

Strasbourg, 21 mai 2021

CAHAI-PDG(2021)06

COMITÉ AD HOC SUR L'INTELLIGENCE ARTIFICIELLE (CAHAI)

GROUPE D'ÉLABORATION DES POLITIQUES (CAHAI-PDG)

L'intelligence artificielle dans le secteur public

www.coe.int/cahai

Table des matières

1. Introduction	3
2. L'IA et le secteur public : vue d'ensemble	3
3. Cas d'utilisation des applications d'IA mis en place par les États membres du Conseil de l'Europe	8
3.1 Exemples d'utilisation dans le domaine judiciaire.....	8
3.2 Exemples d'utilisation dans le domaine fiscal et financier.....	9
3.3 Exemples d'utilisation dans le domaine des soins de santé	10
3.4 Exemples d'utilisation dans le domaine de la lutte contre la criminalité.....	11
3.5 Exemples d'utilisation dans le domaine de la sécurité sociale	13
3.6 Exemples d'applications d'IA employées dans d'autres domaines.....	14
4. Avantages et risques présentés par l'utilisation de l'IA dans le secteur public	16
5. Orientations politiques pour les acteurs publics visant à adopter l'IA	18
5.1 Phase de conception	18
5.2 Phase de passation des marchés	20
5.3 Phase de développement.....	22
5.4 Phase de déploiement.....	26
5.5 Formation continue	30
6. Conclusions	31

1. Introduction

En décembre 2020, le CAHAI a adopté une étude de faisabilité qui expose les possibilités offertes par l'intelligence artificielle (IA) et les défis qu'elle présente en ce qui concerne la protection et la mise en œuvre des droits de l'homme, de la démocratie et de l'État de droit. En plus de fournir une cartographie complète des enjeux dans le contexte de l'IA, l'étude de faisabilité a jeté les bases des activités auxquelles le CAHAI se consacrera en 2021.

Le CAHAI a non seulement identifié les éléments susceptibles de constituer un (des) projet(s) d'instrument(s) juridique(s) sur l'IA, mais a aussi constaté qu'il serait profitable d'élaborer des orientations politiques supplémentaires sur certains thèmes qui ont été abordés dans l'étude de faisabilité. L'un de ces thèmes concerne l'utilisation de l'IA dans le secteur public, compte tenu des avantages et des risques spécifiques qu'elle présente dans ce contexte. Les organismes publics, ou leurs délégataires privés, ont de plus en plus recours à des applications d'IA dans leurs activités quotidiennes, au niveau local comme au niveau national. En outre, les applications d'IA se propagent dans pratiquement tous les domaines d'activité publics, de l'administration fiscale et des administrations de protection sociale aux services répressifs et aux autorités judiciaires. Ces applications peuvent entraîner des gains d'efficacité importants qui peuvent profiter aux citoyens de manière directe et indirecte, mais elles peuvent aussi engendrer certains effets négatifs pour les droits de l'homme, la démocratie et l'État de droit.

Le présent document entend donc cartographier les effets provoqués par l'utilisation de l'IA dans le secteur public et vise à formuler des orientations politiques à l'intention des États membres pour qu'ils puissent s'acquitter de leur obligation de respecter les droits de l'homme et de garantir l'intégrité du processus démocratique ainsi que l'État de droit.

Il convient de noter que le champ d'application de ce document se limitera à la conception, au développement et au déploiement de l'IA par des entités du secteur public et des entités du secteur privé lorsqu'elles exercent des prérogatives de puissance publique. L'utilisation de l'IA par des organismes privés, même si elle peut avoir des répercussions sur la sphère publique, n'entre pas dans le champ d'application du présent document. Par conséquent, bien que l'utilisation sur les médias sociaux de pratiques de microciblage reposant sur l'IA puisse par exemple avoir des effets sur l'intégrité du processus électoral – et donc concerner aussi le secteur public – elle ne sera pas traitée ci-après.

2. L'IA et le secteur public : vue d'ensemble

Plusieurs pays dans le monde ainsi que des organisations internationales, telles que l'Union européenne, l'UNESCO ou l'OCDE, ont compris l'énorme potentiel économique de l'IA, qui est considérée comme une technologie stratégique. La plupart des États membres du Conseil de l'Europe ont déjà adopté des stratégies en matière d'IA ou ont déclaré que l'adoption de l'IA était l'une de leurs priorités dans le cadre de stratégies de transformation numérique plus larges. L'adoption de l'IA dans le secteur public est souvent présentée comme un volet d'une stratégie de ce type.

En effet, la plupart des stratégies nationales prévoient souvent l'utilisation de l'IA dans le secteur public, notamment dans le but d'améliorer les services publics proposés dans l'intérêt des citoyens et de rendre certaines procédures plus efficaces en automatisant les tâches routinières du gouvernement et la coordination entre les différents services de l'administration publique. Par ailleurs, certains États membres voient dans le secteur public un fer de lance du développement et de l'utilisation de l'IA. Certains États membres considèrent aussi que l'IA pourrait contribuer à orienter la prise de décision du gouvernement (dans les domaines de la sécurité publique, de la santé publique ou de l'évaluation des politiques, par exemple).

Selon la cartographie initiale de l'OCDE sur l'IA, sur les 50 pays (dont ceux de l'UE) qui ont adopté des stratégies nationales en matière d'IA, 36 disposent de stratégies spécifiques pour l'utilisation de l'IA dans le secteur public. La plupart de ces stratégies reposent sur des thèmes similaires, qui couvrent le développement économique, la confiance et l'éthique, la sécurité et l'amélioration du vivier de talents. Bien qu'il n'existe pas de législation uniforme relative à l'utilisation de l'IA dans le secteur public, de nombreux gouvernements ont lancé divers projets nationaux ces dernières années dans le but d'utiliser l'IA. On peut observer que l'IA est souvent employée pour améliorer l'efficacité et la prise de décision, favoriser des relations positives avec les citoyens et les entreprises ou résoudre des problèmes spécifiques dans des domaines essentiels tels que la santé, les transports et la sécurité¹.

Les applications d'IA peuvent contribuer à améliorer le rapport coût-efficacité et à favoriser la réalisation d'économies grâce à une automatisation intelligente, ainsi qu'à accroître la force de travail et le capital lorsque l'IA est en mesure de compléter les compétences de la main-d'œuvre existante². Idéalement, le déploiement de systèmes d'IA dans le secteur public devrait également permettre aux fonctionnaires d'abandonner leurs tâches routinières pour se concentrer sur les tâches à forte valeur ajoutée, tout en améliorant la rapidité et la qualité des services publics fournis.

Les statistiques publiées par Statista montrent le degré d'adoption de l'intelligence artificielle dans les différentes organisations du secteur public en fonction des domaines d'activité (voir Fig.2).

¹ Berryhill, J., et al. (2019), « Hello, World: Artificial intelligence and its use in the public sector », *Documents de travail de l'OCDE sur la gouvernance publique*, n° 36, Les Éditions de l'OCDE, Paris, <https://doi.org/10.1787/726fd39d-en>.

² Bernd W. Wirtz, Jan C. Weyerer & Carolin Geyer (2019), « Artificial Intelligence and the Public Sector—Applications and Challenges », *International Journal of Public Administration*, 42:7, 596-615, DOI : 10.1080/01900692.2018.1498103

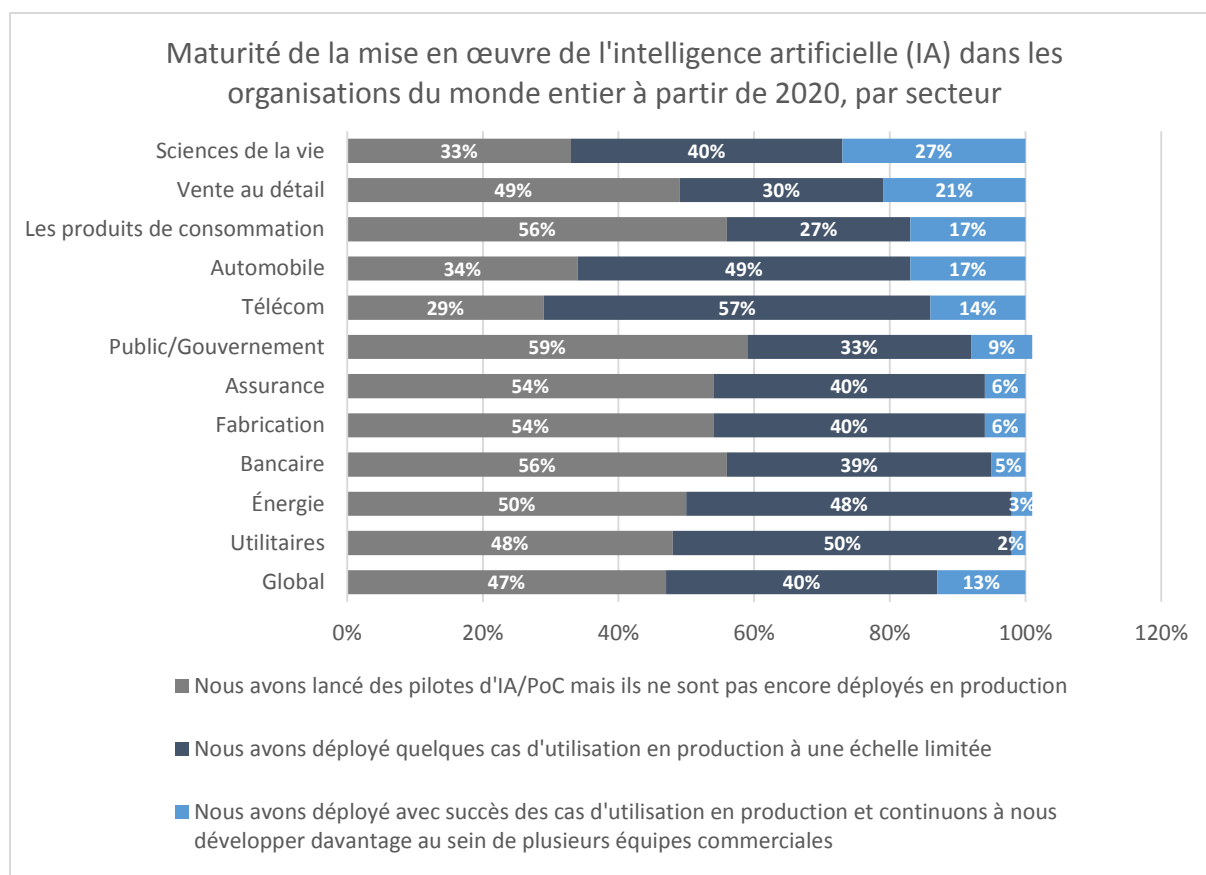
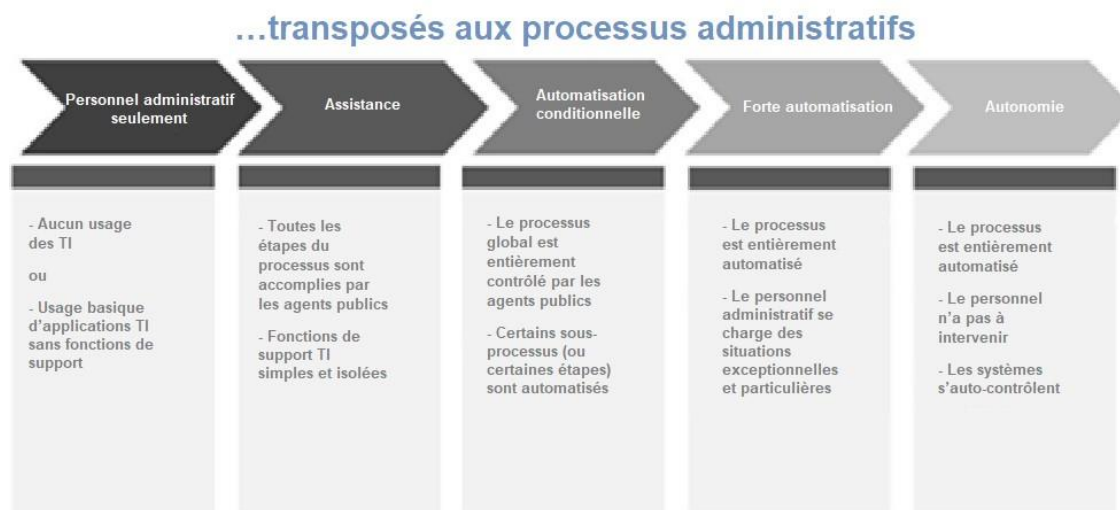


Fig.1. Source ([• Maturity of AI implementation by sector worldwide 2020 | Statista](#))

Cependant les outils utilisant l'IA sont de plus en plus testés ou utilisés par des organisations du secteur public³. Le débat porte principalement sur l'utilisation de systèmes de prise de décision automatisée (*automated decision making*, ci-après ADM) ou algorithmique. À l'heure actuelle, les organismes du secteur public ont recours à ces systèmes essentiellement à un niveau d'automatisation assistée ou conditionnelle. Dans de rares cas, cette automatisation concerne des processus ou des services entiers.

³ Voir par exemple le rapport d'AlgorithmWatch, *Automating Society*, 2020, page 6.



Source : [Table ronde sur l'intelligence artificielle et l'avenir de la démocratie](#), Conseil de l'Europe, 2019

Les fondements juridiques de l'utilisation des systèmes d'ADM peuvent varier selon les États membres. Le droit allemand⁴, par exemple, n'autorise le recours à la prise de décision automatisée que s'il n'existe aucune marge d'appréciation et que cette décision n'implique de répondre que par oui ou par non. Dans tous les cas, il doit être possible de refuser la décision prise automatiquement, d'examiner le processus et d'expliquer comment la décision a été prise. Aux Pays-Bas et en Belgique, la situation est la suivante : le principe de légalité exige qu'il existe une base légale pour toute prise de décision (qui a des conséquences juridiques ou quand les droits de l'homme sont en jeu), qu'elle résulte de l'utilisation de systèmes informatiques ou non. Bien entendu, les normes existantes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit s'appliquent déjà à l'utilisation de l'IA dans les États membres du Conseil de l'Europe. Ainsi, le traitement des données à caractère personnel est régi par la Convention 108, qui complète donc la législation applicable plus particulièrement au contexte du secteur public.

Par ailleurs, il est également reconnu que l'adoption d'applications d'IA comporte certains risques. Compte tenu du rôle déterminant joué par le secteur public dans la société et de l'importance de sa mission, l'atténuation et la prévention de ces risques constituent une tâche essentielle pour les gouvernements qui cherchent à déployer des systèmes d'IA.

La plupart des États membres ont conscience de la nécessité d'inscrire fermement la conception, le développement et le déploiement de l'IA dans un cadre éthique. Dans ce contexte, ils évoquent fréquemment des valeurs et des principes, selon lesquels l'IA doit être centrée sur l'humain, digne de confiance et responsable, et associée à des aspects tels que la transparence et la surveillance humaine. Si la plupart des États membres mentionnent un cadre éthique, certains soulignent aussi spécifiquement la

⁴ https://www.gesetze-im-internet.de/englisch_bdsge/englisch_bdsge.html#p0310.

nécessité de réglementer l'IA et considèrent que ce rôle incombe au secteur public. À cet égard, il est également important de souligner que les obligations relatives aux droits de l'homme sont adressées directement aux États membres et que ceux-ci doivent également les respecter dans le cadre de leur utilisation de l'IA. En outre, ils ont pour responsabilité de préserver les équipements et les infrastructures qui assurent l'existence du système démocratique et de l'État de droit.

Sélection d'exemples de politiques et d'orientations adoptées dans des États membres et non membres en ce qui concerne l'utilisation de l'IA dans le secteur public

Institut Alan Turing, Understanding artificial intelligence ethics and safety – Ce guide propose des orientations exhaustives sur la manière d'appliquer les principes d'éthique et de sécurité de l'IA à la conception et à la mise en œuvre de systèmes algorithmiques dans le secteur public.

Directive canadienne sur la prise de décision automatisée – Le Gouvernement canadien a mis au point une approche fondée sur les risques pour l'adoption de l'IA dans le secteur public, qui répartit les systèmes d'IA en différents niveaux. Les quatre facteurs utilisés pour déterminer le niveau de risque sont l'incidence de la décision concernée : sur les droits des personnes ou des collectivités, sur la santé ou le bien-être des personnes et des collectivités, sur les intérêts économiques des personnes, des entités ou des collectivités et sur la durabilité continue d'un écosystème. En fonction du niveau de risque déterminé, le guide donne des indications sur la meilleure façon de procéder à l'achat de systèmes d'IA du point de vue de la proportionnalité et sur la mesure dans laquelle chaque exigence doit être respectée.

Dans l'étude de faisabilité, plusieurs effets négatifs de l'utilisation de l'IA dans ce contexte ont été abordés. En particulier, l'étude démontre clairement comment les systèmes d'IA peuvent avoir des effets négatifs sur un large éventail de droits civils, politiques et sociaux protégés par la Convention européenne des droits de l'homme et par la Charte sociale européenne, ainsi que sur l'État de droit et la démocratie. Il s'agit notamment du droit à un procès équitable, du droit au respect de la vie privée et familiale, du droit à la liberté d'expression et d'association, du droit à l'égalité et à la non-discrimination ou encore du droit à la sécurité sociale.

Pour autant, l'étude fait également valoir que les systèmes d'IA « peuvent aussi renforcer l'État de droit et la démocratie en améliorant l'efficacité des procédures administratives, en aidant les pouvoirs publics à mieux répondre aux besoins et en libérant du temps pour aborder les questions complexes et importantes. *Les systèmes d'IA peuvent aussi aider les autorités et d'autres acteurs publics à mieux identifier les besoins et préoccupations des citoyens, contribuant ainsi à l'élaboration de politiques plus efficace* ».

L'objectif est donc de trouver une approche adéquate et équilibrée pour le déploiement des systèmes d'IA dans le secteur public, à même d'améliorer l'efficacité de celui-ci et d'aider les acteurs publics à identifier les besoins et les préoccupations de la population, tout en évitant les effets négatifs associés aux systèmes d'IA. Cela suppose non seulement de définir un cadre juridique adapté qui fixe les limites dans

lesquelles les systèmes d'IA peuvent être légalement utilisés, mais aussi de disposer des compétences requises. Les États membres ont par exemple conscience que, s'ils veulent réussir la mise en place de l'IA dans le secteur public, ils devront également investir dans le renforcement des capacités des agents publics. S'ils ne disposent pas des compétences nécessaires, les services publics ne seront pas en mesure de tirer profit des avantages que présente l'IA et ne sauront pas non plus quelles questions essentielles poser pour s'assurer que son utilisation est conforme aux normes relatives aux droits de l'homme, à la démocratie et à l'État de droit.

3. Cas d'utilisation des applications d'IA mis en place par les États membres du Conseil de l'Europe

Divers États membres du Conseil de l'Europe utilisent déjà des systèmes d'IA dans le cadre de leurs services publics. Cette partie présente quelques exemples concrets d'applications d'IA qui ont été déployées ou qui sont actuellement au stade de projet pilote.

3.1 Exemples d'utilisation dans le domaine judiciaire

Hongrie : Projet de reconnaissance vocale et de transcription

Ce projet étudie l'utilisation d'un logiciel de reconnaissance vocale et de transcription dans les tribunaux afin de faciliter et d'accélérer la rédaction des décisions et des procès-verbaux. Cette solution permettrait de réduire la quantité de travail à effectuer manuellement.

Plus précisément, il vise à faciliter le respect des délais liés à l'obligation des tribunaux de mettre par écrit les décisions et les procès-verbaux. Il permettrait en outre d'utiliser plus efficacement le temps de travail du personnel en réduisant le temps consacré à la transcription. En 2018, 726 licences de logiciels de reconnaissance vocale et de transcription ont été achetées pour les juridictions du pays.

Suisse : Utilisation de l'IA dans le système pénal

En Suisse, l'exécution des peines et la justice pénale reposent sur un système de niveaux. En application de ce système, les détenus se voient généralement accorder des périodes de liberté d'une durée croissante à mesure qu'ils purgent leur peine d'emprisonnement. Il est évident que le risque d'évasion et de récidive sont des facteurs déterminants lorsqu'il s'agit d'accorder ces régimes de plus grande liberté. Ces dernières années, et en réaction à des cas de condamnés ayant commis plusieurs actes tragiques de violence et des infractions sexuelles, une approche de la sanction fondée sur les risques (Risk-Oriented Sanctioning) a été adoptée. Aujourd'hui, cette démarche est utilisée dans tous les cantons germanophones de Suisse.

L'objectif premier de l'approche fondée sur les risques est de prévenir la récidive. Lors d'un processus de sélection, les cas sont classés en fonction de la nécessité d'évaluer le risque de récidive des personnes concernées. Sur la base de cette classification, une analyse différenciée de chaque cas est effectuée et donne ensuite lieu à un plan

individuel d'exécution de la peine de l'auteur de l'infraction. Ce tri est effectué par un outil d'ADM appelé Fall-Screening-Tool (outil de sélection des cas, FaST). FaST répartit automatiquement tous les cas dans les différentes catégories en fonction des risques présumés d'évasion et de récidive.

Cette classification est réalisée en utilisant les casiers judiciaires des personnes concernées et se fonde sur des facteurs de risque statistiques généraux, tels que l'âge, les infractions violentes commises avant l'âge de 18 ans, les rapports du juge des mineurs, le nombre de condamnations antérieures, la catégorie des infractions, les peines prononcées, la délinquance polymorphe, la durée de la période sans infraction après leur libération et la violence domestique. Ce processus est effectué de manière entièrement automatisée par l'application d'ADM. Cependant, il est important de noter qu'il ne s'agit pas d'une analyse des risques, mais d'un moyen de filtrer les cas nécessitant une évaluation plus poussée⁵.

Néanmoins, la classification a des effets sur la façon dont les responsables d'une institution donnée prennent des décisions et sur la sélection des évaluations à réaliser. Aucune application d'ADM n'apparaît dans les autres étapes de l'approche de la sanction fondée sur les risques. L'outil FaST n'est donc utilisé que pendant la phase de tri.

3.2 Exemples d'utilisation dans le domaine fiscal et financier

Irlande : Le voicebot du fisc reposant sur l'IA aide les citoyens à s'acquitter de leurs obligations fiscales

Début 2018, l'administration fiscale irlandaise a lancé un projet pilote pour examiner si les technologies de traitement du langage naturel (TLN) reposant sur l'IA pouvaient être utilisées pour améliorer les services proposés aux usagers, réduire les coûts et augmenter l'efficacité. Ce projet a conduit à la mise en place d'un agent numérique virtuel (VDA), ou voicebot, conçu pour traiter un sous-ensemble d'appels de contribuables irlandais relatifs à l'acquittement de leurs obligations fiscales⁶.

Un ensemble de technologies ont été utilisées pour transcrire le discours des usagers, comprendre le texte grâce à un système de TLN afin de pouvoir formuler une réponse, puis convertir cette réponse en discours parlé afin que l'utilisateur puisse l'entendre. L'administration fiscale irlandaise a démontré que la technologie des agents conversationnels vocaux peut assurer un service entièrement automatisé, offrant ainsi une expérience efficace et efficiente aux usagers.

⁵ Treuhardt/Kröger 2018 pp. 24-32

⁶ VDA est un agent conversationnel vocal fonctionnant grâce au moteur de TLN (traitement du langage naturel) d'IBM Watson hébergé sur le cloud. Il utilise d'autres services d'informatique en nuage pour la reconnaissance vocale et la synthèse de texte (Google et AWS/Polly), propose un flux de conversation à suggestions intelligentes, s'appuyant sur le statut actuel de l'utilisateur et des échanges tenus récemment et traite des demandes relatives à l'acquittement des obligations fiscales. Ce système permet aux usagers d'y procéder par téléphone, via l'agent conversationnel, et transfère l'appel à un véritable agent du service des usagers si nécessaire. Enfin, il offre au personnel de l'administration fiscale une interface utilisateur présentant des tableaux de bord synthétiques, des transcriptions des appels passés et des liens vers des enregistrements audio.

Irlande : Utilisation de l'apprentissage automatique pour prévoir les recours à la suite de réévaluations au sein de l'Agence irlandaise d'évaluation (Valuation Office) – projet pilote

Le Valuation Office est un organisme administratif indépendant, qui relève du ministère irlandais du Logement, des Collectivités locales et du Patrimoine. La principale activité de cette structure consiste à fournir aux parties prenantes des évaluations précises et à jour des biens à usage commercial et industriel. Ces évaluations font partie intégrante du système d'évaluation des entreprises en Irlande et sont déterminantes pour un aspect très important des recettes des collectivités locales chaque année.

Le projet de recherche pilote intitulé « Applying Machine Learning to predict appeals after Revaluation at the Valuation Office of Ireland » (application de l'apprentissage automatique pour prévoir les recours à la suite de réévaluations au sein de l'Agence irlandaise d'évaluation) a été mené dans le cadre de l'initiative AutoVal – Automated Commercial Property Valuation using Machine Learning (évaluation automatisée des biens à usage commercial reposant sur l'apprentissage machine)⁷. Ces travaux visaient à prévoir, en utilisant des algorithmes d'apprentissage machine de classification binaire, si un bien à usage commercial ferait l'objet d'un recours devant la juridiction compétente en matière d'évaluation des biens (Valuation Tribunal) après sa réévaluation. Ces travaux de recherche présentent des avantages tangibles pour « signaler » les biens susceptibles de faire l'objet d'un recours et pour réduire le nombre de recours portés devant le Valuation Tribunal à l'issue d'une procédure de réévaluation.

Les modèles appliqués aux données issues de deux programmes de réévaluation ont atteint un degré de précision compris entre 80 et 90 %. L'application des modèles aux données du programme de réévaluation 2019 n'a pas donné d'aussi bons résultats (~35 %) et il a été conclu qu'il fallait réentraîner le modèle en utilisant des données plus récentes.

3.3 Exemples d'utilisation dans le domaine des soins de santé

Russie : Centre de diagnostic et de télémédecine de la ville de Moscou

Le projet pilote d'application de l'IA en radiologie associe deux utilisations différentes de l'IA :

- l'assistance aux médecins dans la détection d'une tumeur sur une radiographie ;

⁷ Les techniques d'apprentissage automatique utilisées après une analyse exploratoire des données sont les suivantes : la régression logistique – un modèle statistique qui, dans sa forme de base, utilise une fonction logistique pour modéliser une variable dépendante binaire ; les forêts d'arbres décisionnels – un modèle ensembliste qui applique de multiples arbres de décision sur divers sous-échantillons de l'ensemble de données et utilise un calcul de la moyenne pour améliorer la précision prédictive et contrôler le surapprentissage par rapport aux arbres de décision uniques ; la méthode du *gradient boosting* – une technique d'apprentissage automatique qui produit un modèle de prédiction sous la forme d'un ensemble de modèles de prédiction faibles, généralement des arbres de décision. Elle construit le modèle par étapes, comme les autres méthodes de boosting, et le généralise pour parvenir à un résultat.

- et le traitement des notes d'observation dictées par le médecin pour remplir le dossier médical du patient. On estime que cette solution permet d'économiser 25 % du temps du médecin et d'établir automatiquement un dossier complet et précis pour chaque patient. Ce système s'appuie sur la boîte à outils Voice2Med produite par Speech Technology Center (Saint-Pétersbourg) et peut être étendu et reproduit à moindre coût.

Ces deux applications d'IA améliorent considérablement le résultat, la précision et l'efficacité du travail des médecins.

Russie : Fournitures médicales dans la région de Tomsk

La région a instauré un régime juridique spécial dans le cadre d'un projet pilote pour autoriser la livraison de fournitures médicales par des drones autonomes dans des régions éloignées sur des centaines de kilomètres. La région s'étend sur 600 km du nord au sud, et sur 780 km d'est en ouest.

Turquie : Applications d'IA permettant de lutter contre la pandémie de covid-19

Des applications d'IA ont été mises en œuvre dans le domaine de la santé dans le cadre de la lutte contre la pandémie de covid-19. La Turquie a ainsi mis en place une plateforme technologique consacrée à la covid-19 pour coordonner les projets de R&D visant à élaborer de nouveaux médicaments et vaccins. Elle accueille actuellement 17 projets différents. Des techniques de recherche reposant sur les mégadonnées sont utilisées pour identifier des candidats-médicaments parmi des milliers de molécules à l'aide de l'IA. Un appel à projets spécifique a également été lancé pour inviter les chercheurs et les entreprises de R&D à proposer des solutions pour combattre la covid-19.

3.4 Exemples d'utilisation dans le domaine de la lutte contre la criminalité

Pays-Bas : Traitement par la police du signalement en ligne des arnaques sur internet

Un système d'IA est utilisé aux Pays-Bas dans le domaine de la répression de la délinquance, pour offrir un service de conseil aux citoyens lorsqu'ils signalent en ligne des cas d'escroquerie sur internet.

Ainsi, le citoyen explique d'abord la situation dans un champ de saisie libre. L'application pose ensuite des questions supplémentaires en utilisant l'analyse de texte. Sur la base du texte et des réponses supplémentaires apportées, le modèle classe le déclarant dans une catégorie et lui donne des conseils et des orientations sur l'action la plus pertinente à engager pour donner suite aux événements.

Pays-Bas : Détection de la fraude au logement, municipalité d'Amsterdam⁸

Ce système d'IA est en cours de test⁹ à Amsterdam pour détecter les fraudes au logement. Compte tenu de la rareté des logements à Amsterdam, la municipalité s'est engagée à empêcher l'utilisation illégale des logements, par exemple en cas de double enregistrement, de fraude aux aides au logement ou de dépassement des limites imposées en matière de location de vacances. L'analyse des données et les modèles de prédiction sont de plus en plus utiles aux services répressifs pour détecter ce type de fraude.

En se fondant sur un grand nombre de facteurs, on calcule à l'avance le risque qu'une fraude soit constatée à une adresse donnée. Ainsi, on augmente la probabilité que les services répressifs détectent en effet une fraude lors de leurs visites inopinées sur place.

Suisse : Utilisation de l'IA pour mettre en place une police prédictive

Dans certains cantons, notamment à Bâle-Campagne, en Argovie et à Zurich, la police utilise des logiciels pour l'aider à prévenir les infractions pénales. Elle s'appuie sur la suite logicielle commerciale « PRECOBS » (Pre-Crime Observation System), qui est uniquement utilisée pour prévoir les cambriolages domestiques. Ce délit relativement fréquent a fait l'objet de travaux de recherche approfondis et les autorités de police disposent en général d'une base de données solide concernant la répartition géographique et chronologique des cambriolages ainsi que les caractéristiques de ces infractions. En outre, il ressort de l'étude des infractions de ce type que leur auteur est souvent un professionnel et qu'il y a donc une probabilité supérieure à la moyenne qu'elles se reproduisent. En outre, des modèles de pronostic correspondants peuvent être créés à partir d'un nombre relativement faible de points de données. PRECOBS repose donc sur la présomption que les cambrioleurs frappent plusieurs fois dans un court laps de temps si leur tentative a porté ses fruits à un endroit donné.

Le logiciel est utilisé pour rechercher, sur la base des rapports établis par la police sur des cambriolages, des modes de fonctionnement, tels que la manière dont les auteurs procèdent et le moment et le lieu où ils agissent. Ensuite, PRECOBS établit une prévision pour les zones où le risque de cambriolage est le plus important dans les 72 heures qui suivent. La police envoie alors des patrouilles ciblées dans ces zones. PRECOBS réalise donc des prévisions sur la base de décisions préalablement saisies et n'utilise pas de méthode d'apprentissage automatique. Bien qu'il soit prévu d'étendre PRECOBS à d'autres délits (tels que le vol de voiture ou le vol à la tire) et de créer ainsi de nouvelles fonctionnalités, il convient de noter que l'utilisation de la police prédictive en Suisse est actuellement limitée à un domaine relativement restreint et clairement défini du travail de prévention de la police¹⁰.

⁸ Pour de plus amples informations, voir (en néerlandais) : Gemeente Amsterdam magazine , Amsterdamse Intelligentie, disponible à l'adresse suivante :

https://assets.amsterdam.nl/publish/pages/922120/magazine_ai.pdf , p. 16.

⁹ Ce prototype est encore en phase de test. Il est développé en interne et par quelques étudiants de l'UvA. La technique d'apprentissage automatique utilisée repose sur les arbres de décisions et les forêts d'arbre décisionnels.

¹⁰ Blur 2017, Leese 2018 pp. 57-72.

Russie : Inspection nationale de la sécurité routière (GIBDD)

La reconnaissance d'images sur les caméras de surveillance de la circulation est progressivement mise en place dans tout le pays. Les caméras détectent automatiquement les infractions au code de la route, notamment le non-respect des voies réservées aux transports en commun, la conduite agressive, l'utilisation d'un téléphone portable au volant ou le défaut de port de la ceinture de sécurité.

Allemagne : Police prédictive, niveau fédéral¹¹

Depuis 2017, le Bundeskriminalamt (Office fédéral de police criminelle, BKA) utilise l'outil d'évaluation des risques RADAR-iTE¹² pour classer les « militants salafistes » en trois catégories correspondant au niveau de la menace qu'ils représentent (forte, manifeste et modérée). Ce système a été mis au point en coopération avec le département de psychologie légale de l'université de Constance. Pour évaluer une personne déjà connue des autorités, l'agent remplit un questionnaire normalisé sur le « comportement observable » du sujet, en s'appuyant sur les données que la police a précédemment recueillies sur cette personne et sur tous les éléments auxquels elle est légalement habilitée à accéder. Une fois que l'agent a pris connaissance des résultats de l'évaluation de la menace que représente la personne, l'agent (ou le service concerné) décide des mesures à prendre.

Russie : Rosfinmonitoring (service russe de surveillance des marchés financiers – RFM)

En 2019, RFM a commencé à tester un système d'IA dans le domaine de la prévention des activités de blanchiment d'argent, de financement du terrorisme et de criminalité financière (LBC). En particulier, en collaboration avec 40 autres États, ce système est utilisé pour traquer le financement de Daech et utilise des modèles comportementaux pour repérer les activités de blanchiment d'argent.

3.5 Exemples d'utilisation dans le domaine de la sécurité sociale

Pays-Bas : Application « Chercheur de travail » de la Caisse d'assurance des salariés (UWV)

L'agence pour l'emploi UWV est chargée d'aider les chômeurs à trouver un emploi rémunéré. Ce soutien consiste à leur proposer des services tels que l'organisation d'ateliers, de webinaires ou de formations en ligne et commence souvent par un entretien personnel dans une succursale d'UWV. Dans la mesure où le nombre de conseillers et les services disponibles sont limités, l'agence pour l'emploi doit définir le délai applicable pour que les chômeurs puissent bénéficier d'un entretien et les

¹¹ AlgorithmWatch, Automating Society Report, 2020, disponible à l'adresse suivante : <https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf>

¹² Bundeskriminalamt, 2017

services les plus adaptés pouvant leur être proposés. C'est dans ce but qu'elle a créé cette application.

Une fois que des allocations chômage ont été accordées au chômeur, un questionnaire lui est soumis dans son espace en ligne. Il répond alors à des questions sur ses antécédents professionnels et sa situation personnelle, et effectue une évaluation de ses perspectives sur le marché du travail. En associant ces réponses aux données relatives aux antécédents professionnels et à la formation du candidat, l'application « Chercheur de travail » fait ensuite une prévision de ses chances de trouver un emploi rémunéré dans un délai d'un an (exprimées en pourcentage) à laquelle s'ajoute un diagnostic des obstacles personnels rencontrés par les chômeurs et des possibilités qui leur sont offertes pour trouver un emploi rémunéré (un chiffre en pourcentage par facteur). Ces pourcentages ne sont pas communiqués au candidat, à moins qu'il n'en fasse la demande ou que cette question soit explicitement abordée lors de son entretien personnel.

3.6 Exemples d'applications d'IA employées dans d'autres domaines

Allemagne : Gestion des crises dans le domaine de la politique étrangère, niveau fédéral¹³

L'Office fédéral des Affaires étrangères utilise l'outil d'analyse de données [PREVIEW](#)¹⁴ (Prediction, Visualization, Early Warning – *Prédiction, visualisation et alerte précoce*) pour déterminer l'évolution des crises internationales. Cet outil analyse les données accessibles au public relatives aux tendances et aux conflits politiques, économiques et sociétaux actuels afin d'identifier les crises qui se font jour. Selon l'Office fédéral des Affaires étrangères, l'IA est utilisée pour traiter les données, qui servent ensuite à produire des infographies et des cartes permettant de mieux comprendre la situation d'un conflit particulier. En outre, les analyses de tendances mettent en avant la manière dont la situation politique et sociétale peut évoluer. Le système PREVIEW est déployé par le département S de l'Office fédéral des Affaires étrangères, qui supervise les mesures en faveur de la stabilité internationale et les interventions en cas de crise. Les résultats obtenus aident également les agents de l'office à déterminer les mesures à prendre par la suite.

Allemagne : Vérification de l'identité des migrants, niveau fédéral¹⁵

Le Bundesamt für Migration und Flüchtlinge (Office fédéral des migrations et des réfugiés, BAMF) a recours à des systèmes de reconnaissance de texte et de reconnaissance vocale pour [identifier les réfugiés](#)¹⁶ depuis 2017. Le personnel de cet

¹³ AlgorithmWatch, Automating Society Report, 2020, disponible à l'adresse suivante : <https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf>

¹⁴ Auswärtiges Amt, 2019

¹⁵ AlgorithmWatch, Automating Society Report, 2020, disponible à l'adresse suivante : <https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf>

¹⁶ Thüer, Köver et Fanta, 2018

organisme peut demander aux demandeurs d'asile de lui donner accès à leur téléphone portable, à leur tablette ou à leur ordinateur portable pour vérifier s'ils disent la vérité sur leur origine. L'office est en mesure d'obtenir toutes les données contenues dans ces appareils et de les traiter grâce à un logiciel. Celui-ci présente à l'agent responsable un bref aperçu de ce contenu, qui comprend également une analyse linguistique du texte extrait des appareils. Selon le BAMF, le logiciel et le matériel nécessaires ont été fournis par la société [Atos SE](#)¹⁷, mais le [magazine VICE a trouvé des éléments établissant](#)¹⁸ la participation de la société MSAB, spécialisée dans les technologies criminalistiques mobiles. Un autre outil déployé par le BAMF vise à identifier les [dialectes dissimulés dans le discours](#)¹⁹. Lorsqu'un demandeur d'asile ne dispose pas d'une pièce d'identité valide, un enregistrement vocal d'une durée de deux minutes dans laquelle il décrit une image dans sa langue maternelle est analysé par un logiciel, qui calcule ensuite le pourcentage de proximité de ce discours avec un certain dialecte.

Irlande : Utilisation de l'IA pour compter les arbres

Dans un premier temps, une étude de faisabilité a été menée avec le CeaDAR (Centre national irlandais d'analyse des données et d'intelligence artificielle appliquées) pour étudier les utilisations potentielles de l'IA et/ou de l'apprentissage automatique en vue d'améliorer la prévision des volumes et de la qualité du bois des forêts irlandaises en utilisant des données multimodales, notamment des données issues de systèmes de lidar, des données d'inventaire, des données historiques sur les récoltes, le climat ou les sols, etc.

Commission européenne : Utilisation de l'IA pour déterminer les risques de fraude et d'irrégularités dans les projets financés par des fonds publics

À l'heure actuelle, le CEPD ne peut citer qu'un seul exemple d'application d'IA utilisée par les institutions, organes ou agences de l'UE, à savoir le système Arachne²⁰, dont les utilisateurs sont les autorités d'audit des États membres de l'UE et le service d'audit de la Commission européenne. Ce système, utilisé dans le domaine du financement des projets publics, vise à identifier les projets, les bénéficiaires, les contrats et les prestataires susceptibles d'être exposés à des risques de fraude, de conflit d'intérêts et d'irrégularités. Il se fonde sur la collecte de données provenant de sources internes et externes et sur l'exploration de données pour le calcul d'indicateurs de risque.

La Pologne a également fourni divers exemples d'utilisation de l'IA dans tous les secteurs mentionnés ci-dessus. Les systèmes d'IA qu'elle emploie visent à détecter les fraudes dans le système fiscal, à informer les citoyens par l'intermédiaire d'un agent conversationnel, à faire des prévisions quant au développement de la pandémie de covid-19, à analyser des messages vocaux et textuels dans le cadre de protocoles d'essai, à calculer les groupes de population pour lesquels l'intervention de l'aide publique serait décisive, à comparer les prévisions concernant l'éducation des élèves

¹⁷ Biselli, 2017

¹⁸ Biselli, 2018b

¹⁹ Biselli, 2018a

²⁰ De plus amples informations sont disponibles sur le site web du Fonds social européen (FSE) – Emploi, affaires sociales et inclusion – Commission européenne (europa.eu)

et à analyser le système de sécurité sociale. Les principales possibilités offertes par les systèmes d'IA utilisés sont la reconnaissance, la prédiction et la fourniture de conseils, grâce à l'application de différentes techniques fondamentales²¹. Les systèmes d'IA utilisés ont un niveau basique d'automatisation et sont soumis à une supervision humaine.

4. Avantages et risques présentés par l'utilisation de l'IA dans le secteur public

Les exemples ci-dessus montrent que des applications d'IA ou, plus précisément, les systèmes experts et les applications reposant sur l'apprentissage automatique, sont employées dans de nombreux domaines d'intervention du secteur public dans plusieurs États membres du Conseil de l'Europe, et à différents niveaux (municipal, régional et national).

Certains systèmes visent à aider directement les citoyens, d'autres assistent les agents publics dans l'exercice de leurs fonctions, et d'autres encore ont une portée universelle.

L'utilisation de l'IA devrait en principe contribuer à accroître la satisfaction des citoyens à l'égard des services publics ; c'est le cas par exemple des systèmes d'IA capables de répondre aux questions fréquemment posées par l'intermédiaire d'agents conversationnels ou d'assistants virtuels ainsi que de proposer une assistance 24 heures sur 24 et 7 jours sur 7. En outre, les systèmes d'IA peuvent contribuer à apporter aux agents publics une satisfaction au travail lorsque ces technologies sont déployées de manière à leur permettre d'exécuter plus rapidement des tâches administratives répétitives et de consacrer plus de temps à leurs interactions avec les citoyens ou à d'autres tâches essentielles, tout en garantissant une utilisation plus rationnelle des ressources publiques.

Sur le plan des valeurs directrices du Conseil de l'Europe, la plupart des applications évoquées précédemment semblent être neutres ou présenter des conséquences potentiellement positives pour les droits de l'homme, l'État de droit et la démocratie. C'est par exemple le cas des applications d'IA visant à détecter les cas de fraude/corruption ou à améliorer l'accès à l'éducation. De façon générale, en renforçant la capacité des citoyens à accéder aux informations, à les partager et à les communiquer, certaines applications d'IA peuvent contribuer au contrôle démocratique des institutions publiques et accroître l'obligation de celles-ci de rendre des comptes.

Les exemples d'applications d'IA présentés ci-dessus appartiennent à diverses catégories : certaines permettent une prise de décision entièrement automatisée, d'autres offrent des conseils en matière de prise de décision, d'autres encore fournissent des informations pertinentes pour prendre des décisions, et d'autres enfin semblent n'avoir aucun lien avec le processus de prise de décision.

Dans tous les cas, l'adoption de solutions d'IA semble être motivée par la promesse évidente d'un accroissement de l'efficacité du processus décisionnel et d'autres

²¹ Apprentissage automatique, vision par ordinateur, apprentissage par renforcement, apprentissage supervisé, systèmes experts, représentation des connaissances, réseaux de neurones, planification automatique, recherche et optimisation et traitement du langage naturel.

processus dans le secteur public, où les indicateurs de l'efficacité sont la rapidité, le coût et la qualité. Ces promesses doivent toutefois être mesurées à l'aune des risques éventuels d'atteintes aux droits de l'homme, à l'État de droit et à la démocratie. Certains des exemples mentionnés ci-dessus montrent que certaines applications d'IA sont encore en phase d'essai, des tests étant effectués à petite échelle avant un déploiement à grande échelle au niveau national. Il est probable que des processus de vérification et de pondération des avantages potentiels par rapport aux risques potentiels aient déjà été engagés dans d'autres États membres ayant déjà mis en place certaines solutions au niveau national. Il serait intéressant d'obtenir d'avantage d'informations de la part des États membres au sujet de ces processus (les audits et les évaluations qui ont pu être conduits sur les effets des applications, par exemple, accompagnés notamment d'un retour d'informations des parties prenantes).

Cela serait d'autant plus souhaitable que les exemples d'utilisation d'applications d'IA dans les États membres énumérés ci-dessus ne sont pas accompagnés d'indications concernant les risques éventuels liés au déploiement de ces systèmes. Cependant, comme l'a noté le CAHAI dans son étude de faisabilité, l'utilisation de l'IA dans le secteur public présente aussi des risques qu'il convient de prendre en compte.

Doivent également être pris en considération les risques résultant d'une mauvaise conception et d'un mauvais développement des systèmes d'IA – tels que l'intégration de biais et de facteurs discriminatoires dans les services fournis, qui aurait pour effet d'accroître l'ampleur des effets préjudiciables de ces systèmes ; les questions de consentement relatives aux ensembles de données utilisés pour construire les systèmes d'IA ; ou la robustesse et la sécurité des systèmes d'IA. Si le recours aux systèmes d'IA peut accroître l'efficacité des institutions publiques, il est possible que ce résultat soit obtenu au détriment de la transparence ainsi que de l'intervention et de la surveillance humaines. En outre, pour se procurer des systèmes d'IA et les déployer, les pouvoirs publics dépendent souvent d'acteurs privés, ce qui risque d'éroder davantage la confiance de la population, car cette situation suscite des interrogations portant sur l'obligation de rendre des comptes, l'indépendance de la surveillance et le contrôle public, qui pourraient être amplifiées par l'utilisation de systèmes d'IA opaques. Un cadre de gouvernance adapté devrait donc permettre aux concepteurs et aux déployeurs de systèmes d'IA d'agir de manière responsable et conformément aux dispositions juridiques pertinentes, tout en laissant ouvertes les voies de recours et d'intervention appropriées des pouvoirs publics quand ce n'est pas le cas.

Pour autant, si la robustesse d'une certaine application d'IA est établie lorsqu'elle est utilisée dans un processus particulier du secteur public, et qu'aucun argument professionnel, éthique, juridique ou politique n'impose de privilégier la version « sans intervention de l'IA » de ce processus, continuer à le faire pourrait tout aussi bien être considéré comme déraisonnable, si cette dernière méthode est à l'origine d'un plus grand nombre d'erreurs.

Cependant, toute discussion concernant ces risques ne présenterait un intérêt que dans le cadre d'une évaluation des applications pratiques, prenant en compte à la fois l'ampleur de la menace potentielle et la probabilité qu'elle se concrétise.

Il est donc conseillé, par principe, de soumettre toutes les applications d'IA déployées dans le secteur public à une procédure d'analyse d'impact suivant une approche fondée sur les risques et traitant des questions relatives aux droits de l'homme, à l'État de droit et à la démocratie, et de renouveler régulièrement cette analyse afin d'évaluer les répercussions réelles de ces systèmes sur leurs utilisateurs. Des garanties de contrôle suffisantes devraient également être prévues, afin de favoriser le respect des principes de transparence, d'équité, de responsabilité et d'obligation de rendre des comptes.

Un certain nombre de recommandations pratiques spécifiques sont formulées ci-après en vue de proposer aux décideurs politiques des orientations appropriées pour les différents stades de la conception, de l'acquisition, du développement et du déploiement des technologies d'IA dans le secteur public, ainsi qu'en ce qui concerne les initiatives de formation continue à mettre en place.

5. Orientations politiques pour les acteurs publics visant à adopter l'IA

Dans cette partie, des orientations politiques concrètes sont formulées à l'intention des acteurs publics qui envisagent de déployer des systèmes d'IA, en accordant une importance particulière aux mesures qui peuvent être prises pour atténuer les risques potentiels concernant les droits de l'homme, la démocratie et l'État de droit. Ces mesures reposent sur un cadre réglementaire sous-jacent adéquat prévoyant des critères de référence clairs applicables à la conception, à l'acquisition, au développement et au déploiement de systèmes fondés sur l'IA. Il est important de noter que ce cadre doit s'appliquer indépendamment du fait que le développement et le déploiement du système d'IA sont réalisés directement par des acteurs publics ou par des acteurs privés agissant pour le compte d'acteurs publics.

5.1 Phase de conception

- Analyser le problème nécessitant une solution

Un système d'IA est un outil destiné à améliorer le bien-être humain, et non une fin en soi. Si l'utilisation de ces systèmes peut être à l'origine de gains d'efficacité importants, ils ne sont pas adaptés pour résoudre tous les problèmes. Il est donc essentiel d'analyser en premier lieu le problème spécifique que l'on souhaite résoudre, puis de définir dans quelle mesure l'IA peut y apporter une solution (et quelle technique d'IA serait la plus adéquate). En effet, il est possible que d'autres approches ou technologies soient plus adaptées pour traiter le problème particulier dont il est question. En outre, l'utilité d'une solution fondée sur l'IA dépend souvent de la disponibilité de données exactes, fiables et représentatives. Dans certains cas, ces données peuvent ne pas exister, ou leur collecte peut s'avérer très coûteuse ou impossible. Ces difficultés potentielles doivent être calculées dans l'analyse coûts-avantages des solutions envisageables.

- Définir les ensembles de données pouvant être utilisés pour le système d'IA

Une fois que le problème a été identifié et que l'on a conclu que l'IA peut contribuer à apporter une solution, on peut procéder à la conception d'une application d'IA spécifique. L'identification du type de données qui doivent être recueillies et analysées

pour résoudre le problème est une étape importante. Les questions relatives à la propriété et à la protection des données doivent être examinées. En outre, il convient d'évaluer si les ensembles de données qui peuvent être utilisés sont suffisamment représentatifs pour donner une vision globale des informations requises. Il sera souvent nécessaire de recourir à ce que l'on appelle des données indirectes – à savoir des points de données qui ne fournissent qu'un paramètre indirect de l'information recherchée.

- Présenter explicitement les hypothèses qui sous-tendent la conception du système

Au cours de la phase de conception d'un système d'IA, un certain nombre d'hypothèses sont formulées. Elles restent souvent implicites, mais ont pourtant une influence sur la manière dont le système d'IA sera construit et donc sur les résultats qui seront obtenus. Ainsi, c'est sur la base d'hypothèses relatives aux causes du problème et à la nature du problème que certaines données seront sélectionnées (ou non), et qu'une solution spécifique sera proposée (ou non). Toutefois, étant donné que ces hypothèses peuvent parfois se révéler erronées ou problématiques (par exemple parce qu'elles reposent sur certains biais injustes), il est important de les présenter explicitement pendant la phase de conception et de les documenter. Si l'on rend ces hypothèses explicites, elles peuvent être rectifiées ou contestées par la suite si elles se révèlent erronées, non pertinentes ou discriminatoires.

- Faire participer les personnes destinées à utiliser le système et tenir compte de leurs capacités

Comme le montrent les exemples ci-dessus, les applications d'IA dans le secteur public sont très diverses. Les utilisateurs de ces applications peuvent être des agents publics (lorsque le système est censé fonctionner en interne pour améliorer l'efficacité de certains processus administratifs, par exemple), mais aussi des citoyens (par exemple lorsque le système est conçu pour interagir directement avec des personnes, comme dans le cas des agents conversationnels). Les choix de conception concernant l'application d'IA doivent tenir compte des personnes destinées à utiliser le système d'IA, ainsi que des compétences et des capacités des utilisateurs.

Pour veiller à ce que le système soit centré sur l'utilisateur, il est conseillé d'impliquer les personnes destinées à utiliser le système d'IA dès la phase de conception et de recueillir leurs observations, afin que le système puisse être construit de manière à répondre au mieux à leurs besoins. Il est par exemple recommandé d'organiser des ateliers participatifs ouverts pour entrer en contact avec les professionnels qui sont en première ligne, et cerner les aspects de leur travail auxquels ils aimeraient consacrer plus de temps et ceux où l'IA pourrait jouer un rôle intéressant. En écoutant attentivement les utilisateurs et en cherchant à comprendre leurs besoins, ainsi qu'en se concentrant sur l'utilisation qui peut être faite de l'IA pour réduire ou éliminer les tâches frustrantes ou répétitives, non seulement le processus gagne en légitimité auprès du personnel, mais il devient également plus facile de découvrir de nouveaux gains d'efficacité potentiels et d'accroître la motivation des agents sur le lieu de travail.

- Dans la mesure du possible, opter pour une conception ouverte et transparente

Dès la phase de conception, il est conseillé d'opter pour un système permettant d'assurer une traçabilité dans la mesure du possible. En particulier lorsque le système est destiné à faciliter des processus décisionnels publics qui peuvent avoir un effet important sur la vie des citoyens, il y a de solides arguments en faveur de l'utilisation de techniques transparentes plutôt que de systèmes de type « boîte noire », de manière à garantir la possibilité de retracer et d'expliquer le processus de décision. En outre, dans la mesure du possible, les logiciels utilisés devraient avoir un code source ouvert. Cela permettrait non seulement à des tiers (tels que des chercheurs ou des organisations de la société civile) de procéder à des évaluations, mais aussi de renforcer la confiance du public dans le système.

- Réaliser une analyse de l'impact sur les droits de l'homme, la démocratie et l'État de droit

Pour atténuer les effets potentiellement négatifs des systèmes d'IA, une mesure importante peut être prise, consistant à réaliser une analyse de l'impact du système sur les droits de l'homme (DH), la démocratie (D) et l'État de droit (E) (analyse d'impact DHDE ou AIDHDE). Cette évaluation peut aider à anticiper les risques potentiels, de sorte que ceux-ci puissent être évités ou atténués autant que possible dès la phase de conception du système. Il convient de noter que le CAHAI-PDG met actuellement au point un modèle d'analyse d'impact. Bien que cette évaluation fasse partie des exigences plus larges de diligence raisonnable en matière de droits de l'homme auxquelles toutes les organisations sont soumises lorsqu'elles prennent des mesures pouvant avoir des répercussions négatives sur les droits de l'homme, elle revêt une importance particulière dans le contexte du secteur public, étant donné que les autorités publiques ont l'obligation directe de protéger les droits de l'homme.

5.2 Phase de passation des marchés

La passation de marchés publics pourrait constituer un facteur important pour l'adoption de l'IA. Par l'intermédiaire de systèmes de marchés publics axés sur l'IA, les acteurs publics peuvent stimuler le développement de solutions d'IA novatrices dans l'intérêt général, tout en ouvrant la voie à des services publics qui utilisent des outils numériques et qui accordent une place centrale aux citoyens. Étant donné que l'IA est une technologie émergente, il peut être difficile pour les États d'établir une voie d'accès au marché pour répondre à leurs besoins²². Puisque de nombreux acteurs publics ne disposent pas aujourd'hui du savoir-faire, des données et des compétences nécessaires pour développer eux-mêmes des systèmes d'IA, la passation de marchés publics est actuellement l'un des principaux moyens que le secteur public utilise pour déployer cette technologie. Cependant, lorsque les systèmes d'IA sont achetés à des acteurs privés plutôt que développés en interne, les normes que les gouvernements doivent respecter afin de s'assurer que leurs systèmes d'IA sont conformes aux droits de l'homme, à la démocratie et à l'État de droit doivent tout de même être respectées. En raison de l'évolution rapide des technologies et de l'absence de normes uniformes

²² Guidelines for AI Procurement (2020). Disponible à l'adresse suivante : <https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement>, dernière consultation le 8 mars 2021.

en matière d'IA, les procédures de passation de marchés existantes pourraient ne pas suffire à garantir que les préoccupations potentielles liées à l'utilisation de l'IA dans le secteur public soient dûment prises en compte. Il est donc essentiel que les acteurs publics tiennent compte des considérations ci-après lorsqu'ils font l'acquisition de systèmes d'IA.

- Examiner la législation et les mesures politiques applicables

Pratiquement tous les États membres ont déjà mis en place une législation relative à la passation de marchés publics dans le domaine des technologies. Outre la législation relative aux marchés publics, il peut également exister d'autres exigences réglementaires ou documents d'orientation qui s'appliquent à l'utilisation des systèmes d'IA (ou des technologies de manière plus générale) par les autorités publiques. Lorsque les acteurs publics cherchent à engager un processus de passation de marché public pour acquérir un système d'IA, il est donc important que la législation et les politiques applicables dans ce contexte soient dûment répertoriées et prises en compte pour ce processus spécifique.

- Adapter les processus de passation de marchés publics si nécessaire et établir des lignes directrices relatives à la passation de marchés publics en matière d'IA

Comme indiqué dans l'étude de faisabilité, les systèmes d'IA peuvent accentuer les risques existants ou en créer de nouveaux. Les organismes publics doivent donc s'appuyer sur leurs politiques existantes en matière de marchés publics et les adapter aux problèmes spécifiques que peut poser l'IA. Cette adaptation peut se faire sous la forme d'une modification de cette législation, mais aussi d'une adoption de lignes directrices relatives aux exigences en matière de passation de marchés publics propres aux systèmes d'IA. Des lignes directrices de ce type ont par exemple été adoptées par le Royaume-Uni. Les exigences relatives à ces marchés publics peuvent être fondées sur des normes et des certifications sociotechniques, et devraient dans l'idéal être élaborées en coopération avec les groupes de parties prenantes concernés. Elles doivent garantir la conformité avec les normes du Conseil de l'Europe en matière de droits de l'homme, de démocratie et d'État de droit, et peuvent ainsi encourager les entreprises privées qui participent aux appels d'offres publics concernant des systèmes d'IA à également adopter ces normes.

La passation du marché public relatif au système d'IA dont il sera fait l'acquisition devra tenir compte, en plus des orientations générales en matière de marchés publics liés à l'IA, des risques spécifiques associés à l'application d'IA en question. Comme on l'a souligné précédemment, tous les systèmes d'IA destinés au secteur public ne présentent pas le même niveau de risque. Il dépend en effet de facteurs tels que la finalité du système et l'utilisation qui en est prévue. En plus de l'AIDHDE susmentionnée, il peut aussi être utile à cet égard de recourir à des outils d'évaluation des risques.

- Garantir une approche pluridisciplinaire et multipartite²³

²³ Comme l'implique la nécessité d'analyser l'impact sur les droits de l'homme, l'État de droit et la démocratie et comme le recommande la Commissaire aux droits de l'homme du Conseil de l'Europe – *Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme – Recommandation*, mai 2019, point 2

Lorsque l'on établit des exigences applicables aux marchés publics propres aux systèmes d'IA, il est conseillé d'adopter une démarche ouverte aux points de vue d'experts de diverses disciplines ainsi qu'aux contributions de différentes parties prenantes. La prise en compte de différentes optiques et perspectives augmente les chances que l'évaluation des possibilités et des risques liés à la passation des marchés publics soit effectuée de manière plus complète et que les effets potentiels sur les différents groupes de parties prenantes soient envisagés.

- Tenir compte des répercussions sur la responsabilité publique

Les pouvoirs publics sont responsables des algorithmes qu'ils utilisent, même si ces derniers sont créés par des tiers et que leur fonctionnement ne peut être expliqué en détail. Il convient donc d'anticiper cette responsabilité au cours de la procédure de passation de marché public. Les gouvernements devraient par conséquent s'assurer qu'ils disposent de toutes les informations nécessaires sur le système d'IA qu'ils achètent, afin de pouvoir en assumer la responsabilité et rendre des comptes sur les conséquences négatives potentielles de son utilisation. Si les acteurs publics peuvent négocier des régimes de responsabilité contractuelle avec l'organisation auprès de laquelle ils font l'acquisition du système d'IA, ceux-ci ne concernent que la relation entre l'acteur public et le contractant, et ne peut conduire les acteurs publics à échapper à leur responsabilité à l'égard du grand public.

C'est pourquoi il était déjà indiqué dans l'étude de faisabilité que les autorités publiques ne devraient pas acquérir de systèmes d'IA auprès de tiers qui ne respectent pas les obligations légales d'information en ce qui concerne leurs systèmes d'IA, ou qui ne sont pas disposés à lever les restrictions à l'échange d'informations (confidentialité ou secrets industriels, par exemple) lorsque ces restrictions entravent le processus d'évaluation d'impact sur les droits de l'homme (y compris la recherche/l'examen externes) et la mise à disposition de ces évaluations au public²⁴. On peut en effet se référer à l'orientation formulée ci-dessus, à savoir l'importance de la réalisation au préalable d'une analyse de l'impact du système sur les droits de l'homme, la démocratie et l'État de droit. Cela permet d'anticiper les risques potentiels, qui doivent être pris en compte dans les exigences relatives à la passation des marchés.

5.3 Phase de développement

Que les systèmes d'IA soient développés en interne ou acquis à l'extérieur, ils doivent répondre à certaines exigences garantissant qu'ils n'ont pas d'effets négatifs sur les droits de l'homme, la démocratie et l'État de droit, afin qu'ils soient véritablement bénéfiques pour la collectivité. Compte tenu des risques qui peuvent survenir dans le cadre de l'utilisation publique des systèmes d'IA – et plus généralement de l'utilisation des systèmes d'IA – les orientations ci-après devraient être prises en compte au cours de la phase de développement des systèmes d'IA.

- Mettre en place des processus de documentation et de journalisation

(Consultations publiques), paragraphe 1 : « L'utilisation des systèmes d'IA par les États devrait être régie par les normes relatives aux marchés publics, appliquées dans le cadre de procédures transparentes, auxquelles tous les acteurs concernés seraient invités à contribuer. »

²⁴ CAHAI(2020)23

Les acteurs publics qui développent des systèmes d'IA (en interne ou par l'intermédiaire d'un prestataire) devraient mettre en place des mesures permettant d'assurer la transparence et la traçabilité du système, telles que la documentation des processus de développement suivis, des choix de sélection et de curation des ensembles de données utilisés et des méthodes d'entraînement et de test de l'algorithme. Ces informations devraient non seulement être documentées, mais dans l'idéal rendues publiques à l'intention des citoyens et des parties prenantes qui souhaitent faire usage de leurs droits d'accès à l'information. En documentant les processus pertinents et en assurant la transparence de ces informations, on pourra renforcer la confiance du public dans l'utilisation des systèmes d'IA.

Il convient également de donner des indications concernant le paramétrage et les choix relatifs aux données d'apprentissage. En cas d'utilisation de méthodes nécessitant de définir des paramètres au préalable ou de recourir à des données d'entraînement, il convient de faire une description de la manière dont le paramétrage a été réalisé et dont les données d'entraînement ont été choisies, et de l'accompagner d'une étude des facteurs discriminatoires potentiels. Au nombre des questions importantes qu'il convient de se poser à cet égard figurent les suivantes : Comment le modèle a-t-il été développé ? Quelles données et quels algorithmes ont été utilisés ? Comment ont-ils été obtenus ? Comment ont-ils été examinés en interne ? Sous quel format les résultats de l'analyse des données sont-ils publiés ? Il convient aussi, en plus de tenir un registre des hypothèses utilisées, de documenter la méthode d'analyse employée et la manière dont sa précision a été mesurée. En outre, il est nécessaire d'archiver les données d'entrée utilisées (données sources/ensembles de données) et de conserver des informations sur la manière dont la qualité des sources des données utilisées a été vérifiée.

Enfin, il convient de noter que des processus de journalisation automatisés peuvent être intégrés au système d'IA pendant la phase de conception et de développement. Ils peuvent faciliter l'archivage des informations nécessaires en assurant l'enregistrement automatique des processus ou des événements qui ont lieu dans le cadre du système d'IA.

- Mettre en place des processus de test et de validation adéquats

Les processus de test et de validation constituent un élément déterminant du cycle de développement d'un système d'IA. Plus le processus est automatisé, et/ou plus le degré de compétence déléguée au système d'IA est élevé, plus il faut attacher d'importance à l'application de méthodes de test adéquates avant l'utilisation du système. Ce principe est particulièrement pertinent lorsque le système est capable d'avoir des effets sur les personnes. Celui-ci ne doit pas seulement être testé sur le plan de la précision et de la robustesse, mais il convient aussi de déterminer si les paramètres et les ensembles de données choisis peuvent avoir un effet négatif et disproportionné sur un groupe donné de la population.

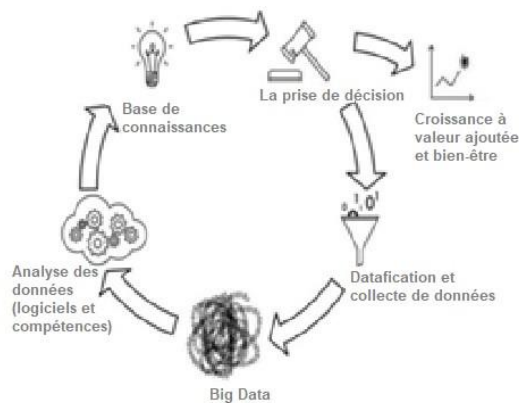
Certaines mesures spécifiques appliquées dans le développement du système peuvent contribuer à mener à bien ce processus. Ainsi, le code est idéalement organisé en modules qui peuvent être évalués séparément ou en les associant. Cela

peut faciliter le test de ces modules pour vérifier leur bon fonctionnement, à la fois séparément et quand ils sont associés.

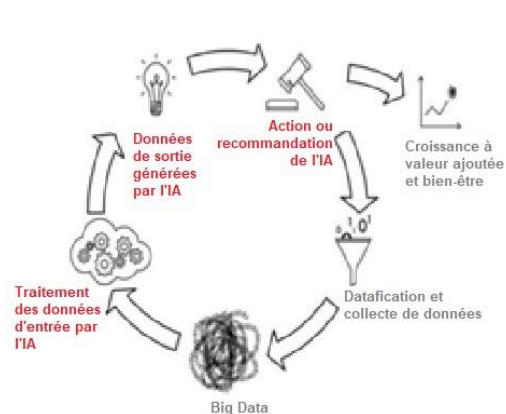
- Instaurer des mécanismes de gouvernance des données

Au cours du cycle de vie d'un système d'IA, qui repose souvent essentiellement sur des données, la protection des données doit être garantie. Une fois développé, le système d'IA est aussi un système algorithmique. En tant que tel, il est utilisé pour traiter des données d'entrée afin d'obtenir des données de sortie particulières, telles que des données de classification, de prédiction ou de recommandation. Si l'IA est alimentée avec des documents numériques portant sur des humains, elle peut formuler des déductions sur les traits de personnalité et les caractéristiques des individus, notamment sur des informations sensibles. Cela peut amener la population à remettre en question leur confiance dans la technologie et dans la manière dont leurs données sont traitées et utilisées. Pour garantir leur confiance légitime, il faut s'assurer que leurs données ne sont pas utilisées d'une manière qui porte atteinte à leur droit à la protection des données à caractère personnel ni à d'autres droits, par la mise en œuvre de mesures de protection technologiques et de mécanismes de gouvernance des données adéquats²⁵. La Convention 108 du Conseil de l'Europe et sa version modernisée, la Convention 108+, offrent d'importants mécanismes de protection des données à caractère personnel qui doivent être respectés dans ce contexte.

Cycle de valorisation des données²⁶



Données et IA – données d'entrée et de sortie²⁷



Hormis la confidentialité des données, la qualité des ensembles de données est l'un des paramètres déterminants dans le développement d'un système d'IA de haute qualité. Les données recueillies ne doivent pas comporter de biais, d'inexactitudes ni d'erreurs. L'intégrité des données est également importante. Le processus d'accès aux

²⁵ COM(2019) 168 final, disponible à l'adresse suivante : [COM\(2019\)168/F1 – FR \(europa.eu\)](https://eur-lex.europa.eu/legal-content/fr/TXT/?uri=CELEX:32019C0168), dernière consultation le 7 mars 2021.

²⁶ OCDE (2015). *Data-Driven Innovation, Big-Data for Growth and Well-Being*, disponible à l'adresse suivante : https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page4, dernière consultation le 11 mars 2021.

²⁷ OCDE (2015). *Data-Driven Innovation, Big-Data for Growth and Well-Being*, disponible à l'adresse suivante : https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page4, dernière consultation le 11 mars 2021.

données doit être géré d'une manière adéquate et des mécanismes de vérification doivent être mis au point pour contrôler le processus de bout en bout. Pour améliorer le processus de gouvernance des données, on peut par exemple poser les questions suivantes²⁸ : Qui sont les personnes concernées et quels sont leurs droits ? Qui est autorisé à recueillir quelles données, à quelles fins, sur quel fondement juridique et pour quelle durée ? Comment les données seront-elles effacées lorsqu'elles ne seront plus nécessaires ? Dans quelles circonstances, le cas échéant, les données peuvent-elles être réutilisées, par qui, dans quel but et sur quel fondement juridique ? Quelles techniques, le cas échéant, sont utilisées pour agréger/anonymiser les données ?

- Tenir compte du risque d'inégalité d'accès ou de traitement

Comme indiqué dans l'étude de faisabilité, l'utilisation des systèmes d'IA risque d'amplifier les biais et les préjugés, qu'il s'agisse de sexisme, de racisme, de discrimination fondée sur l'âge ou d'autres types de discriminations injustes fondées sur divers motifs (y compris les discriminations fondées sur des critères indirects ou intersectionnels). Cela peut se produire pour diverses raisons, par exemple si l'échantillon d'apprentissage est biaisé, si des biais sont introduits dans la conception de l'algorithme ou dans sa fonction d'optimisation, ou s'il est exposé à un environnement d'apprentissage biaisé. Lorsque le système est développé, il est souvent déjà trop tard pour contrer ce risque. Il est donc important de l'anticiper et de l'atténuer dès le départ, en vérifiant par exemple si l'ensemble de données utilisé est suffisamment représentatif et exact, si le système n'a pas de répercussions disproportionnées sur un certain groupe de population et si les hypothèses sur la base desquelles le système est développé ne reposent pas sur des biais injustes. Cet aspect est d'autant plus important que la sous-représentation des groupes vulnérables et marginalisés dans le secteur de l'IA pourrait encore contribuer à amplifier ce risque.

Lorsque les citoyens font partie des personnes destinées à utiliser le système d'IA, il est également important de veiller à l'égalité d'accès au système, en tenant compte notamment de la fracture numérique qui existe à l'heure actuelle ou d'autres obstacles potentiels à l'accessibilité, tels que le handicap, l'âge avancé ou les lacunes de connaissances.

Il est important de noter que les êtres humains ne sont pas non plus exempts de biais et de préjugés qui peuvent avoir une incidence négative sur la prise de décision publique. On peut donc faire valoir que les systèmes d'IA peuvent aussi être utilisés pour tenter de lutter contre les biais humains existants dans le secteur public. Cependant, même lorsque l'IA est utilisée à cette fin, il reste capital de prendre en compte les orientations énoncées précédemment et de veiller à la mise en œuvre de procédures transparentes afin de garantir la confiance du grand public et des agents publics qui déploieront le système d'IA.

- Évaluer les effets sur l'égalité entre les femmes et les hommes

²⁸ Medhora, P. R., Centre pour l'innovation dans la gouvernance internationale (CIGI) (2018), *Data Governance in the Digital Age*, Rapport spécial, disponible à l'adresse suivante : [Data Series Special Reportweb.pdf \(cigionline.org\)](https://www.cigionline.org/Data-Series-Special-Reportweb.pdf), dernière consultation le 7 mars 2021.

et pour atténuer ses effets potentiellement négatifs sur les droits de l'homme, la démocratie et l'État de droit.

- Conserver des cadres de gestion et d'atténuation des risques

Dès la phase de conception et de développement du système d'IA, les acteurs publics ayant l'intention d'y avoir recours devraient anticiper et traiter les risques potentiels que le système d'IA peut engendrer. Toutefois, la gestion et l'atténuation des risques constituent également un élément essentiel qui doit être en place au moment du déploiement du système d'IA. En effet, la gestion des risques est un processus itératif, qu'il est d'autant plus important de conserver que le système d'IA peut être modifié ou apprendre au cours de son déploiement. Le processus de gestion des risques doit être documenté et comporter au moins les étapes suivantes : analyse et évaluation des risques potentiels (prévisibles), de leur probabilité et de leur ampleur ; instauration de mécanismes d'atténuation et de contrôle des risques ; adoption d'un plan de repli au cas où le système ne pourrait plus être utilisé ; et définition d'un plan de communication – adapté aux diverses parties prenantes – sur les risques qui peuvent découler de l'utilisation du système d'IA.

- Réaliser des audits (indépendants) du système d'IA

Les systèmes d'IA qui sont déployés pour effectuer des missions de services public doivent faire régulièrement l'objet d'un audit, et les résultats de ces audits doivent être rendus publics. Lorsqu'une application d'IA peut avoir une incidence négative sur les droits de l'homme, la démocratie ou l'État de droit, il est important que ces audits puissent être réalisés de manière indépendante. Pour garantir la possibilité de mener à bien ces audits, il est impératif de respecter les orientations susmentionnées concernant la traçabilité du système d'IA par la mise en place de processus de documentation et de journalisation. Les résultats des rapports d'audit doivent être rendus publics afin de renforcer la transparence publique (sans toutefois porter atteinte à des intérêts légitimes tels que la sécurité nationale ou les informations protégées par des droits de propriété intellectuelle).

Idéalement, les audits sont réalisés sur la base de normes ou de processus définis d'un commun accord et largement acceptés, de sorte que les rapports établis par les auditeurs – qui présentent les résultats et conclusions de l'audit – puissent faire l'objet de comparaisons et d'évaluations sur la base de normes sous-jacentes communes. Lorsque de telles normes n'existent pas, ou sont insuffisantes pour répondre aux préoccupations spécifiques qui apparaissent dans le contexte de l'IA, il est conseillé d'élaborer de nouvelles normes en coopération avec d'autres pays.

- Mettre en place un registre public des systèmes d'IA déployés dans le secteur public

Les États membres devraient établir des registres publics leur permettant de répertorier les systèmes d'IA utilisés dans le secteur public. Ces registres peuvent contenir des informations de base sur le système, telles que sa finalité, la technique sur laquelle il repose ou le type de données utilisées. La mise à disposition du public de ces registres, pour qu'il puisse les consulter, peut renforcer sa confiance dans l'utilisation des systèmes d'IA dans le secteur public. En outre, ces registres peuvent également promouvoir un débat démocratique inclusif et diversifié autour des systèmes d'IA, et améliorer leur contrôle démocratique. Étant donné que ces registres

pourraient présenter une liste complète des systèmes d'IA utilisés par les acteurs publics, il est particulièrement important de rendre ces informations disponibles pour connaître les systèmes pouvant avoir des effets négatifs sur les droits de l'homme, la démocratie et l'État de droit.

- Mettre en place des mécanismes de retour d'informations

Les utilisateurs de systèmes d'IA – ainsi que les personnes concernées – devraient idéalement avoir la possibilité de donner leur avis sur le déploiement du système. La mise en place d'un mécanisme de retour d'informations permet de recueillir leurs contributions sur les manières d'améliorer le système, par exemple en le rendant plus facile d'utilisation, en améliorant sa précision ou en atténuant les effets négatifs non anticipés. Dans l'idéal, le mécanisme de retour d'informations devrait être accessible à un large éventail de parties prenantes et toute initiative prise pour donner suite à ces contributions devrait être transparente.

- Donner aux citoyens la possibilité d'interagir avec un être humain

Étant donné l'importance de l'accès aux services publics pour les citoyens, dans chaque cas où un système d'IA est utilisé dans ce contexte, il est essentiel que les personnes aient la possibilité d'entrer en contact avec un interlocuteur humain. Il est particulièrement important de leur offrir cette possibilité lorsqu'un service reposant sur l'IA ne fonctionne pas correctement ou n'est pas accessible aux personnes présentant certains handicaps ou ayant des difficultés d'accès au système, ou encore lorsque les résultats produits par le système peuvent avoir un effet significatif sur un individu.

- Assurer la transparence du système et une communication avec les utilisateurs et les citoyens

Une fois déployés, les systèmes d'IA utilisés dans le secteur public doivent être aussi transparents que possible. Il est important d'assurer leur transparence à la fois pour l'utilisateur direct du système (l'agent public qui s'appuie sur les résultats du système et a besoin d'en évaluer le caractère justifiable, par exemple), mais aussi pour le grand public, qui peut en subir les effets. Pour ce faire, il convient de mettre à leur disposition des informations essentielles sur la finalité, les capacités, les limites, la logique et les hypothèses sous-jacentes du système, ainsi que sur les risques potentiels qu'il présente. Les informations fournies doivent être adaptées au public visé et tenir compte de son niveau de connaissances, de son âge et d'autres caractéristiques pertinentes pour veiller à ce que les informations soient communiquées de manière compréhensible. Cette exigence peut aussi, le cas échéant, couvrir les obligations d'information existantes dans le contexte du traitement des données à caractère personnel, ainsi que les exigences spécifiques qui devraient être prises en considération à cet égard (ces informations doivent par exemple être concises et présentées en termes clairs et simples, sous une forme transparente, intelligible et facilement accessible).

Lorsqu'un système reposant sur l'IA est utilisé par le secteur public pour interagir avec des personnes dans le cadre d'une prestation de service public, d'un processus de sélection ou à des fins d'identification, que ce soit directement ou par l'intermédiaire d'intervenants du secteur privé, les utilisateurs ou les personnes ciblées doivent être

informés en termes clairs et accessibles de cette utilisation, du déroulement du processus, de la manière dont les décisions sont prises et des moyens de les faire réexaminer. On peut également se reporter aux obligations existantes en matière d'accès aux documents et aux informations qui incombent aux autorités aux niveaux national et international. Ainsi, conformément à la Convention 108+, les personnes concernées ont le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire.

Les autorités publiques qui utilisent des systèmes d'IA pour prendre des décisions publiques pouvant avoir des conséquences spécifiques pour les citoyens doivent en outre être en mesure d'expliquer au citoyen à la fois les procédures suivies par l'algorithme et (les raisons de) la décision particulière qui a été prise à son égard. En effet, les décisions issues du système devraient pouvoir être expliquées à la personne concernée pour toute décision automatisée. Cela implique également que, dans un premier temps, les organisations gouvernementales ne devraient en principe pas utiliser des algorithmes trop complexes pour qu'ils puissent être expliqués.

- Mettre en place des mécanismes d'obligation de rendre des comptes et de recours

La mise en place de processus et de structures équitables et responsables aide les gouvernements à exploiter le potentiel de l'IA pour transformer les services publics et l'administration et à renforcer la confiance du public dans leur capacité à le faire. Pour que l'obligation de rendre des comptes fonctionne efficacement, les gouvernements doivent non seulement présenter de façon compréhensible la manière dont les systèmes d'IA sont utilisés dans les processus décisionnels, mais aussi donner la possibilité d'évaluer les algorithmes, les données et les processus de conception, par exemple en réalisant des audits. En outre, les acteurs publics devraient chercher à identifier, évaluer, documenter et limiter autant que possible les effets potentiellement négatifs des systèmes d'IA qu'ils déploient. Pour renforcer l'obligation de rendre des comptes et la surveillance mises en place, pour certaines applications qui présentent un risque particulièrement important pour les droits de l'homme, la démocratie et l'État de droit, il peut être judicieux de mettre en place un système d'autorisation préalable devant être accordée par une autorité judiciaire ou par une administration indépendante avant que le système ne puisse être utilisé.

L'obligation de rendre des comptes signifie également qu'un acteur spécifique peut être tenu pour responsable en cas de problème, et qu'une personne potentiellement lésée peut demander réparation. Les voies de recours doivent être clairement présentées et effectivement offertes, et le droit à un recours effectif contre tout préjudice lié à l'IA doit être garanti, l'acteur public qui déploie le système d'IA restant responsable en la matière. Le cas échéant, cela peut également signifier que les régimes de responsabilité doivent être adaptés afin de veiller à ce que les personnes puissent demander une compensation si elles ont subi un préjudice injuste par l'intermédiaire d'un système d'IA utilisé dans le cadre du secteur public.

- Assurer la (ré)évaluation régulière du système d'IA

L'évaluation de la fiabilité des systèmes d'IA et de leur conformité avec les droits de l'homme, la démocratie et l'État de droit est un processus itératif. Il est donc important que l'analyse de l'impact du système d'IA soit régulièrement revue et mise à jour si nécessaire. L'évaluation continue du système se fait dans l'idéal au sein d'une équipe pluridisciplinaire, et avec une large participation des parties prenantes. Les analyses d'impact réalisées périodiquement pourraient par exemple être rendues publiques, en donnant la possibilité aux citoyens, aux organisations de la société civile et aux autres parties intéressées de fournir un retour d'informations à leur sujet.

5.5 Formation continue

Les orientations politiques formulées ci-dessus seront difficiles à mettre en œuvre sans une connaissance et une compréhension adéquates de l'IA, non seulement dans le secteur public, mais aussi auprès du grand public³⁰. Il est donc essentiel que l'utilisation accrue des systèmes d'IA dans le secteur public s'accompagne d'une plus grande sensibilisation aux avantages, aux risques, aux capacités et aux limites de l'IA pour tous ceux qui utilisent l'IA ou qui seront concernés par son application.

- Mettre en place des programmes de formation destinés aux agents publics

Une formation sur l'IA et les mégadonnées devrait être dispensée aux agents publics travaillant dans les institutions et les administrations gouvernementales, le système judiciaire et les services répressifs, ainsi que les organes de contrôle indépendants et les structures nationales des droits de l'homme. Compte tenu de l'importance des données pour le développement et le déploiement des systèmes d'IA, les agents publics devraient également être formés à ce qu'implique la gouvernance des données, à la manière dont la qualité des données peut être évaluée et aux moyens de traiter les données dans le plein respect des règles de confidentialité et de protection des données. En outre, si l'on sensibilise et que l'on forme les agents publics aux possibilités et aux limites de l'IA et des mégadonnées, ils seront non seulement en mesure d'utiliser l'IA de manière plus responsable, mais ils seront également à même de réfléchir aux utilisations de l'IA qui pourraient faire progresser ou faciliter l'exécution de leurs tâches, de façon à améliorer autant que possible les services publics.

- Mettre en place des programmes d'éducation destinés au grand public

Les États membres devraient s'employer à familiariser davantage le grand public avec l'IA en déployant d'importants efforts de sensibilisation, de formation et d'éducation, y compris (et en particulier) dans les établissements scolaires. Il faudrait certes expliquer le fonctionnement de l'IA, mais ne pas oublier de traiter aussi la question de ses éventuels effets – positifs et négatifs – sur les droits de l'homme. Des efforts particuliers devraient être faits pour toucher les groupes marginalisés et les personnes qui, de manière générale, sont moins familiarisées avec l'informatique³¹.

³⁰ Conseil de l'Europe, Commissaire aux droits de l'homme, *Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme – Recommandation*, disponible à l'adresse suivante : <https://www.coe.int/fr/web/commissioner/-/unboxing-artificial-intelligence-10-steps-to-protect-human-rights>, dernière consultation le 11 mars 2021.

³¹ Conseil de l'Europe, Commissaire aux droits de l'homme, *Décoder l'intelligence artificielle : 10 mesures pour protéger les droits de l'homme – Recommandation*, disponible à l'adresse suivante :

6. Conclusions

En analysant différentes sources et différents exemples provenant de certains États membres du Conseil de l'Europe, le présent document montre que l'étude de l'utilisation de l'IA dans le secteur public dans des domaines très différents est un phénomène en développement, compte tenu des avantages qui peuvent être attendus de l'utilisation de cette technologie pour les individus et les administrations publiques. Les systèmes d'IA devraient contribuer à rationaliser les méthodes de fonctionnement du secteur public et à renforcer l'efficacité de l'action administrative, en déchargeant les organisations, dont les ressources sont limitées, des tâches répétitives et sans intérêt, et en ouvrant la voie à des modèles proactifs de prestation de services publics. Toutefois, la prise en compte des avantages potentiels pouvant découler de l'utilisation de ces technologies doit être contrebalancée par une évaluation appropriée des risques d'effets négatifs sur les droits de l'homme, la démocratie et l'État de droit – des risques déjà relevés dans l'étude de faisabilité du CAHAI – dans un secteur qui joue un rôle déterminant dans la protection de ces valeurs au sein des États membres.

Le présent donne aux États membres des orientations supplémentaires pour qu'ils traitent ces risques de manière adéquate, en formulant des recommandations concernant les mesures devant accompagner la conception, l'acquisition, le développement et le déploiement des technologies d'IA dans le secteur public, ainsi que les efforts à entreprendre en matière de formation continue, en vue de rendre le secteur public plus responsable, plus transparent et plus respectueux des normes européennes en matière de droits de l'homme, d'État de droit et de démocratie.