou augmentées) associent des dispositifs fixes ou mobiles (y compris aéroportés) de captation d'images, qui peuvent être de simples caméras de vidéoprotection comme il en existe d'ores et déjà dans de nombreuses communes ou des équipements beaucoup plus performants en termes de définition et de portée, et un système d'IA qui analyse le flux de données et qualifie les objets (au sens large) et les activités. L'outil peut ainsi signaler aux agents humains chargés de sa mise en oeuvre des situations aussi diverses qu'un mouvement de foule, la chute d'une personne dans la rue ou sur une voie ferrée, le stationnement d'un véhicule à un endroit non

autorisé, la divagation d'un animal dangereux, la présence d'un attroupement ou d'une personne répondant à un signalement donné, un bagage abandonné, une arme à feu dans la main d'une personne[301], un comportement suspect quelconque[302]… Il peut aussi, le cas échéant, détecter des émotions à travers des gestes ou des expressions du visage. De façon moins intrusive, le système peut produire de simples statistiques permettant de mieux connaître certains phénomènes ou d'orienter des contrôles (exemple de la comptabilisation des personnes portant le masque dans les transports en commun : décret n° 2021-269 du 10 mars 2021).

Le projet envisagé pour assurer la sécurité des Jeux Olympiques de Paris 2024 en est une illustration. Les foules à gérer seront considérables et le risque d'attentat terroriste, d'agressions et de mouvements de panique sera particulièrement élevé. Le déploiement de plusieurs milliers de « caméras intelligentes » serait de nature à dissuader certains comportements criminels ou délinquants (puisqu'il serait porté à la connaissance du public), à faciliter l'interpellation des fauteurs de troubles, ou encore à retrouver un enfant qui s'est égaré dans la foule.

Le domaine de la **surveillance des frontières** se prête également à l'utilisation de ces SIA, qui sont à même de détecter, à partir d'images prises par des avions, des hélicoptères, des drones[303] ou des tours de contrôle, des embarcations transportant des migrants ou des tentatives de franchissement illégal de frontière.

Ces SIA présentent, sur le travail humain, l'avantage d'analyser en continu un plus grand nombre d'images avec une acuité globalement supérieure, y compris, le cas échéant, de nuit ou dans des conditions météorologiques difficiles et à un moindre coût. Ils ne sont en revanche pas sans risque du point de vue de la protection des droits et libertés fondamentaux, à trois égards : une première ingérence, diffuse, porte sur l'influence que le déploiement de tels dispositifs est susceptible d'exercer sur les comportements et l'équilibre psychologique des citoyens ; la deuxième, de nature « pathologique », porte sur les risques de détournement de finalités, dont l'ampleur défend des dispositifs techniques de « bridage » et de sécurisation (y compris contre les cyberattaques) ; la troisième porte sur les décisions et actions susceptibles d'être prises ou entreprises par des agents humains sur la base des informations fournies par le SIA, notamment si celles-ci déclenchent l'organisation

[301] V. par ex. l'outil développé par la société américaine ZeroEyes composé d'anciens militaires, qui vise à identifier en temps réel les armes à feu apparentes dans des bâtiments publics et à alerter les forces de l'ordre en quelques secondes. Le modèle est optimisé pour reconnaître les armes classiquement utilisées dans les fusillades de masse, récurrentes aux États-Unis.

[302] Une entreprise japonaise a conçu un outil (AI Guardman) destiné à détecter les vols à l'étalage à partir du langage corporel des auteurs, en comparant les gestes effectués avec des « modèles suspects » prédéfinis (regards alentours pour s'assurer de l'absence de témoin visuel, prise d'un produit rapidement glissé dans le sac…).

[303] V. le projet Roborder expérimenté au Portugal, en Grèce et en Hongrie, qui vise, à partir des données collectées par des capteurs optiques et thermiques embarqués dans divers moyens de transport maritimes, terrestres et aériens, à détecter des activités suspectes ou des accidents (pollution maritime…) aux frontières.

d'un contrôle sur place ou l'activation, automatique ou manuelle, d'autres fonctionnalités emportant une ingérence supérieure dans les droits et libertés, notamment l'identification des personnes physiques (V. *infra*).

Il y a lieu de relever que, quand bien même les images font-elles l'objet, dès leur captation et avant leur transmission à un agent, d'un traitement destiné à empêcher toute identification par un humain (floutage…), elles constituent des données à caractère personnel soumises aux règles applicables en fonction de la finalité du traitement – en particulier le titre III de la loi du 6 janvier 1978, pris pour la transposition de la directive « police-justice » pour ce qui concerne les caméras utilisées à des fins de police administrative ou judiciaire (V. sur ce point : CE, 22 décembre 2020, *Association La Quadrature du Net*, n° 446155, T.). La portée concrète de ces règles et, notamment, la question des droits des personnes (droit d'information et d'opposition…) et celle de la nécessité d'une base légale dédiée, a donné lieu à un projet de position de la CNIL soumis à consultation publique[304].

L'usage le plus intrusif en la matière est, évidemment, celui qui consiste à **identifier les personnes physiques dans les lieux sous surveillance, notamment par l'utilisation de leurs caractéristiques biométriques** et, singulièrement, la reconnaissance faciale. Cette fonctionnalité a connu de très importants progrès dans la période récente[305].

Concrètement, cette identification est réalisée en confrontant ces caractéristiques, telles qu'elles ressortent des images, aux gabarits biométriques figurant dans un ou des fichiers auxquels les services ont accès (par exemple, le fichier « traitement des antécédents judiciaires »). Ce type de SIA, qui fonctionne selon le même principe que l'authentification par reconnaissance des empreintes digitales ou du visage à la demande d'une personne (par exemple, pour déverrouiller un smartphone), est susceptible d'être utilisé aussi bien pour des contrôles d'accès à des installations (portiques) que pour le repérage « diffus » de personnes dans des rassemblements.

Plusieurs expérimentations ont été conduites en la matière (test fictif avec des volontaires au carnaval de Nice en 2019, fluidification des flux de passagers par Aéroport de Paris et pour le tournoi de Roland Garros 2020…) dont certaines ont donné lieu à des interventions de la CNIL (contrôle d'accès des élèves dans deux lycées de Provence-Alpes-Côte d'Azur ou à l'entrée du stade Saint Symphorien du FC Metz lors des matchs de football). D'usage courant en Chine, ces dispositifs ont été mis en œuvre par plusieurs polices britanniques (Londres, outil « *AFR Locate* » en Galles du Sud[306]…)

---

[304] Projet de position relative aux conditions de déploiement des caméras dites « intelligentes » ou « augmentées » dans les espaces publics.

[305] Selon le rapport *Index 2021* de AI100, en 2017, le taux de faux négatif (cas dans lequel la machine indique à tort qu'une personne sur une image n'est pas celle qui est recherchée) était encore compris, selon les types d'image, entre 20 et 50% ; en 2020, il est inférieur à 1% pour des photos normées, de type passeport, et de moins de 4% pour des photos diverses.

[306] Un arrêt de la Cour d'appel d'Angleterre et du Pays de Galles du 11 août 2020 (C1/2019/2670) a annulé un jugement de la High Court qui avait admis l'utilisation de ce dispositif, et jugé ce dernier illégal en raison du pouvoir discrétionnaire des officiers de police dans la détermination des personnes recherchées, de la relative indétermination des zones de

et aux États-Unis, aboutissant à quelques réussites spectaculaires comme à des erreurs tout aussi impressionnantes – qui ont depuis conduit de nombreux États ou localités (comme San Francisco) à interdire leur usage alors que, par ailleurs, des firmes privées continuent de les développer et vendent leurs prestations aux forces de l'ordre. La Roumanie a mis en œuvre, à partir de caméras de vidéoprotection installées dans les aéroports de Bucarest, un dispositif combinant la reconnaissance faciale (à partir du fichier des personnes recherchées et des personnes suspectées de terrorisme) et une solution d'analyse intelligente des images ayant pour objet de déceler les comportements suspects et d'alerter les agents.

Le projet de règlement de l'UE entend prohiber l'usage de systèmes d'identification biométrique en temps réel par les forces de l'ordre, sauf exceptions limitativement énumérées : recherche ciblée de victimes potentielles, prévention d'une menace réelle et précise à une infrastructure critique, à la vie, à la santé ou à la sécurité d'une personne ou d'une menace terroriste, ainsi que la détection, la localisation, l'identification et la poursuite d'un auteur, suspect ou condamné à un crime grave.

4. L'appui opérationnel aux agents

On a vu, au point 1 de la présente fiche, que les SIA pouvaient être utilisés afin d'acquérir une meilleure connaissance des phénomènes et, en conséquence, d'optimiser le dimensionnement, le pré-positionnement et le déploiement des moyens dont disposent l'administration et l'autorité judiciaire.

Des SIA peuvent également être mis en œuvre dans l'activité opérationnelle des agents, dans les configurations les plus diverses. On se bornera ici à mentionner quelques cas d'usage à titre illustratif.

L'intervention des services d'incendie et de secours peut être optimisée et sécurisée grâce à la vision par ordinateur, en améliorant la reconnaissance des lieux, des personnes et des choses pour les pompiers en dépit de la fumée et des flammes (utilisation de filtres…). Les SIA peuvent également être utilisés pour prévoir la survenance imminente de dangers.

déploiement, des insuffisances de l'analyse d'impact sur la vie privée et de l'incapacité des services à démontrer l'absence de biais ou la mise en œuvre de mesures destinées à les limiter. En revanche, la Cour d'appel a écarté le grief tiré de la disproportion de l'atteinte aux intérêts individuels au regard de la contribution de l'outil à l'intérêt général.

P-Flash[307]

Un modèle relativement prometteur de prévision des embrasements généralisés éclairs (*flashover*), dénommé P-Flash, a été développé par le *National Institute of Standards and Technology* sur la base de données de température provenant des détecteurs de chaleur et de fumée des bâtiments.

L'apprentissage machine est utilisé pour extrapoler les données fournies par ces équipements lorsqu'ils cessent de fonctionner. Faute de pouvoir simuler réellement un nombre suffisant d'incendies pour collecter les données correspondantes, les chercheurs ont utilisé un outil de simulation virtuelle (jumeau numérique) pour réaliser plus de 5000 tests, en modifiant les paramètres influant sur le déclenchement et la dynamique du phénomène. Ils ont pu également utiliser des données réelles portant sur treize sinistres. L'expérimentation montre que l'outil a permis de prévoir les *flashovers* une minute à l'avance dans 86 % des simulations ; les 14 % des simulations restantes portaient essentiellement sur des « faux positifs » (fausses alertes), et non sur des faux négatifs qui mettraient en danger la vie des pompiers. L'étude a révélé certaines limites, en particulier pour les départs de feux dans les pièces fermées.

La fiabilisation et le déploiement d'un tel outil sur le terrain permettrait d'ajuster la tactique d'intervention et d'avertir les pompiers en opération de l'imminence d'un *flashover*.

Il peut être recouru aux SIA basés sur le traitement du langage naturel pour améliorer le **suivi des contrôles** réalisés en collectant rapidement les données figurant dans les rapports de contrôle afin de les mettre à disposition des agents chargés d'apprécier la réalité de la mise en conformité, voire en automatisant en partie le suivi par le déclenchement d'une alerte en cas de divergence entre les données disponibles et le grief ou la recommandation formulés. A titre d'exemple, un projet baptisé SIANCE, développé avec l'appui du dispositif des entrepreneurs d'intérêt général (EIG), vise à améliorer la visibilité de l'Autorité de sûreté nucléaire (ASN) sur l'état du parc et de la radioprotection. Il inclut un apprentissage supervisé entraîné sur les « lettres de suite » de l'ASN notifiées à l'issue des inspections.

Le contrôle automatisé des documents peut aussi servir à garantir la **fiabilité des rapports de police**, comme ambitionne de le faire l'outil VériPol en Espagne. Fruit d'une collaboration entre l'Université de Cardiff, celle de Madrid et la police nationale espagnole, ce système expérimental, déployé à Malaga et en Murcie, a été entraîné sur la base de 1122 rapports dont 534 authentiques et 588 frauduleux afin d'identifier les faux rapports de police (en général, il s'agit de rapports plus courts, moins circonstanciés, centrés sur les biens plutôt que sur les personnes…) et, ainsi, prévenir les erreurs judiciaires et évincer les auteurs du service.

Les activités de contrôle des personnes, par exemple dans le cadre de la **police de l'immigration**, peuvent également justifier le déploiement d'outils de traduction automatique ou encore à des SIA de reconnaissance vocale afin d'identifier l'origine

[307] E.Y. Fu, W.C. Tam, R. Peacock, P. Reneke, G. Ngai, H. Leong and T. Cleary, *Predicting Flashover Occurrence using Surrogate Temperature Data*, *Proceedings of the AAAI Conference on Artificial Intelligence*

d'une personne à partir de la langue parlée ou de l'accent. L'homologue de l'OFPRA en Allemagne a par exemple déployé un SIA permettant de distinguer, parmi les demandeurs d'asile prétendant venir de Syrie, les locuteurs arabes dont l'accent le confirme.

Dans ce même domaine, un usage controversé des SIA a consisté, dans le cadre du projet iBorderCtrl soutenu par l'Union européenne, inspiré du système Avatar (« *Automated Virtual Agent for Truth Assessments in Real time* ») canadien, à aider les agents de la police aux frontières à **détecter les mensonges** des passagers débarquant dans des aéroports, aux frontières extérieures de l'Union. Concrètement, le système, version améliorée du polygraphe, analyse 38 « micromouvements » et « microexpressions » du visage afin de déceler des signes de nervosité supposés accompagner des déclarations mensongères en réponse aux questions posées et d'orienter le passager, selon le cas, vers une file d'attente en vue d'un contrôle allégé ou, au contraire, vers un point de contrôle complet incluant l'analyse de données biométriques. Ce système, qui affiche un taux de faux positif encore élevé (un quart des personnes suspectées d'avoir menti disaient la vérité), est en cours d'expérimentation en Grèce, en Hongrie et en Lettonie.

On peut imaginer d'autres cas d'usage en la matière, par exemple pour **l'évaluation de l'âge** des « mineurs non accompagnés », qui repose aujourd'hui sur des tests osseux dont le résultat est confronté à un atlas réalisé entre 1931 et 1942 sur des enfants américains de famille aisée (atlas de Greulich et Pyle), dont on ne peut exclure qu'il ne corresponde pas tout à fait à la morphologie de jeunes algériens ou sénégalais en 2022, et dont on sait en tous les cas qu'ils ne permettent pas une distinction claire entre 16 et 18 ans, alors qu'il s'agit évidemment de la tranche d'âge critique et la plus représentée parmi les mineurs non accompagnés allégués[308].

*

En dépit des controverses qu'ils suscitent, les usages « sécuritaires » et de contrôle des SIA méritent d'être encouragés, pour autant qu'ils soient assortis de très solides garde-fous.

D'une part, les services concernés ne doivent pas être privés des outils que ceux qu'ils pourchassent ou contre lesquels ils défendent notre société emploient aujourd'hui avec une efficacité redoutable. Qu'il s'agisse de cybercriminalité ou tout simplement de criminalité et de délinquance recourant au numérique, l'État ne peut se priver des armes que ses adversaires utilisent, à la différence près que celles de l'État sont étroitement contrôlées.

D'autre part, devant l'ampleur des données à traiter, le recours à des systèmes d'IA restaure une efficacité réelle des autorités de contrôle, qui peuvent redéployer les moyens économisés vers d'autres missions non automatisables (en l'état), sur le terrain. Limiter l'usage des systèmes d'IA, c'est encourager la privatisation de la sécurité : d'une part, en laissant des prestataires privés développer des outils qui ne

[308] V. Byoung-Dai Lee et Mu Sook Lee, *Automated Age Assessment Using Artificial Intelligence : The Future of Bone Age Assessment*, Korean J Radiol, mai 2021 ; 22(5): 792–800.

seront pas conçus avec les garanties et garde fous encadrant l'action des services de l'État ; d'autre part, la renonciation à l'aide des systèmes d'IA conduirait à déléguer de plus en plus à des prestataires privés, faute d'effectifs suffisants et employables judicieusement, comme le montre le recours croissant à de sociétés de sécurité privée, par exemple lors des grands évènements.

Là comme ailleurs, maintenir les agents de l'État au cœur de l'action ou leur valeur ajoutée personnelle est délivrée, parce qu'on les aura débarrassés de tâches répétitives que des systèmes d'IA accomplissent au mieux, la concentrant sur des gestes utiles, est le principal résultat à attendre. Loin de justifier leur cantonnement ou leur abandon, la maîtrise humaine des activités de sécurité publique, de renseignement et de contrôle passe par un recours accru à ces outils.

Fiche n° 4 : Justice

L'activité juridictionnelle est habituellement identifiée comme l'un des secteurs prometteurs de l'action publique pour le déploiement de l'intelligence artificielle. Elle est aussi l'un de ceux dans lesquels, sur les plans médiatique et marketing, le fantasme a d'emblée pris le pas sur la raison, à travers la notion de « justice prédictive », présentée abusivement comme la capacité d'un système d'IA à deviner par avance le sens des décisions de justice.

Le succès relatif initial de cette promesse commerciale formulée par certaines entreprises spécialisées dans le numérique juridique (« *legaltechs* ») à l'adresse d'avocats, d'entreprises voire de particuliers, repose notamment sur la convergence intuitive entre le raisonnement mathématique et le raisonnement juridique. Elle procède toutefois d'une double méprise.

Sur le plan terminologique, elle repose sur une traduction littérale erronée de l'expression anglo-saxonne « *predictive justice* » qui signifie en réalité « justice prévisible ». La prévisibilité du droit est un enjeu majeur, pleinement intégrée à la construction jurisprudentielle à travers, notamment, l'importance du précédent ; la prédictibilité des décisions de justice est quant à elle une illusion présomptueuse et dangereuse, surtout lorsqu'elle vise à anticiper sur la position personnelle de tel ou tel juge, pris individuellement. C'est d'ailleurs la raison pour laquelle la loi n° 2019-222 du 23 mars 2019 a interdit toute réutilisation des données d'identité des magistrats et membres du greffe ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées[309].

La seconde méprise tient à ce que les outils de « justice prédictive » n'ont évidemment pas la capacité de lire l'avenir, pas plus que l'humain. Ils se bornent en réalité à y projeter les solutions du passé. Trivialement, un système d'IA ne dit pas quelle est la bonne solution à un problème de droit, mais fournit à l'utilisateur la probabilité que des milliers de juges du passé (ceux dont les décisions ont été utilisées pour l'apprentissage de l'algorithme le cas échéant) apportent la réponse A ou la réponse B à une question donnée.

Le scepticisme croissant qu'a produit le discours marketing, en l'absence de concrétisations significatives dans le quotidien des tribunaux et des avocats, a d'ailleurs conduit à un recul progressif de cette notion de « justice prédictive » au profit de celle de « jurimétrie », portée par des acteurs institutionnels comme le Conseil national des barreaux (et inspirée de l'anglais « *jurimetrics* » qui recouvre les méthodes d'analyse quantitative et statistique du droit), ou encore de « justice algorithmisée »[310].

[309] V. notamment le 4° alinéa de l'art. L. 10 du code de justice administrative et le 3° alinéa de l'art. L. 111-13 du code de l'organisation judiciaire.

[310] A. Basdevant, A. Jean et V. Storchan, *Mécanisme d'une justice algorithmisée*, Ed. Fondation Jean Jaurès, 2021.

L'effervescence irrationnelle autour du phénomène de la « justice prédictive » ne saurait toutefois justifier une réaction de rejet, de dénigrement ou de défiance. D'une part, elle ne doit pas occulter l'absolue nécessité pour le monde de la justice de développer une réflexion prospective sur l'impact de ces innovations technologiques sur l'avenir du juge et de son office. D'autre part, et, à plus court terme, elle ne doit pas dissuader les juridictions d'exploiter le potentiel concret des systèmes d'IA dans l'éventail des activités qu'elles déploient, à commencer par les tâches les plus répétitives et modestes auxquelles des moyens humains précieux sont inopportunément consacrés. Une étude réalisée par le laboratoire d'innovation et d'intelligence artificielle de l'école de droit de l'Université de Buenos Aires et le Bureau du Procureur public de la ville a estimé que, sur 169 tâches réalisées par ce Bureau en matière de contentieux administratif et fiscal, 54 seraient entièrement automatisables, 74 ne le seraient pas et 41 partiellement[311].

Les règles utilisées et les procédures suivies dans les juridictions étant relativement normées, elles constituent un terrain favorable au déploiement de systèmes-experts. En outre, les algorithmes d'apprentissage-machine – essentiellement dans le domaine du traitement du langage naturel, puisqu'il s'agit en règle générale de traiter des données de textes – peuvent être développés en mettant à profit la structuration naturelle de certaines données, à commencer par les décisions de justice elles-mêmes, qui suivent en général un format relativement standardisé.

Les trois principales fonctionnalités susceptibles d'être mobilisées sont l'exploitation massive de données (pour l'établissement des faits, par identification de suspects, victimes, détection des corrélations, recoupements… ; pour l'identification des règles de droit applicables au litige ; l'assistance à l'appréciation et à la qualification juridique des faits…) ; la transformation des données et la génération de contenus (anonymisation, traduction, production de décisions…) et, enfin, l'automatisation de process, tant pour le fonctionnement interne des tribunaux que pour la relation aux usagers[312].

Dans ce champ de l'action publique également, la réflexion sur l'IA doit impérativement s'inscrire dans une réflexion plus vaste sur la **numérisation de l'activité juridictionnelle**, ce qui suppose à la fois :

- de doter les magistrats, les agents de greffe et l'ensemble des collaborateurs y participant des équipements informatiques adaptés à leurs missions, alors que l'informatisation de la justice est parfois incomplète ou en trop fort décalage avec l'état de l'art ;
- de privilégier les processus dématérialisés et, plus encore, nativement numériques (c'est-à-dire avec des documents et données produites dès le départ en format numérique et qui n'ont jamais eu de matérialité physique, par

[311] V. L. Cevasco, J. G. Corvalan et E. M. Le Fevre Cervini, *Artificial Intelligence and Work – Building a New Employment Paradigm*.

[312] V. sur cette typologie et des statistiques sur les projets européens en cours ou réalisés : *Study on the use of innovative technologies in the justice field – Final Report* [Etude sur l'utilisation des technologies innovantes dans le domaine de la justice – Rapport final, commandée par la Commission européenne, septembre 2020.

opposition aux documents « scannés »), comme c'est le cas, par exemple, avec le dépôt des recours et mémoires dans Télérecours et Télérecours citoyens ;

- de moderniser rapidement les systèmes d'information dont l'archaïsme empêche toute exploitation performante des données et des métadonnées (notamment par un système d'IA) et toute interopérabilité avec d'autres systèmes.

Le développement optimal de systèmes d'IA publics et privés dans le domaine du droit suppose en outre d'achever, dans les délais prescrits par l'arrêté du 28 avril 2021, la mise en ligne des décisions de justice dans les conditions prévues par la législation issue de la loi pour une République numérique, en sus de la mise à disposition, dans un format exploitable, des données afférentes aux autres sources du droit, nationales, européennes et internationales.

La fonction de juger

S'agissant de la fonction de juger, au sens strict, la mise en œuvre de systèmes d'IA est théoriquement concevable dans les deux volets classiques que sont la prise de décision automatisée (« juge robot ») et l'aide à la décision de justice (l'IA comme « assistant de justice », intervenant pour aider à identifier, analyser et résoudre des questions posées dans un litige, notamment à travers une fonctionnalité de recommandation de solutions).

Elle peut en outre emprunter aux deux branches classiques de l'IA : les systèmes-experts, consistant à programmer tout ou partie des raisonnements juridiques sous la forme d'un arbre de causalité (« si délai de recours de deux mois à compter de la notification + décision notifiée avec voies et délais de recours le 2 mars 2021 + requête introduite le 10 mai 2021, alors requête tardive et rejet pour irrecevabilité ») ; et l'apprentissage-machine, consistant à nourrir l'algorithme des décisions de justice rendues par le passé afin de dégager des raisonnements et des appréciations qui pourront ensuite être utilisés pour la résolution de litiges futurs. Il pourra s'agir, par exemple, de l'identification et de la pondération des critères ou indices mobilisés par le juge pour fixer le montant d'une prestation ou d'une indemnité, le quantum d'une sanction, le droit d'obtenir telle ou telle décision favorable…

L'automatisation de la fonction de juger ne saurait être celle de la décision de justice elle-même.

En l'état de la loi, la garantie humaine dans la fonction de juger est expressément prévue au premier alinéa de l'article 47 de la loi du 6 janvier 1978, qui exclut qu'une décision de justice impliquant une appréciation sur le comportement d'une personne puisse avoir pour fondement (exclusif ou non) un traitement automatisé de données à caractère personnel destiné à évaluer certains aspects de la personnalité de la personne. Cette disposition ne couvre qu'une partie minoritaire du contentieux (contentieux pénal, contentieux de l'asile…). Toutefois, il est certain que l'ensemble des dispositions régissant le fonctionnement des juridictions, voire la Constitution elle-même, font obstacle à ce qu'une décision de justice soit rendue

sur le fondement exclusif d'un traitement algorithmique – autrement dit, que le SIA prenne lui-même la décision de justice.

Le robot-juge est en général considéré comme présentant les atouts suivants[313] :

- L'uniformité territoriale de la jurisprudence (pour autant que les textes applicables soient identiques sur le territoire considéré) et l'égalité des citoyens devant la justice, quelle que soit la juridiction à laquelle ils s'adressent. Le résultat est censé être identique à situation identique (ou ne différant que sur des aspects non pertinents pour la résolution de la question) ;
- Une plus grande sécurité juridique en raison d'une prévisibilité accrue du droit et la suppression (ou la réduction) de « l'aléa judiciaire », par la neutralisation de paramètres susceptibles d'influencer le sens de la décision ou sa motivation[314] ;
- La prise en compte d'un nombre accru de paramètres pertinents et la capacité, à première vue paradoxale pour des outils reposant sur la catégorisation, à produire des jugements « sur-mesure », épousant plus finement les contours particuliers de chaque litige ;
- La célérité de la justice : l'IA calcule plus vite que l'humain, et ne souffre pas de la fatigue ;
- La réduction du coût de la justice (baisse d'effectifs, économie des coûts inhérents à des formations longues, possibilité de traiter des litiges de masse à coût quasi constant…) ;
- Une baisse du volume du contentieux lui-même, les parties étant davantage susceptibles de régler leurs litiges à l'amiable si elles disposent d'une évaluation de leurs chances de succès.

En contrepoint, les limites et risques de l'automatisation pure et simple sont nombreuses et dirimantes[315]. Elles tiennent :

- à l'acceptabilité sociale d'une justice, civile et plus encore pénale, qui ne serait plus rendue directement par l'homme, tout en l'étant juridiquement et symboliquement au nom du peuple français ;
- à la disparition de la fonction « cathartique » de l'oralité dans le procès, qui permet de s'exprimer et d'être écouté par la partie adverse et par des juges : cette fonction, qu'elle se déploie lors de l'instruction ou de l'audience, et *a*

[313] V. sur cette question J.-P. Buyle et A. Van den Branden, *La robotisation de la justice*, in L'intelligence artificielle et le droit, sous la coordination de H. Jacquemin et A. de Streel, Larcier, pp. 259 à 317 ; A. Van den Branden, *Les robots à l'assaut de la justice. L'intelligence artificielle au service des justiciables*, Bruylant, février 2019.

[314] V. sur ce point : L. Pécaut-Rivolier et S. Robin, Justice et intelligence artificielle, préparer demain – épisode I, *Dalloz* Actualité, 14 avril 2020 : « *Nombreuses sont les études qui démontrent que des données fort variées et parfois étonnantes – le petit-déjeuner du juge, sa fatigue, l'influence médiatique, son égocentrisme à ses préjugés divers – peuvent influer sur la décision prise »* (et les renvois aux études mentionnés).

[315] V. notamment sur ces questions : A. Garapon et J. Lassègue, *Justice digitale. Révolution graphique et rupture anthropologique*, Paris, PUF, 2018 ; Y. Meneceur, *L'intelligence artificielle en procès – Plaidoyer pour une réglementation internationale et européenne*, Bruylant, 2020.

fortiori, dans une éventuelle phase de médiation juridictionnelle, fait partie intégrante de la mission de régulation et de résolution des conflits dévolue aux tribunaux ;

- à une forme de « fossilisation » des raisonnements et appréciations, dès lors que le modèle ne se nourrirait que de données du passé (et des flux de données qu'il génère lui-même sur cette base, créant ainsi un effet dit « performatif » de la prise de décision juridictionnelle automatisée), sans place pour la créativité contentieuse, l'infléchissement ou le revirement de jurisprudence (ou les solutions d'espèce basées sur des considérations nouvelles) et en contradiction avec le caractère nécessairement vivant du droit ;
- aux biais affectant les données d'entraînement et aboutissant à des décisions prises sur la base de motifs discriminatoires ou inopérants ;
- à la question de l'acceptabilité de l'erreur-machine pour des décisions susceptibles d'avoir un impact important sur la société, sur la vie des personnes ou l'économie ;
- à l'impossibilité d'utiliser dans ce cadre des modèles non explicables, qui ne seraient en conséquence pas en mesure de fournir une motivation permettant aux parties de comprendre les raisons du verdict juridictionnel ;
- à la difficulté d'« encapsuler » dans une fonction mathématique, même très complexe, des raisonnements et, surtout, des appréciations dont les déterminants ne sont pas entièrement exprimés dans les décisions. L'intime conviction peut se forger sur des ressentis difficilement objectivables mais parfaitement admissibles. Il n'est en outre pas rare que le juge consente à quelques concessions avec la « perfection logique » au profit de solutions jurisprudentielles plus simples, plus compréhensibles pour les justiciables et, s'agissant du juge administratif, plus opérationnelles pour les administrations ;
- aux sérieux doutes qu'on peut nourrir, au moins en l'état actuel des connaissances et des données disponibles, quant à la performance de tels traitements algorithmiques compte tenu de la quantité et de l'exploitabilité des données brutes. Certes, la rédaction des décisions de justice est relativement normalisée ; mais elles sont loin d'être parfaitement homogènes et des subtilités rédactionnelles peuvent aisément échapper à la machine. Certains contentieux ne donnent lieu qu'à un nombre très réduit de décisions, insuffisantes pour constituer une base d'apprentissage fiable. En outre, le travail de labellisation des données d'entraînement suppose une formation juridique suffisante. Certaines expériences[316] affichent certes un taux de performance (exprimé en pourcentage de décisions dont la machine a pu anticiper le sens) relativement « flatteur » (entre 70 et 80% dans les deux cas mentionnés), mais, d'une part, tout biais ne peut être exclu dans la sélection des décisions à des fins d'apprentissage et de test, par élimination de celles qui sortent de la norme et, d'autre part, il s'agit de

[316] V. par ex. : N. Aletras, D. Tsarapatsanis, D. Preotiuc-Pietro et V. Lampos, *Predicting judicial decisions of the European Court of Human Rights : a Natural Language Processing perspective*, 2016 ; D. M. Katz, M. J. Bommarito et J. Blackman, *Predicting the Behavior of the Supreme Court of the United States : A General Approcach*, Cornell University, 2014.

juridictions admettant les opinions dissidentes, de sorte que la machine a pu tenir compte dans ses « prédictions » de l'identité des membres des formations de jugement. D'autres expérimentations se sont avérées peu concluantes[317] ;

- au risque élevé d'obsolescence accélérée des modèles en raison des évolutions rapides du cadre juridique, du fait de la multiplicité des sources de droit et de l'inflation normative, qui entraîne des coûts de maintenance voire de développement élevés ;
- à terme, à la disparition du savoir-faire des juges, dépossédés par la machine, avec des conséquences difficilement réversibles.

Il ne semble pas que la voie de la robotisation de la justice ait été empruntée avec succès par d'autres pays occidentaux à ce jour. L'annonce fracassante, en 2019, de l'automatisation intégrale des jugements rendus dans de petits litiges civils (dont l'enjeu serait inférieur à 7000 euros) en Estonie ne semble, pour l'heure, pas avoir été suivi du déploiement d'un outil opérationnel[318]. En Chine, le Parquet populaire de Shanghai Pudong expérimente un SIA entraîné sur la base de 17 000 affaires entre 2015 et 2020, qui formule des réquisitions de peines sur la base d'une description des données du litige, dans sept catégories d'infractions courantes (vol, fraude, jeux d'argent, conduite dangereuse, violences volontaires, entrave aux dépositaires de l'autorité publique…) ; le taux d'exactitude annoncé est de 97%, ce qui signifie un taux d'erreur judiciaire (au stade des réquisitions) de 3%...

La stratégie publique devrait concentrer ses efforts sur le développement encadré de **l'assistance juridique par des systèmes d'IA**, qui apparaît comme une formule à la fois plus prometteuse et plus acceptable socialement[319], à condition d'être encadrée.

Aucune disposition ni aucun principe n'empêche une juridiction de se doter d'outils numériques d'aide à la décision, incluant des systèmes d'IA, sous réserve de l'assortir de garanties appropriées tenant notamment à la réalité et à l'autonomie de l'intervention humaine dans la prise des décisions. La proposition de règlement de la Commission européenne identifie à cet égard le domaine de « l'administration de la justice » comme susceptible de donner lieu à la mise en service de systèmes d'IA à hauts risques, et prévoit elle-même que les systèmes d'aide à la décision de justice (consistant à rechercher et interpréter les faits et la loi, et à appliquer la loi à un ensemble concret de faits) relèvent de cette catégorie.

---

[317] V. notamment l'expérimentation d'un outil de « justice prédictive » dans les cours d'appel de Douai et de Rennes, qui ne s'est pas avéré plus performant que les moteurs de recherche classiques.

[318] Cette initiative n'est d'ailleurs pas recensée dans l'étude sur l'utilisation des technologies innovantes dans le domaine de la justice réalisée à la demande de la Commission européenne en septembre 2020.

[319] V. en ce sens J. Ulenaers, *The Impact of Artificial Intelligence on the Right to a Fair Trial : Towards a Robot Judge ?*, Asian Journal of Law and Economics, Vol. 11, no. 2, 2020, pp. 20200008.

Certains avantages attachés à la prise de décision juridictionnelle intégralement automatisée, sont susceptibles d'être retirés, à un degré moindre, de cette modalité d'intervention des systèmes d'IA dans l'activité juridictionnelle. En particulier, la confrontation du jugement humain avec le résultat de l'analyse machine peut être de nature à atténuer l'effet des biais affectant le premier et jouer un rôle régulateur dans des contentieux où la sensibilité personnelle et les convictions du juge sont susceptibles d'exercer un influence sensible sur le sens du jugement (contentieux des étrangers et de l'asile, contentieux pénal…). Ils peuvent aussi servir à révéler au juge des déterminants de l'appréciation ou de la décision qu'il n'avait pas de lui-même identifié ou formalisé. Les risques associés au déploiement de tels outils ne doivent pas être sous-estimés. Mal conçus ou mal utilisés, ils peuvent induire en erreur le juge et aboutir à des résultats exactement inverses à ceux qui étaient recherchés et, partant, au déni de justice. En la matière, les résultats produits par un système d'IA ne peuvent se concevoir que comme un élément d'appréciation parmi d'autres pour le juge et il ne saurait être exigé de ce dernier qu'il justifie dans sa décision des raisons qui l'ont conduit à ne pas suivre une recommandation de la machine. Tout biais d'automatisation, par lequel le juge s'en remettrait de fait à une telle recommandation et se placerait ainsi sous la dépendance de l'outil et, indirectement, de ceux qui l'ont conçu, constituerait une atteinte inacceptable à l'indépendance du juge et une forme déguisée de robotisation complète dont les limites et inconvénients ont été précédemment décrits.

Afin de limiter ce risque et, en particulier, d'éviter le biais d'ancrage (difficulté à faire abstraction et à se distancier de sa première impression), il serait préférable que le juge conduise sa propre analyse du dossier avant de la confronter à la recommandation d'un système d'IA, et non l'inverse (de nombreux rapporteurs publics préfèrent aussi se forger leur propre opinion en examinant le dossier « à blanc », avant de se plonger dans la note du rapporteur et, le cas échéant, celle du réviseur).

L'introduction de tels systèmes pour la qualification des faits et la formulation de recommandations quant à la résolution des litiges pourrait être envisagée, à titre expérimental, pour des contentieux de faible importance et présentant un caractère provisoire ou reposant, comme l'algorithme d'apprentissage machine, sur une logique probabiliste (évaluation du bien-fondé des moyens afin de déterminer s'il est sérieux pour l'admission des pourvois en cassation ou l'octroi du sursis à exécution des décisions du juge administratif, pour déterminer s'il existe un doute sérieux quant à la légalité de la décision administrative, pour apprécier les chances de succès d'une requête en vue de l'octroi de l'aide juridictionnelle…).

L'assistance à la fonction de juger pourrait emprunter la voie de systèmes-experts, dont les règles sont connues et intégrées dans le modèle, pour la résolution de questions n'appelant pas, ou peu, d'appréciation. Tel serait le cas, par exemple de l'analyse de la tardiveté éventuelle d'un recours, qui suit des règles essentiellement objectives, ou du respect de la condition de ressources pour l'attribution de l'aide juridictionnelle.

A l'heure actuelle, on recense peu de projets de SIA portant sur la fonction de juger, et encore moins de réalisations concrètes. La Suède a engagé une réflexion exploratoire sur l'utilisation de l'apprentissage machine pour la prise de décision juridictionnelle, sans envisager à ce stade la conception d'un outil. L'expérience présentée comme la plus aboutie est celle de l'outil *Prometea*.

Prometea

Créé par le bureau du parquet de Buenos Aires, cet outil, entraîné sur 2400 jugements et 1400 avis juridiques préalables du Parquet, comporte plusieurs fonctionnalités dont l'une consiste à recommander une solution à des litiges portant sur l'attribution d'un logement ou d'aides sociales. En renseignant le numéro de l'affaire, la machine proposerait en vingt secondes et dans 96% des cas la solution « correcte » (celle qui a été donnée dans un cas passé substantiellement identique, identifié parmi les 300 000 documents en base) et produirait automatiquement le projet de décision correspondant, en réduisant de 99% le taux d'erreurs de typographie.

Ce système a été adapté et déployé, sous le nom de *PretorIA*, à la Cour constitutionnelle de Colombie. Il repose sur un modèle explicable d'apprentissage supervisé, entraîné sur 1200 affaires labellisées par dix juristes, et des systèmes-experts. L'outil analyserait en cinq secondes les 2700 décisions de *tutela*[320] soumises à la Cour chaque jour (au lieu de 202 jours/homme de travail), avec un taux de succès de 95% sur 13 critères d'examen (sur les 33 définis par la Cour et donnant lieu à entraînement du modèle sur 1200 affaires) ; la production d'un document par *PretorIA* prendrait environ une minute là où il nécessitait deux heures et quarante minutes auparavant.

La cour de Moron (province de Buenos Aires) utiliserait aussi l'outil *Prometea* pour la détermination du lien de causalité dans les accidents de la route[321].

La **rédaction des décisions de justice** constitue également un champ intéressant pour le déploiement des SIA. Certaines d'entre elles apparaissent largement sinon complètement automatisables dès lors qu'elles ne font intervenir aucune appréciation[322] : à titre d'exemple, une décision de non-admission d'un pourvoi en cassation devant le Conseil d'État se compose de données qui existent dans les

[320] En Colombie, l'action de *tutela* est la voie de droit permettant à tout justiciable de réclamer devant n'importe quel juge et à tout moment la protection de ses droits constitutionnels fondamentaux menacés par l'action ou l'abstention d'une autorité. Les décisions de *tutela* sont automatiquement transmises à la Cour constitutionnelle qui choisit discrétionnairement celles qui méritent une révision afin d'unifier la jurisprudence. Plus de la moitié des décisions concernent le droit à la santé.

[321] J. G. Corvalan et E. M. Le Fevre Cervini, *Prometea experience. Using AI to optimize public institutions*, CERIDAP, 2/2020. L'outil permettrait ainsi de rédiger mille décisions en matière de droits au logement en 45 jours, contre 174 sans lui.

[322] Sous réserve de la pratique consistant à requalifier les moyens au regard du contrôle qu'exerce le juge de cassation sur l'appréciation portée par les juges du fond (contrôle de qualification juridique des faits ou absence de contrôle de l'appréciation souveraine, hors dénaturation).

systèmes d'information, qu'il s'agisse de l'identité des parties, des dates d'enregistrement des productions, des conclusions et des moyens, des textes invoqués, et de mentions systématiques (citation de l'article L. 821 du code de justice administrative, énonciation qu'aucun des moyens soulevés n'est de nature à permettre l'admission…). A défaut de structuration des données par le requérant lui-même (par exemple, par un formulaire dans lequel l'avocat aux Conseils renseignerait les conclusions, d'une part, et le libellé synthétique de chaque moyen d'autre part), un SIA serait susceptible d'identifier lui-même les éléments pertinents dans les mémoires et de les injecter telles quelles dans le projet de décision, sans reformulation, et sans autre intervention des membres de la formation de jugement qu'une relecture de vérification. La production des ordonnances de série, des ordonnances de rejet pour défaut d'avocat obligatoire voire des ordonnances pour incompétence et d'autres causes d'irrecevabilité se prêterait également à l'« industrialisation » sans autre intervention humaine que le déclenchement et la validation par l'apposition de la signature (électronique le cas échéant).

Une telle réflexion pourrait être mise en regard de la capacité des requérants à générer eux-mêmes des requêtes par le biais d'un système d'IA, susceptibles de créer des afflux contentieux sans précédents, alternativement aux actions collectives (action de groupe et action en reconnaissance de droit).

Par ailleurs, la **mise en forme automatique** des projets de jugements et des minutes, laquelle ne requiert du reste qu'un algorithme basique étranger à tout apprentissage machine, permettrait de ne plus mobiliser de précieuses ressources humaines pour l'accomplissement d'une tâche parfaitement standardisée.

La recherche documentaire juridique

Cette activité essentielle à la fonction de juger emprunte la plupart du temps la voie d'une recherche en « plein texte » ou d'une requête dans un moteur de recherche fonctionnant à l'aide de mots-clés, dans des bases de données propres aux juridictions (bases de jurisprudence, comme Jurinet ou Ariane) ou dans celles qu'offrent le secteur privé (éditeurs juridiques « traditionnels », *legaltechs*…).

Le développement de l'intelligence artificielle dans le domaine du traitement du langage naturel a permis de développer la « recherche sémantique ». Celle-ci permet, à partir d'une question formulée en langage courant, d'avoir accès à des sources documentaires et données utiles eu égard non seulement aux termes employés mais aussi au contexte dans lequel ils le sont, en se fondant sur la fréquence statistique des associations entre mots et expressions ou le rapprochement des écrits portant sur un même domaine. Le système peut « qualifier » le besoin ainsi exprimé et renvoyer des réponses susceptibles de mieux satisfaire la requête qu'une simple recherche par mots-clés. Tel est le cas de l'algorithme du moteur de recherche de Google, que de nombreux magistrats utilisent d'ores et déjà en complément, voire en substitution, des bases de données spécialisées, pour accéder à des contenus rendus librement accessibles au public. Certaines *legaltechs* proposent également des moteurs de recherche sémantique dédiés à la recherche juridique. En optimisant l'exploitation d'un volume documentaire qui ne cesse de croître en raison de la diversité croissante des sources

du droit et de l'intensité de la production juridictionnelle et doctrinale, ces systèmes peuvent constituer un auxiliaire précieux pour le juge, ce qui plaide pour leur développement et leur utilisation à grande échelle.

Des systèmes d'IA peuvent également identifier et extraire d'un ensemble de données non structurées et non indexées des informations utiles[323], avec un « score de confiance », la liste des sources sur lesquelles la réponse est fondée et des suggestions de requêtes alternatives. Les résultats peuvent être actualisés automatiquement en fonction de l'enrichissement de la base de connaissances. Ces outils permettent également de retrouver quasi-instantanément une pièce dans un dossier volumineux. Autant de fonctionnalités qui permettent d'accroître la productivité des métiers du droit.

L'instruction, le traitement « administratif » des dossiers contentieux et la mise à disposition des décisions

Outre le travail de résolution des litiges soumis au juge, sur la base de recherches documentaires, le traitement d'un dossier contentieux se compose d'une succession de tâches de nature plus administrative, de l'enregistrement d'un recours à la notification et la publication de la décision et l'archivage du dossier, en passant par l'élaboration d'un calendrier d'instruction, la communication des écritures, le traitement et le rattachement de productions et de documents à des dossiers en instance (lorsqu'ils ne sont pas directement versés dans le dossier par les parties *via* une application), le lancement, le suivi et la gestion des mesures d'instruction, l'organisation de séances d'instruction, l'établissement des rôles, la convocation aux audiences et leur organisation, et le compte rendu de l'audience et du délibéré. Les SIA peuvent contribuer à décharger les magistrats et les greffes d'une partie des tâches matérielles afférentes à cette activité afin de leur permettre de se recentrer sur les tâches à plus forte valeur ajoutée.

La fonction de **catégorisation et de détection des similitudes** qui peut être assignée à un système d'IA permet non seulement d'identifier des précédents utiles (ce qui se rapporte à la fonction de recherche juridique précédemment mentionnée), mais aussi de rapprocher des affaires en cours afin d'améliorer et d'accélérer leur traitement contentieux. Tel est l'objet du projet expérimental de détection des séries contentieuses porté par le Conseil d'État et retenu dans le cadre du deuxième appel à manifestation d'intérêts organisé par Etalab. Il repose sur un apprentissage supervisé à partir de la labellisation des conclusions et des moyens dans un grand nombre de requêtes contentieuses, de manière à entraîner la machine à distinguer les unes et les autres et à mettre en relation des conclusions connexes ou identiques

[323] V. par ex. le projet « Avvocatura 2020 » en Italie ou encore l'outil, basé sur un réseau de neurones de traitement du langage naturel, développé par le ministère de la justice britannique afin de détecter rapidement des récurrences dans les centaines de rapports d'inspection des établissements pénitentiaires (rédigés par l'administration pénitentiaire, les commissions de surveillance indépendantes et l'Ombudsman des prisons et de la probation, équivalent du Contrôleur général des lieux de privation de liberté), notamment pour l'analyse des incidents, l'identification de facteurs géographiques ayant une incidence sur les établissements, orienter les inspections et mieux allouer les moyens.

(même décision attaquée par exemple) ou une argumentation similaire (nombre important de requêtes stéréotypées ou articulant des moyens analogues contre des décisions de même nature…).

La Croatie (outil « Speech-to-text » en service dans les juridictions croates depuis décembre 2018), l'Estonie, l'Allemagne, ou encore la Hongrie développent ou ont développé des outils de **création automatisée des procès-verbaux d'audiences** utilisant l'apprentissage machine (reconnaissance vocale). L'enregistrement et la retranscription automatique – ainsi que, le cas échéant, des traitements subséquents comme le rapprochement avec les écritures ou d'autres pièces du dossier afin de détecter d'éventuelles incohérences – seraient pertinentes dans les contentieux où l'oralité occupe une place importante (audiences pénales, contentieux des étrangers et de l'asile[324]…). Les outils de reconnaissance vocale et de retranscription textuelle peuvent aussi être mobilisés lors des séances orales d'instruction et des audiences d'instruction.

Les progrès rapides de la **traduction automatique simultanée** poseront inévitablement, à terme, la question de **l'interprétariat** dans le contentieux des étrangers, notamment devant la Cour nationale du droit d'asile (et, en amont, devant l'Office français de protection des réfugiés et apatrides - OFPRA), et l'ensemble des contentieux, civils et pénaux, dans lesquels des étrangers non francophones sont parties. Le Domstolsverket suédois, qui est une administration chargée d'apporter aux juridictions des ressources et un appui logistique, développe à cette fin un outil « Automatic transcription » qui retranscrit et traduit en anglais les voix enregistrées.

La **traduction en français des pièces produites dans les affaires contentieuses** peut également être facilitée par les SIA. Toutefois, il semble préférable de maintenir le principe de la production de pièces en langue française et de ne pas mettre la traduction ou le contrôle de sa fidélité au texte original à la charge de la juridiction. C'est à la partie qui produit d'y procéder, sous sa responsabilité, et à la partie adverse de contester la pertinence de la traduction, les deux parties pouvant recourir à des systèmes d'IA qui seront, à terme, certifiés.

En revanche, il serait opportun de confier à des SIA opérés par les juridictions, et (ré-)entraînés à partir de données issues de textes et d'actes juridiques, le soin d'assurer la **traduction des décisions de justice** (notamment en anglais), à des fins de rayonnement institutionnel mais aussi de convergence dans l'application du droit européen (en améliorant l'accessibilité des décisions juridictionnelles françaises pour des juridictions étrangères). La Suède est en train de concevoir un tel outil pour les arrêts de la Cour suprême.

[324] En l'état, le Conseil d'Etat, juge de cassation en matière d'asile, ignore tout des échanges intervenus à l'audience devant la Cour nationale du droit d'asile, hormis le cas dans lequel la décision de la CNDA elle-même fait état des propos tenus ; or de nombreuses appréciations et qualifications font l'objet d'un contrôle entier qui s'accommode mal d'une telle asymétrie d'informations.

Enfin, la tâche qui donne lieu au plus grand nombre de projets de SIA dans le monde judiciaire européen est celle qui consiste à **pseudonymiser voire anonymiser les décisions de justice** afin de réduire ou supprimer l'atteinte à la protection de la vie privée des protagonistes du dossier au moment de leur publication. Des projets en ce sens sont en cours de développement en Autriche, au Danemark, en Suède, en Finlande (outil « Anoppi »), en Estonie (outil « Texta »), en République tchèque, en Croatie, en Espagne ou encore au Luxembourg. En France, Etalab a développé, à la demande du Conseil d'État, un outil de pseudonymisation fondé sur l'apprentissage machine. De son côté, la Cour de cassation met au point un dispositif analogue de pseudonymisation adossé à un SIA recourant à l'apprentissage supervisé, auquel le Conseil d'État est également associé.

Pseudonymisation automatisée des décisions judiciaires en France

Afin d'assurer la mise en œuvre de l'« open data des décisions de justice » décidée par la loi pour une République numérique, la Cour de cassation a lancé en 2019 la conception d'un SIA reposant sur l'apprentissage machine supervisé, permettant d'automatiser la pseudonymisation des décisions de justice rendues publiques, c'est-à-dire la suppression des éléments permettant l'identification directe des parties et des tiers qui y sont mentionnés. Elle s'est dotée d'une équipe-projet composée de data scientists et d'un développeur, coordonnée par une magistrate de la Cour, et mobilisant une équipe d'une quinzaine d'annotateurs parmi les agents de catégorie C de la Cour, encadrée par une directrice des services de greffe, pour labelliser les mentions à occulter, constituer le jeu de données d'entraînement et assurer le contrôle-qualité des résultats de l'outil.

A date, le taux de performance (vrais positifs) est de l'ordre de 98% sur les noms et prénoms et de 90% pour les autres données identifiantes. Le travail se poursuit pour tendre vers un taux de 100%. Les facteurs de réussite identifiés du projet tiennent à la disponibilité d'une quantité importante de données (plusieurs millions de décisions), l'accès direct à ces décisions, le temps important qui a été consacré au nettoyage des données, la constitution d'une équipe qualifiée en interne – de préférence au recours à un prestataire privé, l'engagement personnel et l'animation dynamique d'une équipe compétente d'annotateurs, le travail sur l'ergonomie de l'outil d'annotation et l'acquisition d'un processeur graphique (GPU).

Le perfectionnement et la » formation continue » du SIA, notamment pour tenir compte de l'évolution des décisions et de l'apparition de nouveaux vocables et entités, l'accompagnement des agents affectés à l'annotation et de la montée en charge de leur activité, et la lourdeur de la réalisation de l'analyse d'impact relative à la protection des données à caractère personnel font partie des défis soulevés par cet outil.

L'exécution des décisions de justice

Les SIA pourraient être mobilisés pour dynamiser l'exécution des décisions de justice.

L'automatisation du recouvrement des amendes pénales, dont le taux reste extrêmement faible, pourrait constituer un chantier prioritaire en la matière. Le ministère de la justice finlandais a développé un outil (« *Robot Process*

Automation ») permettant de relier plus facilement les paiements aux sanctions auxquelles ils correspondent, et de traiter les défauts de paiement comme les surpaiements et les remboursements (en cas d'annulation de la sanction par exemple).

A plus long terme, on pourrait imaginer une forme de contrôle automatisé de l'exécution de décisions (à titre d'exemple, il serait aisé de vérifier qu'une disposition que le juge a enjoint à l'administration d'abroger l'a été effectivement). A l'heure actuelle, il appartient à la partie ayant eu gain de cause de prendre l'initiative ; mais comme le montre la possibilité reconnue au juge administratif de prononcer des injonctions et des astreintes d'office, ce dernier se voit reconnaître une place de plus en plus active dans le contrôle de l'effectivité de ses décisions.

La relation avec les usagers du service public de la justice et l'accès au juge

La numérisation de l'activité juridictionnelle, en particulier le développement des téléprocédures, a considérablement accru les possibilités de suivi, par les parties, de l'avancement du traitement de leur dossier. Pour autant, les greffes restent régulièrement sollicités pour répondre à des questions récurrentes qu'un robot conversationnel pourrait traiter.

Un « *chatbot* » de ce type est en cours de déploiement dans les juridictions suédoises, à l'initiative du Domstolsverket. Le ministère de la justice autrichien conçoit également un *chatbot* qui assiste les parties dans le suivi de leur dossier sur téléphone portable.

Un tel outil n'a évidemment pas vocation à se substituer aux avocats dans la délivrance de conseils juridiques. Il pourrait éclairer les justiciables sur la juridiction qu'ils doivent saisir pour contester une décision donnée, les modalités de cette saisine, l'état d'avancement du dossier, les échéances à venir, les modalités de contestation de la décision rendue…

Les juridictions, les ordres d'avocats et le Conseil national des barreaux permettraient de fédérer les acteurs de la justice autour de la construction d'une interface.

Le pilotage de l'activité juridictionnelle

Au-delà des questions, communes à l'ensemble des autorités publiques (de gestion des locaux, des fluides ou des approvisionnements par exemple), et de la production automatisée de statistiques[325], le cas échéant sous la forme de représentations graphiques qui en facilitent la compréhension et l'explicitation (permettant par exemple d'identifier les goulots d'étranglement dans les chambres et de réaffecter rapidement des dossiers pour rééquilibrer les stocks), les SIA peuvent être un

[325] V. par ex. le projet italien « *Aut Dedere, Aut Judicare* » dont le but est de fournir des statistiques dans le domaine de la coopération judiciaire pénale par l'exploitation automatisée de données dans les productions judiciaires (mandats d'arrêts, extraditions, transfèrements…).

puissant instrument à la main des juridictions, en particulier de leurs chefs, pour orienter l'activité juridictionnelle en fonction des priorités objectives.

L'analyse des décisions rendues permet d'identifier d'éventuelles divergences de jurisprudence. Tel est l'objet de l'un des quinze projets sélectionnés dans le cadre du deuxième appel à manifestation d'intérêts organisé par Etalab et la DITP, porté par la Cour de cassation et accompagné par un chercheur en traitement du langage naturel de l'INRIA, qui consiste à rapprocher ses décisions et celles des cours d'appel traitant d'un sujet donné afin de détecter des interprétations divergentes de la loi. Les principaux défis consistent à automatiser le « titrage » des décisions (mots-clés), à définir le concept de divergence et à mobiliser des ressources suffisantes pour l'annotation des décisions.

Déployé à l'échelle d'un ordre de juridiction, un tel outil permettrait de piloter la politique jurisprudentielle des cours suprêmes en leur donnant une plus grande visibilité sur les questions qui méritent une intervention jurisprudentielle, qu'il s'agisse d'un arbitrage entre des positions divergentes adoptées par les juges du fond, de précisions ou d'un revirement de jurisprudence dans le cas où celle-ci a produit des conséquences excessivement fâcheuses.

Les SIA peuvent aussi contribuer à **l'allocation optimale et fluide des ressources** en fonction d'une évaluation plus fine et objective des besoins, de la définition de la carte judiciaire nationale au pilotage d'une juridiction voire d'une chambre.

L'**affectation des dossiers** entre les chambres des tribunaux, des cours et du Conseil d'État se prêterait également à une automatisation sur la base des dizaines voire centaines de milliers d'affectations qui en révèlent la logique, du reste essentiellement thématique, en tenant compte également, pour les contentieux transverses, de l'état des stocks des chambres, afin d'orienter efficacement les flux.

Fiche n° 5 - Travail et emploi

L'essentiel de la littérature sur le recours à des systèmes d'intelligence artificielle (SIA) dans le monde du travail se concentre sur son **impact sur les conditions de travail, les salaires et le niveau d'emploi**. Dès lors que les SIA sont capables de prendre en charge les tâches les plus simples et les plus répétitives, et permettent de démultiplier celles que les robots peuvent accomplir sur une ligne de production, ils peuvent en effet conduire à une réallocation de la main d'œuvre. Toutefois, si l'ampleur de la destruction des emplois à moins forte valeur ajoutée fait l'objet de multiples études, l'impact net de l'IA elle-même, en tant que nouvelle « révolution industrielle », ne fait l'objet d'aucun consensus scientifique.

Ce débat, tant politique qu'économique, et qui traverse nombre de sociétés, dont la France, éclipse un autre volet des liens entre l'IA d'une part et le travail et l'emploi d'autre part : l'utilisation de l'IA par les administrations pour rendre plus accessibles et plus efficientes les politiques publiques qu'elles déploient dans ce domaine. Or, les SIA peuvent permettre de répondre à plusieurs enjeux auxquels sont confrontées les politiques de travail et d'emploi dont, notamment :

1/ **L'information des acteurs sur leurs droits et obligations**, qui permet à la fois d'assurer l'application du droit et la sécurisation de leurs démarches, alors que la complexité, notamment juridique, d'un certain nombre de dispositifs rend leur appropriation plus difficile, par les salariés comme par une partie des employeurs ;

2/ **La meilleure connaissance de l'écosystème** et notamment des comportements et des besoins des acteurs (entreprises, salariés, organismes de formation…), l'anticipation des évolutions du marché du travail, le suivi de cohortes, l'analyse de parcours professionnels afin d'affiner les modèles de retour à l'emploi ;

3/ **La personnalisation de l'accompagnement des demandeurs d'emploi** et notamment de ceux qui sont les plus éloignés de l'emploi, ce qui implique une meilleure allocation des moyens financiers et humains, en priorité pour les situations les plus complexes, et l'amélioration de l'appariement (« *matching* ») entre demandeurs d'emploi et offres de formation ;

4/ **Le ciblage des contrôles**, *via* l'identification des entreprises à risque au sein d'un secteur d'activité ou l'analyse d'accords permettant d'identifier des écarts à la norme, par exemple en matière d'égalité professionnelle, d'accidents du travail, de travail dissimulé ou de recours illégal aux contrats courts.

Toutefois, la mise en œuvre de SIA publics dans le champ des politiques du travail et de l'emploi, tout utile et souhaitable qu'elle soit, ne peut se faire sans une **prise en compte des craintes et critiques, fondées ou non, qu'elle suscite**. Cinq séries de craintes peuvent être soulignées.

- La première tient à ce que l'IA conduirait à supprimer des postes dans des administrations ayant déjà fortement contribué aux efforts de rationalisation

consécutifs aux différents plans d'économies et de réorganisation des services administratifs.
- La deuxième tient à ce que l'IA porterait en soi une déshumanisation du lien et la fin de l'interaction directe avec l'usager, alors que l'accompagnement des publics les plus fragiles, davantage sujets à l'illectronisme, parmi lesquels comptent les chômeurs les plus éloignés de l'emploi, nécessite un contact direct et régulier.
- La troisième tient à ce qu'un SIA mal entraîné ou entraîné sur des jeux de données mal calibrés, reproduisant des biais, conduirait à figer voire amplifier les discriminations déjà observées sur le marché de l'emploi.
- La quatrième tient à ce que la technicité, voire l'inintelligibilité, des SIA conduirait à la dépossession des acteurs des politiques du travail et de l'emploi, insuffisamment au fait des tenants et aboutissant de tels systèmes, dans ces domaines traditionnellement marqués par le dialogue social et, partant, par le poids des partenaires sociaux (bureaucratisation, étatisation…).
- La cinquième tient à ce qu'une prise de décision automatisée sans intervention humaine pourrait conduire à des conséquences graves pour les publics visés en cas d'erreur, par exemple s'agissant d'une radiation de Pôle emploi ou d'un refus d'indemnisation.

Ces craintes ne peuvent être balayées d'un revers de main dès lors qu'elles constituent autant d'obstacles au développement des SIA publics et à l'adhésion des acteurs à cette démarche : elles impliquent que les décideurs publics en tiennent compte, dans l'identification des projets prioritaires, dans la méthode mise en œuvre pour les déployer et dans le contrôle en continu des résultats obtenus. La feuille de route AMDAC du MTEI a ainsi prévu de travailler à une « cartographie des controverses » et questionnements engendrés par l'IA, de détourer les contours d'un comité d'éthique ouvert et de bâtir une doctrine pour une IA éthique dans la sphère travail.

En l'état des cas d'usage, les craintes exprimées doivent être tempérées. En effet, les SIA publics aujourd'hui développés dans le champ des politiques du travail et de l'emploi ont été avant tout pensés comme de nouveaux services proposés à l'usager ou à l'agent, avec l'objectif d'améliorer le service rendu.

1. L'IA au service de l'accessibilité et de l'application du droit du travail

S'agissant de l'accessibilité du droit du travail, deux projets développés par le ministère du travail, de l'emploi et de l'insertion (MTEI), reposant sur le traitement automatisé du langage naturel (TALN), peuvent être mis en lumière.

Le premier est le **code du travail numérique**, qui a pour objet de permettre « *en réponse à une demande d'un employeur ou d'un salarié sur sa situation juridique, l'accès aux dispositions législatives et réglementaires ainsi qu'aux stipulations conventionnelles, en particulier de branche, d'entreprise et d'établissement, sous réserve de leur publication, qui lui sont applicables* » (article 1er de l'ordonnance n° 2017-1387 du 22 septembre 2017 relative à la prévisibilité et la sécurisation des relations de travail).

La particularité de ce SIA est qu'il permet d'**apporter une réponse juridiquement qualifiée et compréhensible à l'usager**, *via* une brique fonctionnelle de TALN qui traduit en langage « accessible » les dispositions et stipulations applicables à sa situation, en fonction, notamment, de la branche à laquelle est rattachée son entreprise (le CTN intègre, en plus des dispositions du code du travail, 30 000 textes conventionnels et un fonds de 2 500 réponses rédigées par les services de l'inspection du travail). Les résultats de recherche reposent sur une analyse sémantique des questions posées et sont améliorés en fonction des comportements des utilisateurs grâce à un modèle supervisé de *learning-to-rank*, qui permet au système de proposer des réponses toujours plus pertinentes en fonction des réactions des utilisateurs (sur le modèle des moteurs de recherche comme Google ou Bing).

Le code du travail numérique propose également différents **simulateurs** pour, par exemple, estimer le montant d'une indemnité de licenciement ou de précarité de fin de CDD ou la durée d'un préavis de licenciement. La boîte à outil propose également des modèles de lettres et des documents personnalisables (réclamation de congés payés, affichage obligation relatif au harcèlement sexuel, réclamation de congés payés, rupture de période d'essai à l'initiative de l'employeur ou du salarié, etc.).

Au 25 mai 2021, le code du travail numérique, dont le premier prototype a été mis en ligne en avril 2018, avait dépassé le cap des cinq millions de visites. **Cette réussite tient à plusieurs facteurs** :

- Le projet apporte une **réponse à un besoin clairement identifié** : l'accessibilité d'un droit du travail souvent complexe, en raison de ce qu'il procède de multiples sources, notamment conventionnelles, et dont la primauté diffère selon les matières, tandis que les réponses apportées aux usagers sont fractionnées au sein de multiples SI ;
- Le projet a bénéficié de l'**écosystème du développement des SIA publics et mobilisé des ressources diverses**, avec un pilotage ministériel reposant sur la fabrique numérique des ministères sociaux et de l'expertise métier (DGT) et l'appui de plusieurs dispositifs interministériels (entrepreneurs d'intérêt général, coaching de beta.gouv, financement du FTAP, appui d'Etalab…) ;
- Le projet a été **mis en œuvre de manière progressive** : la mise en ligne de premiers prototypes à partir d'un premier lot de 50 questions/réponses a précédé le développement du site tel qu'il est accessible actuellement[326].

Toutefois, il ne faut pas nier que **la démarche a suscité un certain nombre d'interrogations, voire de critiques**.

En premier lieu, son **articulation avec d'autres services d'information en ligne**, comme service-public.fr, qui délivre des informations standardisées mais exhaustives, a été questionnée[327]. Il convient toutefois de noter que le code du

[326] https://www.fabrique.social.gouv.fr/startups/code-du-travail-numerique/

[327] G. Koubi, *Le « code du travail numérique », un portail de services : renseignements et consultations enchevêtrés ?*, *in* Le Droit Ouvrier, février 2020.

travail numérique n'a vocation à remplacer ni Légifrance, ni service-public.fr, mais constitue une source d'information supplémentaire répondant à un besoin précis. La diversité des sources documentaires peut d'ailleurs apporter une certaine sécurité pour les entreprises et les salariés, et d'éventuelles contradictions peuvent être signalées afin d'amener les services compétents à s'interroger sur la cohérence des textes et la doctrine administrative.

En deuxième lieu, le CTN a fait naître une certaine **incertitude sur la portée juridique des informations qu'il délivre**. Le II de l'article 1er de l'ordonnance précitée dispose que « *L'employeur ou le salarié qui se prévaut des informations obtenues au moyen du "code du travail numérique" est, en cas de litige, présumé de bonne foi* ». Cette disposition vise à prendre en compte le risque d'erreur des informations délivrées, ces erreurs pouvant procéder d'un mauvais paramétrage du SIA lui-même mais aussi des *inputs* envoyés par l'usager (ex. la branche professionnelle dont il relève).

La portée de cette présomption de bonne foi demeure à préciser. L'administration devra sans doute en tenir compte dans les procédures de sanction des employeurs induits en erreur par le CTN, sous réserve, éventuellement, du cas où l'erreur commise par l'outil était à ce point grossière qu'elle ne pouvait abuser l'entreprise et aurait même dû l'inciter à la signaler aux services compétents. Au-delà, on peut s'interroger sur l'incidence d'une erreur dans le contentieux du licenciement, par exemple sur l'appréciation de l'existence d'une faute du salarié se prévalant d'informations fournies par le CTN et l'ayant conduit à adopter un comportement que son employeur lui reproche, ou dans le contentieux des accidents du travail, sur l'appréciation de la faute inexcusable de l'employeur. La disposition précitée, dont la rédaction diffère de celle qui fonde la possibilité de se prévaloir à l'encontre de l'administration d'une circulaire erronée (art. L. 312-3 du CRPA), ne paraît en revanche pas impliquer, contrairement à certaines analyses[328] et prévisions évoquant la perspective d'un « quasi-rescrit »[329], que les informations fournies par le CTN soient opposables à l'administration, ni, *a fortiori*, dans les relations entre les employeurs et les salariés et au juge. Autre question épineuse : reviendra-t-il à l'usager du CTN de prouver que les informations qu'il a renseignées étaient justes et qu'il n'a pas induit le SIA en erreur ?

Il sera d'autant plus nécessaire de trancher ces questions rapidement que le CTN ne comprend pas encore (si tant est que ce soit possible…) l'ensemble des accords d'entreprises alors que, depuis les ordonnances de 2017, ces derniers prévalent sur les conventions de branche dans un grand nombre de domaines.

En troisième lieu, le CTN n'échappe pas à une forme d'**effet déceptif**, dès lors que la réalisation n'atteint pas tout-à-fait les objectifs très ambitieux, et peut-être exagérés, qui avaient pu entourer son lancement. Le CTN n'est pas un « distributeur automatique de solutions » opposables à l'administration mais un « distributeur

[328] CFDT.fr, *Le code du travail numérique, quelle valeur de l'information ?*, janvier 2020.

[329] Rapport de la Commission des affaires sociales de l'Assemblée nationale sur le projet de loi d'habilitation à prendre par ordonnances les mesures pour le renforcement du dialogue social, 6 juillet 2017.

automatique d'informations »[330], dont l'apport essentiel est, grâce au TALN, de mieux comprendre les requêtes de ses usagers et d'améliorer l'accessibilité du droit. Ainsi, comme le soulignaient Laurence Pécaut-Rivolier et Stéphane Robin : « *Il ne s'agit pas là de fournir un avis (et encore moins de rendre une décision) sur une situation ou un contentieux donné, mais, plus modestement, de guider l'utilisateur (employeur ou salarié) vers les textes ou règlements pertinents.* »[331]

A l'inverse de cet effet déceptif, ce cas d'usage nous démontre, une fois de plus, que le remplacement des administrations dans leur rôle de contact et de conseil à l'usager par des SIA demeure encore au stade du fantasme. Ce n'est d'ailleurs pas l'objet du CTN que de remplacer les pôles renseignements des DIRECCTE dans le traitement des situations qui nécessitent une relations humaine et une attention particulière[332]. Ce n'est pas plus dans l'esprit de ses créateurs que de remplacer les professionnels du droit (ou les syndicats) dans le conseil aux justiciables[333], mais de faire en sorte de faire du numérique « *le vecteur d'une plus grande justice* » dès lors que « *la compétence reconnue au professionnel ne saurait justifier une inaction dont pâtirait la partie la plus faible dans la relation de travail* »[334].

Le ministère du travail a également lancé un projet d'**analyse des accords d'entreprise** afin de mieux identifier les textes applicables (on sait que de multiples avenants successifs sont signés, notamment à l'issue des NAO), de mieux exploiter leurs métadonnées (par exemple dans le cadre du Bilan annuel de la négociation collective) et de faciliter leur anonymisation avant publication.

La principale brique fonctionnelle utilisée, qui repose sur le TALN, et bénéficie des acquis techniques et fonctionnelles du CTN, utilisera un algorithme hybride, avec un réseau de neurones entraînés sur 10 millions de documents et un jeu de règles paramétrables. Ce chantier est piloté par une *task-force* ministérielle, composée de la DARES, de la DGEFP, de la DGT et de la DNUM. La première livraison est prévue pour le premier trimestre 2022, avec un portail entreprise, le lancement du nouvel SI intitulé « D@ccord » était prévu pour le dernier trimestre 2022.

2. L'IA au service des politiques d'emploi

Le cas d'usage de l'IA le plus connu, et aussi le plus simple à représenter, dans le cadre des politiques d'emploi, est **l'appariement entre offres et demandeurs d'emploi**, qui se rapproche des méthodes d'identification des candidats utilisées

---

[330] E. Netter, Le « code du travail numérique », *Intelligence artificielle, gestion algorithmique et droit du travail. Les travaux de l'AFDT*, *Dalloz*, 2020, 224719608X. ⟨hal-02924955⟩.

[331] V. l'article de L. Pécaut-Rivolier, S. Robin, « Justice et intelligence artificielle, préparer demain – épisode III2 », *Dalloz* actualité, 17 avril 2020.

[332] Ce point fait partie des enseignement de la démarche de lancement du CTN : https://www.fabrique.social.gouv.fr/startups/code-du-travail-numerique/kickoff/

[333] Cf. Y. Struillou : « le code du travail numérique : un vrai code… et qui est numérique ! », IA et droit du travail, *Dalloz*, 2020.

[334] *Idem*.

dans le recrutement[335]. Mais l'utilisation de l'IA dans le domaine de l'emploi ne se limite pas à ce cas d'usage et connaît un développement remarqué ces dernières années, notamment en France : Pôle Emploi s'est doté d'une stratégie « Intelligence Emploi » et développe plusieurs cas d'usage depuis 2018, avec l'appui financier du FTAP et des fonds européens.

- L'IA au service de la qualité du service rendu aux demandeurs d'emploi

Les SIA peuvent en premier lieu être utiles pour **personnaliser les programmes d'accompagnement** auxquels peuvent être éligibles les demandeurs d'emploi (une formation professionnelle par exemple).

Pôle emploi déploie deux outils en cours d'évaluation ou d'expérimentation. Le premier formule, en passant régulièrement en revue le dossier d'un demandeur et détectant les changements qui y surviennent, des recommandations personnalisées (allant de la suggestion d'actualisation du CV au terme d'une formation à la suggestion du recours à une aide disponible en passant par des analyses critiques…) ; elles sont transmises au demandeur et au conseiller, et fournissent ainsi une base de dialogue lors des entretiens entre eux. Un deuxième outil procède à des analyses critiques de CV et formule des suggestions d'amélioration (par exemple au vu d'expériences, de rajouter des qualités ou des capacités qui ne sont pas décrites).

La construction d'un outil d'appréciation de l'employabilité, à l'étude chez Pôle Emploi, s'avère beaucoup plus délicate. En Estonie, l'outil OTT du Fonds d'assurance chômage, entraîné sur plus de 100 000 dossiers de personnes sans emploi, évalue les chances de succès des demandeurs d'emploi dans différents parcours professionnels, procède à une catégorisation de ces derniers afin de concentrer les moyens sur ceux qui ont le plus besoin d'accompagnement, et formule des recommandations cohérentes avec leurs qualifications.

Outre qu'il ne pourrait être qu'une aide à la décision individuelle, un tel SIA soulève de redoutables problèmes de biais potentiels, y compris le biais d'automatisation. En Pologne, un SIA divisait les chômeurs en trois catégories, sur la base de données recueillies lors d'un entretien initial puis d'un test informatisé, afin de déterminer le niveau de soutien à apporter au demandeur d'emploi. Or, il s'est révélé inefficace et potentiellement discriminatoire. Le garde-fou prévu, à savoir le contrôle par l'humain de la catégorisation, s'est révélé inefficace, moins d'1% des décisions prises par l'algorithme ayant été remis en question par excès de confiance dans le SIA, manque de temps, ou crainte de répercussions de la part des superviseurs : le SIA, censé être un outil d'aide à la décision, avait donc fini par décider seul. Ce système a été jugé contraire à la Constitution et officiellement abandonné en 2019.

L'utilisation de l'IA peut également permettre d'**automatiser la demande et d'accélérer le versement de l'indemnisation chômage**. Le Fonds estonien d'assurance-chômage a ainsi développé une série d'usages de l'IA afin d'automatiser

[335] Sur les outils utilisés par les entreprises, V. « Intelligence artificielle et Management des ressources humaines : pratiques d'entreprises », F. Chevallier et C. Dejoux, *in* Enjeux numériques, n° 15, septembre 2021.

l'indemnisation des demandeurs d'emploi. Une première plateforme, EMPIS, fournit une automatisation du back-office qui, notamment, pré-remplit les formulaires, afin d'éviter aux demandeurs d'emploi de le faire, par la collecte de données provenant de différentes bases. Le temps d'inscription au Fonds serait passé d'une heure à 10 minutes. Une seconde plateforme, TETRIS, évalue l'éligibilité aux prestations d'invalidité, ce processus devant, depuis une réforme de 2016, être achevé en 30 jours, à partir du croisement de 13 bases de données. 90% des dossiers seraient terminés une semaine avant leur date d'échéance.

Enfin, l'IA peut être mise au service d'une **relation agents/usagers (chômeurs comme employeurs) simplifiée**. L'utilisation des *chatbots* est désormais bien identifiée (V. par exemple le projet de l'office de l'emploi flamand - VDAB).

Pôle Emploi a par ailleurs développé un outil de détection du niveau de priorité des mails et de traitement automatique des demandes (dans les cas définis comme éligibles par le collège IA de Pôle Emploi), afin d'accélérer les réponses et de libérer du temps-agent. Ce projet a d'ailleurs été développé suite au témoignage d'un agent qui avait identifié le temps de traitement des mails comme un irritant du quotidien : comme l'indiquait la Cour des comptes dans son rapport annuel 2020, le volume des mails a explosé sous l'effet de la transformation numérique (5,1M de mails traités en 2015, 33,8M en 2018).

A la suite du déploiement généralisé de l'outil, Pôle emploi estime que 20% du temps de travail environ a été ainsi dégagé, les agents étant dispensés de tâches répétitives, et pouvant se concentrer sur l'accompagnement des demandeurs. Aucune réponse n'est automatique : les mails sont triés et hiérarchisés, et une réponse est suggérée pour certains d'entre eux, mais il n'appartient qu'à l'agent de choisir la réponse. On notera que le déploiement du SIA a été précédé d'une mise à niveau des matériels et logiciels (le SIA ne peut fonctionner que sur un équipement numérique homogène) et que l'association et l'écoute des agents ont été la clef de réussite du processus.

Dans le même ordre d'idée, mais aussi au titre de l'objectif d'accessibilité du droit du travail et de l'emploi pour les demandeurs, Pôle emploi a déployé sur son site un moteur de recherche d'un niveau de performance analogue à celui des grands moteurs publics de recherche. Efficace à 86% (c'est-à-dire que la bonne réponse à la question posée est donnée parmi les trois premières possibilités dans 86% des cas), il assure ainsi que la première information sur le site est beaucoup plus pertinente et satisfaisante. Elle a, de ce fait, réduit de 15% le volume de mail adressé aux conseillers, leur laissant le temps gagné disponible pour une action plus utile et personnalisée.

- L'IA au service d'une politique de l'emploi plus performante

Une autre utilisation de l'IA porte sur l'**analyse « prédictive » de l'évolution du taux de chômage ou de la probabilité de devenir chômeur de longue durée**, à partir de l'analyse de variables socio-économiques (sexe, âge, nationalité), d'informations sur l'aptitude au travail et l'adéquation à certaines catégories d'emploi (éducation, santé, *soft skills*), et les opportunités (développement du marché du travail régional).

Ce type de modèles statistiques sont développés depuis les années 1990 dans de nombreux pays, mais l'IA permet de les doter de nouvelles fonctionnalités[336]. Au Danemark, le modèle combine des données provenant de dossiers administratifs et du recueil en ligne d'informations comportementales, sur une base volontaire : il oriente les travailleurs sociaux, qui gardent la main sur l'orientation des demandeurs d'emploi vers les programmes d'accompagnement. En Belgique (Flandres), la principale innovation est l'utilisation des « données de clic » qui recensent l'activité des demandeurs d'emploi sur le site Internet du service public de l'emploi. L'outil estonien OTT précédemment mentionné prévoyait d'évaluer le risque qu'un demandeur d'emploi devienne chômeur de longue durée, mais ce cas d'usage a été abandonné en raison de sa complexité et du manque d'utilité pour les services.

Dans la même veine, l'IA sert à **identifier les besoins futurs en termes de formation et de compétences**, selon les évolutions probables du marché du travail[337], et ce alors que l'amélioration de l'adéquation entre les compétences et les besoins du marché du travail demeure une priorité des agendas de lutte contre le chômage[338].

L'IA peut en outre permettre d'**accompagner les entreprises faisant face à des difficultés de recrutement**. Pôle Emploi développe ainsi des SIA basés sur des algorithmes prédictifs et de l'apprentissage machine pour modéliser un historique de données sur les offres d'emploi (ex. motifs de clôture) pour prédire leur attractivité et, à terme, faire des suggestions aux employeurs pour les rendre plus attractives.

Se développent également des outils permettant d'**identifier les entreprises susceptibles d'embaucher à court ou moyen terme**. C'est le principe de *La bonne boîte*, service développé par Pôle Emploi. Le SIA attribue un potentiel d'embauche aux entreprises, sans qu'elles aient au préalable déclaré vouloir embaucher. Les demandeurs d'emploi sont alors encouragés à envoyer des candidatures spontanées. L'outil s'appuie sur les déclarations préalables à l'embauche, dont la base de données est mise à jour chaque mois. Le site compte 400 000 visites par mois, et le taux de « réussite » de l'algorithme varie entre 70 et 80%. Ce service a été développé avec *La bonne alternance*, qui ne concerne que les entreprises susceptibles de recruter en contrat d'alternance ou de professionnalisation.

Les SIA sont aussi utilisés pour améliorer la **détection des fraudes ou des offres d'emploi illégales**. Les SIA basés sur le TALN peuvent ainsi détecter des motifs illégaux ou frauduleux dans une offre d'emploi : ce cas d'usage est développé en France, par Pôle Emploi.

---

[336] *Profiling tools for early identification of jobseekers who need extra support*, OCDE, 2018.

[337] OIT, *The feasibility of using big data in anticipating and matching skills needs*, 2020.

[338] Dans cette logique, la Suisse développe un SIA assistant les décisions d'affectation territoriale (entre cantons) des demandeurs d'asile autorisés à travailler puis des réfugiés, afin de maximiser leurs chances de trouver un emploi et de s'intégrer, en lieu et place de la répartition aléatoire qui prévalait jusqu'alors. L'objectif affiché est d'augmenter le taux d'emploi des personnes sous protection internationale de 30%.

Il convient enfin de souligner que si le service public a nettement pris conscience des atouts des SIA pour assumer ses missions, le secteur privé (de l'intérim, du recrutement, de la prestation de service…) s'est engouffré dans ce nouveau monde avec des moyens beaucoup plus importants et sans être tenu, au regard de ses missions, par les mêmes exigences de respect du principe de légalité ou des principes fondamentaux du service public. C'est l'un des domaines dans lequel l'écart de compétitivité entre privé et public peut devenir vite nuisible et fausser les décisions. Il est donc justifié que l'État y maintienne le niveau d'investissement qu'il a consenti (Pôle emploi a par exemple reçu à cette fin 30 millions d'euros du fonds de transformation de l'action publique).

Fiche n° 6 : Education

L'usage de l'IA dans le système éducatif, qui a aujourd'hui dépassé le stade de l'expérimentation, remodèle les fondements de l'enseignement et de l'apprentissage. Elle promet une éducation plus inclusive, adaptée aux besoins particuliers des élèves et aux rythmes d'apprentissage de chacun, et une transformation du rôle de l'enseignant, libéré de certaines tâches répétitives, indispensables à l'acquisition de connaissances (exercices, corrections…). L'IA apporte alors un commencement de réponse à la problématique suivante, identifiée de longue date : si le tutorat individuel est la meilleure méthode d'enseignement, il n'est pas possible de le proposer à tous les élèves[339].

Limiter l'utilisation de l'IA dans ce champ aux systèmes de tutorat intelligents est toutefois réducteur. En effet, par l'analyse massive de données, l'IA peut également **améliorer l'administration du système éducatif lui-même**. C'est par exemple l'ambition de la plateforme Parcoursup, qui vise à améliorer un processus déjà à l'œuvre avant sa création, à savoir l'automatisation de l'admission en première année de l'enseignement supérieur, dans un environnement davantage contraint par la hausse du nombre de bacheliers.

Wayne Holmes, enseignant-chercheur à l'*University College London*, consultant à l'UNESCO, auteur de plusieurs ouvrages sur l'IA dans l'éducation, distingue ainsi quatre domaines d'utilisation de l'IA dans le champ éducatif : l'IA pour les étudiants (l'enseignement et le soutien), l'IA pour les enseignants et l'IA pour le système éducatif[340].

---

[339] UNESCO, *IA and education : Guidance for policy-makers*, 2021.

[340] W. Holmes, M. Bialik, Ch. Fadel, *Artificial Intelligence in Education – Promises and Implications for Teaching and Learning*, Center of Curriculum Redesign, USA, 2019.

Différents types de SIA éducatifs[341]			
Enseignement à distance	Soutien aux élèves	Soutien aux professeurs	Soutien au système éducatif
- Systèmes de tutorat intelligents (y compris les générateurs - Tutorat basé sur le dialogue - Applications d'apprentissage des langues (y compris la détection de la prononciation)	- Environnements d'apprentissage exploratoires - Évaluation formative de l'écriture - Orchestrateurs de réseaux d'apprentissage - Applications d'apprentissage des langues - Apprentissage collaboratif - Évaluation continue - Compagnons d'apprentissage - Recommandation de cours -Aide à l'autoréflexion (analyse de l'apprentissage, tableaux de bord métacognitifs) - Agent conversationnel	- Diagnostic - Évaluation sommative de l'écriture, notation des essais - Surveillance des forums d'étudiants - Assistants pédagogiques - Génération automatique de tests - Notation automatique - Recommandation de contenu de ressources éducatives libres - Détection du plagiat - Détection de l'attention et des émotions	- Exploration des données éducatives pour l'allocation des ressources - Diagnostic des difficultés d'apprentissage (par exemple, la dyslexie) - Enseignants synthétiques - Outil de recherche sur l'apprentissage

D'autres catégorisations sont possibles. Ainsi, une analyse de 150 cas mentionnés par des études scientifiques (revues par les pairs)[342] distingue les systèmes adaptés et personnalisés d'enseignement, l'évaluation, le profilage et la prédiction et les tutorats intelligents. La répartition des cas entre ces quatre catégories montre une prédominance des usages de l'IA en faveur du profilage et de la prédiction, les autres cas se répartissant de manière équilibrée entre les trois catégories restantes :

[341] Parlement Européen, Recherche pour la commission CULT - L'utilisation de l'intelligence artificielle dans l'éducation, mai 2020. A partir de W. Holmes et al., *op. cit*.

[342] O. Zawacki-Richter, V. I. Marín, M. Bond, Fr. Gouverneur. 2019. « Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where Are the Educators ? », *International Journal of Educational Technology in Higher Education* 16 (1): 39. https://doi.org/10.1186/s41239-019-0171-0.

Applications IA dans les études évaluées par les pairs[343]		
Applications (en nombre / en pourcentage)		
Systèmes adaptatifs et personnalisation (enseignement du contenu des cours ; recommandation de contenu personnalisé ; soutien aux enseignants et à la conception de l'apprentissage ; utilisation des données académiques pour surveiller et guider les étudiants)	27	18 %
Évaluation (notation automatisée ; retours ; évaluation de la compréhension, de l'engagement et de l'intégrité académique des étudiants ; évaluation de l'enseignement)	36	24 %
Profilage et prédiction (décisions d'admission et programmation des cours ; abandon et maintien dans le cursus ; modélisation de profils d'étudiants et réussite scolaire)	58	39 %
Systèmes de tutorat intelligents (contenu de cours ; diagnostic automatisé des points forts ; création de matériels d'apprentissage ; facilitation de la collaboration)	29	19 %
Total	150	100 %

Malgré ces potentialités bien identifiées, **les stratégies dédiées à l'IA dans le système éducatif sont rares**, à l'inverse de la littérature scientifique, nombreuse sur les applications de l'IA dans ce domaine, les transformations qu'elle implique pour le métier d'enseignant ou les principes éthiques qui doivent encadrer son développement. Les réflexions menées par plusieurs acteurs publics ainsi que leurs aboutissements doivent néanmoins être mentionnés, même s'ils ne constituent pas des stratégies opérationnelles de politique publique en tant que telles :

1/ Le 24 novembre 2021, **l'UNESCO a adopté, à l'unanimité de ses 193 États membres, le premier accord mondial portant sur l'IA, la « Recommandation sur l'éthique de l'intelligence artificielle »[344]**. Ce processus, lancé lors de la 40e session de sa Conférence générale en novembre 2019, a été appuyé par un groupe d'experts *ad hoc*, chargé de formaliser un projet d'instrument normatif mondial, soumis à des consultations multipartites puis à la négociation intergouvernementale à compter de la fin 2020.

Cette recommandation souligne notamment que les enjeux éthiques du déploiement de l'IA dans le champ éducatif impliquent une attention particulière, dès lors que « *vivre dans des sociétés en cours de numérisation exige de nouvelles pratiques éducatives, une réflexion éthique, une pensée critique, des pratiques de conception responsables et de nouvelles compétences, au vu des conséquences sur le marché de l'emploi, l'employabilité et la participation citoyenne* ». L'éducation et la recherche constitue le 8ème domaine stratégique développé par cette recommandation, qui insiste sur la nécessité de transmettre au grand public les connaissances nécessaires à la maîtrise de l'IA et d'encourager l'acquisition des savoirs fondamentaux « préalables » à l'éducation à l'IA. L'utilisation de SIA dans le

[343] *Op. cit.*, note 341.

[344] https://unesdoc.unesco.org/ark:/48223/pf0000379920_fre.page=15.

domaine éducatif fait l'objet d'une mention à son point 104, sans toutefois que l'adhésion à leur déploiement soit clairement affirmée[345].

Cette recommandation est l'aboutissement d'un investissement continu sur plusieurs années de l'UNESCO sur le recours à l'IA dans les champs de l'éducation, des sciences et de la culture. Dès 2015, la Déclaration de Qingdao matérialisait l'engagement des États membres de l'UNESCO à exploiter les technologies émergentes pour réaliser l'ODD n° 4 d'une éducation inclusive de qualité. En 2019, le Consensus de Beijing sur l'IA et l'éducation[346] formulait des recommandations aux États en la matière, notamment l'utilisation de l'IA pour permettre un apprentissage personnalisé et autonomiser (et non remplacer) les enseignants. Il s'en est suivi la publication de plusieurs travaux visant à donner des outils aux acteurs éducatifs[347] ou à recenser des cas d'usages de l'IA dans le champ éducatif[348].

Si cette recommandation a donc le mérite de tracer une perspective de consensus mondial relatif à l'utilisation de l'IA et à son encadrement éthique, deux limites majeures doivent être notées. D'une part, elle ne revêt **aucun caractère contraignant** – ce qui n'a pu que faciliter son adoption. D'autre part, ce consensus mondial a été adopté sans qu'y prennent part deux États majeurs dans le domaine de l'IA, en raison de leur mise en retrait de l'organisation : les États-Unis et Israël.

2/ En second lieu, il convient de mentionner les **travaux des institutions de l'Union européenne**. Le Parlement européen, d'abord, s'est saisi de cette question avec l'adoption, le 19 mai dernier, d'une résolution relative à l'IA dans les domaines de l'éducation, de la culture et de l'audiovisuel (P98TA (2021) 0238). Cette résolution appelle notamment la Commission à **classer les SIA éducatifs parmi les utilisations à haut risque**, et à leur appliquer des « exigences plus strictes en matière de sécurité, de transparence, d'équité et de responsabilité ». C'est l'option retenue par la Commission dans le projet de règlement IA, qui indique que les SIA utilisés dans

---

[345] « *Les États membres devraient également faire en sorte que les technologies de l'IA autonomisent les élèves et les enseignants et améliorent leur expérience, tout en gardant à l'esprit que les dimensions relationnelles et sociales et la valeur des formes traditionnelles d'enseignement sont essentielles dans les relations enseignant/élève et élève/élève, et qu'elles devraient être prises en compte lorsque l'on examine l'adoption de technologies de l'IA dans l'éducation. Les systèmes d'IA utilisés dans l'enseignement devraient être soumis à des exigences strictes en matière de suivi, d'évaluation des capacités ou de prédiction des comportements des apprenants. L'IA devrait soutenir le processus d'apprentissage sans réduire les capacités cognitives, ni recueillir de données sensibles, dans le respect des normes pertinentes en matière de protection des données personnelles. Les données communiquées pour acquérir des connaissances qui sont collectées pendant les interactions entre l'apprenant et le système d'IA ne doivent pas être utilisées abusivement, détournées ou exploitées à des fins criminelles, y compris dans un but commercial.* »

[346] https://unesdoc.unesco.org/ark:/48223/pf0000368303

[347] A noter en particulier la publication d'un guide à destination des décideurs politiques en 2021 et de rapports sur l'impact de l'IA sur le développement des compétences () et le développement durable (https://unesdoc.unesco.org/ark:/48223/pf0000366994).

[348] Sans que cette recension ne se limite aux SIA publics : https://unesdoc.unesco.org/ark:/48223/pf0000370307.locale=en

l'éducation ou la formation professionnelle, destinés à être utilisés pour déterminer l'accès ou l'affectation aux établissements d'enseignement ou évaluer les étudiants relèvent des SIA à haut risque (Annexe III, point 3).

L'IA et l'utilisation des données dans le domaine de l'éducation et de la formation font par ailleurs partie des priorités du **second Plan d'action en matière d'éducation numérique (2021-2027)**, présenté par la Commission européenne en septembre 2020, et précisant ainsi une orientation déjà retenue par le premier Plan (2018). Dans ce cadre, le Plan prévoit notamment la création d'un groupe d'experts chargé d'élaborer des lignes directrices éthiques sur l'utilisation de l'IA et des données dans le cadre de l'éducation et de la formation, en vue de les partager avec les enseignants et les éducateurs, et le soutien à des activités au titre d'Horizon Europe, programme européen pour la recherche et l'innovation.

S'agissant de la France, le recours à l'IA dans le champ éducatif est une des composantes de la stratégie nationale pour l'IA. Dans son volet « Eduquer à l'IA », quatre objectifs sont poursuivis : sensibiliser, démystifier l'IA et former à l'IA en éducation ; former à un usage raisonné des ressources pédagogiques et des SIA ; assister les enseignants et aider les cadres au pilotage en améliorant la personnalisation des apprentissages et la gestion des parcours des élèves ; contribuer à la formation des citoyens et à l'employabilité des jeunes et des adultes. C'est dans ce cadre que des **partenariats d'innovation** avec des laboratoires et des start-ups ont été noués pour développer des assistants digitaux et que des formations pour les élèves et les agents du MENJS ont été proposées. Pour accompagner ces évolutions, un **comité d'éthique pour les données d'éducation** a été mis en place en 2019. Son premier avis, rendu en juillet 2020, n'aborde pas directement les enjeux spécifiques liés à l'utilisation de ces données par l'intelligence artificielle, bien que ces problématiques se recoupent largement. Il en va ainsi, notamment, de l'utilisation des données par les acteurs privés qui représentent la grande majorité des développements de SIA éducatifs[349].

Malgré des opportunités désormais bien identifiées et documentées grâce aux très nombreuses applications existantes, **le déploiement de SIA éducatifs fait l'objet de critiques (ou de craintes) qu'il convient de bien appréhender**, dans l'objectif de les lever. Une partie de ces critiques révèle certains malaises profonds qui traversent notre système éducatif (manque de moyens, remise en cause de la place de l'enseignant), mais aussi l'attachement à la promesse de la promotion républicaine par l'école (accès de tous à l'éducation, égalité entre élèves). Comme l'indique la résolution du Parlement européen mentionnée *supra*, les SIA éducatifs « peuvent déterminer le parcours éducatif et professionnel d'une personne et ont par conséquent une incidence sur la capacité de cette personne à assurer sa propre subsistance. Lorsqu'ils sont mal conçus et utilisés, ces systèmes peuvent mener à des violations du droit à l'éducation et à la formation ainsi que du droit à ne pas subir de discriminations, et perpétuer des schémas historiques de discrimination ». Il faut toutefois **se garder de faire de l'utilisation de l'IA le bouc-émissaire facile** de

[349] Comme le montrent des cartographies des usages de l'IA dans l'éducation : V. par ex. https://cartographieia.ca/. Le poids économique de l'IA éducative est estimé à 6 Md$ en 2024.

politiques publiques plus ou moins assumées : par exemple, Parcoursup n'est responsable ni de l'incapacité matérielle de l'enseignement supérieur à absorber l'augmentation régulière du nombre de néo-bacheliers (+17% de candidats entre 2018 et 2020), ni des critères de sélection mis en œuvre par les établissements.

Enfin, si cette fiche traitera pour l'essentiel de l'utilisation de l'IA dans le champ de la formation initiale, il importe également d'intégrer le recours aux SIA dans le cadre de la **formation tout au long de la vie**, qui est particulièrement propice aux parcours d'apprentissage flexibles, comme le rappelle également le consensus de Beijing sur l'IA et l'éducation, adopté sous l'égide de l'UNESCO en 2019.

1. L'enseignement adapté et personnalisé

Un des principaux apports de l'IA dans le champ éducatif est le développement d'un **enseignement adapté aux profils des élèves, selon un parcours d'apprentissage personnalisé**. Cet enseignement, lorsqu'il est entièrement personnalisé et repose sur une interaction élève/machine, intervient alors en complémentarité avec les temps collectifs d'apprentissage qui, par l'interaction, en raison des savoirs qu'ils permettent d'acquérir (et notamment de savoir-être, comme le respect des règles collectives), sont irremplaçables. Toutefois, par la personnalisation des apprentissages (rythme, exercices…) selon la progression de chaque élève et par la détection du comportement de l'élève (expressions faciales, postures corporelles…), l'IA peut permettre de répondre à la demande croissante d'une instruction inclusive davantage adaptée aux particularités et individualités des élèves.

Comme l'indique le rapport sur l'intelligence artificielle dans les domaines de l'éducation, de la culture et de l'audiovisuel de la Commission de la culture et de l'éducation du Parlement européen, « *le véritable objectif de l'IA dans les systèmes éducatifs devrait être de parvenir à une éducation aussi individualisée que possible, qui propose aux élèves des parcours académiques personnalisés en fonction de leurs forces et de leurs faiblesses et mette à leur disposition du matériel pédagogique adapté à leur profil, tout en préservant la qualité de l'éducation et le principe d'intégration de nos systèmes éducatifs* »[350].

En France, **le ministère chargé de l'Éducation nationale s'est lancé dans le déploiement de plusieurs SIA éducatifs visant à l'acquisition des savoirs fondamentaux** (français et mathématiques) au cycle 2 (CP, CE1 et CE2). Développés dans le cadre du Partenariat d'innovation en intelligence artificielle (P2IA), grâce à un marché public innovant, ces SIA, qui sont au nombre de six (3 en français et 3 en mathématiques), consistent en des services numériques d'assistance aux professeurs, permettant de différencier les activités des élèves. Ces outils sont actuellement en phase de pré-industrialisation et la phase d'exploitation démarrera après des contrôles d'accès aux ressources et de conformité (analyses juridiques).

A titre d'illustration, la solution « MathIA » propose des exercices personnalisés aux élèves, et adapte le niveau des questions aux résultats de chacun. Le professeur peut suivre sur un tableau de bord les activités de ses élèves et leur progression,

[350] Rapport 2020/2017(INI), 19 avril 2021.

constituer des groupes de niveaux et choisir des parcours « clés en main » proposés par l'outil. Kaligo combine lui deux principales applications : la lecture d'une part, l'outil analysant si la lecture de l'enfant est correcte avant de le corriger s'il y a lieu, et la dictée d'autre part, grâce à l'analyse de la graphie et la correction des fautes. Un tableau de bord permet à l'enseignant de suivre le niveau de ses élèves et, éventuellement, de mettre en place des stratégies de remédiation adaptées.

Dans le même esprit, la banque de ressources numériques éducatives propose des activités différenciées basées sur des parcours générés par IA en appui de l'enseignement des langues et cultures de l'Antiquité (LCA). A noter également qu'a été retenu dans le cadre des « Challenges Education » de BPI France, qui finance des projets d'innovation numérique, le projet « IAFLUENCE » de la DEPP du MENJS : cet outil de correction automatique de la fluence a pour objectif d'enregistrer l'élève et de l'évaluer. Il devrait être entraîné sur la base des voix enregistrées de 15 000 élèves.

En raison toutefois de la sensibilité des questions éducatives, le développement de telles technologies en la matière implique de déployer des **efforts d'explication, tant vis-à-vis des parents que des professeurs et des équipes éducatives**, et de prévenir toute visée finaliste dans le développement de l'IA.

Un certain nombre de **points de vigilance** doivent donc faire l'objet d'une attention particulière.

En premier lieu, il convient de souligner que **les SIA éducatifs ne peuvent être utiles que s'ils sont bien appropriés par les équipes enseignantes**, et pleinement intégrés dans leur pédagogie. Deux principales critiques peuvent à ce titre émerger : d'une part, la perte de contrôle de l'enseignant sur le suivi des élèves, d'autre part la perception de l'IA comme un gadget alors que les établissements ont des outils numériques obsolètes (ordinateurs, photocopieuses, etc.). Or, le développement de SIA éducatifs doit systématiquement être accompagné par les professeurs, qui auront nécessairement la charge d'assister les élèves dans leur utilisation : plus qu'ailleurs, l'adhésion de l'agent public est donc une condition du déploiement de l'IA.

Cela implique d'abord que ces technologies soient conçues comme complémentaires des méthodes pédagogiques plus « classiques » : **l'IA n'est pas là pour remplacer l'enseignant, mais pour le décharger de tâches et l'aider à analyser la progression et les difficultés de chacun de ses élèves**. C'est par exemple le parti pris de l'assistant vocal d'apprentissage de l'anglais en primaire « *Captain Kelly* », expérimenté dans une centaine d'écoles en 2020 dans le cadre du plan d'action pour une meilleure maîtrise des langues vivantes étrangères, et déployé nationalement d'ici fin 2021 (budget du marché : 850 K€), et qui propose aux enseignants des exercices à intégrer dans leur cours. C'est également l'objectif de l'agent conversationnel « Jules », l'assistant numérique développé par le CNED dans le cadre de la mesure *Devoirs Faits*, qui apporte des réponses aux questions des collégiens et propose des contenus personnalisés, afin de répondre aux obstacles à la compréhension : avec la multiplication de son usage par les élèves doit augmenter la pertinence des réponses apportées.

Les enseignants doivent donc être associés à leur conception et à leur déploiement, dans le cadre de réseaux nationaux et académiques de référents IA. Les équipes projet chargées de concevoir les SIA éducatifs devraient donc systématiquement

intégrer des professeurs, y compris en exercice. Les enseignants doivent également être sensibilisés d'une part aux outils qu'ils seraient amenés à utiliser, en formation initiale comme en formation continue, d'autre part à leur utilisation et à leur intégration dans leur programme pédagogique, enfin à leurs limites.

A ce titre, il faut souligner que le ministère de l'Éducation nationale a pris en compte cet enjeu et bien identifié l'adhésion des équipes enseignantes comme facteur clé du succès des SIA éducatifs. Les six SIA développés dans le cadre du Partenariat d'innovation IA (P2IA), ont été co-construits par des laboratoires (EdLab), des start-ups (EdTech) et des équipes pédagogiques (délégations académiques au numérique éducatif DANE, IEN chargés de la mission numérique départementale, équipes de circonscription, enseignants référents pour les usages du numérique ERUN), et déployés dans des écoles volontaires. Le budget du projet est de 16 M€. La construction du prototype expérimental, puis de la solution pré-industrialisée s'est faite par un passage du laboratoire à la classe et de la classe au laboratoire, permettant de préparer le passage à l'échelle. Le MEN propose en outre des formations académiques, des supports d'auto-formation (type MOOC) en partenariat avec l'INRIA et des webinaires, à destination des cadre, des formateurs et des professeurs[351]. Le succès de ces premiers outils, qui sont en phase d'industrialisation, a conduit le MEN à prévoir de nouveaux partenariats d'innovation (école, collège, lycée français, mathématiques, langues vivantes), budgétés à hauteur de 36 M€.

En second lieu, tout développement d'assistants éducatifs reposant sur l'IA doit s'accompagner d'une **réflexion sur l'impact du numérique dans le développement cognitif et relationnel des enfants**. Mettre dans les mains des plus jeunes des outils numériques à des fins de familiarisation est évidemment un avantage, voire une nécessité, dès lors que l'accès à nombre de services est désormais dématérialisé et que la compétence numérique est devenue essentielle sur des segments étendus du marché du travail. L'avantage des SIA éducatifs, dès lors qu'ils facilitent les apprentissages et une éducation plus inclusive, est, lui, indéniable. Toutefois, l'impact des écrans sur les plus jeunes est discuté, comme le recul des temps d'interaction humaine, et la mise en place d'un apprentissage hybride doit intégrer, dès le déploiement d'outils numériques auprès des élèves, qui doit rester modéré, un suivi particulier de leurs conséquences en termes de motricité, de développement relationnel ou encore de troubles de l'attention.

En troisième lieu, le déploiement de SIA éducatifs doit anticiper les **inégalités d'accès au numérique** et la persistance de zones blanches, qui hypothèquent l'utilisation de SIA nécessitant une connexion Internet et peut hypothéquer le déploiement de POC. Le SIA « *Captain Kelly* » est à ce titre intéressant dès lors qu'il prend la forme d'une application mobile, qui, une fois téléchargée, ne nécessite plus de connexion Internet.

2. La prévention du décrochage et de l'échec scolaires

[351] A noter, l'effort entrepris dans le projet ERASMUS+ AI4T (France, Italie, Irlande, Luxembourg et Slovénie) pour sensibiliser et former l'ensemble des enseignants aux enjeux et possibilités de l'IA en éducation sur la base du soutien initié en 2020 au MOOC IAI Class'Code.

L'IA peut être mobilisée pour la **prévention du décrochage et de l'échec scolaires**, *via* l'identification des élèves ou étudiants en difficulté ou qui peuvent l'être probablement, par l'utilisation de traitements massifs de données, des modèles d'arbres de décision ou de réseaux bayésiens. Existent ainsi des SIA de détection de la dyslexie ou de l'analphabétisme (via l'analyse de l'écriture) mais aussi de détection des profils en risque de décrochage.

Une étude portant sur l'écart d'attribution des diplômes dans l'enseignement supérieur à distance, dans les filières sciences, technologies, ingénierie et mathématiques, a étudié l'impact des systèmes de soutien aux étudiants issus de minorités ethniques en risque d'échec et montré la réussite potentielle de telles orientations : après identification de ces étudiants *via* un système d'analyse « prédictive » et intervention d'un soutien pédagogique spécifique, l'étude a montré que ces étudiants ont eu des taux de réussite plus élevés de plusieurs points de pourcentage[352].

Le système *OU Analyse*, développé par l'université ouverte du Royaume-Uni, permet ainsi l'identification des étudiants qui risquent d'échouer à leur prochain devoir : les étudiants identifiés sont alors orientés vers des tuteurs et une équipe de soutien. Est également testé un outil de recommandation d'activités personnalisées adaptées aux étudiants afin d'améliorer leur niveau.

L'IA peut également intervenir **en amont de l'orientation en première année de l'enseignement supérieur**, pour évaluer les chances de réussite des étudiants au regard de leurs résultats au lycée. L'université de Nantes a ainsi développé l'outil ApLLy (Auto-positionnement en Ligne des Lycéens[353]) qui permet aux lycéens, anonymement, d'autoévaluer leurs chances de réussir dans la filière STAPS, au regard de leurs résultats scolaires et de leur pratique sportive, et de les orienter, le cas échéant, vers un autre cursus (par exemple en apprentissage « métiers du sport »). Il a aussi été développé pour les bacheliers souhaitant s'orienter en psychologie et en droit. Dans le cadre du développement de cet outil, l'université souhaite désormais construire un moteur d'analyse automatique des données, permettant d'identifier les variables corrélées ou déterminantes de la réussite en L1 et leur pondération et d'améliorer la précision de ses recommandations *via* l'affinage des profils de réussite (datavisualisation).

De nombreux SIA permettant d'aider à **l'inclusion des élèves rencontrant des difficultés particulières** existent. Il s'agit toutefois, pour l'essentiel, d'applications privées, dont l'utilisation par les pouvoirs publics reste à préciser (et à encourager). Ainsi, la *Global Digital Library* propose des livres en libre accès pour les écoles et, par l'assistant vocal de Google, permet que les ouvrages soient lus aux personnes illettrées. De multiples outils permettent également la traduction automatique (*speech-to-text*) en langue étrangère d'un cours, afin de permettre aux élèves ne parlant pas la langue de poursuivre leurs apprentissages. Enfin, les outils permettant

[352] M. Hlosta, C. Herodotou, M. Fernandez, V. Bayer (2021). *Impact of Predictive Learning Analytics on Course Awarding Gap of Disadvantaged students in STEM*. In: 22nd International Conference on Artificial Intelligence in Education, AIED 2021, Lecture Notes in Artificial Intelligence, Springer.

[353] https://aplly.univ-nantes.fr/form/staps/32081fff-8d61-4019-82b3-90022cc629f3/welcome

de pallier les difficultés rencontrées par les élèves en situation de handicap doivent également être mentionnés (par exemple pour transcrire un cours parlé en texte fluide et ponctué pour les étudiants sourds et malentendants, ou pour mettre en surbrillance les termes importants d'un texte afin d'aider l'enfant à focaliser son attention). A ce titre, le robot *Kaspar*, développé par l'Université du Hertfordshire, est entraîné pour interagir avec les enfants autistes et les aider à acquérir des compétences sociales de base telles que l'imitation ou, par le jeu, des gestes du quotidien (tenir une brosse à dent ou une cuillère). Un projet similaire, porté par l'Agglomération de Saint-Quentin-en-Yvelines et la DSDEN des Yvelines, a été retenu dans le cadre des « Challenges Éducation » de la BPI.

3. La modernisation de l'administration de l'éducation

Comme dans d'autres champs, le recours à l'IA est perçu comme un **levier de modernisation administrative**.

En premier lieu, les SIA permettent d'**automatiser des tâches répétitives et à faible valeur ajoutée** pour les agents publics que sont les enseignants. Ainsi, selon l'OCDE, les enseignants du secondaire ne passent que 44% de leur temps de travail à enseigner, 15% à préparer leurs cours, 11% à corriger et noter leurs élèves, et 8% à accomplir des tâches administratives et de gestion, mais seulement 3% à communiquer avec les parents[354].

A ce titre, **les SIA correcteurs de copies font débat**. Ils ne sont pas (encore ?) déployés en France, mais leur utilisation grandissante à l'étranger (au moins 21 États américains les utilisent, du collège à l'université[355]) permet un premier retour d'expérience. Si le gain de temps possible pour les enseignants est évident, et si l'amélioration du service aux élèves l'est tout autant (certains outils permettent aux élèves d'obtenir un retour immédiat, de corriger leur copie et de soumettre une nouvelle copie), deux séries de critiques ont émergé.

La première est pédagogique. La correction des copies ne se limite pas à l'attribution d'une note, mais doit s'accompagner de commentaires et d'une appréciation générale afin de permettre aux élèves de comprendre leurs erreurs et de progresser. Des SIA plus développés voient donc le jour, à l'image de *Writing Pal*, développé par l'Arizona State University, qui fournit des commentaires individualisés pour améliorer les techniques de rédaction des étudiants (structures rhétoriques, expression d'idées complexes, etc.), ces derniers pouvant corriger leurs rédactions pour appliquer immédiatement les conseils prodigués.

La seconde porte sur leur fiabilité et leur crédibilité, dès lors que des biais et des erreurs ont été remarqués dans les SIA développés. Ainsi, un SIA mal entraîné peut attribuer une valeur trop forte à l'utilisation de certains termes (techniques ou recherchés), au détriment de la construction des phrases, de la substance du propos ou de la créativité rédactionnelle. Or, dès lors que de tels biais sont connus par les étudiants, le système peut être facilement trompé et attribuer une note élevée à une

[354] JRC, *Emerging technologies and the teaching profession*, 2020.

[355] https://analyticsindiamag.com/artificial-intelligence-grade-essay-student/

composition absurde. Une telle critique ne doit toutefois pas faire oublier que toute correction humaine comporte également des biais, qui ne sont pas plus acceptables parce que procédant d'un cerveau humain mal éclairé… Dès lors, tout SIA de correction doit être déployé avec prudence, être validé par des tests statistiques de grand ampleur permettant de comparer leurs résultats à ceux de correcteurs humains et, dans un premier temps, être déployé comme un « second correcteur » permettant de fiabiliser la correction humaine[356].

En second lieu, les SIA peuvent être utilisés comme **outils d'appui à la formation des enseignants ou de test d'innovations pédagogiques**. Le projet « ISTIA »[357] du rectorat de l'académie de Versailles retenu dans le cadre des « Challenges Education » de BPI France, a ainsi pour but d'élaborer un SIA permettant de proposer des offres personnalisées de formation aux agents, au regard des besoins exprimés par les agents, de leur parcours professionnel et de formation, et des besoins de l'institution. Plus innovants sont les SIA de simulation de classes qui leur permettent de tester leur enseignement sur des élèves virtuels : c'est notamment l'objet de TeachLivE, développé par l'University of Central Florida, qui permet aux enseignants de s'entraîner et de tester leurs pratiques pédagogiques avec des avatars (représentant des élèves ayant ou non des comportements spécifiques : un élève qui s'ennuie, un élève qui ne répond pas, un élève qui conteste les consignes, etc.), avant de les mettre en œuvre face à leurs élèves.

En troisième lieu, les SIA peuvent **délivrer des projections afin d'aider à la décision et au pilotage des politiques éducatives**, par exemple pour déterminer l'évolution de la carte des établissements à 10, 20 ou 50 ans, ou le nombre de professeurs nécessaires pour les épreuves de rattrapage du baccalauréat.

En quatrième lieu, les SIA constituent des **outils d'aide au pilotage des politiques éducatives en ce qu'ils permettent de mieux en connaître les impacts**. Cela est vrai, de manière générale, dans la mesure où ils permettent le traitement et l'analyse de grands jeux de données.

Ils peuvent également avoir pour objet de faciliter le suivi de cohortes d'étudiants et l'analyse de leurs trajectoires après leur sortie du système scolaire ou universitaire. Ainsi, le SIA « iTrack-Skills », *chatbot* qui recueille et analyse les réponses des diplômés, développé par l'Organisation internationale du travail et testé (phase pilote) au Monténégro et en Tanzanie (et bientôt au Bangladesh), vise à améliorer les études de suivi auprès des diplômés des formations techniques et professionnelles. Le déploiement de cet outil vise à répondre au défi d'assurer une participation élevée des diplômés à ces enquêtes, dès lors que les faibles taux de réponse à ces études réduisent leur fiabilité. Les informations recueillies par « iTrack-Skills » doivent, *in fine*, et selon ses promoteurs, être utilisées pour ajuster les programmes, améliorer l'adéquation des compétences et renforcer l'orientation professionnelle.

---

[356] V. à ce propos la grille d'analyse fondée sur la probabilité et la sévérité du préjudice sur les élèves, proposée par Vuorikari, R., Punie, Y. and Cabrera Giraldez, M., *Emerging technologies and the teaching profession*, 2020.

[357] Outil Interactif d'aide à la Structuration Individualisée de votre parcours de formation Académique

La plateforme Parcoursup

Parcoursup est, depuis 2018, la plateforme nationale de préinscription en première année de l'enseignement supérieur. Les candidats y formulent dix vœux non hiérarchisés. La répartition des élèves, dans le cadre de la procédure nationale, est mise en œuvre par un algorithme national déterministe. Lorsque le nombre de candidats à une formation excède ses capacités d'accueil, une commission propose au président de l'établissement, pour chaque formation, un classement pédagogique, sur la base de critères d'examen qu'elle définit358, et, le cas échéant, en recourant à un algorithme local. **L'algorithme national calcule alors l'ordre d'appel dans lequel les candidats se voient proposer une admission** dans la formation en question (en fonction du classement pédagogique et d'autres critères, comme le taux minimum de boursiers, ce qui revient à modifier le résultat obtenu par l'algorithme local) : lors de cette phase d'admission, Parcoursup propose, tous les jours, des propositions aux candidats classés, qui doivent confirmer leur vœu. Pendant la procédure, les lycéens sont accompagnés par leur professeur principal. En « back-office », d'autres applications, fondées sur des **algorithmes d'apprentissage**, simulent l'ensemble du processus pour le sécuriser et prévenir les erreurs, comme des surréservations (appel de trop de candidats par rapport au nombre de places).

La mise en place de cette plateforme a suscité, et suscite encore, un vaste débat, portant à la fois sur la transparence des algorithmes utilisés (seul l'algorithme national, accompagné de documents explicatifs, est public) et des critères définis par les établissements, et sur l'angoisse suscitée par les listes d'attente et les refus essuyés par les lycéens. Une telle angoisse n'est pas étonnante, dès lors que **Parcoursup constitue un nouveau rite de passage**, et la première plateforme sur laquelle le jeune bachelier prend une décision, en responsabilité, qui l'engage.

Pour accompagner ce changement, **l'effort d'explication et de transparence** mené par l'équipe projet doit être salué. Il est d'autant plus aisé que la plateforme a été développée et est administrée en interne, garantissant ainsi une maîtrise des technologies employées (l'équipe est constituée de 35 ETP, dont 25 sur la maîtrise d'œuvre). De multiples documents d'informations ont été mis en ligne, ainsi que le code source national, des données statistiques. La plateforme indique également aux candidats quels sont les attendus des formations : exigences relatives aux connaissances, à l'intérêt, à l'engagement, au savoir-être… Un comité éthique et scientifique a également été mis en place auprès de la MESRI : totalement indépendant, il définit seul les thèmes de ses travaux. Les trois rapports qu'il a remis depuis 2019, au regard de leur caractère très opérationnel, démontrent en réalité qu'il constitue davantage une instance de réflexion et de proposition (comme pourrait l'être une inspection générale) qu'un comité d'éthique.

Au final, **si Parcoursup demeure un système complexe, cela ne tient en réalité pas tant à la plateforme elle-même qu'à ce qu'il révèle de notre système d'enseignement supérieur** : inadéquation de l'offre et de la demande de formation, déficit de transparence ou inéquité des critères retenus par les universités, manque d'appropriation voire rejet de la réforme ORE par certaines équipes enseignantes, etc.

[358] Article D. 621-1-13 du code de l'éducation.

4. L'amélioration de la vie étudiante

L'IA peut, enfin, s'inscrire dans une stratégie d'**amélioration de l'accompagnement des étudiants dans leur quotidien sur les campus, *via* des assistants digitaux de type « conciergerie »**. A titre d'exemple, l'université du Staffordshire a développé un *chatbot* et un *voicebot* permettant de répondre aux questions des étudiants 24 heures sur 24[359]. L'application, qui s'installe sur les téléphones mobiles, a pour ambition de fournir des informations sur les horaires des cours ou les services disponibles sur le campus, de permettre de contacter des tuteurs personnels, de conseiller des livres, et de signaler les étudiants ayant besoin d'un soutien supplémentaire. L'application *Genie* de la Deakin University[360] joue également ce rôle de « conciergerie » en répondant à une foultitude de questions, notamment celles des étudiants de première année, portant sur les horaires, les devoirs, le plan du campus, le solde de sa carte de l'université…

En Géorgie (États-Unis), un SIA permet depuis 2016 de répondre aux questions de routine posées par les étudiants sur le forum en ligne d'un cours d'IA (quand rendre le prochain devoir, où trouver des lectures) : il permet de « compléter » l'équipe des assistants d'enseignement pour **améliorer le soutien pédagogique**.[361] Cet outil a toutefois été critiqué dans la mesure où les étudiants n'étaient pas au courant que les réponses étaient formulées par un SIA. Une telle démarche peut par ailleurs être étendue dès lors que le SIA est suffisamment alimenté : la plateforme « *Ada goes to school* » du Bolton College permet aux enseignants de créer des chatbots spécialisés par matière, connectés à de vastes bibliothèques numériques.

[359] Staffordshire University, *Introducing Beacon, a digital friend to Staffordhire University students*, janvier 2019.

[360] IT News, « Deakin's Genie assistant tackles 12,000 conversations a day », septembre 2019.

[361] G. Tech, *Artificial Intelligence Course Creates AI Teaching Assistant*, mai 2016.

Fiche n° 7 : Protection sociale

Le recours aux SIA dans le domaine de la protection sociale peut être réparti en deux types d'application : le premier concerne le rapport aux assurés (interaction avec les usagers) ; le second la liquidation des prestations et la lutte contre le non-recours aux droits ; le troisième le « back office » (traitement des données au service des stratégies de contrôle).

1. L'interaction avec les usagers

Les **robots conversationnels**, qu'il s'agisse de chatbots ou de voicebots, représentent le cas d'usage de l'IA le plus répandu dans le champ de la protection sociale. En France, ces SIA sont déjà utilisés par la CNAV (ARIA – Assistance retraite IA, qui assure 500 000 interactions par mois) ou par l'URSSAF Caisse nationale. Cette dernière a notamment mis en place, durant la première vague de l'épidémie de Covid-19, un « chatbot de crise », qui a traité environ un million de questions. L'intérêt de ces outils est d'être disponibles 7 jours sur 7, 24 heures sur 24, et, en période de pic d'activité, de permettre aux équipes d'absorber les appels téléphoniques pour les demandes les plus complexes, qui ne peuvent être traitées par un outil automatisé.

La pertinence de ces outils repose sur leur conception en lien étroit avec les métiers, les agents en contact avec les cotisants ayant vocation à être mobilisés pour la rédaction des scripts, qui doivent être mis à jour à intervalles réguliers, et l'entraînement de l'outil. Le passage à l'échelle doit également être conditionné par un taux de satisfaction élevé (autour de 80% à l'URSSAF Caisse nationale, par exemple), ce qui implique un pilotage serré et une possibilité de poursuivre l'entraînement tant que le taux-cible n'a pas été atteint. Le voicebot dédié au chèque emploi associatif, développé par l'URSSAF Caisse nationale, lancé en 2018 avec la DITP et soutenu dans le cadre de l'AMI-IA, traite aujourd'hui de bout en bout 30% des appels, dont 100% des questions les plus simples (dites « de niveau 1 »).

L'acceptabilité de ces outils ne semble pas poser de problème majeur à ce jour, ni vis-à-vis des usagers (qui sont prévenus qu'un SIA traite leur demande et dont l'appel est redirigé vers un agent humain dès que le SIA ne sait pas répondre), ni vis-à-vis des agents, habitués, dans les caisses de sécurité sociale, à ce que les systèmes d'information les déchargent de certaines tâches.

Ces robots conversationnels sont largement utilisés à l'étranger : pour les demandes liées aux prestations en cas d'accident du travail en Argentine, pour les demandes des employeurs sur le régime des travailleurs domestiques en Uruguay, pour les demandes liées aux prestations de garde d'enfants en Autriche, etc.

2. La liquidation des droits

La plupart des caisses ont recours à des systèmes permettant la liquidation automatique des droits, le SIA étant chargé de faire les contrôles réglementaires (éventuellement, en analysant les documents numérisés transmis par l'assuré pour identifier, par exemple, *via* une brique de reconnaissance de caractères, le numéro

d'allocataire) et en enclenchant le versement de la prestation. Ces technologies très matures ne posent, aujourd'hui, aucune difficulté, et présentent des taux d'erreur très marginaux. Des assistants digitaux de type RPA peuvent également traiter les dossiers incomplets en allant chercher, dans un fichier X, les informations manquantes, pour les recopier dans le logiciel Y, et ainsi liquider la prestation. L'utilisation de ces SIA, en chantier à la CNAF, implique néanmoins un accompagnement particulier des agents, dont les missions sont réorientées vers des tâches à plus forte valeur ajoutée, comme le traitement des cas les plus complexes[362].

L'apport de ce type d'outils est double : ils permettent d'automatiser le versement des aides sociales et de réduire les délais de traitement. A Trelleborg, en Suède, où une telle automatisation a été mise en place à partir de 2015, les délais de traitement des dossiers des personnes en situation de vulnérabilité économique ont été réduits de 10 à 1 jour. Au regard de la fragilité des personnes concernées, il a été décidé que les décisions de refus restent traitées par les travailleurs sociaux eux-mêmes.

Les SIA peuvent également contribuer à l'amélioration de l'accès aux droits, en identifiant, *via* des outils de *datamining*, les personnes susceptibles de bénéficier d'une prestation mais qui n'ont formulé aucune demande. Un tel modèle a ainsi été déployé par l'EDSC (Emploi et Développement social Canada) pour identifier les personnes âgées à bas revenu ayant droit à un supplément de revenu garanti. Une expérimentation est en cours à la CNAF, pour identifier le non-recours au RSA, à la prime d'activité et à l'ASF et établir des listings d'appel qui sont envoyés aux CAF.

3. Le contrôle des fraudes

Les SIA sont désormais couramment utilisés pour détecter la **fraude aux prestations** (V. aussi la fiche n° 2 relative aux SIA de contrôle et de sanction). Ils peuvent intervenir dès **l'élaboration des plans de contrôle**. La CNAF a ainsi eu recours à un SIA pour identifier les facteurs de risque et attribuer un score de risque aux dossiers. Les équipes des CAF gardent toutefois la main sur leur stratégie de contrôle (nombre de dossiers à contrôler, variable à privilégier, etc.). L'URSSAF Caisse nationale a également développé plusieurs preuves de concept pour aider les inspecteurs dans leur métier de contrôle, en leur fournissant une liste des entreprises jugées à risque : en 2019, s'agissant des TPE, le taux de redressement moyen était de 13% lorsque le contrôle était basé sur les suggestions du SIA, et de 7% pour les contrôles non fondés sur le SIA ; s'agissant des PME, l'écart varie quasiment du simple au double. Ces cas d'usage se développent également à l'étranger, l'Agence flamande pour l'enfance et la famille ayant recours depuis 2016 à un SIA pour identifier les services de garde d'enfants qui nécessitent une inspection approfondie, *via* un algorithme d'apprentissage automatique supervisé.

A l'étranger, la Norvège a déployé un outil visant à orienter les contrôles sur les étudiants déclarant vivre en-dehors du foyer familial et sollicitant le versement de prestations ou l'octroi d'un prêt par le fonds compétent. Selon les informations

[362] V. sur l'importance de la relation humaine, en tant que principe fondamental du travail social : Haut Conseil du travail social, *Travail social et intelligence artificielle,* juin 2019.

transmises, alors que seulement 5,5% des étudiants tirés au hasard bénéficiaient indûment du soutien du fonds, cette proportion était de 11,6% pour le groupe des étudiants sélectionnés à l'aide du SIA.

La mise en place de tels systèmes ne va toutefois pas sans oppositions et interrogations sur leur plus-value et les conditions de leur mise en œuvre. Ainsi, aux Pays-Bas, le système SyRi (pour « system risk indication ») avait pour objet d'orienter les enquêtes sur les fraudes fiscales ou sociales *via* l'analyse de certains quartiers, sans que les habitants soient au courant qu'ils faisaient l'objet d'une analyse. Mais, en ciblant en premier lieu les quartiers pauvres, le système n'a fait qu'alimenter les critiques sur ses biais discriminatoires, et, en 2020, sur la requête de plusieurs ONG, la Cour du district de La Haye a jugé qu'il méconnaissait le principe de respect de la vie privée garanti par l'article 8 de la CEDH, dès lors que l'équilibre avec l'objectif de lutte contre la fraude n'avait pas été trouvé (en raison notamment de son opacité). En outre, les erreurs commises par ce type de SIA peuvent avoir des conséquences catastrophiques pour les allocataires qui se voient demander de rembourser des prestations auxquelles ils avaient pourtant droit. En Australie, 400 000 personnes ont bénéficié d'une compensation à la suite de demandes de remboursement illégales.

Fiche n° 8 : Santé

De tous les champs de l'action publique dans lesquels il est possible de recourir à l'IA, celui de la santé est peut-être celui qui suscite le plus d'espoirs et de fantasmes :

1/ Parce qu'il touche directement à la vie, ce qui emporte mécaniquement une cohorte de **questions éthiques**. Il n'est à ce titre pas négligeable que l'émergence de nouvelles techniques dans le champ médical ait donné lieu à la structuration de cette réflexion éthique – dite « bioéthique » – dès la fin des années 1970.

S'agissant de l'IA en santé, les discussions sur ses implications en matière d'éthique sont nourries. Le Comité consultatif national d'éthique (CCNE) se penche régulièrement sur les enjeux liés aux outils numériques et, en juin 2021, l'OMS a publié son premier rapport mondial sur l'IA en santé, qui identifie six principes directeurs relatifs à sa conception et à son utilisation (protéger l'autonomie de l'être humain ; promouvoir le bien-être et la sécurité des personnes ainsi que l'intérêt public ; garantir la transparence, la clarté et l'intelligibilité ; encourager la responsabilité et l'obligation de rendre des comptes ; garantir l'inclusion et l'équité ; promouvoir une IA réactive et durable).

2/ Parce qu'il s'agit d'un secteur en forte expansion depuis deux siècles, ensuite, qui, touchant à un des fondements de notre contrat social, la socialisation du risque, et dans un contexte d'**extension du domaine du soin** (par l'effet du vieillissement de la population, d'une meilleure détection des pathologies, de la découverte de nouveaux traitements, etc.), pose une double question. Celle des droits des citoyens à accéder aux soins en premier lieu : le progrès sanitaire porte le risque d'une inégalité croissante entre patients, praticiens et établissements, si l'innovation en santé n'est pas accessible à tous. Mais il porte aussi la promesse d'une égalisation des conditions d'accès aux soins pour celles et ceux qui vivent dans les « déserts médicaux ». Celle de la soutenabilité des finances sociales en second lieu : la généralisation d'un suivi standardisé, reposant sur de nouvelles techniques, dont certaines sont algorithmiques, est porteuse d'un paradoxe : elle peut se traduire par une extension de l'accès aux soins à moindre coût, mais aussi par l'adoption de solutions coûteuses à produire et à mettre en œuvre et à la plus-value discutée.

3/ Parce que, pour les professions médicales et paramédicales, comme dans d'autres secteurs, l'irruption de l'IA en santé constitue un **changement ambivalent entre extension des possibles, augmentation des capacités et fantasmes du remplacement**. Les réactions des soignants face à l'IA diffèrent en fonction des spécialités, des âges et des profils, mais nous pouvons distinguer : l'optimisme raisonnable, qui repose sur une bonne compréhension des avantages et des limites des SIA de santé ; l'optimisme béat pouvant conduire à surestimer la valeur ajoutée des SIA ou leur impact sur l'organisation, avec le risque d'une prise en main éphémère ou superficielle des outils ; la curiosité, propre à chaque innovation en santé, notamment chez les praticiens qui assument en parallèle des fonctions de recherche, et pour lesquels le déploiement de l'IA en santé peut constituer un

facteur secondaire d'attractivité ; l'indifférence, résultant d'un déficit d'acculturation aux enjeux de l'IA en règle générale et de l'IA en santé en particulier ; la défiance, de la part de professionnels de santé réfractaires, convaincus que le mouvement à l'œuvre conduira au remplacement de l'homme par la machine et à une dégradation de la qualité des soins..

4/ Parce que le recours à l'IA en santé implique également de penser **l'évolution de la relation entre le patient et le professionnel de santé**, détenteur du savoir médical mais aussi accompagnant du processus de soins. L'IA peut en effet participer de l'autonomisation du patient. C'est notamment le cas des outils de surveillance qui favorisent un suivi à distance, permettant d'améliorer l'observance mais conduisant dans le même temps à une diminution des contacts et donc à distendre la relation patient/médecin.

Toutefois, l'irruption de l'IA demeure pensée dans le cadre de certains fondamentaux de cette relation entre professionnel de santé et patient, notamment l'information de ce dernier sur les décisions envisagées, leurs raisons et leurs conséquences (serment d'Hippocrate). Ainsi, l'article L. 4001-3 du code de la santé, issu de la loi du 2 août 2021 relative à la bioéthique, consacre le **droit du patient à être informé du recours à un SIA** dont l'apprentissage a été réalisé à partir de données massives (ce qui exclut les systèmes basés sur les règles) et **l'obligation pour le concepteur du traitement algorithmique de s'assurer de l'explicabilité** de son fonctionnement pour les utilisateurs (III)[363].

L'avis 129 du CCNE relatif à la révision de la loi de bioéthique (2018) avait proposé l'inscription au niveau législatif du principe de garantie de supervision humaine de toute utilisation du numérique en santé et estimé « *nécessaire que toute personne ayant recours à l'intelligence artificielle dans le cadre de son parcours de soins, en soit préalablement informée afin qu'elle puisse donner son consentement libre et éclairé* ».

5/ Parce que, dans un **contexte de concurrence potentielle entre établissements** induite par une part de tarification à l'activité (hors parenthèse Covid) et de difficultés de recrutement, à tout le moins dans certains établissements, et dans certaines spécialités, la tentation de déployer des SIA pour se conformer à l'obligation d'un soin moderne et novateur est une réalité, tout comme l'espoir de convertir ce déploiement en économies de personnel ou de consommables.

6/ Parce que la question de la **protection des données personnelles**, dont les données de santé constituent une catégorie éminemment sensible, se pose avec une acuité particulière. Dans certains cas, notamment pour certaines pathologies rares, en raison de l'existence de petites cohortes, les précautions prises (absence d'accès aux mentions écrites et défacialisation, par exemple, pour un SIA de qualification de coupes d'IRM) n'écartent pas totalement le risque d'identification des patients.

[363] V. sur une analyse de cet article : C. Crichton, « L'intelligence artificielle dans la révision de la loi bioéthique », *Dalloz*, septembre 2021.

En parallèle, un certain nombre de projets d'IA se heurtent à la qualité des données annotées, qui n'ont pas été recueillies dans la perspective d'une analyse par un SIA, mais simplement dans le cadre de la délivrance d'une prestation de santé. Ainsi, les données relatives à une personne hospitalisée pour un problème respiratoire ne feront pas nécessairement mention du cancer qui l'affecte par ailleurs. Dans certains cas, on relève jusqu'à 30% d'erreurs dans la description des pathologies associées aux malades. Dès lors, corriger ces erreurs passe par le croisement des données avec d'autres sources, comme celles correspondant aux médicaments administrés.

La France, État-providence centralisé, dispose néanmoins d'un atout déterminant, la masse des données médico-administratives, figurant au Système national des données de santé (SNDS), créé par la loi de modernisation de notre système de santé (LMSS) du 26 janvier 2016, et qui rassemble de nombreuses bases (SNIIRAM, PMSI, etc.). En effet, les SIA en santé reposant sur l'apprentissage automatique nécessitent un nombre phénoménal de données : l'INSERM indique à ce titre que les applications d'apprentissage profond exigent 50 000 images dans le cas des mélanomes et 128 000 dans celui des rétinopathies.

Par conséquent, la plupart des SIA demandent un niveau de planification et d'investissement discriminant, posant la question de l'interopérabilité des solutions, de l'hébergement des données en santé, et de la maturité des directions chargées des systèmes d'information hospitaliers (SIH). Pour soutenir les projets innovants comportant des enjeux importants en termes de protection de la vie privée, la CNIL, dans le cadre de son « bac à sables » données personnelles, a sélectionné, en mai 2021, 12 projets dans le champ de la santé qui bénéficieront d'un accompagnement renforcé ou personnalisé.

7/ Parce que l'IA en santé est un **secteur économique en forte croissance**. Son poids économique était estimé en 2020 à 4,9 milliards de dollars[364], avec une évolution de près de 50% par an. La promesse d'une innovation de rupture, d'une captation du marché ou d'un rendement marginal pousse nombre de dirigeants du domaine, d'investisseurs ou d'entrepreneurs vers le secteur. En France, le nombre de startups dans le domaine de l'IA était de 191 en 2020[365], dont plus de la moitié sont situées en Ile-de-France.

L'ensemble des acteurs de l'IA en santé partagent des problématiques communes :

- L'obligation de respecter un cadre réglementaire strict, qu'il soit relatif à la sécurité du patient ou à la validation scientifique ;
- Un besoin de compétences diverses pour concevoir, produire, distribuer, suivre et développer leur gamme de produits, et donc de coopération entre des profils médicaux et informatiques (notamment des développeurs conscients des impératifs architecturaux des SI hospitaliers) ;

[364] Institut ReportLinker, Artificial *Intelligence in Healthcare Market by Offering, Technology, Application, End User and Geography - Global Forecast to 2027*, octobre 2021.

[365] BPI, Panorama des startups santé françaises utilisant l'IA, Juillet 2020.

- Une compétition régie par des règles microéconomiques particulières : effet de réputation, asymétrie d'information pour les praticiens connaissant mal les produits, taille critique et économies d'échelles relatives au traitement massif de données, importance des innovations de rupture ou itératives ;
- Une nécessaire internationalisation, que ce soit pour mettre en commun ou comparer des jeux de données, être reconnus par la communauté scientifique ou pour accéder à de nouveaux marchés. Il convient ainsi de noter que 90% des 90 lauréats du concours BPI 2020 ont initié des collaborations nord-américaines, posant la question de l'utilisation et du stockage de ces données;
- Une absence relative d'accompagnement public par le ministère chargé de la santé, en dépit des déclarations relatives au « virage numérique », l'initiative étant dans les faits laissée au ministère de l'Économie.

Ce dernier point ne doit pas conduire à occulter **l'intervention d'une multitude d'acteurs publics**. La direction générale de l'offre de soins (DGOS) et celle de la recherche, des études, des évaluations et des statistiques (DREES) ou encore l'Agence du numérique en santé jouent un rôle de promotion, de cadrage ou d'accompagnement. Les programmes expérimentaux dits « article 51 » nés de la LFSS 2018 au niveau régional et national, le fléchage de 1,5 milliard d'euros sur quatre ans par le dispositif « Ma santé 2022 » ou l'accompagnement de la BPI et de la DGE montrent la multiplicité des initiatives, qui laissent néanmoins une forte marge de manœuvre aux acteurs locaux. Au niveau déconcentré, les ARS peuvent porter ce sujet, mais il s'agit avant tout d'une réaction à ou d'un accompagnement des initiatives privées ou d'établissements publics de santé, et d'une promotion globale de l'innovation en santé.

De ce fait, les projets d'IA en santé émergent par la collaboration entre des équipes soignantes, des chercheurs et des entrepreneurs, dans la tradition de collaboration structurée depuis 1958, et sont financés à la fois sur fonds publics et privés. A ce titre, se pose la question du retour financier et médical sur investissement pour les centres hospitaliers. S'il semble logique que les établissements puissent bénéficier de la solution à la mesure de leur participation, certains modèles conduisent les établissements à avoir à payer, après un certain temps, des solutions qu'ils ont cofinancées et dont ils ont permis l'émergence grâce à l'utilisation des données de santé dont ils disposaient. La structuration juridique et organisationnelle de ces coopérations est primordiale pour éviter un « make *and* buy » injustifié. A ce titre, il faut noter que l'émergence de « pépinières » de start-ups directement adossées aux centres hospitaliers peut favoriser un équilibre des relations contractuelles.

Et à l'étranger ?

Une rapide comparaison internationale illustre les similarités et divergences entre les politiques de l'IA en santé : similarités, car la plupart des pays « numérisés » font face à l'extension quantitative et qualitative de la sphère du soin et disposent d'un tissu médico-économique favorisant l'apparition de solutions soumises à des normes cliniques internationales ; divergences, car l'organisation des systèmes de soins et le droit des données restent le produit des cultures nationales.

Ainsi, le **système de santé britannique**, qui se rapproche du nôtre par la prééminence d'établissements publics et le financement direct d'établissements privés non lucratif, connaît lui aussi des collaborations locales entre établissements, unités de recherche et entreprises en devenir. Mais il s'est distingué par le déploiement d'un « *IA Lab* » menant des appels à projets nationaux avec un cahier des charges exigeant en termes d'interopérabilité avec l'équivalent de leur PMSI, les dossiers patients informatisés et l'architecture informatique des hôpitaux.

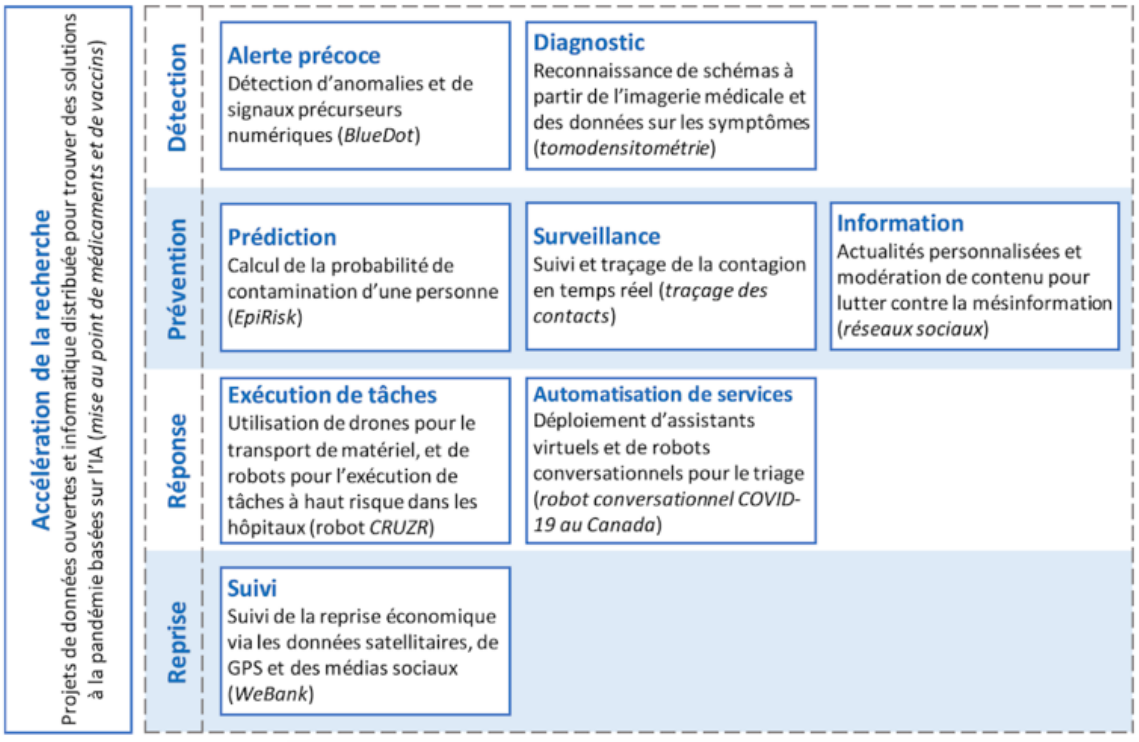
Le Royaume-Uni s'est également lancé dans la constitution d'une base de données anonymisées, à disposition des chercheurs et structures publiques ou privées, sous la houlette du *NHS Digital*. Si cette collecte automatique contenait une clause d'*opt-out* pour les patients, le délai de son application a fait débat, ainsi que sa conformité aux règles de la protection des données personnelles, équivalentes au RGPD. L'*Information Commissioner's Office*, homologue de la CNIL, a concentré ses critiques sur le premier point.

Territoire pionnier dans l'application de SIA en santé, les **Etats-Unis** se caractérisent par un environnement encore plus concurrentiel entre établissement de santé, les poussant à mettre en avant des solutions innovantes. Mais au-delà de cette compétition privée, le *HSS* et les *Centers for Medicare et Medicaid services* ont initié un paiement dédié pour l'utilisation de certaines solutions basées sur l'IA, comme *Viz ContaCT*, qui détecte les risques de thrombose aux urgences. Ce financement s'inscrit dans une transition du paiement à l'acte vers un paiement au parcours de soins (*fee for service to bundled payment*).

Enfin, **la Chine** fait face à un croisement sans commune mesure entre problématiques de santé (maladies chronique, vieillissement, inégalités d'accès aux soins) et exigence sociale d'accès aux soins, à laquelle l'IA en santé peut apporter nombre de solutions. Le droit chinois des données personnelles, autorisant leur utilisation sous réserve d'autorisation par des administrations dédiées, a permis de développer de nombreuses applications. L'avantage compétitif chinois en IA propre au financement et à la masse des données interroge, à terme, sur la conciliation entre l'efficacité des SIA et l'adaptation à d'éventuels standards internationaux.

Le déploiement de l'IA en santé pose enfin de nombreux défis en termes juridiques, de collaboration entre établissements de santé et entreprises, de sécurisation des données de santé (V. sur ce point, les difficultés rencontrées par le Health Data Hub, mentionnées dans la 4ème partie de l'étude), ou de validation scientifique. Pour autant, il est indéniable aujourd'hui que l'IA constitue une réponse (certes partielle) à de multiples enjeux de santé publique, tels que la lutte contre les maladies chroniques (grâce à des diagnostics plus rapides, plus précoces et de meilleure qualité) ou l'accompagnement des médecins face à l'accroissement continu des sollicitations et des données produites. Pour chacun de ces acteurs, il importe donc de trouver un juste milieu entre optimisme et pessimisme, de ne pas surestimer l'impact de l'IA et d'**inscrire son utilisation dans une mutation globale du système de santé**, impactant les prises en charge et les frontières des responsabilités des différents acteurs.

Si ces points incitent à la réflexion (plus qu'à la prudence), l'engouement autour des SIA en santé s'explique par la diversité des applications possibles, qu'est venue illustrer la pandémie de Covid-19. L'OCDE[366] a ainsi identifié les différents usages de l'IA aux différents stades de la crise sanitaire :



Indépendamment de cette classification, nous reviendrons sur cinq grands types d'usage de l'IA en santé.

366 « Utiliser l'intelligence artificielle au service de la lutte contre le Covid-19 », *in* Les réponses de l'OCDE face au coronavirus, avril 2020.

1. L'aide au diagnostic

L'aide au diagnostic vise à repérer une pathologie ou à en déterminer la typologie, et laisse entrevoir une amélioration qualitative et quantitative de la prise en charge.

Les Hospices civils de Lyon, dans le cadre de leur guichet d'innovation en IA, qui prodigue un soutien financier et opérationnel aux projets de SIA, développent ainsi un projet d'optimisation de la prise en charge endoscopique diagnostique et thérapeutique des tumeurs colorectales. Il en est attendu une amélioration importante de la détection des tumeurs superficielles, dès lors que les compétences humaines ont été identifiées comme variables et insuffisantes (tumeurs manquées, mauvaises prédictions de l'histologie, options thérapeutiques non adaptées…).

La détection des arrêts cardiaques lors des appels d'urgence

Le projet « Corti », développé au Danemark avec des fonds européens, vise à améliorer la rapidité et la précision du diagnostic des arrêts cardiaques extrahospitaliers dans le cadre des urgences médicales (type SAMU). Un assistant numérique vocal, entraîné par apprentissage automatique, analyse l'interaction médecin-patient et aide le preneur d'appel à identifier les signes traduisant le risque d'arrêt cardiaque. Les évaluations disponibles font état d'une réduction de 43% du nombre des arrêts cardiaques non détectés et d'une reconnaissance des signes par le SIA 25% plus rapidement que par le preneur d'appel humain.

Les solutions d'aide au diagnostic en **imagerie médicale** touchent à la détection du cancer du sein et du cancer du poumon, ou permettent l'analyse précise de radiographies thoraciques et cérébrales. Ces solutions peuvent souligner les points de fractures osseuses (logiciel *Azmed* utilisé dans plusieurs CHU, par exemple au services des urgences du CHU de Nice), les carcinomes, le diamètre d'une artère, ou hiérarchiser les images en fonctions du risque global.

Ces SIA d'aide au diagnostic dans le champ de l'imagerie médicale présentent quelques particularités prononcées qu'il convient de souligner :

- Ils se concentrent sur une spécialité réputée évolutive et technophile, avec la prise en main régulière de nouveaux scanners, PACS, suivis de produits de contraste ;
- Ils s'inscrivent dans un contexte de raréfaction de la ressource médicale, notamment dans le secteur public ;
- Ils posent des défis particuliers en termes d'anonymisation des données (défacilialisation des IRM, extraction des inscriptions sur l'image), de compatibilité de formats et de connexion aux PACS (*Picture archiving and communication system*, permettant de gérer et d'archiver les images).

L'aide au diagnostic est également permise par des algorithmes de **modélisation des facteurs de risques grâce à l'analyse de grandes cohortes**. Ainsi, dans le champ de la santé mentale, l'INSERM et l'université de Bordeaux, en collaboration avec des universités québécoises, ont élaboré un algorithme d'apprentissage automatique d'identification des principaux facteurs prédictifs des comportements suicidaires au

sein de la population étudiante[367]. Ses concepteurs espèrent s'appuyer sur le classement des 70 facteurs potentiels identifiés, selon leur importance, pour mieux identifier les étudiants à risque et les orienter vers une prise en charge adéquate.

Plus généralement, les SIA ayant recours au *datamining* permettent d'identifier toutes les anomalies possibles dans une zone donnée : on peut ainsi citer l'outil de détection des anomalies de la rétine, développé par le laboratoire de Traitement de l'information médicale de l'université de Brest, qui a été élaboré en combinant l'analyse d'images et l'analyse du contenu de dossiers médicaux[368].

On peut également citer l'outil développé par l'Université Cornell pour détecter la survenance de l'état dépressif de patients bipolaires en analysant les changements dans la saisie de messages sur un smartphone, ou encore l'exploitation de l'historique des mots-clés entrés dans un moteur de recherche pour diagnostiquer le risque de survenance d'un cancer du pancréas ou du poumon (Microsoft).

2. L'aide à la prescription et la décision médicale post-diagnostic

L'aide à la prescription est généralement obtenue par le **croisement de données multiples**. Nombre de Centres de lutte contre le cancer (CLCC) s'intéressent ainsi aux combinaisons de biomarqueurs et de cibles thérapeutiques pour réaliser des prédictions relatives à l'espérance de vie ou aux effets des traitements. Un partenariat entre Unicancer et le HDH a été signé en juillet 2021 pour permettre la collection de bases de référence en cancérologie, à partir des données des CLCC, des structures hospitalières et des structures de ville.

Le projet européen Desiree, auquel participent l'INSERM et l'AP-HP, vise à accompagner les cliniciens dans le traitement et le suivi des patientes atteintes de cancers du sein afin d'**adapter les protocoles** et améliorer la qualité des soins. Il s'appuie sur un raisonnement symbolique, *via* l'intégration des recommandations de bonnes pratiques, mais peut également apprendre des succès ou échecs rencontrés dans des dossiers médicaux déjà résolus. Une évaluation sur 138 décisions (prises sans le système, puis rejouées avec l'aide de Desiree), a montré que les décisions prises après utilisation du système étaient, dans 75% des cas, meilleures que celles prises sans : https://pubmed.ncbi.nlm.nih.gov/32972655/

On peut également citer les résultats obtenus par des équipes de cancérologie digestive de l'AP-HP, qui ont élaboré un SIA, fondé sur l'apprentissage profond, permettant de prédire la réponse thérapeutique à une radio-chimiothérapie préopératoire chez les patients atteints d'un cancer du rectum, permettant de proposer un traitement conservateur plutôt qu'une ablation complète. Dans le domaine de la transplantation rénale, des équipes de Necker et Saint-Louis ont coordonné l'élaboration du premier algorithme de prédiction du risque de perte de

[367] Scientific report, *A machine learning approach for predicting suicidal thoughts and behaviours among college students*, Article number: 11363 (2021).

[368] Science Direct, Automatic detection of referral patients due to retinal pathologies through data mining, 2016.

greffon, afin d'optimiser le suivi des patients et le développement de nouveaux médicaments immunosuppresseurs.

L'aide à la prescription vise également à garantir un **bon usage des médicaments**. Le projet *Synapse Medicine*, issu d'une alliance entre les CHU de Bordeaux et de Rennes, l'Hôpital européen George Pompidou, l'INSERM et une start-up bordelaise, propose ainsi une plateforme de données actualisées en continu sur les interactions médicamenteuses. Faciliter l'accès à une information fiable dans un contexte de production soutenue des informations scientifiques, qui a, de plus, connu une accélération inédite lors de la pandémie de Covid-19, permet ainsi de lutter contre le défaut ou l'excès d'informations, et d'analyser objectivement et en temps réel des prescriptions parfois divergentes, pour sécuriser les traitements.

Dans le champ de la **pharmacovigilance**, l'IA peut également permettre d'automatiser le contrôle des ordonnances (grâce à des briques de traitement automatique du langage naturel) ou de surveiller la réaction à des produits de santé en particulier. L'Agence nationale de sécurité du médicament et des produits de santé (ANSM) a ainsi mis en place un dispositif spécifique de surveillance des vaccins contre le Sars-CoV-2, et la nouvelle application nationale de pharmacovigilance intègre un module d'IA permettant une catégorisation et un pré-codage des effets indésirables déclarés, pour fluidifier l'expertise réalisée par les CRPV.

3. La robotique médicale

La **chirurgie assistée par ordinateur** se réfère à plusieurs types d'outils : la formation des chirurgiens et l'aide à la préparation d'une opération, *via* des logiciels de simulation ; l'aide à la réalisation d'une intervention, pour améliorer la précision des gestes ou permettre des opérations à distance.

En amont de l'opération, l'outil est une aide à la planification. Par l'analyse d'un certain nombre de documents, notamment d'imagerie médicale, le SIA peut produire des images 3D de la zone qui sera opérée, pour que le chirurgien anticipe avec la plus grande précision les gestes qu'il sera amené à réaliser. Certains logiciels peuvent aussi simuler l'opération en amont, pour former et entraîner le praticien sur des patients virtuels.

Pendant l'opération, le robot chirurgical peut améliorer la précision du geste et contrôle l'acte interventionnel. C'est un apport déjà ancien de la robotique, le premier robot, Anthrobot, ayant été utilisé pour la première fois au milieu des années 1980 : sa fonction était de donner des instruments au chirurgien par commande vocale. Ces robots sont généralement très onéreux, pour des apports qui font parfois débat. Le robot *Da Vinci* permet ainsi d'assister le chirurgien pour la réalisation d'une prostatectomie totale. 40% de ces opérations auraient été réalisées avec cette technique en 2015. La HAS a donné en 2016 un avis favorable à son remboursement par la sécurité sociale en soulignant, néanmoins, que cette technique n'apportait pas de progrès majeurs par rapport aux techniques déjà existantes.

Il faut enfin souligner le développement de robots chirurgicaux entièrement autonomes, sans assistance humaine. La première mondiale a été réalisée en 2016, pour une opération des intestins, le soutien humain étant encore nécessaire pour certaines phases de l'opération. Début 2022, un robot américain a réalisé la première opération sans aucune intervention humaine.

La robotique médicale trouve une seconde application dans les **prothèses dites « intelligentes »**, qui se différencient de leurs homologues classiques par leur capacité à agir en répondant à l'environnement et aux intentions du patient. Ces prothèses, équipées d'une caméra et/ou d'un capteur, hiérarchisent les objets rencontrés et identifient des formes jamais rencontrées en associant le mouvement humain et une série de données (image, texture, angle, lumière).

Enfin, les **robots d'assistance, correspondant un champ d'IA plus large, pourraient à terme représenter** une évolution majeure dans la dispensation des soins, **en** interagissant avec les humains, facilitant le recueil d'information, le suivi de l'observance et l'accompagnement dans la rééducation.

4. L'aide à la gestion du parcours patient

Des différents cas d'usage de l'IA en santé, l'aide au parcours patient reste le plus difficile à cerner, car l'ensemble de ces usages peuvent impacter les différentes étapes du parcours médical d'un individu, sans que son incidence soit précisément quantifiable. Pour reprendre l'exemple d'une aide au diagnostic d'une fracture périphérique aux urgences, celle-ci n'aura d'impact sur le parcours patient que si l'on considère que l'IA fait mieux que le professionnel, permet à un autre professionnel de santé de faire cette analyse et/ou au patient d'être pris en charge plus rapidement dans une filière plus adaptée.

Il convient également de préciser en introduction que **la distinction entre les SIA d'aide au diagnostic, d'aide à la décision médicale post-diagnostic et d'appui au parcours patient est souvent artificielle dans le cas de l'IA en santé**, nombre de systèmes s'attachant à traiter ensemble de ces trois aspects. Ainsi, le programme PsyCARE vise à la fois l'amélioration de la détection et de l'intervention en cas de psychose grâce aux croisements d'analyses biologiques, imageries cérébrale ou phénotypes digitaux (3% de la population étant concernée par des psychoses et 30% de ces patients ne répondant pas ou peu aux traitements), l'identification de cibles thérapeutiques, et l'adaptation de l'environnement du patient *via* des programmes thérapeutiques adaptés au patient dès les premiers stades du trouble (applications d'entraînement cognitif personnalisé et de suivi de l'engagement du patient).

Toutefois, certains SIA se concentrent non sur le parcours patient en tant que parcours thérapeutique, mais au sens de prise en charge par un service de soins. Ils consistent alors à **anticiper les flux et adapter, à terme, l'organisation des services** pour répondre aux besoins des patients. Le centre hospitalier de Valenciennes a ainsi développé un bac de données et un système visant à prédire l'afflux des patients aux urgences, en fonction des flux passés, de données épidémiologiques, des actualités ou encore des prévisions météo. Le SIA indique des anticipations à un jour, cinq jours et dix jours. D'autres modules complémentaires sont en cours de développement,

pour, d'une part, anticiper les besoins en ressources humaines (médecins, infirmier, aides-soignants) en fonction des flux attendus et, d'autre part, améliorer l'orientation des patients et l'attribution des lits, en fonction de l'état du patient, des lits disponibles et des lits prévus comme bientôt disponibles.

5. L'appui aux fonctions administratives et médico-économiques

Les potentialités offertes par le traitement automatique du langage sont telles que les SIA de synthèse automatique (d'un dossier patient par exemple, V. le logiciel *LERUDI* qui vise à filtrer et hiérarchiser l'information pour appuyer les décisions du SAMU et dans les services d'urgences[369]), de traduction, de rédaction (lettres de sortie, rapports d'imagerie), d'analyse d'ordonnances, etc. sont déjà largement déployés dans les services de santé.

Se sont également développés les SIA d'appui au **codage de l'activité**, qui visent à traiter les codifications PMSI (le Programme de médicalisation des systèmes d'information au centre du suivi de l'activité médio-économique dans le cadre de la T2A) de manière automatisée. Ainsi, les Hospices civils de Lyon déploient « ODIAR », un outil de fouille intelligente des dossiers patients informatisés pour optimiser le codage et sa valorisation. Le projet de recherche DimIA vise, à partir des données du dossier patient informatisé de chaque patient et des CRH à coder les diagnostics principaux et associés, les comorbidités, les actes et les soins effectués lors d'un séjour.

Se basant sur l'analyse textuelle, ces initiatives ont pour objectif de libérer du temps aux praticiens (qui préfèrent se concentrer sur leurs tâches de soins et/ou de recherche), aux techniciens de l'information médicale (chargés de retranscrire manuellement cette activité), d'optimiser la classification de ces séjours, et de fiabiliser la chaîne de facturation de l'activité à l'Assurance maladie.

*

Le potentiel de l'IA en santé, qui permet aux professionnels de santé d'optimiser leur temps de travail, de prioriser la prise en charge des patients et de se consacrer aux cas les plus graves, est difficilement contestable. Plus globalement, elle constitue un facteur d'attractivité pour les professionnels, dans un cadre de pénurie croissante. Si la valeur ajoutée des SIA est réelle, il convient toutefois de penser son impact global pour le patient, les praticiens et les établissements de santé, dans le cadre d'une chaîne de soins repensée. L'utilisation de l'IA ne se décrète pas, elle se construit en impliquant la multitude des acteurs de la santé. D'où l'importance comme ailleurs d'un optimisme réel mais lucide : l'IA permet de traiter en masse de l'information et de disposer d'un « filet de sécurité » à travers une surveillance adaptative ou une seconde lecture, sans remplacer l'humain du point de vue technique, organisationnel ou éthique. Pour autant, son appropriation progressive semble nécessaire pour dépasser les appréhensions et déceptions, et habituer l'hôpital à ces changements organisationnels.

[369] P. Gayet, J. Charlet, N. Janin *et al.*, « Une synthèse du dossier médical pour décider aux urgences : le projet LERUDI », Ann. *Fr. Med. Urgence* 7**,** 166–173 (2017).

Annexe 10 : Note sur le cadre juridique actuel des systèmes d'IA dans la sphère publique

Il n'existe, en droit positif, ni définition juridique de l'« intelligence artificielle », ni régime juridique propre aux systèmes d'intelligence artificielle (SIA).

Le droit applicable à un SIA résulte, d'une part, de la combinaison des *corpus* juridiques applicables à ses composantes (traitements de données, algorithmes, infrastructures et équipements, logiciels…), et, d'autre part, des exigences supérieures, de niveau constitutionnel et conventionnel, dont la portée concrète n'est toutefois révélée que par le juge au fur et à mesure de ses interventions, encore rares dans ce domaine.

I - Le droit des données à caractère personnel encadre la légalité des SIA y recourant

Tout SIA utilisant des données à caractère personnel constitue un traitement (ou un ensemble de traitements) de données à caractère personnel régi par la loi n° 78-17 du 6 janvier 1978. Ils relèvent, selon le cas, du règlement général sur la protection des données personnelles (RGPD), du titre III de cette loi pour les traitements relevant de la directive dite police-justice[370], et du titre IV de la même loi pour les traitements ne relevant pas du champ du droit de l'Union (notamment ceux qui intéressent la défense et la sûreté de l'État, comme les traitements des services de renseignement).

Il n'est pas utile de rappeler le contenu de ces textes, notamment du RGPD. On insistera ici uniquement sur les rares occurrences de dispositions qui intéressent plus particulièrement l'intelligence artificielle ou soulèvent une difficulté d'application aux SIA.

Il y a lieu d'observer à titre liminaire que l'expression même d'intelligence artificielle ne figure dans aucun de ces textes, et que ni le RGPD, ni la directive police-justice n'évoque expressément les algorithmes. Le considérant 15 du RGPD et le considérant 18 de la directive police-justice rappellent le principe de neutralité technologique de la protection des données, c'est-à-dire son application à tout traitement de données à caractère personnel, quelle que soit la technique utilisée.

1.1. L'encadrement de la prise de décision automatisée

Les dispositions qui intéressent le plus l'IA sont celles qui encadrent la **prise de décision automatisée**, y compris le profilage, défini par le droit européen comme

[370] Directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation de ces données, et abrogeant la décision-cadre 2008/977/JAI du Conseil.

« *toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne* »[371].

S'agissant des traitements relevant du RGPD, l'article 47 de la loi du 6 janvier 1978, appliquant l'article 22 du règlement, interdit en principe qu'une décision qui produit des effets juridiques ou affecte de manière significative une personne soit prise sur le seul fondement d'un traitement automatisé destiné à prévoir ou évaluer certains aspects personnels de la personne. Cette prohibition est assortie de deux séries d'exceptions, dont la portée a été éclairée par la décision du Conseil constitutionnel (décision n° 2018-765 DC du 12 juin 2018) :

- d'une part, pour les **décisions nécessaires à la conclusion ou à l'exécution d'un contrat** ou **en cas de consentement explicite** de la personne concernée, sous réserve de mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée, incluant le droit d'obtenir une intervention humaine, d'exprimer son point de vue et de contester la décision, ainsi que le droit d'obtenir sur demande les règles définissant le traitement et les principales caractéristiques de sa mise en œuvre (sous réserve des secrets protégés par la loi) ;
- d'autre part, pour les **décisions administratives individuelles** comportant une mention explicite informant leurs destinataires que celles-ci sont fondées sur un traitement algorithmique, à condition :

1° que les règles définissant le traitement et les principales caractéristiques de sa mise en œuvre soient communiquées sur demande (ce qui exclut toute prise de décision automatisée lorsqu'un secret protégé par la loi s'oppose à cette communication) ;

2° qu'aucune donnée dite sensible ne soit traitée ;

3° que le responsable de traitement s'assure de la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard ; le Conseil constitutionnel en a déduit que la prise de décision automatisée était exclue sur la base d'algorithmes « *susceptibles de réviser eux-mêmes les règles qu'ils appliquent, sans le contrôle et la validation du responsable du traitement* » ;

[371] Selon les lignes directrices du CEPD sur le profilage, ce dernier consiste à recueillir des informations sur une personne (ou un groupe de personnes) et à évaluer leurs caractéristiques ou leurs comportements afin de les placer dans une certaine catégorie ou un certain groupe, notamment pour analyser et/ou faire des prédictions sur leur capacité à effectuer une tâche, leurs intérêts ou leur comportement probable.

4° que la décision puisse faire l'objet d'un **recours administratif**, lequel ne peut être examiné sur le seul fondement d'un traitement automatisé de données à caractère personnel (V. sur ce point la décision, cons. 70, qui ajoute qu'en cas de recours contentieux, le juge « *est susceptible d'exiger de l'administration la communication des caractéristiques de l'algorithme* »).

Dans tous les cas, le Conseil constitutionnel a rappelé que les dispositions autorisant le recours à des traitements algorithmiques pour la prise de décision n'avaient ni pour objet ni pour effet « *d'autoriser l'administration à adopter des décisions sans base légale, ni à appliquer d'autres règles que celles du droit en vigueur* ».

Ce régime juridique appelle plusieurs observations.

D'une part, il est plus strict que ce que commande l'article 22 du RGPD. En effet, ce dernier admet la prise de décision automatisée sur la base d'un traitement de données sensibles, si un motif d'intérêt public important le justifie et sous réserve de garanties appropriées. En outre, si l'exigence de maîtrise du traitement algorithmique par le responsable de traitement est au nombre des « *mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée* » dont le RGPD impose l'adoption, le texte n'implique pas nécessairement la prohibition de la prise de décision automatisée sur la base d'algorithmes qui ne seraient pas explicables.

D'autre part, la portée même de cette interdiction est incertaine. Selon le commentaire aux cahiers du Conseil constitutionnel de la décision de 2018, la prise de décision automatisée ne peut reposer sur des « *algorithmes auto-apprenants* », notion que la décision ne mentionne qu'en écho à l'argumentation des requérants et en utilisant des guillemets. Dans ses observations, le Gouvernement indiquait que le texte interdisait de fonder exclusivement une décision administrative « *sur un algorithme qui n'est pas explicable, par exemple parce qu'il est auto-apprenant, c'est-à-dire capable de réviser de lui-même les règles qu'il applique en fonction de l'objectif qui lui est assigné* », ce qui limite la prise de décision automatisée aux « *algorithmes appliquant aux données qui leur sont fournies une suite d'opérations logiques ou mathématiques programmées par leur concepteur* ». Dans cette logique, seuls des systèmes basés sur les règles pourraient fonder exclusivement une prise de décision automatisée. L'exclusion des algorithmes d'apprentissage machine entraînerait une importante restriction du champ de la prise de décision automatisée dont on peine à comprendre la logique. D'une part, un système-expert d'une grande complexité peut s'avérer moins explicable qu'un algorithme d'apprentissage machine simple. Le projet de règlement européen sur l'IA prend justement soin de ne pas différencier les obligations qu'il prévoit en fonction des différentes approches de l'intelligence artificielle au regard des risques qu'elles comportent. D'autre part, s'il fallait comprendre l'expression « *susceptibles de réviser eux-mêmes les règles qu'ils appliquent, sans le contrôle et la validation du responsable du traitement* » comme renvoyant uniquement aux modèles qui continuent d'évoluer à la faveur d'un apprentissage continu ou incrémental, leur exclusion aboutirait au résultat paradoxal qu'une prise de décision automatisée serait possible à l'aide d'un modèle « figé », dont les performances et, en particulier, l'exactitude, sont en principe plus faibles.

Enfin, à la lettre, la loi de 1978, comme le RGPD, ne se préoccupe ni de l'ampleur, ni du caractère favorable ou défavorable des effets juridiques produits par une décision automatisée. En effet, en s'en tenant à la lettre du texte, toute décision produisant des effets juridiques, significatifs ou non, favorables ou non, entre dans le champ d'application de ce régime, alors que, parmi les décisions sans effet juridique, seules celles qui affectent significativement une personne en relèvent. Paradoxalement, ce régime juridique peut donc pénaliser la personne concernée en empêchant ou en compliquant l'édiction automatisée et à bref délai d'une décision dont elle serait bénéficiaire. On observera aussi que le texte ne tient pas compte des effets produits sur les tiers.

L'angle mort de l'ensemble de ce régime de prise de décision automatisée porte sur les systèmes d'aide à la décision (décision assistée par l'IA) dont la performance, la fiabilité, la commodité et la récurrence de l'usage peuvent créer un fort « **biais d'automatisation** », c'est-à-dire la tendance à accorder aux analyses et aux recommandations de la machine une confiance excessive et, le cas échéant, supérieure à son propre jugement d'humain (voire à celui d'un tiers). Dans ses lignes directrices sur la prise de décision automatisée, le Comité européen de la protection des données (CEPD) considère que seule une intervention humaine tenant compte d'autres paramètres ou, à tout le moins, réexaminant réellement la situation individuelle de la personne permet de considérer que la décision n'est pas prise uniquement sur le fondement du traitement automatisé ; le fait d'entériner de façon systématique les résultats du traitement ne suffit pas.

Le recours à la prise de décision automatisée peut être plus restreint dans certains secteurs. Ainsi :

- s'agissant des traitements de données à caractère personnel régis par la directive « police-justice », l'article 95 de la loi du 6 janvier 1978, transposant l'article 11 de cette directive dans le sens le plus strict et en cohérence avec la jurisprudence du Conseil constitutionnel (n° 2003-467 DC du 13 mars 2003), interdit par principe qu'une décision qui produit des effets juridiques ou affecte de manière significative une personne dans ce domaine soit prise sur le seul fondement d'un traitement automatisé destiné à prévoir ou évaluer certains aspects personnels de la personne. Il proscrit en outre le profilage entraînant une discrimination sur la base des catégories particulières de données (« données sensibles ») ;
- le premier alinéa de l'article 47 de la loi de 1978 interdit l'édiction d'une décision de justice impliquant une appréciation sur le comportement d'une personne sur le fondement d'un traitement automatisé de données à caractère personnel destiné à évaluer certains aspects de la personnalité de cette personne. A la lettre, cette prohibition s'étend ainsi à la simple prise de décision assistée ;
- l'article 4-3 de la loi n° 2016-1547 du 18 novembre 2016 de modernisation de la justice du XXIème siècle fait obstacle à ce qu'un service en ligne de conciliation, de médiation judiciaire ou d'arbitrage ait « *pour seul fondement un traitement algorithmique ou automatisé de données à caractère personnel* ». Il fait en outre obligation au service qui recourt à un tel traitement d'en informer les parties, en leur fournissant sur demande les règles définissant ce traitement ainsi que les principales caractéristiques de sa mise en œuvre, et de recueillir leur

consentement. L'utilisation d'algorithmes dont le responsable de traitement n'aurait pas la maîtrise est prohibée. Ces dispositions peuvent s'appliquer à des médiateurs publics (médiateur de l'énergie par exemple) ;

- l'article L. 4001-3 du code de la santé publique, issu de la loi n° 2021-1017 du 2 août 2021 relative à la bioéthique, encadre le recours, par le professionnel de santé qui dispense un acte de prévention, de diagnostic ou de soin, à un dispositif médical comportant un traitement de données algorithmique dont l'apprentissage a été réalisé à partir de données massives. Il doit en être lui-même informé et pouvoir accéder aux données du patient utilisées dans le traitement et aux résultats. Il doit aussi s'assurer que le patient est informé du recours à ce dispositif et de l'interprétation qui résulte de son utilisation. Enfin, les concepteurs du traitement algorithmique doivent s'assurer de l'explicabilité de son fonctionnement pour les utilisateurs. Ce régime s'applique aussi bien à la prise de décision assistée qu'à la prise de décision automatisée – sous réserve de la compatibilité de cette dernière avec la règle déontologique selon laquelle le médecin « *ne peut aliéner son indépendance professionnelle sous quelque forme que ce soit* » (art. R. 4127-5 du code de la santé publique) et avec le principe selon lequel le patient *« prend, avec le professionnel de santé (…) les décisions concernant sa santé* ».

2. Les difficultés posées par la soumission des SIA au RGPD

La soumission des SIA portant sur des données personnelles au RGPD n'est pas sans poser certains problèmes. Alors que le législateur européen ne semble pas envisager de modifier ce texte, ceux-ci pourraient appeler des solutions inventives dans la mise en œuvre du RGPD.

Parmi ces difficultés :

- l'exigence de **minimisation des données** posée par le RGPD entre en tension avec tant la démarche de conception d'un SIA (dont la performance est en général d'autant plus grande que les données d'apprentissage de qualité sont nombreuses) qu'avec la logique même de l'apprentissage machine (qui vise à brasser un maximum de données pour en tirer des conclusions que sa puissance seule rend accessibles), même si la recherche investit dans des IA plus « frugales » en données. Si cette exigence s'apprécie au regard de la finalité poursuivie, la question de savoir si la recherche d'une performance toujours supérieure peut justifier le recueil et l'exploitation d'une quantité toujours plus importante de données reste posée ;
- la **définition large des données « à caractère personnel » à travers la condition d'identifiabilité :** la démultiplication des sources d'information et, justement, l'amélioration des capacités d'exploitation par des outils d'IA, rendent de plus en plus aisément identifiables les personnes concernées à partir de données non directement identifiantes. Or dès lors que cette condition est remplie, le RGPD devient en principe applicable en toutes ses dispositions. Il peut en résulter une forme de disproportion entre la modestie de l'occurrence et de la gravité du risque d'atteinte à la vie privée et la lourdeur des contraintes juridiques pesant sur le responsable de traitement. Se pose par exemple la question de

l'application du RGPD aux traitements algorithmiques dépourvus de toute fonctionnalité de reconnaissance faciale et procédant à la suppression quasi-instantanée de données susceptibles de permettre l'identification des personnes dont le visage est capté, « un instant de raison » ;

- la **minimisation de la durée de conservation** peut faire obstacle à un audit ou un contrôle efficace du modèle, qui suppose de disposer des données originellement utilisées pour son entraînement ;
- le **principe de limitation des finalités** (avec l'encadrement des traitements subséquents poursuivant des finalités compatibles avec la finalité initiale) et la **granularité du consentement libre et éclairé** des personnes concernées (finalité par finalité) limitent les possibilités d'élargissement ou d'infléchissement de l'objet du SIA en construction, alors que la phase de conception comme la phase de déploiement peuvent révéler la nécessité ou la pertinence d'une révision des finalités. La recherche réitérée du consentement ou même la nécessité d'informer les personnes concernées afin de les mettre à même d'exercer leur droit d'opposition peuvent s'avérer très pénalisants, au point de tenir en échec des projets répondant à un motif d'intérêt public éminent ;
- le **caractère potentiellement théorique, sinon fictif, des droits des personnes** (notamment celle de transparence, si les informations portées à la connaissance de la personne concernée sont trop techniques ou touffues pour être intelligibles, dans un contexte de complexification des algorithmes) ou les difficultés opérationnelles que soulèverait leur interprétation large (ex: possibilité donnée à la personne concernée de faire rectifier non seulement ses données à caractère personnel qui sont factuellement inexactes, mais aussi, comme l'indique le CEPD, les résultats d'un algorithme de profilage qui ne lui conviennent pas) ;
- le risque d'un **doublonnement entre l'étude de risque** qu'exigerait le règlement IA **et l'analyse d'impact sur la protection des données (AIPD)** prévue par le RGPD.

II - Les citoyens bénéficient d'un droit d'information relativement étendu sur les traitements algorithmiques utilisés par l'administration

Indépendamment du droit dont dispose toute personne concernée d'accéder à ses propres données à caractère personnel, utilisées pour l'apprentissage d'un système d'IA, deux séries de dispositions générales assurent une certaine transparence du fonctionnement de ces systèmes, sans préjudice d'éventuelles dispositions particulières.

1. Le droit à la communication des documents administratifs

L'article 15 de la Déclaration de 1789 consacre le **droit à la communication des documents administratifs**, qui peut faire l'objet de limitations liées à des exigences constitutionnelles ou justifiées par l'intérêt général, pourvu que l'atteinte qui y est ainsi portée n'est pas excessive au regard de l'objet poursuivi (décision n° 2020-834 QPC du 3 avril 2020).

Dans ce cadre, l'article L. 311-2 du code des relations entre le public et l'administration garantit à toute personne le droit d'obtenir communication des documents qui présentent un lien suffisamment direct avec la mission de service public dont est investie une personne morale de droit public ou de privé. Tel est le cas des fichiers, des codes-sources, et de l'ensemble de la documentation achevée relative à un SIA utilisé dans le cadre d'une telle mission.

Le caractère préparatoire des documents (par exemple, dans la phase de conception du SIA) fait temporairement échec au droit d'accès. En outre, les secrets protégés par la loi mentionnés à l'article L. 311-5 de ce code s'opposent à la communication des documents qui en sont couverts. Les principales exceptions susceptibles d'être opposées à ce titre sont le secret de la défense nationale (pour les SIA relevant de ce secteur), la sûreté de l'État, la sécurité publique, la sécurité des personnes et la sécurité des systèmes d'information des administrations, ainsi que la recherche et la prévention des infractions de toute nature. Le législateur peut ajouter d'autres exceptions au cas par cas, sous réserve de respecter le principe de proportionnalité (V. à propos du secret des délibérations des équipes pédagogiques, la décision du 3 avril 2020 précédemment mentionnée, à propos de Parcoursup). Le Conseil constitutionnel tient compte notamment du rôle joué par le SIA et de la place qu'occupe l'intervention humaine dans le processus, du caractère obligatoire ou non du recours à l'IA ou encore de la disponibilité d'autres informations permettant de comprendre les ressorts des décisions prises.

La protection de la vie privée, mentionnée au 1° de l'article L. 311-6 du CRPA, est également susceptible de s'opposer à la communication non seulement des données à caractère personnel utilisées pour l'entraînement du modèle mais aussi, le cas échéant, des codes-sources s'il apparaît possible, par rétro-ingénierie, de retrouver ces données.

Enfin, s'agissant d'algorithmes développés par des entreprises privées pour les besoins de l'administration, le **secret des affaires** est susceptible de faire obstacle à la communication de certaines informations présentant une valeur commerciale. Cette exception ne paraît toutefois pouvoir jouer que de façon marginale, eu égard à l'importance de l'enjeu de transparence des algorithmes et aux exigences de l'article 15 de la Déclaration de 1789. Il y a lieu en outre de rappeler que les droits de propriété intellectuelle dont le code-source est grevé peuvent être transférés à l'administration dans le cadre de l'option B de l'article 38 du cahier des clauses administratives générales des marchés relatifs aux technologies de l'information et de la communication[372]. Dans ce cas, ces droits de propriété intellectuelle détenus par l'administration ne sauraient fonder un refus de communication sur le fondement de l'article L. 311-4 du CRPA, pas plus qu'ils ne font obstacle à la qualification d'informations publiques et à l'application subséquente du régime de réutilisation prévue par le CRPA (CE, 8 février 2017, *Société Notrefamillecom*, n° 389806). En revanche, lorsqu'un tiers à l'administration est titulaire des droits de propriété intellectuelle sur le système d'IA, l'article L. 311-4 du CRPA ne permet la

[372] Tel a été le choix, par exemple, de la région Occitanie, dans de récents marchés publics dédiés au développement de SIA.

communication des documents grevés de tels droits qu'avec l'accord de leur titulaire (CE, 8 novembre 2017, *Association spirituelle de l'Eglise de scientologie Celebrity Centre*, n° 375704).

2. L'information renforcée des destinataires de décisions individuelles fondées sur un traitement algorithmique

Les personnes destinataires de décisions individuelles fondées sur un traitement algorithmique bénéficient d'une **information renforcée et, pour partie, spontanée**.

D'une part, la **prise de décision automatisée**, c'est-à-dire l'édiction d'une décision sur le seul fondement d'un traitement algorithmique, donne lieu à une information particulière des personnes concernées sur son existence et « *des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée* (V. le f) du paragraphe 2 de l'article 13 et le g) du paragraphe 2 de l'article 14 du RGPD, et son article 15 qui permet, au titre du droit d'accès, d'obtenir les mêmes informations.

D'autre part, la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique a prévu dans le code des relations entre le public et l'administration (CRPA) une série de garanties d'informations bénéficiant à **toute personne faisant l'objet d'une décision individuelle fondée sur un traitement algorithmique**, y compris s'il n'en constitue pas le fondement exclusif et que, par suite, le régime de la prise de décision automatisée n'est pas applicable. Ce champ d'application large se déduit à la fois de la différence de rédaction entre le CRPA et l'article 47 de la loi de 1978, et sur la disposition de l'article R. 311-3-1-2 du CRPA qui exige de porter à la connaissance de la personne le « degré de contribution » du traitement à la prise de décision – ce qui implique qu'il puisse s'agir d'une décision automatisée ou simplement assistée par le système.

Ce principe d'information, qui s'applique sous réserve des secrets protégés par la loi, s'apparente à une fusée à trois étages :

- Premier étage : doivent être publiées en ligne, de manière spontanée, les « règles » qui définissent les « principaux » traitements algorithmiques des administrations d'au moins 50 agents, lorsqu'ils fondent des décisions individuelles (art. L. 312-1-3 du CRPA) ;
- Deuxième étage : toute décision individuelle prise sur le fondement d'un traitement algorithmique doit le mentionner et rappeler les droits dont bénéficie son destinataire (art. L. 311-3-1). La jurisprudence n'a pas éclairé les conséquences d'une lacune, alors que l'article 47 de la loi de 1978 prévoit expressément la « nullité » d'une décision automatisée ne comportant pas une telle mention ;
- Troisième étage : si l'intéressé en fait la demande, l'administration doit lui communiquer les règles définissant le traitement, et les principales caractéristiques de sa mise en œuvre. Les informations à communiquer, sous une forme intelligible, sont énumérées à l'article R. 311-3-1-2 du CRPA : il s'agit du degré et du mode de contribution du traitement algorithmique à la prise de décision, de la nature des données traitées et leurs sources, des paramètres du

traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé, ainsi que les opérations effectuées par le traitement.

Ces dispositions n'excluent pas la possibilité pour le législateur de régir par des dispositions spéciales le droit d'accès aux documents relatifs aux traitements algorithmiques (CE, 12 juin 2019, *Université des Antilles*, n° 427916, 427919, T.), dès lors que la restriction qui en découlerait poursuit un objectif d'intérêt général. Il en va ainsi, par exemple, pour la procédure nationale de pré-inscription pour l'accès aux formations initiales du premier cycle de l'enseignement supérieur dite « Parcoursup » : la protection du secret des délibérations des équipes pédagogiques au sein des établissements, afin d'assurer leur indépendance et l'autorité de leurs décisions, justifie certaines restrictions au droit d'accès (décision n° 2020-834 QPC du 3 avril 2020, *Union nationale des étudiants de France*).

Il y a lieu d'observer que ces dispositions ne concernent qu'un fragment très minoritaire des usages actuels de l'intelligence artificielle, dans la mesure où, le plus souvent, les usages existants ou envisagés à court terme ne conduisent pas à l'édiction d'une décision individuelle.

III - Les règles applicables aux systèmes d'information de l'administration

Il ne sera ici question, brièvement, que de quelques règles propres à ces systèmes.

L'article 16 de la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique prescrit aux administrations de veiller à « *préserver la maîtrise, la pérennité et l'indépendance de leurs systèmes d'information* », et d'encourager l'utilisation des logiciels libres et des formats ouverts lors du développement, de l'achat ou de l'utilisation, de tout ou partie, de ces systèmes d'information.

En ce qui concerne la cybersécurité, le II de l'article 9 de l'ordonnance n° 2005-1516 du 8 décembre 2005 relative aux échanges électroniques entre les usagers et les autorités administratives et entre les autorités administratives prescrit à toute autorité administrative qui met en place un système d'information de déterminer les fonctions de sécurité nécessaires pour protéger ce système. Celles-ci doivent respecter le référentiel général de sécurité approuvé par un arrêté du 13 juin 2014. En outre, les opérateurs de services essentiels sont soumis à des prescriptions particulières qui découlent de la loi n° 2018-133 du 26 février 2018 portant diverses dispositions d'adaptation au droit de l'Union européenne dans le domaine de la sécurité.

S'agissant de l'interopérabilité, l'article 11 de la même ordonnance renvoie à un référentiel général d'interopérabilité le soin de fixer les règles techniques adéquates, en déterminant notamment les répertoires de données, les normes et les standards qui doivent être utilisés par les autorités administratives. Le référentiel en vigueur a été approuvé par un arrêté du 20 avril 2016.

IV - Les normes supérieures encadrent l'utilisation des SI

Une fois adopté, le règlement sur l'IA – le cas échéant complété par l'instrument juridiquement contraignant en gestation au sein du Conseil de l'Europe – donnera aux systèmes d'IA un cadre juridique complet et harmonisé au sein des Etats membres de l'Union européenne.

Dans l'intervalle, outre les règles précédemment mentionnées, le recours aux SIA sera encadré par les exigences constitutionnelles et conventionnelles.

1. Les exigences constitutionnelles peuvent interdire ou encadrer le recours à certains SIA

Le Conseil constitutionnel a eu peu d'occasions d'éclairer cette problématique. Dans sa décision n° 2003-467 DC du 13 mars 2003, il a jugé que l'exclusion du recours exclusif à un algorithme pouvait constituer une garantie nécessaire à la conformité à la Constitution de traitements particuliers de données à caractère personnel (en l'occurrence, dans le domaine de la police). Sa décision du 12 juin 2018 déjà mentionnée traite de la prise de décision automatisée. Sa décision n° 2019-796 DC du 27 décembre 2019 apporte des éclairages sur le dispositif de collecte automatisée des données à caractère personnel rendues publiques par les personnes concernées sur les réseaux sociaux et les sites internet. Enfin, sa décision du 3 avril 2020 précédemment mentionnée porte sur le droit d'information sur les algorithmes utilisés par l'administration.

Sous réserve de cette dernière décision, c'est donc **au regard des exigences constitutionnelles s'imposant aux traitements de données à caractère personnel** qu'il a examiné la conformité à la Constitution de dispositions impliquant le recours à l'intelligence artificielle, en vérifiant que le législateur n'avait pas opéré une conciliation déséquilibrée entre les objectifs d'intérêt général poursuivis et le droit au respect de la vie privée, ainsi que la liberté d'expression et de communication, eu égard aux garanties dont le dispositif en cause était assorti. La plupart de ces garanties font directement écho au droit des données à caractère personnel : destruction rapide des données collectées insusceptibles de traitement ou s'avérant sans intérêt (principe de minimisation) ; droit des personnes à la rectification de leurs données (principe d'exactitude) ; niveau de grade des agents mettant en œuvre le traitement, spécialement habilités à cette fin et tenus au secret professionnel (mesures techniques et organisationnelles de sécurité) ; limitation de la sous-traitance à la conception de l'outil, la sous-traitance étant exclue pour la collecte et le traitement (ce qui, cependant, laisse de côté la question des données d'entraînement) ainsi que la conservation.

- Le principe de légalité et la garantie des droits

Il va de soi que le principe, rappelé par le Conseil constitutionnel à propos de dispositions législatives autorisant la prise de décision automatisée, selon lequel de telles dispositions n'habilitent pas l'administration à adopter des décisions sans base légale, ni à appliquer d'autres règles que celles du droit en vigueur, vaut pour

l'ensemble des SIA. L'administration doit donc toujours s'assurer de la conformité des décisions qu'elle édicte aux règles s'imposant à elles, qu'elle utilise ou non un SIA.

La complexité de traitements algorithmiques utilisés par l'administration et la difficulté dans laquelle elle peut se trouver d'en expliquer les résultats, y compris au juge, peuvent soulever une interrogation à cet égard. Il peut s'avérer difficile, pour le citoyen mais aussi pour le juge, d'apprécier la fiabilité et la valeur de cette information, prise en compte par l'administration pour prendre sa décision.

- *Le droit au respect de la vie privée*

La disposition en cause dans sa décision n° 2019-796 DC du 27 décembre 2019 (LFI pour 2020), prévoyait le déploiement, à titre expérimental et pour une durée de trois ans, d'une assistance à la recherche de la fraude fiscale par l'emploi d'un système d'IA fondé sur la collecte et l'exploitation automatisées des contenus accessibles publiquement sur les réseaux sociaux (« *scraping* ») et de nature à révéler des signes extérieurs de richesse (au sens de l'article 168 du CGI) ou des activités irrégulières ou dissimulées.

L'analyse du Conseil constitutionnel repose sur le fait que la collecte de données s'opère de manière massive et indifférenciée sur les plateformes par des moyens « *informatisés et automatisés* » en vue d'agréger les données en opérant des recoupements et des corrélations. Le juge constitutionnel raisonne alors classiquement en vérifiant que n'est pas déséquilibrée la conciliation entre l'intérêt général justifiant le recours au traitement – ici, la poursuite de l'objectif de valeur constitutionnelle de lutte contre la fraude et l'évasion fiscales, grâce au renforcement des pouvoirs des moyens de contrôles des administrations fiscales et douanières – et l'atteinte portée à la vie privée, eu égard aux nombreuses garanties dont le dispositif est assorti.

Son contrôle de nécessité et de proportionnalité apparaît particulièrement strict :

- d'une part, le Conseil constitutionnel s'assure que seules les infractions les plus graves ou celles qui sont spécifiquement commises grâce à internet justifient qu'il soit procédé à la collecte automatisée massive de données. Il a d'ailleurs censuré, à ce titre, les dispositions qui auraient permis de recourir à ce traitement à l'encontre des personnes ayant omis une déclaration, puisque l'infraction est déjà connue et que sa répression est possible sans recourir à un tel outil ;
- d'autre part, le Conseil constitutionnel tire de l'exigence de proportionnalité rappelée par la loi qui lui était déférée l'obligation pour le pouvoir réglementaire, sous le contrôle du juge, de veiller à ce que les algorithmes utilisés par ces traitements ne permettent de collecter, d'exploiter et de conserver que les données strictement nécessaires à ces finalités. Il estime ainsi, au titre des garanties nécessaires, que le dispositif législatif n'est pas « auto-portant » et qu'il doit être précisé par voie réglementaire afin d'encadrer le choix et la portée des algorithmes utilisés ;
- enfin, tout en admettant la conformité à la Constitution de ce dispositif, le Conseil constitutionnel s'est réservé la possibilité de procéder à un nouvel examen à l'issue de l'expérimentation, à la lumière des résultats produits par cette collecte

automatisée de données. Ce faisant, il a entendu marquer que l'atteinte portée à la vie privée devait être justifiée par la contribution effective du dispositif aux motifs d'intérêt général poursuivis, et non de manière théorique.

- La liberté d'expression et de communication

Toujours dans sa décision du 27 décembre 2019, le CC a souligné que l'atteinte à la liberté d'expression et de communication résultait de ce que, en effectuant cette collecte, le système d'IA était susceptible de dissuader d'accéder aux plateformes, préalablement identifiées comme constituant un vecteur de la liberté en cause[373]. Le Conseil constitutionnel a ainsi relevé que la loi limitait les données susceptibles d'être collectées aux plateformes ouvertes au public et aux contenus volontairement rendus publics, à l'exclusion de toute donnée sensible.

- Le principe d'égalité

Même si le Conseil constitutionnel n'a pas eu à se prononcer sur ce point dans sa décision de 2019, il va de soi que tout SIA opéré par l'administration, qu'il implique ou non un traitement de données à caractère personnel, est soumis au principe d'égalité.

La portée de cette évidence reste toutefois à clarifier, au regard de la problématique de l'équité et de celle des biais algorithmiques discriminatoires, développées dans la troisième partie de l'étude.

2. En l'état, la portée des exigences conventionnelles ne se distingue pas de celle des normes constitutionnelles

On retrouve, dans la convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales et, dans le champ du droit de l'Union, dans la charte des droits fondamentaux de l'Union européenne, des exigences comparables à celles qui viennent d'être mentionnées, qu'il s'agit du droit au respect de la vie et de la protection des données à caractère personnel, de la liberté d'expression ou du principe d'égalité.

La jurisprudence des cours européennes, très rare en matière d'IA, n'aborde le sujet qu'au travers du prisme de la protection des données personnelles, laissant de côté à ce stade tout le pan de l'IA qui ne les implique pas directement. **La jurisprudence CEDH repose classiquement sur le principe de nécessité de l'ingérence** (le recours à l'IA traitant de données personnelles ne doit être envisagé que s'il est nécessaire et proportionné, eu égard à l'effectivité des garanties – y compris le droit au recours effectif). Les États prenant les devants dans le déploiement de ces techniques ont aux yeux de la cour une responsabilité éminente pour définir de bonnes pratiques et de bons usages. La cour reconnaît la nécessité de ne pas renoncer à des techniques (dans la compétition avec le crime organisé notamment) mais les regarde avec

[373] Décision n° 2009-580 DC du 10 juin 2009, dite « Hadopi I », à propos d'un dispositif limitant l'accès technique à Internet : la liberté d'expression et de communication implique la liberté d'accéder aux services sur Internet, eu égard, notamment « *à l'importance prise par ces services pour la participation à la vie démocratique et l'expression des idées et des opinions* ».

circonspection. Les principaux droits impactés par l'IA dans la jurisprudence sont le droit à la vie privée, la dignité, le droit au recours, la liberté de communication et la non-discrimination[374].

Le droit positif conventionnel est donc peu fourni.

L'Assemblée parlementaire du Conseil de l'Europe a toutefois exprimé le souhait qu'un instrument juridiquement contraignant encadrant le recours à l'IA soit élaboré à ce niveau. A la différence de la proposition de règlement de la Commission européenne, elle semble moins se préoccuper de la création d'un cadre de confiance susceptible de stimuler l'innovation technologique que des risques résultant de l'utilisation des SIA. En effet, si elle « convient » que l'IA peut avoir des retombées très positives, elle s'est surtout dite « *fermement convaincue de la nécessité d'instaurer un cadre réglementaire général pour l'IA, définissant des principes spécifiques fondés sur la protection des droits de l'homme, la démocratie et l'État de droit* ».

Il importe que le Gouvernement français, avec ses partenaires européens, s'assure de la convergence de cet instrument avec les règles fixées par le règlement IA, en anticipant sur l'interprétation que les juridictions seront susceptibles d'en donner.

Dans l'intervalle, les instances européennes et internationales ont produit une doctrine éthique abondante et largement convergente, qui demeure toutefois peu opérationnelle et, en tout état de cause, dépourvue d'effet juridique.

Au plan européen, la **CEPEJ** (commission européenne pour l'efficacité de la justice auprès du Conseil de l'Europe) a adopté, les 3 et 4 décembre 2018, une charte des principes éthiques relatifs à l'utilisation de l'IA dans les systèmes judiciaires, adoptée par les 47 Etats membres du Conseil de l'Europe. Les cinq principes retenus sont ceux de respect des droits fondamentaux, de non-discrimination, de qualité et de sécurité, de transparence, neutralité et intégrité intellectuelle, de maîtrise par les utilisateurs. Non-contraignante, promouvant des principes aussi généraux que consensuels (ce qui a, au demeurant, facilité son adoption), cette charte sert de base pour les programmes de coopération bilatérale de la Cepej, dont la majorité porte désormais sur la digitalisation et l'IA dans les systèmes judiciaires.

Au plan international, **l'UNESCO** a également adopté une recommandation lors de sa conférence générale de novembre 2019, comportant des principes éthiques généraux qui devraient être soumis fin 2021 à la conférence générale. Ayant la vocation de devenir un instrument universel, la recommandation se fonde sur une définition volontairement large de l'IA[375], et énonce le principe de respect des droits

---

[374] Principalement : GC 4 décembre 2008, *S et Marper c/Royaume uni*, n° 30562/04 et 30566/04) ; 13 février 2020, *Gaugrham c/ Royaume uni*, n° 45245/15) ; 30 janvier 2020, *Breyer c/ Allemagne*, n° *50001/12) ;* 12 janvier 2016, *Szabo et Vissy c/ Hongrie*, n° 37138/14).

[375] « *Systèmes capables de traiter les données et l'information par un processus s'apparentant à un comportement intelligent, et comportant généralement des fonctions de raisonnement, d'apprentissage, de perception, d'anticipation, de planification ou de contrôle.* »

humains, de reconnaissance de l'importance de la protection de l'environnement et des écosystèmes, de recherche de la diversité et de l'inclusion, de contribution à la vie dans des sociétés pacifiques justes et interdépendantes ; ces objectifs, que l'IA doit prendre en considération, doivent être atteints en respectant les principes de proportionnalité et d'innocuité (ce dernier comportant la nécessité de fonder l'IA sur des bases scientifiques et d'assurer la prévalence du choix humain pour les décisions graves ou irréversibles, de sécurité et de sûreté, d'équité et de non-discrimination, de durabilité, de respect de la vie privée, de surveillance et de décision humaine, de transparence et d'explicabilité, de responsabilité et de redevabilité, de sensibilisation et d'éducation, de bonne gouvernance et collaboration multipartite et adaptative).