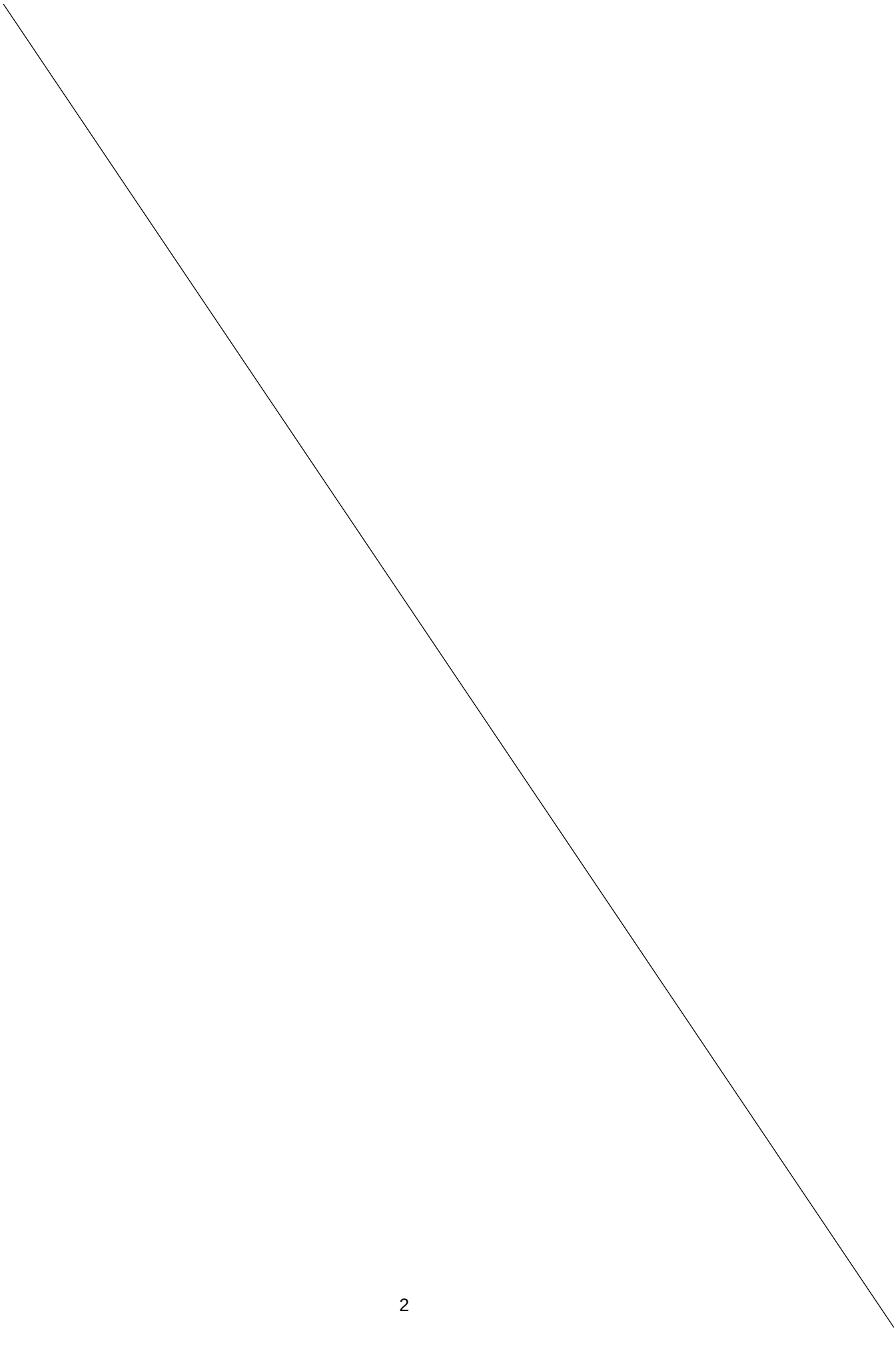
Sujet n° 2

Vous êtes Ingénieur(e) de l'agriculture et de l'environnement (IAE) en poste en direction régionale de l'alimentation, de l'agriculture et de la forêt. Votre directeur, alerté par une utilisation de l'intelligence artificielle (IA) générative par les agents de la DRAAF, vous demande de rédiger une note à son attention.

Cette note synthétisera ce que sont les IA génératives, leurs risques et intérêts dans le cadre professionnel en fonction publique, la réglementation en vigueur, et des propositions d'actions à mettre en œuvre à destination des agents de la DRAAF. En complément de cette note, une affiche des règles de bon usage à destination des agents de la DRAAF est attendue.

Liste des documents



DOCUMENT 1

Article internet bigmedia.bpifrance.fr du 11 avril 2024

Intelligence artificielle générative : de quoi parle-t-on ?

Capable de générer des images, des vidéos, voire de la musique, l'intelligence artificielle (IA) générative reproduit la capacité cognitive humaine de manière globale et polyvalente. Entre aubaine pour les entreprises et source de questionnement au niveau éthique, on fait le point sur l'une des avancées technologiques majeures de la décennie.

Le pape François vêtu d'une longue doudoune blanche Balenciaga, Emmanuel Macron en plein sprint à l'occasion d'une manifestation à Paris ou encore Barack Obama mangeant une glace sur la plage en compagnie d'Angela Merkel. La génération d'images par des intelligences artificielles comme Midjourney est bluffant de réalisme. Contrairement à l'IA spécialisée, qui se concentre sur des tâches spécifiques et limitées, cette dernière aspire à une compréhension et des compétences comparables à celles d'un être humain dans une grande variété de domaines. Une avancée technologique majeure non-exempte de quelques points d'interrogation.

Qu'est-ce que l'IA générative ?

L'intelligence artificielle générative (IA générative) est une catégorie d'IA qui se concentre sur la création de données, de contenu ou de choses artistiques, de façon indépendante. Elle diffère de l'IA classique, qui se concentre, quant à elle, sur des tâches spécifiques telles que la classification, la prédiction ou la résolution de problèmes. L'IA générative vise à produire de nouvelles données qui ressemblent à celles créées par des êtres humains, que ce soit sous forme de texte, d'images ou encore de musique par exemple.

Comment fonctionne l'intelligence artificielle générative ?

L'IA générative fonctionne en utilisant des modèles d'apprentissage automatique (modèle Machine Learning) pour créer du contenu de manière autonome. L'une des techniques les plus couramment utilisées en IA générative est l'utilisation de réseaux de neurones artificiels : les réseaux génératifs adverses (GAN) et les réseaux de neurones récurrents (RNN). Les GAN reposent sur une architecture composée de deux réseaux de neurones dits concurrents : un générateur qui crée une image et la transmet à un discriminateur qui détermine si l'image est réelle ou synthétique. Quant aux RNN, ils sont le plus souvent utilisés pour générer du texte ou de la musique.

Exemples d'intelligence artificielle générative

- ChatGPT : un chatbot développé par OpenAI, capable de générer du texte de haute qualité et de répondre à des questions ;
- DALL-E : créé par OpenAI, cet outil permet de générer des images à partir de requêtes textuelles, d'effectuer des retouches sur des images et de créer des variations à partir d'images existantes ;
- Midjourney : une intelligence artificielle génératrice d'images très puissante qui rivalise avec DALL-E et Stable Diffusion pour la création d'images synthétiques réalistes ;
- Stable Diffusion : une technique d'entraînement de réseaux génératifs permettant de produire des images haute résolution de grande qualité ;
- Bard : concurrent de ChatGPT, c'est une intelligence artificielle (IA) développée par Google, conçu pour engager des conversations et générer différents types de texte.

Quelle utilisation de l'IA générative ?

Les applications de l'IA générative sont variées. On peut les utiliser pour créer de l'art généré par ordinateur, des textes automatiques, des modèles 3D, des musiques, des vidéos et bien plus encore. Ils sont également utilisés dans des domaines tels que la génération de contenus pour les jeux vidéo, la création de visuels pour la publicité et même la génération de médicaments. L'intelligence artificielle générative permet donc de créer tout type de contenu, que ce soit du texte, des images, des vidéos ou de la musique.

Texte

Démocratisé par ChatGPT fin 2022, la génération de texte via une intelligence artificielle a explosé en seulement quelques mois. Gagner du temps, être plus efficace ou demander moins de ressources, tous les prétextes sont bons pour avoir recours à l'IA génératrice de texte. Si l'ordinateur n'a pas (encore) remplace l'humain dans son travail, ce que certaines IA sont capables de rédiger est assez bluffant. Ces générateurs de texte ont appris à jongler avec des milliards de mots sur le web, utilisant des algorithmes d'apprentissage automatique pour transformer une simple requête d'utilisateur en un article complet.

Images

Un générateur d'image avec intelligence artificielle est un type de logiciel qui va être à même de créer des images à partir de données brutes.

Vidéos

Plusieurs modèles d'intelligence artificielle permettent de créer des vidéos à partir de quelques mots tapés dans une barre de recherche. Les outils de génération de vidéos IA Fliki, Flexclip, Pictory, Elai ou encore Lumen5 progressent à une vitesse folle et sont à même de générer un film ou un clip vidéo en quelques secondes.

L'intelligence artificielle générative en entreprise

L'IA générative offre de nombreuses possibilités aux entreprises, notamment :

- Création de contenu : elle peut générer automatiquement du texte, des articles, des rapports, des emails et même des messages marketing, permettant de gagner du temps dans la production de contenu ;

- Conception graphique : l'IA générative peut être utilisée pour la création d'images, de logos et de designs, facilitant ainsi la création visuelle pour les entreprises ;

- Optimisation des opérations : l'IA générative peut aider à optimiser les opérations en générant des modèles de prévision, en planifiant la logistique ou en optimisant les chaînes d'approvisionnement ;

- Support client : elle peut automatiser les réponses aux requêtes des clients, offrant ainsi un support sans faille via des chatbots ;

- Créativité : elle peut stimuler la créativité en générant des idées ou en collaborant avec les équipes créatives ;

- Sécurité : elle peut être utilisée pour détecter les menaces de sécurité en analysant des données.

L'usage de l'IA générative dans les TPE et PME

Un besoin de pédagogie sur le fonctionnement des IA génératives

La toute dernière étude Bpifrance Le Lab est sans appel : la révolution des IA génératives n'a pas encore eu lieu dans les TPE et PME. 72 % des dirigeants ayant répondu à cette enquête n'utilisent pas ces outils par choix ou n'en perçoivent pas l'usage. En revanche, un « ***shadow* AI**» est susceptible de s'être développé par endroit : c'est-à-dire un usage masqué, non déclaré au supérieur par les collaborateurs.

Si les dirigeants de TPE et PME commencent tout juste à s'emparer des outils d'intelligence artificielle générative, cela reste encore bien marginal. Les raisons résident dans le manque de connaissance des usages potentiels de ces dernières. Si ces outils peuvent être un accélérateur pour certaines tâches d'écriture, reporting, analyse et construction de différents documents, les dirigeants ne semblent pas en saisir toutes les applications et le niveau de bouleversement qu'ils impliquent. Pour certains métiers, ne pas s'approprier ces outils génèrera un retard qui pourrait leur être fatal. Selon Xavier Lazarus, managing partner du fonds d'investissement Elaia : « *Ne pas embrasser totalement ce changement de paradigme, c'est non seulement prendre le risque de rester à l'écart du progrès, mais également de perdre toute pertinence à l'avenir*. »

Selon le cabinet Roland Berger, 32 % des emplois dans le secteur privé en France sont directement exposés à une automatisation. Et pour tous les autres, cela représente autant d'opportunités manquées.

Moins d'un dirigeant sur cinq utilise les IA génératives…

Selon l'enquête Bpifrance Le Lab, 15 % des dirigeants de TPE et PME déclarent utiliser les IA génératives. L'utilisation varie selon les secteurs et atteint 24 % parmi les patrons du secteur des services. Les moins familiers avec les IA génératives sont les chefs d'entreprise des secteurs de la construction (4 %) et des transports (5 %), alors que les dirigeants de l'industrie et du commerce sont proches de la moyenne (respectivement 12 % et 11 % à les utiliser de manière régulière ou occasionnelle). L'utilisation des IA génératives par les dirigeants de PME (20 %) est légèrement supérieure à celle des dirigeants de TPE (15 %). Si ces résultats semblent faibles, ils sont en réalité en ligne avec le reste des Français, qui sont aussi 16 % à utiliser les IA Génératives selon une **étude de l'Ifop / Talan** datée de mai 2023. 13 % des dirigeants déclarent néanmoins vouloir s'en servir prochainement, démontrant au moins une curiosité envers ces technologies.

… mais leurs collaborateurs utilisent ces IA dans l'ombre

Dans la grande majorité des dirigeants qui déclarent ne pas vouloir utiliser les IA génératives, 14 % d'entre eux ne les utilisent pas par choix ou en ont interdit l'usage en entreprise. 58 % n'ont d'ailleurs pas l'intention des les utiliser à court terme. Pourquoi ? 10 % des dirigeants craignent une mauvaise utilisation des

outils d'IA par leurs collaborateurs, à l'image du partage de données confidentielles ou un manque de vérification des réponses données par les outils. 14% des répondants en ont d'ailleurs interdit l'usage au sein de leur société. Mais selon la même **étude Ifop/Talan** de mai 2023, parmi les 16% des Français déclarant utiliser les IA, 44 % les utilisent dans le cadre professionnel et privé, et 68 % d'entre eux déclarent ne pas le dire à leur supérieur.

L'utilisation de ces outils en entreprise peut donc être plus répandue qu'on ne le croit, selon Stéphane Amarsy, fondateur de The Next Mind, avec une utilisation « *dans l'ombre* » (en anglais « *shadow AI* »). Au lieu d'interdire, les dirigeants gagneraient peut-être à ouvrir un dialogue afin de mieux comprendre les usages qui sont fait de ces outils.

L'IA générative : évolution ou révolution ?

Chat GPT, QuillBot, DALL-E... Quelques mots clefs tapés dans l'un de ces outils d'IA générative suffisent désormais à un internaute pour générer une image, texte, une vidéo, voire un programme informatique. Si ce phénomène de création n'est pas nouveau, il est aujourd'hui accessible à portée de clics. À l'origine du phénomène, Chat GPT, lancé en collaboration ouverte fin 2022 par la société Open AI, qui a permis au plus grand nombre de se familiariser gratuitement avec l'**intelligence artificielle** et ses multiples possibilités.

Si Chat GPT 3.5, accessible gratuitement en ligne, n'est qu'une IA spécialisée se limitant au texte, sa version 4, en service payant, est une intelligence artificielle générative. Chat GPT 4 dépasse en effet le simple stade du **deep learning** (procédé d'apprentissage automatique utilisant des réseaux de neurones possédants plusieurs couches de neurones cachées, NDLR) en étant capable de résonner comme un cerveau humain pour produire n'importe quel type de contenu. En ce sens, la démocratisation de ces outils est une révolution. Y compris pour les entreprises, puisqu'elles entrevoient un levier potentiel majeur d'amélioration de la productivité. Néanmoins il convient de garder en tête que ces modèles d'IA s'appuient sur des techniques développées depuis le milieu des années 2010.

Une régulation de l'IA générative possible ?

L'être humain réduit à l'état d'avatars et exploité comme source d'énergie pour alimenter la Matrice ? Le futur dystopique du film Matrix faisait froid dans le dos à sa sortie en 1999. Presque 25 ans plus tard, la réalité n'a heureusement pas dépassé la fiction, mais le développement de l'IA générative n'est pas sans poser des questions d'ordre éthique. Comment être certain que l'intelligence artificielle, qui envahit déjà de nombreux pans de notre quotidien, ne soit pas biaisée ? Capable de réfléchir par elle-même, est-elle à même de duper celui qui s'en sert pour le manipuler ?

Une faille globale (erreur, discrimination...) au sein d'un outil d'intelligence artificielle pourrait avoir un impact mondial et causer un préjudice pour des millions de personnes. Doit-on brider l'innovation et légiférer ou au contraire laisser le champ des possibles aux ingénieurs ? **L'Union Européenne, à travers l'IA Act (Artificial Intelligence Act)**, travaille depuis plusieurs mois déjà sur un projet de **réglementation de l'intelligence artificielle**. L'objectif est de garantir la sécurité et le droit des utilisateurs tout en permettant le développement des différents outils. Si les approches diffèrent, les Etats-Unis, avec un projet piloté par la National Telecommunications and Information Administration, et la Chine, à travers la Cyberspace Administration of China, souhaitent également une régulation rapide de l'IA.

DOCUMENT 2

Article internet vie-publique.fr du 5 avril 2024

IA : quel potentiel et quels risques dans les services publics ?

Société | Institutions

Publié le 5 avril 2024 | 15 minutes |

Par : Lucie Cluzel-Metayer - professeure de droit public, université Paris Nanterre

Mieux cibler les populations précaires, mieux déterminer l'éligibilité des chômeurs à l'allocation chômage ou mieux rédiger les réponses en ligne aux usagers... Le potentiel de l'IA pour améliorer les politiques publiques est important. Les risques sont également identifiés et un cadre juridique national et européen se construit.

Qu'est-ce qu'un système d'IA ?

L'intelligence artificielle (IA) semble être le nouvel outil incontournable de l'action publique. Qu'il s'agisse d'attribuer des aides sociales, de mieux cibler les fraudeurs, de surveiller la population dans l'espace public ou de communiquer avec les usagers sur internet, **les autorités publiques utilisent de plus en plus l'IA**. En France, 908 millions d'euros du plan France Relance sont d'ailleurs dédiés à la transformation numérique des services publics, prenant notamment la forme d'un recours croissant à l'IA.

Définir un système d'IA (SIA) n'est pas chose simple. On considère en principe qu'il s'agit d'un système basé sur des algorithmes, fonctionnant avec une certaine autonomie. Ce système est capable d'établir des prévisions, de formuler des recommandations, ou de prendre des décisions influant sur l'environnement, en d'autres termes, d'effectuer des tâches relevant habituellement de l'intelligence humaine. **Certains SIA, comme les systèmes d'IA générative tels que ChatGPT, ou encore les systèmes de reconnaissance faciale, mobilisent des données en masse (*Big data*) et sont extrêmement performants.** Ce sont des IA de *machine learning*, qui s'améliorent en fonction des données qui les alimentent et sont en grande partie autonomes des humains qui les développent. Les administrations qui utilisent depuis longtemps des algorithmes simples (dits aussi "déterministes" car entièrement programmés par l'être humain comme c'est le cas de Parcoursup, par exemple), mobilisent aussi de plus en plus ces IA de *machine learning*.

Les potentialités de l'usage de l'IA dans les services publics expliquent son important essor, mais les risques, en particulier pour les libertés fondamentales, ne sauraient être sous-estimés. Ce qui justifie, d'ailleurs, que des règles commencent à être adoptées, au niveau national comme au niveau européen.

Les opportunités et les cas d'usage de l'IA dans les services publics

Les opportunités que représente le recours à l'IA pour les services publics sont multiples. Du côté de l'administration, l'IA permet de soulager les agents de tâches fastidieuses et répétitives, comme le traitement de millions de vœux d'étudiants pour accéder à l'enseignement supérieur. **En plus d'accélérer le temps de la décision, l'automatisation promet d'assurer une meilleure allocation des moyens matériels et humains, en les dédiant à des tâches que les algorithmes ne peuvent pas traiter.**

Du côté des usagers, l'utilisation de l'IA peut s'avérer précieuse pour prendre des décisions plus adaptées à chaque situation, par l'exploitation massive des données. Grâce au profilage, l'IA peut par exemple permettre de mieux cibler les populations précaires afin de leur attribuer des aides automatiquement, comme c'est le cas du tarif social énergie mis en place en Belgique. **D'une certaine manière, l'IA peut ainsi contribuer à rétablir l'égalité des droits.**

L'IA permet, en somme, de mieux appréhender les réalités économiques et sociales. Elle est, en cela, un puissant levier d'amélioration des politiques publiques par les connaissances qu'elle produit.

En tant qu'**outil d'aide à la décision et à l'action publique**, l'IA offre des fonctionnalités variées. Elle peut servir à identifier et authentifier des personnes : par exemple, PARAFE, système de reconnaissance faciale, commence à être utilisé dans les aéroports français pour fluidifier le passage aux frontières.

L'IA peut aider à déterminer l'accès aux droits et aux services publics : France Travail utilise ainsi un système d'IA pour déterminer l'éligibilité des chômeurs à l'allocation-chômage d'aide au retour à l'emploi, ainsi que le montant de cette aide et sa durée. En matière de sécurité publique, la vidéosurveillance assistée par l'IA a été autorisée par le législateur dans le cadre de la loi sur les Jeux olympiques et paralympiques de 2024 pour faciliter la détection des évènements anormaux dans l'espace public. De même, **en matière de lutte contre la fraude fiscale, l'IA est utilisée par l'administration pour mieux cibler les "anormalités".**

Aujourd'hui, **l'IA générative** (capable de remplir des tâches générales de création de données nouvelles - images, vidéos, textes - de manière indépendante, à partir de données d'apprentissage) **est expérimentée au sein des services publics pour aider les agents dans la rédaction des réponses en ligne aux usagers.** Sans prétendre à l'exhaustivité, mentionnons également l'utilisation de l'IA pour accroître l'efficacité de la gestion des services et des territoires, comme en témoigne l'avènement des "Smart Cities", qui ambitionnent d'améliorer les infrastructures et les services urbains et ainsi, la qualité de vie des citoyens.

Les risques de l'utilisation de l'IA dans les services publics

Si les potentialités ne sont pas négligeables, les risques ne sauraient être, pour autant, sous-estimés. **Il existe des risques, d'abord, que l'utilisation de l'IA porte atteinte à certaines libertés fondamentales**.

L'efficacité de l'IA dépend de l'exploitation massive de données, qui sont souvent des données personnelles. Le risque d'atteinte à la vie privée est dès lors très important, surtout quand l'IA permet une identification directe de l'individu dans l'espace public, comme c'est le cas avec la reconnaissance faciale, ou encore lorsqu'elle exploite des données que les individus n'ont pas conscience de divulguer, comme c'est le cas de l'IA utilisée dans le cadre du contrôle fiscal, qui est autorisée à fouiller les réseaux sociaux.

Aussi, **lorsque l'IA se déploie sur l'espace public, son usage présente des risques non seulement pour la vie privée, mais aussi pour d'autres libertés.** Le cas de la vidéosurveillance augmentée (VSA) est, à ce titre, intéressant en ce que l'usage de l'IA, entraînée sur des millions d'heures de flux d'images de personnes et capable d'analyser en temps réel les images filmées par les quelque 90 000 caméras installées sur notre territoire, permet le développement d'une surveillance généralisée des personnes.

Son déploiement impacte dès lors, non seulement le droit au respect de la vie privée, mais aussi la liberté d'aller et venir, la liberté d'expression, de manifestation et de conscience, dès lors que pour que ces libertés puissent s'exercer librement sur l'espace public, la préservation de l'anonymat est essentielle. **Le fait de se sentir surveillé peut avoir un effet dissuasif** (ce que les Anglo-Saxons nomment le *chilling effect*) sur l'exercice de ces libertés. La VSA peut également conduire à une intériorisation de nouvelles normes sociales, correspondant à des "comportements anormaux" (comme marcher dans le sens contraire au sens commun, ou chuter à terre) dont les auteurs devront répondre devant les forces de l'ordre, alors même que ces comportements ne sont pas des infractions pénales et ne sont pas édictés par des autorités élues.

L'IA peut, en outre, présenter d'importants biais discriminatoires parce que les algorithmes reproduisent la subjectivité des données qui les alimentent, voire les accroissent en raison de leur échelle de déploiement. L'algorithme Compas utilisé dans le système judiciaire américain pour décider ou non de la libération des détenus en fonction du risque de récidive, par exemple, est nourri de données en apparence neutres et parfaitement légales. Elles véhiculent en réalité des discriminations raciales résultant de l'exploitation de décisions anciennes, empreintes de préjugés raciaux : l'algorithme reproduisant des biais préexistants, les détenus d'origine afro-américaine ont des scores plus élevés de risque de récidive que les détenus de type caucasien.

Aussi, **l'utilisation des SIA peut conduire à stigmatiser certaines populations**, plus surveillées que d'autres. Le Défenseur des droits a ainsi alerté sur le fait que l'IA de lutte contre la fraude aux prestations sociales, utilisé par la Caisse nationale des allocations familiales, cible plus fréquemment les bénéficiaires percevant les prestations sociales les plus précaires, parce qu'au titre des "facteurs de risque" susceptibles de générer un signalement, figuraient l'absence ou l'irrégularité d'emploi, et l'absence ou la faiblesse de ressources. Le département de la Seine-Saint-Denis a saisi à nouveau le Défenseur des droits de cette question en décembre 2023, ce qui montre que le sujet est loin d'être épuisé.

La discrimination peut donc être générée par le choix des données et par le paramétrage de l'algorithme. Elle peut également être le fait de l'IA elle-même. Il ne faut pas oublier que le fonctionnement d'une IA de *machine learning* n'est pas explicable, même par les développeurs qui l'ont programmée. Lors de la phase d'apprentissage, c'est l'IA qui propose des résultats à partir de corrélations statistiques issues des milliards de données qui l'ont entraînée. Par exemple si, pour détecter le maraudage, l'IA détermine que statistiquement, les personnes qui sont statiques dans l'espace public sont des personnes de couleur, l'IA va opérer une corrélation "maraudage-personne de couleur". Cela va conduire mécaniquement à surveiller davantage, en conditions réelles, ces personnes, ce qui pose naturellement des problèmes aigus de discriminations.

À ces risques pour les libertés fondamentales s'ajoutent ceux liés à l'environnement numérique. **Avec la numérisation croissante des services publics, leur surface d'attaque numérique, que l'on peut définir comme l'ensemble des éléments physiques et numériques qui pourraient être compromis pour faciliter une cyberattaque, s'est naturellement étendue.** Ainsi, en septembre 2022, après plus d'un mois d'attaque informatique contre un établissement hospitalier, un groupe de hackers a diffusé des données médico-administratives des patients et personnels, l'hôpital ayant refusé de payer la rançon demandée. À la faveur de la pandémie, les attaques d'établissements de santé ont d'ailleurs doublé entre 2020 et 2021.

Notons, pour finir, que **le coût énergétique de l'exploitation de l'IA, en forte hausse avec la diffusion massive de ses applications**, ne saurait être sous-estimé dans le contexte de réchauffement climatique que nous connaissons.

L'adoption d'un cadre juridique

Les risques de l'IA n'ont pas échappé aux législateurs qui commencent à adopter des réglementations pour tenter d'encadrer son utilisation, au niveau national et européen.

D'abord, le RGPD et la loi informatique et libertés encadrent l'IA en tant que traitement automatisé de données à caractère personnel. Ces textes posent des obligations assez contraignantes pour les responsables de traitements, ainsi que certaines interdictions. Par exemple, selon l'article 22 du RGPD, "la personne concernée a 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". Cela signifie qu'en principe, l'administration ne peut imposer à une personne de faire l'objet d'une décision administrative entièrement automatisée. **Mais le principe est assorti d'importantes exceptions, offrant aux États membres une large marge de manœuvre.** L'interdiction ne s'applique pas, notamment, lorsque la décision "est autorisée par le droit de l'Union ou le droit de l'État membre auquel le responsable du traitement est soumis et qui prévoit également des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée". Ainsi, dès lors que le système d'IA est autorisé par une loi, ou même un règlement et que des garanties pour le respect des libertés sont assurées (ce que le juge pourra contrôler), **la décision administrative peut effectivement être exclusivement fondée sur l'algorithme**.

Les textes restent très permissifs en ce qui concerne les applications administratives : la loi informatique et libertés révisée prévoit ainsi la possibilité de prendre des décisions administratives individuelles fondées sur des algorithmes, et même celle de prendre des décisions entièrement automatisées si des données sensibles ne sont pas en jeu et si l'administration est capable d'expliquer le fonctionnement de l'algorithme mobilisé.

Ensuite, la loi pour une République numérique de 2016 impose des exigences particulières de transparence concernant l'utilisation des algorithmes par les administrations. Elle pose en effet le principe que ces derniers doivent être considérés comme des documents administratifs (article L. 300-2 du code des relations entre le public et l'administration) ce qui implique que les règles définissant le traitement ainsi que les principales caractéristiques de sa mise en œuvre puissent être communiquées à toute personne en faisant la demande (L. 311-1 du même code) et publiées en ligne si l'algorithme utilisé fonde une décision individuelle (L. 312-1-3).

L'approche par les risques, retenue par le règlement européen sur l'IA qui vient d'être adopté (accord, décembre 2023) et qui entrera en vigueur en 2026, **s'inscrit dans la même logique de soumettre certains usages de l'IA à des exigences fortes en matière de transparence**, en particulier. Le règlement distingue en effet les IA selon les risques qu'elles présentent pour la santé, la sécurité, la démocratie, les libertés fondamentales et l'État de droit.

Certains usages sont purement et simplement interdits : c'est le cas des systèmes de notation sociale, de certains systèmes de reconnaissance des émotions dans le domaine du travail et de l'éducation, des systèmes capables de manipuler les comportements humains ou d'exploiter certaines vulnérabilités ou encore des outils dits de "justice prédictive" visant à profiler les personnes pour évaluer le risque qu'elles commettent des infractions pénales. C'est également le cas des systèmes d'IA d'identification biométrique à distance en temps réel (VSA) dans les espaces publics à des fins répressives. **L'interdiction est cependant assortie d'une exception notable, qui permet aux autorités d'utiliser ce type d'IA "dans la mesure où cette utilisation est strictement nécessaire".** La recherche de victimes, la prévention d'une menace spécifique, substantielle et imminente pour la vie ou la sécurité physique des personnes ou la prévention d'attaques terroristes, mais aussi la recherche d'auteurs d'infractions graves, pourront, par exemple, justifier le recours à la reconnaissance faciale.

Dans la majorité des cas, les systèmes d'IA utilisés par les pouvoirs publics seront autorisés, mais appartiendront à la catégorie des IA à haut risque justifiant l'application de règles spécifiques destinées à en prévenir les dangers. Des mesures d'ordre technique et procédural devront être adoptées pour assurer la traçabilité de l'utilisation de l'IA (système de gestion des risques, gouvernance des données, documentation technique, mesures d'enregistrement de l'activité de l'IA), la robustesse des dispositifs en matière de cybersécurité, la transparence vis-à-vis des utilisateurs et le contrôle humain. L'idée générale est d'ailleurs de "garder la main" pour éviter toute déshumanisation, c'est-à-dire d'être capable d'interpréter les résultats proposés par l'IA, de tester les SIA régulièrement, de les modifier au besoin, de prendre de la distance par rapport aux solutions proposées voire, de renoncer à leur utilisation. L'accès aux informations relatives à l'identité du fournisseur et aux caractéristiques et performances du système permettra également aux utilisateurs – en l'occurrence aux autorités administratives – d'imputer la responsabilité des dysfonctionnements.

Ainsi, les systèmes d'IA destinés à affecter les étudiants dans les établissements d'enseignement, à évaluer l'éligibilité et le montant des prestations sociales ou encore, à identifier les comportements anormaux dans l'espace public, seront soumis à ces règles de transparence et de contrôle renforcés. **Ces systèmes devront, de surcroît, être enregistrés dans une base de données de l'Union européenne** après une évaluation de

leur conformité à ces exigences. Les citoyens pourront donc enfin avoir accès à un inventaire des systèmes d'IA utilisés par les autorités publiques.

La prise de conscience de l'importance d'encadrer l'utilisation de l'IA dépasse aujourd'hui le cadre de l'Union européenne. Les États-Unis ont adopté un *Executive Order* (E.O. 14110) à ce sujet, tandis qu'une convention cadre sur l'IA est en préparation au sein du Conseil de l'Europe, pour garantir que les usages de l'IA soient pleinement compatibles avec les droits de l'homme, la démocratie et l'État de droit. Dans ce cadre, la révision du manuel "L'Administration et vous" pour intégrer l'IA aux principes directeurs d'une "bonne administration" (Conseil de l'Europe, 2024), est le signe que la spécificité de l'utilisation de l'IA par les pouvoirs publics est enfin effectivement prise en considération.

DOCUMENT 3

Extrait du dossier de veille technologique lab.securite-sociale.fr - Octobre 2024

Qu'est-ce que la souveraineté numérique ?

D'après la chaire Digital, gouvernance et souveraineté de Sciences Po, la souveraineté numérique désigne la **capacité des Etats d'agir dans le cyberespace**, en faisant **respecter leurs règles** aux différents acteurs du monde virtuel.

En France, le terme souveraineté numérique a émergé dans le débat public au début des années 2010, pour mettre en exergue la forte dépendance de la France et de l'UE vis-à-vis des entreprises numériques étrangères (principalement américaines), et par conséquent la nécessité de construire une stratégie française et européenne pour s'en émanciper.

Cette émancipation relève de deux domaines : la **sphère juridique** au travers de laquelle l'Etat pourrait parvenir à réguler les leaders technologiques contemporains, et la **sphère économique et industrielle** pour rattraper le retard technologique qui place la France et l'Europe en situation de dépendance.

La souveraineté numérique polarise néanmoins les débats, qui mettent en opposition les **objectifs politiques de long terme** (émancipation et autonomie) face aux **contraintes opérationnelles court terme** (garantir le fonctionnement de l'économie très dépendante des services numériques existants).

Par ailleurs, cette dichotomie se matérialise par l'écart entre le discours des représentants politiques et la réalité opérationnelle : si la **PFUE** (Présidence française du Conseil de l'Union Européenne) affirme en 2023 placer la **souveraineté numérique** dans le **top trois de ses sujets prioritaires** (aux côtés de la crise énergétique et de la protection de l'environnement), la stratégie de l'UE relève aujourd'hui davantage de l'association d'une **multitude de stratégies nationales isolées** (ex : Plan France 2030).

Que signifie la souveraineté numérique pour les politiques français ?

Selon les discours portés par les politiques français, le concept de souveraineté économique fait appel aux trois composantes classiques de la souveraineté - la **puissance, le peuple et l'État** - en association avec l'UE

Dans cette optique, la souveraineté numérique européenne permettrait d'aboutir selon le président Macron à une Europe "*puissante dans le monde, pleinement souveraine, libre de ses choix et maître de son destin*". Ainsi, la souveraineté numérique est interprétée comme le pouvoir d'une autorité à établir unilatéralement les règles régissant l'espace numérique, tout en garantissant la **liberté stratégique de l'Europe**, se rapprochant ainsi de **l'autonomie** et de **l'indépendance**.

Concernant le **titulaire de ces pouvoirs**, le gouvernement français distingue le modèle européen des modèles américain et chinois. Dans le modèle américain, la souveraineté technologique est accaparée par les grandes entreprises, tandis que dans le modèle chinois, elle est sous l'emprise de l'État autoritaire (illustré par le cas emblématique d'Alibaba). Face à ces deux extrêmes, le ministre Bruno Le Maire propose un **modèle européen démocratique**, où la souveraineté technologique est entre les mains du **peuple européen**.[2].

En ce qui concerne l'exercice de cette souveraineté numérique, le gouvernement français reconnaît que celle-ci ne peut être pleinement réalisée au niveau national sans l'implication de l'Union européenne. Une **coopération entre les États membres** est ainsi nécessaire pour assurer une concurrence équitable et protéger les entreprises européennes face aux géants du numérique américains et chinois. L'Union Européenne, du fait de sa taille, son dynamisme et ses ressources financières, est considérée comme l'échelon pertinent pour faire face à ces enjeux.

Enfin, la souveraineté numérique « *ne peut être l'apanage de seuls États* »[3]. Un rapport déposé à l'Assemblée nationale en 2021 souligne que les entreprises, grâce à leurs innovations, et les citoyens, à travers leurs usages, doivent également participer à sa construction. Dans ce modèle promu par la PFUE, les **entreprises jouent un rôle crucial** dans la construction de la souveraineté numérique, tandis que les **citoyens** contribuent à travers la génération de données exploitables par ces entreprises du numérique.

La notion de souveraineté numérique revêt une véritable dimension stratégique pour les différents Etats du monde. Être souverain, au sens strict du terme c'est ne plus dépendre d'autrui sur le plan économique, technique et géopolitique. Dans le contexte de globalisation de notre économie, une réelle interdépendance existe entre les Etats rendant l'indépendance totale irréaliste et la souveraineté se traduit par la volonté de ne pas se soumettre à la puissance d'autres puissances, notamment celles des entreprises américaines. Il s'agit là d'un enjeu majeur pour l'Union Européenne, qui aspire à un modèle supranational démocratique dont la souveraineté technologique restera éternellement détenue par et pour le peuple européen.

Cependant, cela reste encore loin d'être réalisé au sein de l'UE, qui accuse un retard majeur face aux géants du numérique chinois et aux américains. Pour être capable de se doter de l'instrument de puissance que représente l'IA, elle doit donc réussir à se doter des inducteurs et des freins évoqués précédemment pour répondre aux nombreux enjeux techniques sous-jacents.

Par ailleurs, si l'on s'en tient à la définition proposée par le juriste en relations internationales, Serge Sur, une **puissance** est une **« entité capable de faire, faire faire, empêcher de faire ou refuser de faire ».**

La souveraineté numérique apparaît alors comme un instrument éminemment efficace pour maintenir son rang de grande puissance géopolitique, et ce à plusieurs égards : pour l'accès aux données, la protection des données, l'indépendance technologique et la cybersécurité. Ces différents points sont détaillés au sein de la revue des enjeux techniques développée *infra*.

Développer des solutions d'IA souveraine

Les développements en cours pour une IA souveraine

Le développement des IA souveraines en France et en Europe permet de répondre aux défis posés par l'IA exposés plus haut.

En termes de protection des données et de conformité règlementaire au RGPD notamment, les IA souveraines offrent davantage de garanties grâce à **un hébergement et un entrainement de données local**, et en permettant ainsi un contrôle national des données sensibles des citoyens et des entreprises.

En s'appuyant sur des technologies locales ou même open source (en France, des projets comme le P16 visent à développer des bibliothèques logicielles open source pour l'IA, favorisant ainsi une collaboration entre les secteurs public et privé) la dépendance des entreprises vis-à-vis des géants technologiques étrangers se réduit et renforce l'autonomie des états tout en assurant une protection renforcée des infrastructures techniques.

Les IA souveraines intègrent **des mesures de sécurité avancées** pour protéger contre les cyberattaques et les accès non autorisés.

Enfin, en terme de transparence et d'éthique, en France, les IA souveraines mettent l'accent sur **la transparence des algorithmes et l'explicabilité des décisions**, ce qui est crucial pour maintenir la confiance des utilisateurs et respecter les normes éthiques1.

En résumé, les IA souveraines permettent de protéger les données, de renforcer l'indépendance technologique, d'améliorer la sécurité, de stimuler l'innovation locale et de garantir la transparence et l'éthique. Ces aspects sont essentiels pour répondre aux enjeux de souveraineté numérique en France et en Europe.

Albert et Mistral : des exemples de LLM souverain

Ces deux solutions font l'objet d'une description dans le dossier « IA, cas d'usage et panorama des solutions ».

Albert

Développée par la Direction interministérielle du numérique (DINUM) en France, Albert est un chatbot conçu pour assister les agents des services publics français. Il a été présenté comme la première IA 100 % souveraine en Europe. Albert a été développé pour répondre aux besoins spécifiques des administrations françaises, en utilisant des données locales et en respectant les régulations nationales et européennes en matière de protection des données. Cela permet de garantir que les informations sensibles restent sous contrôle national et que l'IA fonctionne de manière transparente et sécurisée.

DOCUMENT 4

Extrait du guide de stratégie d usage de l intelligence artificielle en matière de gestion des ressources humaines dans la fonction publique de l état disponible sur fonction-publique.gouv.fr - Juin 2024

Parmi les différentes formes d'IA existantes, ces dernières années, **une application particulière d'IA est montée en puissance et se trouve au cœur de l'actualité du monde du travail : l'IA générative**, même si d'autres formes d'IA sont déjà utilisées depuis longtemps. Toutes les formes d'IA étant fondées sur les données, **« les gouvernements doivent garantir qu'ils ont accès à des données suffisantes, de qualité et impartiales avant de pouvoir tirer pleinement et éthiquement parti de ces techniques »**[2].

Au niveau européen, après les travaux de l'Alliance européenne pour l'IA, ainsi que les orientations de la stratégie européenne pour l'IA (2018) et le Livre blanc sur l'IA (2020), **un règlement sur l'IA** négocié depuis 2021 et adopté le 21 mai 2024 (mais qui n'entrera en vigueur qu'en 2026) donne une vision européenne de l'IA fondée sur **l'éthique et la gestion des risques d'utilisation de cette technologie, qui, concernant en particulier la formation professionnelle, l'emploi et les ressources humaines, sont considérés comme élevés.**

Concernant la fonction publique en particulier, la stratégie du ministère de la Transformation et de la Fonction publiques pour le déploiement de l'IA dans la fonction publique a été initiée le 5 octobre 2024 avec l'expérimentation de l'IA générative, via la plateforme « Services publics + » mise en place par la DITP et la création de l'« AllIAnce » publique-privée[6], en complément des travaux de la DINUM depuis le mois de juin 2023 pour développer une IA générative souveraine. Par ailleurs, la **feuille de route des SIRH** de l'Etat[7] du 8 mars 2023 pour la période 2023-2027 comporte depuis fin mars 2024 un nouvel objectif d'interroger le recours à l'IA. Cet objectif est transverse et concerne l'ensemble des domaines en GRH (compétences, dématérialisation, pilotage, etc.).

« L'expérimentation d'un outil d'IA d'aide à la formulation de réponses à des questions d'usagers sur la plateforme Services publics + a permis de diviser le délai de réponse par deux : de 6 jours, on est passé à 3 jours, avec **une appréciation croissante de la part des agents** qui utilisent les réponses générées »

Grégoire TIROT, Chargé de la préfiguration du service Innovation et Participation citoyenne à la Délégation interministérielle à la transformation publique (DITP).

« L'IA est aujourd'hui utilisée dans les fonctions publiques pour de la reconnaissance de données et de motifs de comportements **à des fins de prévision et de production de supports d'interaction avec l'humain** ».

Daniel GERSON, Chef de projet Gestion et emploi publics au sein de la Direction de la gouvernance publique de l'OCDE.

Pour préciser les opportunités d'usage de l'IA en GRH de la fonction publique d'Etat, une enquête a été conduite auprès du réseau interministériel de prospective RH entre décembre 2023 et janvier 2024 et a permis d'aboutir aux conclusions suivantes :

• Les **principaux bénéfices** attendus de l'usage de l'IA en GRH sont : gain de temps pour les managers et les services RH ; élargissement des champs d'analyse ; optimisation de l'offre de services RH ; anticipation des évolutions ; facilitation du travail des managers et des services RH ; personnalisation de l'offre de services RH ; amélioration de la situation des agents ; meilleur accompagnement des agents ; harmonisation des processus RH ; articulation des différents outils et processus RH ; développement de l'approche prospective dans la GRH.

• Les **principaux risques** liés à l'usage de l'IA en GRH sont : sensibilité des données utilisées par l'IA, au sens du Règlement général pour la protection des données (RGPD) européen ; déresponsabilisation managériale ; RPS et tensions sociales ; persistance des biais cognitifs issus des programmations algorithmiques défaillantes ; perte d'attractivité des postes liée à la déshumanisation des services RH.

L'utilisation de l'IA de façon générale, et a fortiori dans le monde du travail, nécessite (i) **de connaître et de maîtriser les données utilisées et qu'elles le soient de façon légale et éthique,** (ii) **de connaître les capacités des outils d'IA** (qui peuvent être limités en termes de raisonnement logique ou reproduire ou créer des biais et des discriminations), (iii) **de choisir les tâches effectuées par des outils d'IA et celles effectuées par des humains** (l'IA devant aider le travail humain et non le remplacer), (iv) **d'accompagner les agents** dans l'appropriation et l'utilisation des outils d'IA et (v) **de comprendre, de contrôler et de maîtriser les résultats produits** (qui peuvent être biaisés, limités et plausibles mais pas nécessairement corrects) **et leur utilisation**.

Ces objectifs doivent structurer **la gouvernance de et avec (et non par) l'IA** dans les organisations de travail, dont le principe directeur doit être la **responsabilité humaine** de l'utilisation de l'IA.[14]

Les éléments exposés précédemment conduisent la DGAFP à proposer, dans une logique de symétrie des attentions, une stratégie d'usage de l'IA en GRH dans la fonction publique d'Etat.

Découlent des enjeux précités des **principes** regroupés en 3 thématiques :

• **Protection de l'agent et de l'administration** : utilisation légale et transparente des données (privées et sensibles) ; solidité/ robustesse et sécurité de l'IA ;

• **Contrôle par l'humain** : concertation et autonomie (fixation du cadre d'usage en collaboration avec les représentants du personnel ; déclaration d'utilisation de l'IA par les agents ; liberté des agents d'utiliser des outils d'IA ou leurs résultats) ;

• **Déontologie** : transparence, justice, équité, diversité, inclusion, non-discrimination.

Pour appliquer ces principes dans l'utilisation de l'IA en GRH, **un cadre d'usage** selon les 3 axes suivants apparaît nécessaire.

Intégration et utilisation de l'IA :

- **Identifier les responsables** des technologies et méthodologies d'IA ;
- **Identifier les fonctions** pouvant exploiter l'IA (direction, comptabilité et finances, informatique, communication, logistique ...) ;
- **Identifier les missions** pouvant être assurées avec l'IA ;
- **Définir les objectifs** (éventuellement chiffrés) à atteindre pour la structure et ses RH en utilisant l'IA : amélioration de la productivité, résolution des problèmes complexes... ;
- **Choisir des outils et logiciels d'IA** fiables et les connaître afin de maîtriser les résultats et ainsi sécuriser la responsabilité des agents publics et des administrations ;
- **Définir des méthodologies d'exploitation** de l'IA et des données ;
- **Créer une obligation déclarative des agents** dès lors qu'ils ont recours à l'IA ;
- **Etablir une cartographie des risques** liés à l'utilisation de l'IA en GRH publique pour mettre en place **un contrôle et un audit internes** de la stratégie.

DOCUMENT 5

Extrait des recommandations de l'intelligence artificielle générative en entreprise du cabinet Derriennic Associés - derriennic.com - Juin 2023

Principaux Risques	Préconisations
Confidentialité : Toute donnée fournie à un modèle d'IA générative est susceptible d'être transmise au fournisseur de l'IA, qui peut ensuite s'en servir pour entrainer l'IA, en améliorer la performance. En résulte un risque de divulgation d'informations sensibles ou de données personnelles au public puisque les données communiquées par les salariés sont susceptibles d'être reproduites dans les résultats générés à l'avenir pour d'autres utilisateurs, hors de l'entreprise.	La Commission Européenne dans ses lignes directrices **interdirait à son personnel tout partage** avec l'IA **d'information qui ne relève pas déjà du domaine public** et toute donnée personnelle. *En complément, les mesures suivantes pourraient être initiées : favoriser la sélection d'outils d'IA :ne permettant pas la réutilisation des données des utilisateurs à des fins de réentrainement du modèle, à des fins d'amélioration du système d'IA ;assurant la confidentialité des données communiquées ou permettant un paramétrage de la confidentialité ;sensibiliser le personnel aux problématiques de confidentialité et de respect du secret des affaires.*
Erreurs : Les lacunes du modèle d'IA ou le défaut de qualité ou de mise à jour des données d'entrainement peuvent conduire à des réponses erronées ou biaisées.	L'IA n'est qu'un outil, une aide, dont l'utilisateur doit conserver la maitrise. La Commission Européenne dans ses lignes directrices appellerait son personnel à *« toujours **évaluer de manière critique toute réponse produite** par un modèle d'IA générative disponible en ligne **afin de détecter les biais** potentiels **et les informations factuellement inexactes** ».*
Violation de droits de PI puisqu'à ce jour, il n'existe aucune obligation de **transparence des éditeurs sur les sources des données utilisées à des fins d'entrainement** *On rappellera les risques de contrefaçon ou de parasitisme pour les utilisateurs dans la mesure où les résultats des IA génératives sont susceptibles de reproduire tout ou en partie des contenus utilisés à des fins d'entrainement, sans disposer des droits et autorisations nécessaires.*	La Commission Européenne dans ses lignes directrices appellerait son personnel à « *ne **jamais reproduire directement les résultats d'un modèle d'IA générative dans des documents publics**, comme lors de la création de textes de la Commission, notamment ceux qui sont juridiquement contraignants* ». La Commission envisagerait donc une utilisation uniquement à des fins internes, sans communication au public. A noter que, selon nous, cette recommandation, volontairement stricte à des fins de sécurité juridique, pourrait être adaptée en fonction (i) de la nature des résultats générés, (ii) de l'usage des IA et de la destination de leurs résultats, et (iii) des garanties apportées par l'éditeur. *A titre d'exemple : concernant les IA génératives de codes, le risque pourrait être minoré par le recours à des outils de scans et d'administration des licences open source, afin d'identifier les composants soumis à licence libres et corrélativement de s'assurer du respect des licences concernées, de leur éventuel caractère contaminant, voire de la modification du code généré aux fins de suppression des composants concernésune réutilisation y compris à des fins commerciale serait possible dans l'hypothèse où l'entreprise disposerait de garanties de la part de l'éditeur en cas de recours de tiers en lien avec les contenus générés, ce que certains éditeurs d'IA génératives (notamment Adobe) semblent désormais prêts à proposer à leurs clients.*
Indisponibilité : Les systèmes d'IA génératives étant le plus souvent accessibles en ligne, il peut y avoir un délai de réponse long ou une indisponibilité de l'outil à n'importe quel moment.	La Commission Européenne dans ses lignes directrices appellerait son personnel à **ne jamais s'appuyer sur ces outils pour les tâches critiques et urgentes**.

DOCUMENT 6

Extrait des discours de vœux 2025 de la ministre du MASA Annie Genevard et de la secrétaire générale du MASA Cécile Bigot-Dekeyzer

Discours de la ministre Annie Genevard, vœux 2025

Un mot sur un sujet particulier, qui mérite que je m'y arrête. L'intelligence artificielle présente des opportunités pour faciliter le travail des agents, renforcer son efficacité et faciliter la vie des usagers agricoles. Une feuille de route sera finalisée d'ici la fin du trimestre, avec des cas d'usage identifiés, et je serai attentive à ce que les enjeux sociaux et environnementaux de l'IA soient bien pris en compte.

Discours de la secrétaire générale du MASA Cécile Bigot-Dekeyzer, vœux 2025

Nous allons aller encore plus loin en 2025 avec l'élaboration d'une feuille de route ministérielle en matière d'Intelligence artificielle. L'IA présente des opportunités tant pour les services que les usagers : il faudra définir les cas d'usage présentant une réelle valeur ajoutée, veiller à la qualité et la sécurité des données, et être attentifs aux enjeux sociaux et environnementaux associés à ces technologies.

DOCUMENT 7

Article internet vie-publique.fr du 11 février 2025

Intelligence artificielle : le cadre juridique européen de l'IA en sept questions

Société | International

Dernière modification : 11 février 2025 | 14 minutes | Par : La Rédaction

Le règlement européen sur l'intelligence artificielle (IA) du 13 juin 2024 est paru au Journal officiel de l'Union européenne du 12 juillet. Le Conseil de l'Europe a adopté le 17 mai 2024 un traité international visant à garantir une IA respectueuse des droits fondamentaux. Retour sur l'IA dans l'UE en sept questions avec Vie-publique.

1 Existe-t-il une législation européenne sur l'IA ?

Le **règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle**, premier acte législatif sur l'intelligence artificielle (IA), est paru au *Journal officiel de l'Union européenne (JOUE)* du 12 juillet 2024.

Le texte a été adopté par 523 votes pour, 46 contre et 49 abstentions au Parlement européen le 13 mars 2024.

Ce cadre réglementaire de l'IA poursuit les objectifs suivants :

- veiller à ce que les **systèmes d'IA mis sur le marché** soient **sûrs** et respectent la **législation en vigueur en matière de droits fondamentaux**, les valeurs de l'UE, l'**État de droit** et la **durabilité environnementale** ;
- promouvoir l'adoption d'une **IA axée sur l'humain** et **digne de confiance** ;
- garantir un **cadre juridique uniforme** afin de **faciliter les investissements et l'innovation** ;
- renforcer la **gouvernance** et l'**application effective de la législation existante** en matière d'**exigences de sécurité applicables aux systèmes d'IA** et de **droits fondamentaux** ;
- améliorer le **fonctionnement du marché intérieur** pour des applications d'**IA légales et sûres**, et empêcher la fragmentation du marché.

Plus spécifiquement, le règlement établit :

- des règles harmonisées concernant la mise sur le marché, la mise en service et l'utilisation de systèmes d'IA dans l'UE ;
- l'**interdiction de certaines pratiques** ;

- des exigences spécifiques applicables aux systèmes d'**IA à haut risque** ;
- des règles harmonisées en matière de transparence applicables :
  - aux systèmes d'IA destinés à interagir avec des personnes,
  - aux systèmes de reconnaissance des émotions et de catégorisation biométrique,
  - aux systèmes d'IA générative utilisés pour générer ou manipuler des images ou des contenus audio ou vidéo ;
- des mesures visant à soutenir l'innovation, particulièrement pour les petites et moyennes entreprises (PME).

Cette approche doit tenir compte des résultats bénéfiques sur les plans sociaux et environnementaux que peut apporter l'IA, mais aussi des nouveaux risques ou des conséquences négatives que peut engendrer cette technologie.

La cohérence est assurée avec la Charte des droits fondamentaux de l'UE, mais aussi avec le droit dérivé de l'UE en matière de **protection des données**, de **protection des consommateurs**, de **non-discrimination** et d'**égalité entre les femmes et les hommes**. Le règlement complète le droit existant en matière de non-discrimination en prévoyant des exigences qui visent à réduire au minimum le risque de discrimination algorithmique, assorties d'obligations concernant les **essais**, la **gestion des risques**, la **documentation** et le **contrôle humain** tout au long du cycle de vie des systèmes d'IA.

Le règlement sur l'IA sera pleinement applicable à partir du 2 août 2026.

2 Quelle est la définition de l'intelligence artificielle donnée par le règlement sur l'IA ?

Afin de garantir la **sécurité juridique**, de faciliter la **convergence internationale** et une large acceptation, tout en offrant la souplesse nécessaire à la **prise en compte des évolutions technologiques rapides** de ce domaine, la **notion de "système d'IA"** est définie par le règlement. Elle est alignée sur les travaux des organisations internationales qui œuvrent dans ce domaine.

Une des caractéristiques essentielle des systèmes d'IA est leur **capacité d'inférence**, que le règlement définit comme "*le processus consistant à générer des sorties telles que des **prédictions**, du **contenu**, des **recommandations** ou des **décisions**, qui peuvent **influencer l'environnement physique ou virtuel**, et la capacité des systèmes d'IA à inférer des modèles ou des algorithmes, ou les deux, à partir d'entrées ou de données*".

Les techniques qui permettent l'inférence lors de la construction d'un **système d'IA** comprennent :

- des approches d'**apprentissage automatique**, qui apprennent à partir des données la manière d'atteindre certains objectifs ;
- des approches fondées sur la **logique** et les **connaissances** qui font des inférences à partir des **connaissances encodées** ou de la **représentation symbolique de la tâche à résoudre.**

Ces techniques vont au-delà du traitement de données de base en ce que la capacité d'inférence d'un système d'IA lui permet l'**apprentissage**, le **raisonnement** ou la **modélisation**.

Les systèmes d'IA étant conçus pour fonctionner à **différents niveaux d'autonomie**, ils bénéficient d'un certain degré d'indépendance dans leur action par rapport à une ingérence humaine et de capacités à fonctionner sans intervention humaine.

Le règlement sur l'IA définit aussi :

- un **système d'identification biométrique à distance** comme "*un système d'IA destiné à identifier des personnes physiques sans leur participation active, généralement à distance, en comparant les données biométriques d'une personne avec celles qui figurent dans une base de données*" ;
- un **système d'identification biométrique à distance en temps réel** comme "*un système d'identification biométrique à distance dans lequel l'acquisition des données biométriques, la comparaison et l'identification se déroulent sans décalage temporel important et qui comprend non seulement l'identification instantanée, mais aussi avec un léger décalage afin d'éviter tout contournement des règles*".

3 Quelles sont les pratiques d'IA interdites par le règlement ?

Le **règlement** sur l'IA **interdit** les **pratiques suivantes** en matière d'intelligence artificielle :

- système d'IA ayant recours à des **techniques subliminales** au-dessous du seuil de conscience d'une personne pour **altérer substantiellement son comportement** et de manière à causer un préjudice physique ou psychologique (**manipulation du comportement humain** pour contourner le libre arbitre) ;
- système d'IA exploitant les **éventuelles vulnérabilités dues à l'âge ou au handicap** d'un individu pour altérer substantiellement son comportement et de manière à causer un préjudice physique ou psychologique ;
- systèmes d'IA destinés à **évaluer ou à établir un classement de la fiabilité de personnes** en fonction de leur **comportement social** ou de **caractéristiques personnelles** et pouvant entraîner un traitement préjudiciable de personnes, dans certains contextes, injustifié ou disproportionné. L'accord trouvé entre le Parlement et les États membres précise l'**interdiction des systèmes de catégorisation biométrique** utilisant des **caractéristiques sensibles** (opinions politiques, religieuses, philosophiques, orientation sexuelle…) et la **notation sociale basée sur le comportement social ou les caractéristiques personnelles** ;
- systèmes d'IA pour mener des évaluations des risques des personnes physiques **visant à évaluer ou à prédire le risque qu'une personne physique commette une infraction pénale**, uniquement sur la base du profilage d'une personne physique ou de l'évaluation de ses traits de personnalité ou caractéristiques. Cette interdiction ne s'applique pas aux systèmes d'IA utilisés pour étayer l'évaluation humaine de l'implication d'une personne dans une activité criminelle ;
- systèmes d'IA créant ou développant des **bases de données de reconnaissance faciale** par le moissonnage non ciblé d'images faciales provenant d'internet ou de la vidéosurveillance ;

- **reconnaissance des émotions** sur le **lieu de travail** et les **établissements d'enseignement**, sauf pour des raisons médicales ou de sécurité ;
- systèmes d'**identification biométrique à distance en temps réel** dans des espaces accessibles au public à des fins répressives, sauf dans les cas suivants :
  - recherche ciblée de **victimes potentielles spécifiques de la criminalité** (enfants disparus, traite, exploitation sexuelle),
  - prévention d'une **menace spécifique, substantielle et imminente pour la vie ou la sécurité des personnes** ou la prévention d'une **attaque terroriste**,
  - identification, localisation ou poursuite à l'encontre des auteurs ou des suspects de certaines **infractions pénales punissables d'une peine d'une durée maximale d'au moins quatre ans**.

L'utilisation de systèmes d'identification biométriques à distance en temps réel doit :

- tenir compte de la situation donnant lieu au recours au système et de la **gravité** ou de l'**ampleur du préjudice en l'absence de son utilisation** ;
- tenir compte des **conséquences sur les droits et libertés** de toutes les personnes concernées (gravité, probabilité, ampleur) ;
- être subordonnée à une **autorisation préalable octroyée par une autorité judiciaire ou administrative** compétente.

L'annexe III du règlement dresse une liste des **systèmes d'IA à haut risque**. Les eurodéputés ont inclus dans le règlement une **analyse d'impact obligatoire sur les droits fondamentaux**, également applicable au secteur bancaire et des assurances. Les systèmes d'IA utilisés pour influer sur le résultat d'élections et le comportement des électeurs sont classés à haut risque. Les citoyens auront le droit de :

- déposer des **plaintes** concernant les systèmes d'IA ;
- recevoir des **explications sur les décisions fondées sur des systèmes d'IA à haut risque** ayant une incidence sur leurs **droits**.

Le non-respect des règles pourra entraîner des **amendes** pouvant aller jusqu'à **35 millions d'euros ou 7% du CA mondial**, en fonction de la taille de l'entreprise et de l'infraction.
Le 4 février 2025, la Commission a publié les **lignes directrices sur les pratiques interdites en matière d'IA** telles qu'elle sont définies par le règlement. Elles donnent un aperçu des **pratiques d'IA jugées inacceptables** en raison de leurs **risques potentiels pour les valeurs européennes** et les **droits fondamentaux**. Ces lignes directrices fournissent, afin d'aider les parties prenantes à comprendre la législation sur l'IA et de garantir une application cohérente et uniforme dans toute l'UE :

- des **interprétations**, non contraignantes, sur l'interprétation des interdictions par la Commission ;
- des **explications juridiques** ;
- des **exemples pratiques**.

4 Quelles sont les autorités françaises compétentes pour veiller au respect des droits fondamentaux par l'IA ?

Selon l'article 77 du règlement sur l'IA, chaque État membre doit identifier les autorités ou organismes publics nationaux qui supervisent ou font respecter les obligations au titre du droit de l'Union visant à protéger les droits fondamentaux, y compris le droit à la non-discrimination, en ce qui concerne l'utilisation des systèmes d'IA à haut risque visés à l'annexe III dudit règlement.

Les autorités françaises identifiées sont :

- la **Direction générale de la concurrence, de la consommation et de la répression des fraudes (DGCCRF)** ;
- la **Commission nationale de l'informatique et des libertés (CNIL)** ;
- le **Défenseur des Droits**.

Cette liste pourra être complétée.

5 Comment le règlement sur l'IA soutient-il l'innovation ?

Les systèmes d'intelligence artificielle connaissent une **évolution rapide** qui nécessite la mise en place de nouvelles formes de **contrôle réglementaire** et d'un **espace sûr pour l'expérimentation**. Le règlement doit garantir une **innovation responsable** et l'intégration de garanties et de mesures d'atténuation des risques appropriées. Pour garantir un cadre juridique propice à l'innovation, les États membres devront mettre en place des ***"bacs à sable réglementaires"*** sur l'IA pour faciliter l'innovation sous un contrôle réglementaire strict avant la mise sur le marché ou en service de ces systèmes.

L'objectif est :

- de **favoriser l'innovation** en créant un **environnement contrôlé d'expérimentation** au stade du **développement** afin d'assurer la conformité des systèmes d'IA innovants avec le règlement ;
- de **renforcer la sécurité juridique pour les innovateurs** ainsi que le contrôle et la compréhension des **possibilités**, des **risques émergents** et des **conséquences** de l'utilisation de l'IA ;
- d'accélérer l'**accès aux marchés** en supprimant les obstacles pour les PME et les jeunes entreprises.

6 Au-delà du règlement, comment l'UE prend-elle en compte l'intelligence artificielle ?

L'UE n'a pas attendu le règlement sur l'IA du 13 juin 2024 pour s'intéresser à cette question et à ses enjeux.

En 2017, le Conseil européen souligne la nécessité de faire preuve d'un **sens de l'urgence face aux tendances émergentes, notamment l'IA** [Réunion du Conseil européen (19 octobre 2017) – conclusions].

En 2018, la Commission présente un premier **plan coordonné sur l'intelligence artificielle** (communication de la Commission). Il s'agit d'un engagement commun visant à renforcer le potentiel européen pour être compétitif à l'échelle mondiale, notamment au travers de stratégies nationales des États membres.
Dans ses conclusions de 2019 sur le plan coordonné, le Conseil de l'UE souligne la nécessité d'assurer le **respect intégral des droits des citoyens européens**. Le Conseil demande que la législation en vigueur fasse l'objet d'un réexamen [Intelligence artificielle b) Conclusions sur le plan coordonné dans le domaine de l'intelligence artificielle - Adoption].

Le Parlement européen a travaillé sur plusieurs résolutions relatives à l'IA dans divers domaines :

- **aspects éthiques** [2020/2012(INL)] ;
- **régime de responsabilité civile** (résolution du Parlement européen du 20 octobre 2020) ;
- **droits de propriété intellectuelle** (résolution du Parlement européen du 20 octobre 2020) ;
- **droit pénal** et son **utilisation par les autorités policières et judiciaires** dans les **affaires pénales** [2020/2016(INI)] ;
- **éducation**, **culture** et **audiovisuel** [2020/2017(INI)].

Ces textes contiennent des recommandations à destination de la Commission, mais aussi des **propositions législatives** sur les différents thèmes abordés.

Le 19 février 2020, la Commission publie un Livre blanc sur l'intelligence artificielle qui réaffirme le **double objectif européen** de **promouvoir l'adoption de l'IA** et de tenir compte des **risques associés à certaines utilisations** de cette technologie.

En 2020, le Conseil européen demande à la Commission de "*présenter une **définition claire et objective des systèmes d'intelligence artificielle à haut risque***" [Réunion extraordinaire du Conseil européen (1er et 2 octobre 2020) - conclusions].

Dans des conclusions publiées en octobre 2020, le Conseil de l'UE rappelle que "*si les technologies numériques, y compris l'IA, présentent de plus en plus d'opportunités et d'avantages, leur conception, leur développement, leur déploiement et leur* ***utilisation abusive*** *pourraient aussi* ***comporter des risques pour les droits fondamentaux, la démocratie et l'État de droit***" [Conclusions de la présidence – La charte des droits fondamentaux dans le contexte de l'intelligence artificielle et du changement numérique].

En avril 2021, la Commission présente des mesures sur l'IA :

- une communication sur la promotion d'une approche européenne ;
- un réexamen du plan coordonné en matière d'intelligence artificielle ;
- une proposition de cadre réglementaire sur l'IA et une analyse d'impact.

7 Quelles différences entre le texte sur l'IA de l'UE et la convention-cadre sur l'IA adoptée par le Conseil de l'Europe ?

Le 17 mai 2024, le Conseil de l'Europe a adopté la **convention-cadre sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit**. Il s'agit d'un traité international, adopté par les ministres des affaires étrangères des **46 pays membres du Conseil de l'Europe**.

Cette convention-cadre est le premier texte d'ampleur internationale **juridiquement contraignant dans le domaine de l'IA**. Ses dispositions devront être transposées dans le droit national des États signataires.
Le texte vise à garantir une IA respectueuse des droits fondamentaux, entre autres face aux risques :

- de pratiques discriminatoires ;
- de remise en cause des processus démocratiques ;
- d'atteinte à la vie privée ;
- d'utilisation de l'IA par certains États à des fins répressives.

La convention-cadre établit des **règles relatives au respect des droits fondamentaux** à toutes les étapes du cycle de vie de systèmes d'IA.

La convention-cadre est un traité international incluant les États membres de l'UE, mais aussi onze États non membres, dont les États-Unis, le Japon et le Canada. Ce texte est compatible avec le règlement sur l'IA de l'UE qui concerne plus spécifiquement l'Union européenne.
La convention-cadre permet ainsi d'être un texte juridiquement contraignant pour des signataires hors UE parmi les acteurs majeurs dans la recherche et le développement des systèmes d'IA.

DOCUMENT 8

Article internet ibm.com 22 - Décembre 2023

Qu'est-ce que le biais d'IA ?

Le biais de l'IA, également appelé biais de machine learning ou biais d'algorithme, fait référence à l'apparition de résultats biaisés en raison de préjugés humains qui faussent les données d'entraînement ou l'algorithme d'IA d'origine. Cela entraîne des résultats faussés et des conséquences potentiellement néfastes.

Lorsque les préjugés ne sont pas traités, ils entravent la capacité des individus à participer à l'économie et à la société. Le biais réduit la précision de l'IA, et donc son potentiel.

Les entreprises sont moins susceptibles de bénéficier de systèmes qui produisent des résultats biaisés. Les scandales résultant des biais de l'IA pourraient favoriser la méfiance des personnes de couleur, des femmes, des personnes en situation de handicap, de la communauté LGBTQ ou d'autres groupes marginalisés.

Les modèles sur lesquels reposent les efforts d'IA absorbent les préjugés de la société qui peuvent être discrètement intégrés dans les monceaux de données sur lesquels ils sont entraînés. La collecte de données biaisées par l'histoire et qui reflètent les inégalités sociétales peut nuire aux groupes historiquement marginalisés dans des cas d'utilisation tels que le recrutement, la police, la notation de crédit et bien d'autres. Selon le Wall Street Journal, « alors que l'utilisation de l'intelligence artificielle se généralise, les entreprises ont encore du mal à lutter contre les biais omniprésents ».[1]

Exemples et risques concrets

Lorsque l'IA commet une erreur due à un biais (comme le fait de refuser des opportunités à certains groupes de personnes, de les identifier de manière erronée sur des photos ou de les punir injustement), l'organisation fautive subit des dommages à sa marque et à sa réputation. Dans le même temps, les personnes appartenant à ces groupes et la société dans son ensemble peuvent subir un préjudice sans même s'en rendre compte. Voici quelques exemples très médiatisés de disparités et de biais dans l'IA et des torts qu'ils peuvent causer.

Dans le domaine des soins de santé, la sous-représentation des données relatives aux femmes ou aux groupes minoritaires peut fausser les algorithmes prédictifs de l'IA.[2] Il a ainsi été constaté que les systèmes de diagnostic assisté par ordinateur (CAO) fournissaient des résultats moins précis pour les patients afro-américains que pour les patients blancs.

Si les outils d'IA peuvent rationaliser l'automatisation de la numérisation des CV lors d'une recherche pour aider à identifier les candidats idéaux, les informations demandées et les réponses exclues peuvent entraîner des résultats disproportionnés entre les groupes. Par exemple, si une offre d'emploi utilise le mot « ninja », elle pourrait attirer plus d'hommes que de femmes, même si cela ne constitue en aucun cas une exigence de l'emploi.[3]

Pour tester la génération d'images, Bloomberg a demandé la création de plus de

5 000 images d'IA et a constaté que « le monde selon Stable Diffusion est dirigé par des PDG blancs. Les femmes sont rarement médecins, avocates ou juges. Les hommes de couleur commettent des crimes, tandis que les femmes de couleur servent des hamburgers. »[4] Midjourney a mené une étude similaire sur la génération d'art par l'IA, en demandant des images de personnes exerçant des professions spécialisées. Le résultat montrait à la fois des personnes jeunes et âgées, mais les personnes âgées étaient toujours des hommes, ce qui renforce les préjugés sexistes sur le rôle des femmes sur le lieu de travail.[5]

Les outils de police prédictive basés sur l'IA utilisés par certains organes de la justice pénale sont censés identifier les zones où les actes criminels sont le plus susceptibles de se produire. Cependant, ces outils s'appuient souvent sur des données historiques d'arrestations, qui peuvent renforcer les schémas existants de profilage racial et de ciblage des communautés minoritaires relativement disproportionné.[6]

DOCUMENT 9

Formation pour les agents de la fonction publique sur mentor.gouv.fr

Découvrir les IA génératives

Numérique et système d'information et de communication

Proposée par **Direction interministérielle du numérique**

Produite par **Direction interministérielle du Numérique et Institut national de recherche en sciences et technologies du numérique**
Contact : **campus.numerique@modernisation.gouv.fr**

Vous souhaitez acquérir les bases sur les IA génératives ? Cette formation est faite pour vous !

Plongez dans l'univers des intelligences artificielles génératives et découvrez comment les utiliser efficacement dans votre quotidien professionnel.

- Comprendre l'intérêt et les usages des IAG dans votre travail
- Maîtriser les bonnes pratiques et la rédaction d'instructions efficaces

Voir moins

Session

Accès permanent
Durée en ligne **01h**
Public cible : **Tout public**
Accompagnement : **À réaliser en autonomie, à son rythme et dans l'ordre prédéfini.**
Structure organisatrice : **Dinum et Inria Academy**

Contenu | **Métier(s)** | **Termes de la licence**

L'intelligence artificielle générative est un sujet dont on parle de plus en plus, mais il n'est pas toujours facile de s'y retrouver avec la multitude d'informations disponibles.

Cette formation est l'occasion de découvrir, de manière simple et concrète, comment les IA génératives fonctionnent et comment elles peuvent transformer vos pratiques professionnelles.

Vous apprendrez à comprendre leur potentiel, à identifier leurs usages dans votre travail quotidien, et à adopter les bonnes pratiques pour en tirer le meilleur parti.

En un mot, vous repartirez avec les clés pour maîtriser ces outils innovants et les intégrer efficacement dans vos missions.