

Cantens, Thomas

Working Paper

Comment pensera l'État avec ChatGPT? Les douanes comme illustration de l'intelligence artificielle générative dans les administrations publiques

FERDI Document de travail, No. P330

Provided in Cooperation with:

Fondation pour les études et recherches sur le développement international (FERDI),
Clermont-Ferrand

Suggested Citation: Cantens, Thomas (2023) : Comment pensera l'État avec ChatGPT? Les douanes comme illustration de l'intelligence artificielle générative dans les administrations publiques, FERDI Document de travail, No. P330, Fondation pour les études et recherches sur le développement international (FERDI), Clermont-Ferrand

This Version is available at:

<https://hdl.handle.net/10419/276988>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Comment pensera l'État avec ChatGPT ?

Les douanes comme illustration de l'intelligence artificielle générative dans les administrations publiques*

THOMAS CANTENS

➔ THOMAS CANTENS, Unité Recherche et Politiques, Bureau du Secrétaire général de l'Organisation Mondiale des Douanes (OMD), chercheur associé au Centre d'Études et de Recherches sur le Développement International (CerdI)
Email : thomas.cantens@wcoomd.org

Résumé

Cet article examine les enjeux propres à l'intelligence artificielle générative (IAG) dans les administrations publiques et les effets de sa nature langagière sur les interactions homme-machine. Au niveau organisationnel, les avantages de l'IAG sont la réduction des coûts par l'internalisation des tâches, l'uniformisation et la conformité du langage administratif, l'accès à un vaste corpus de connaissances et, potentiellement, un changement de paradigme dans la détection de la fraude. L'article tente de déconstruire les raisons de la défiance envers l'IAG. Les risques associés à l'IAG dans le secteur public sont souvent similaires à ceux associés à l'IA conventionnelle, mais si certains risques restent pertinents, d'autres le sont moins car la technologie de l'IAG en limite l'explicabilité donc aussi les usages dans les administrations publiques. Au niveau individuel, la nature « langagière » de l'IAG est une rupture profonde, par rapport à la nature « numérique » des systèmes d'IA mis en œuvre dans les administrations publiques jusqu'à présent. L'impact transformateur de l'IAG sur la production intellectuelle de l'État fait naître la crainte de remplacement ou plutôt d'asservissement des fonctionnaires aux machines. L'article plaide pour le développement de la vigilance et de la pensée critique comme compétences spécifiques pour les fonctionnaires qui sont devenus très spécialisés et qui devront penser avec une machine qui, elle, est éclectique par nature.

Mots-clés: intelligence artificielle générative ; administrations publiques ; ChatGPT; douanes.

ORCID ID: <https://orcid.org/0009-0001-9750-7769>

* L'auteur remercie Vianney Dequiedt et Céline de Quatrebarbes pour les échanges sur cet article et leurs relectures. L'auteur remercie également Kunio Mikuriya, Secrétaire Général de l'Organisation Mondiale des Douanes pour son soutien à la recherche dans ce domaine et Mourad Arfaoui et Mariya Polner pour les discussions régulières sur ce sujet.

Bureaucratic administration means fundamentally domination through knowledge. This is the feature of it which makes it specifically rational.

Max Weber, 1978, *Economy and Society*, University of California Press, p. 225

Dans la manufacture et le métier, l'ouvrier se sert de son outil; dans la fabrique il sert la machine. Là le mouvement de l'instrument de travail part de lui; ici il ne fait que le suivre.

Karl Marx, 1867, *Le Capital*, Livre I, chapitre 15.

Une administration-fiction

Avertissement : à ce jour, aucune administration publique n'a indiqué intégrer l'IAG dans ses pratiques de travail.

Niamey, Niger, direction générale des douanes, juin 2025. Etim, en poste aux affaires internationales, allume son ordinateur pour écrire une note de politique générale sur le rôle des douanes dans la mise en œuvre des politiques environnementales. Etim connaît bien son code des douanes et il a une expérience de terrain, mais ses études en économie ne l'ont pas préparé aux questions environnementales. Il a quelques intuitions, et en mémoire la catastrophe du Probo Koala en 2006 qui a affecté ses collègues ivoiriens, mais il craint d'omettre des points essentiels dans sa note au ministre, en préparation de la réunion régionale sur « commerce, douanes et environnement » qui se tiendra à Abuja au Nigeria anglophone voisin. Etim n'est pas inquiet, sur son poste informatique, il dispose de son assistant personnel de recherche, un agent conversationnel d'intelligence artificielle générative (IAG), AdminGPT, qui puise ses informations dans un gigantesque corpus de textes académiques, administratifs, réglementaires, de conventions internationales et de lois nationales, de rapports d'experts, et ce dans toutes les langues. Cette IAG a été installée au Niger l'année dernière, projet conjoint de l'Union Africaine et de la Banque mondiale, d'agences des Nations unies, de l'Organisation mondiale du commerce et de l'Organisation mondiale des douanes. Ses collègues des pays de la sous-région disposent du même accès à cette connaissance vraie, choisie et vérifiée, augmentée de sources de presse reconnues. Etim initie sa conversation avec AdminGPT par une question générale, le rôle des douanes dans les politiques environnementales. La réponse ne le surprend pas, elle est simple et claire, elle balaye le sujet à 360° ; tout au plus n'aurait-il pas intégré dans sa réflexion certains produits comme les pesticides. Son attention se porte sur la Convention de Basel qu'il connaît moins, il poursuit sa discussion avec l'IAG dans cette direction. Le côté fiscal l'intéresse aussi, il y reviendra dans un deuxième embranchement de la conversation. Une semaine plus tard d'intenses discussions avec l'IAG, d'allers-retours sur Google et Google Scholar, de lecture des rapports eux-mêmes et d'articles d'auteurs distingués par l'IAG, de confrontation de ces idées avec ce qu'il connaît des orientations politiques de son gouvernement et des capacités de son administration, Etim a terminé sa note de quelques pages, synthétisant les enjeux politiques, l'avancement des autres pays, discutant des

enjeux techniques pour les douanes nigériennes, et enfin proposant une position de l'administration accompagnée d'options techniques sur le terrain. Etim a appris, sans nul doute, il a mobilisé sa connaissance de terrain pour interroger l'IAG, il a aussi eu le temps de réfléchir, en profondeur, sur des points critiques pour son administration. Le ministre est satisfait, il a reçu des notes similaires de l'administration des impôts et des services environnementaux, toutes les fonctionnaires du pays ayant accès à AdminGPT. Sa perspective sur les questions environnementales est maintenant relativement exhaustive, l'IAG s'est alimenté sur le continent africain mais aussi ailleurs en Amérique du Sud et en Asie. Le ministre demande à ses services de partager ces notes politiques avec les autres pays participant à la réunion régionale. Les fonctionnaires nigériens vont s'activer à les passer au traducteur automatique en langue anglaise intégré à AdminGPT, puis ils confieront le résultat à l'IAG une dernière fois en lui demandant d'en améliorer le style. De leur côté, les collègues nigériens ont procédé identiquement et diffusé leur propre position en anglais, et en français avec leurs pays voisins.

Au même moment, à Niamey, Aboubakar, enquêteur à la brigade nationale de surveillance, travaille sur un trafic de pièces détachées automobiles en provenance d'Europe. Il a rassemblé une masse de documents, articles de presse, rapports d'enquête, documents d'expert, factures commerciales, tarifs douaniers des pays voisins, données de trafic maritime, rapports de police, et déclarations en douane. Bertrand n'utilise pas le même agent d'IAG qu'Etim, le sien est national et isolé, partagé uniquement avec les autres services de sécurité intérieure. Son profil l'autorise à sélectionner des sources de données extérieures à la douane, ses requêtes sont tracées et lorsque les textes collectés contiennent des données nominatives comme des procès-verbaux par exemple, ils sont stockés de façon temporaire pour les besoins de l'enquête. Son rapport final une fois rédigé fermera immédiatement son accès à des informations non douanières. L'IAG digère son corpus de textes et lui renvoie un graphe d'acteurs, des importateurs et des revendeurs nigériens mais aussi européens et asiatiques, ces acteurs sont reliés entre eux et à des objets, des conteneurs, des navires, et les liens du graphe sont des actions, des relations juridiques, des achats, des convoyages. Aboubakar commence son enquête, il observe le graphe, le manipule en trois dimensions et cherche des liens suspects ; plus tard, il reviendra solliciter l'IAG lorsque ses requêtes auprès des entreprises concernées lui auront fourni de nouveaux documents.

Dans son bureau de douane à Agadez au nord du pays, Aminata, inspectrice vérificatrice, a un doute sur le classement tarifaire des bridges dentaires, un produit relativement rare au passage de cette frontière désertique. La description de la marchandise sur la déclaration en douane est très technique, Aminata interroge AdminGPT et lui demande assistance. Les résultats de l'IAG ne sont pas toujours fiables, mais l'IAG la guide dans sa consultation de son épais tarif papier, à elle de faire son travail de douanière.

Retour à Niamey, Ibrahim et Kimba partagent les mêmes locaux de la direction de l'informatique, mais le premier y est chef de projet informatique et le second analyste de données. Ibrahim conduit le basculement du système de dédouanement vers un nouveau logiciel en fin d'année. Ibrahim compte lui aussi avec AdminGPT, il lui a demandé les méthodes de gestion les plus appropriées compte tenu du temps imparti, du caractère très critique du système pour les recettes et de l'implication d'experts internationaux, et quels

seraient les indicateurs de suivi et de réussite. Lorsque le projet prendra du retard, l'IAG lui conseillera des techniques de compression de temps, et, à la fin du projet, l'IAG lui indiquera, pas-à-pas au fil de ses questions, comment tirer les leçons de l'exécution du projet. Kimba est douanier mais il a acquis une robuste connaissance en science des données grâce à un projet régional de formation. Le directeur général lui a confié une tâche un peu confidentielle : identifier les « mauvaises pratiques » via le système de dédouanement, ces pratiques qui sont des niches de corruption. Kimba commence par consulter AdminGPT, il lui demande d'aller chercher les données historiques des déclarations en douane dans le système de dédouanement SYDONIA, tout y est tracé, et il explique ce qu'il cherche, des « patterns », des groupes d'inspecteurs aux pratiques similaires. Kimba connaît bien le terrain, il saura interpréter les résultats de la machine. Il demande à l'IAG de lui écrire du code en R pour analyser les données, puis de les représenter dans un graphique de *clusterisation* des inspecteurs. L'IAG lui répond, Kimba parcourt le code et le juge imparfait, l'IAG a ignoré un paramètre, le temps mis à dédouaner une marchandise qui peut être un signe de corruption ; il lui demande de refaire son code en tenant compte de celui-ci et de lui proposer différentes modélisations. Le code et les choix méthodologiques générés par la machine semblent maintenant plus pertinents à Kimba qui a gagné du temps. Ses missions auprès du directeur général sont si diverses qu'il ne peut se rappeler de tous les modèles possibles, ni surtout des instructions R pour les mettre en œuvre, sans compter les erreurs de frappe qui font bugger le programme. Kimba exécute lui-même le code de l'IAG, il tient à conserver la main sur cette étape, garder intact le plaisir du moment où du code s'exécute sur un ordinateur et se découvre le résultat. Expérimenté et soucieux d'envisager tous les scénarios, Kimba changera quelques paramètres encore, lui-même, pour ajuster ses modèles, puis il écrira son rapport détaillé et argumenté, AdminGPT ne l'ayant pas encore convaincu de sa pertinence sur ce dernier moment de l'analyse.

Mme B. est importatrice de bazins « riches », elle importe au Togo, au Mali mais c'est sa première fois au Niger, c'est une « grande commerçante » comme on dit, alors elle se rend sur le site Web des douanes nigériennes pour poser ses questions à ChapChapImport, leur nouveau robot, une IAG qui se fonde sur le corpus des lois nigériennes et des notes de la douane nigérienne. Ses réponses sont précises, elles n'engagent pas l'administration légalement mais Mme B. apprécie de pouvoir converser de façon fluide avec le robot. Sait-elle seulement qu'il s'agit d'un robot ? Peu lui importe au final, elle le trouve aimable, précis et elle a ses renseignements à toute heure. Une adresse de courriel est toujours à sa disposition au cas où elle voudrait obtenir un engagement légal de la part de l'administration.

M. X est un élégant fraudeur, fin lecteur des revues scientifiques, il ne veut pas dissimuler des cartons dans le double-fond d'un conteneur, il mise sur les données, et, derrière son ordinateur, il est en train de parier contre l'intelligence artificielle. Lui aussi est entré en conversation avec ChapChapImport, et il le teste. Pendant quelques heures, M. X lui fournit des descriptions complexes de sa marchandise, des blousons, pour voir ce que le patient ChapChapImport lui renvoie comme position tarifaire. Pour ses blousons, il vise au final le 61.05, la position des chemises qui payent des droits de douane moins élevés que les blousons. M. X. tord sa description commerciale dans tous les sens, change quelques mots, jusqu'à ce que le robot doute et propose plusieurs positions, toutes avantageuses par

rapport à celle des blousons. Va pour cette description qui créera le trouble, en espérant que le douanier n'aura pas la mauvaise idée d'affronter la poussière et le soleil pour ouvrir son conteneur. Au moment d'enregistrer sa déclaration dans le système des douanes nigériennes, M. X pense à Mallarmé et son coup de dé, un coup d'intelligence artificielle jamais n'abolira le hasard.

Introduction

Toute ressemblance avec des faits existants n'est pas fortuite. Le caractère purement fictionnel de cette vision futuriste d'une administration ayant intégré l'IAG est réduit à un seul point : le financement d'un agent conversationnel d'IAG par les bailleurs. Le reste, les problématiques et les usages de l'IAG existent déjà : remplacer les *chatbots* des administrations¹, dépourvus de capacité d'apprentissage et n'ayant qu'une interprétation limitée du langage humain par règles et mots-clefs² (Conseil de l'Europe 2023) ; assister à la rédaction, au résumé³, à la correction de l'orthographe et de la syntaxe, à l'amélioration du style⁴ (Huang et Tan, 2023 ; Yang et al., 2023 ; Xames et Shefa, 2023), voire à la prise de notes pendant une visioconférence⁵ ; assister la recherche préliminaire, dans la production d'une synthèse des idées et des enjeux sur un domaine, dans la phase préliminaire de l'instruction juridique d'un dossier, pour la recherche universitaire, les professions juridiques et de conseil et le journalisme (Ophal et al. 2023) ; assister à l'analyse de données, choisir les données, écrire le code informatique, réaliser des graphiques et analyser des faits saillants (Cheng et al. 2023) ; assister à la gestion de projets informatiques dans l'administration (Minelle et Stolfi 2023) ; assister au recueil de la preuve digitale (Rodriguez et al. 2022, Henseler and van Beek, 2023) et représenter des corpus de documents sous forme de graphes à des fins de lutte contre la fraude (Graham et al. 2023). D'autres exemples d'usages non liés à de la production textuelle sont fournis par le Conseil de l'Europe (2023) (sans toutefois partager de référence à des expérimentations existantes) : la sélection

¹ Les Chatbots existent depuis longtemps dans les administrations, y compris les douanes, pour le BREXIT (<https://www.thebotforge.io/case-studies/brexit-customs-information-chatbot/>), les douanes finlandaises (<https://tulli.fi/en/about-us/contact-information/information-for-private-customers/customs-chatbot-hippu>), les douanes indiennes (<https://www.icegate.gov.in/>), les douanes américaines (<https://www.cbp.gov/employee-resources/family/employee-assistance-program/enhanced-services/tess-faqs>) ; de même pour les administrations fiscales <https://www.ciat.org/ciatblog-podra-utilizarse-chat-gpt-en-las-administraciones-tributarias/?lang=en>. Toutefois, leurs capacités semblent limitées à des domaines spécifiques tels que la fourniture d'informations générales sur le dédouanement.

² Le gouvernement français prévoit l'utilisation de ChatGPT pour assister les fonctionnaires chargés de répondre aux usagers. <https://www.gouvernement.fr/upload/media/content/0001/05/0a63326de41e9a36d7878966030e3ce55e98bddf.pdf>

³ Voir Babl and Banl (2023) en science médicale.

⁴ Une certaine vigilance s'applique ici à l'amélioration d'un texte. Dans un texte relatif à ce sujet et traitant des critiques contre l'IA, quelques phrases finales ont été ajoutées au texte lui-même, par le GAI, fournissant une perspective positive sur l'avenir des solutions qui s'appliqueront aux problèmes actuels. Ces phrases n'étaient pas en contradiction avec le sens du texte, mais il était surprenant de voir que les dernières phrases de la section étaient un peu plus optimistes que dans le texte initial...

⁵ <https://techmonitor.ai/technology/ai-and-automation/microsoft-to-integrate-chatgpt-into-teams>

automatique de CV, l'examen de contrats, l'éligibilité à des services sociaux, l'assistance aux services de comptabilité.

Depuis le succès de ChatGPT, apparemment unique parmi les applications mises en ligne ces dernières années⁶, de nombreux professionnels s'interrogent sur le renouveau de leurs pratiques : juristes⁷, journalistes⁸, professions médicales⁹, professions éducatives¹⁰, enseignants-chercheurs¹¹, financiers¹², analystes de données¹³, designers¹⁴.

Les fonctionnaires sont pour l'instant les invisibles de cette réflexion publique sur l'intégration de l'IA dans les pratiques professionnelles, sinon pour des positions défensives : lutter contre les usages malveillants de l'IA¹⁵, établir des règles internes d'usage qui relèvent plus d'un rappel au bon sens¹⁶, ou définir les risques de l'IA et leurs paradigmes dans les administrations (Council of European Union, 2023).

Dans ce contexte, au vu de l'administration-fiction précédente, ce papier se questionne donc sur le déploiement de l'IA dans les administrations : quels bénéfices-risques et limites propres à l'IA, et pour quels effets sur le rapport fonctionnaire – machine, compte tenu du caractère langagier, textuel, spécifique à l'IA par rapport aux autres formes d'IA exclusivement numériques installées jusqu'à présent dans les administrations publiques ?

Une première section analyse les bénéfices spécifiques attendus de l'IA. Une deuxième section tente de déconstruire une partie des motifs de défiance à l'égard de l'IA,

⁶ ChatGPT est supposé avoir atteint 100 million d'utilisateurs après 1 mois, alors qu'il a fallu 9 mois à TikTok qui détenait le record de progression (<https://www.businessinsider.com/chatgpt-may-be-fastest-growing-app-in-history-ubs-study-2023-2?r=US&IR=T>). Ces nombres sont à prendre avec précaution, aucun chiffre officiel n'a été confirmé par OpenAI.

⁷ Pierce and Goutos (2023) et <https://www.reuters.com/legal/legal-ai-race-draws-more-investors-law-firms-line-up-2023-04-26/>

⁸ Ophal et al. 2023.

⁹ Exemple, Wagner et Ertl-Wagner (2023).

¹⁰ Les références récentes à l'usage de l'IA en milieu éducatif sont parmi les plus nombreuses. Voir <https://blogs.worldbank.org/education/how-use-chatgpt-support-teachers-good-bad-and-ugly> Voir aussi par exemple Benteshi et al. (2023) pour l'évaluation automatique des examens, incluant l'IA.

¹¹ ChatGPT a été cité comme co-auteur dans certains documents de recherche, voir <https://www.nature.com/articles/d41586-023-00107-z>

¹² Chen, Z., Zheng, L. N., Lu, C., Yuan, J., & Zhu, D. (2023).

¹³ Cheng et al., 2023.

¹⁴ https://www.lemonde.fr/pixels/article/2023/06/15/comment-les-metiers-creatifs-de-l-image-adoptent-la-generation-par-intelligence-artificielle_6177766_4408996.html.

¹⁵ A priori, ces usages malveillants sont les risques de fraude à l'identité pour du phishing, des fraudes en ligne, la génération et la diffusion de contenus à but politique, de désinformation ou de terrorisme, la création automatique de code malveillant (Europol, 2023). INTERPOL a publié un *AI Toolkit* diffusant les bonnes pratiques pour intégrer l'IA dans les administrations policières, mais, publié en juin 2023, il ne tient pas compte de l'IA, décrite comme "science fiction" (voir Document 6, p. 11, <https://www.interpol.int/How-we-work/Innovation/Artificial-Intelligence-Toolkit>).

¹⁶ La Commission européenne a diffusé une note interne pour l'usage de ChatGPT, essentiellement prévenant les fonctionnaires des risques d'erreur et leur déconseillant de l'utiliser dans des situations urgentes et pour des dossiers sensibles. L'auteur a eu accès à cette note qui n'a pas été rendue publique mais uniquement annoncée <https://www.euractiv.com/section/artificial-intelligence/news/eu-commission-issues-internal-guidelines-on-chatgpt-generative-ai/>

en montrant notamment que les risques attribués à l'IA sont ceux précédemment attribués à l'IA et que si certains risques demeurent pertinents, d'autres le sont moins du fait des usages de l'IA eux-mêmes contraints par les limites inhérentes à la technologie utilisée. La troisième section du papier explore ce que penser avec un outil au caractère langagier signifierait et changerait dans la place du fonctionnaire au cœur de la production intellectuelle de l'État, au même titre que l'introduction des machines dans les usines au 19^e siècle a chamboulé les pratiques de la production matérielle. De nombreux exemples développés dans le document, comme la fiction administrative, sont basés sur les besoins identifiés dans les administrations douanières par l'auteur au cours de plusieurs années d'expérience sur le terrain et de recherche dans les États membres de l'Organisation mondiale des douanes (OMD). En raison de la diversité des missions douanières, la plupart des besoins discutés dans ce document et des contraintes associées aux valeurs du service public sont communs à de nombreuses administrations publiques, si ce n'est à toutes.

Quels bénéfices spécifiques à l'IA ?

Bien qu'elle soit qualifiée d'IA, comme nous le pressentons chacun à son contact via ChatGPT, il y a quelque chose qui se passe de nouveau dans l'IA. Elle est capable de créer du texte, des images, des vidéos, du son, à partir de requêtes formulées dans les langues courantes, mais, bien entendu, elle ne pense, ni ne lit, ni n'écrit. Cette évolution n'est pas technique¹⁷, elle repose sur les réseaux de neurones et les modèles de langage¹⁸ appliqués à de très larges corpus de textes¹⁹ et qui existent depuis plus d'une vingtaine d'années. En revanche, son caractère général et son accessibilité impressionnent. Ses modèles sont dits "*general-purpose*" ou "*foundation models*" pour signifier qu'ils n'ont pas d'objectif spécifique. L'IA peut, au moins en première instance d'une réflexion, assurer une partie des fonctions humaines que sont le résumé et l'amélioration de textes, la synthèse de connaissances et la création d'images et de textes littéraires et d'images. Enfin, à la différence de l'IA spécialisée, l'IA ne requiert pas de compétence en informatique ou en statistique pour interagir avec elle. Il n'est pas non plus nécessaire de l'intégrer dans des processus technologiques complexes comme c'est le cas des IA numériques, ce qui constitue souvent un obstacle à leur déploiement hors laboratoire, le fameux « passage à l'échelle »²⁰ (Davenport and

¹⁷ <https://www.zdnet.com/article/chatgpt-is-not-particularly-innovative-and-nothing-revolutionary-says-metas-chief-ai-scientist/>

¹⁸ Les modèles de langage analysent les documents et calculent les probabilités d'association entre les mots (pour plus d'informations sur le fonctionnement de ChatGPT, voir <https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286>).

¹⁹ Le corpus d'entraînement de ChatGPT est encore confidentiel, mais il est estimé à plusieurs centaines de gigaoctets de documents (environ 600 Go), fournis par WebText (un corpus standard de pages web) augmenté de pages Wikipédia, d'articles académiques, de blogs. Ce corpus représente probablement plus de 500 milliards de mots (<https://scoms.hypotheses.org/1059>)

²⁰ Dans les administrations douanières, au cours des dernières années, les experts de l'OMD ont observé que ce passage à l'échelle, le déploiement des IA en situation réelle, était particulièrement complexe pour les administrations qui disposent déjà de systèmes informatiques complexes.

Ronanki 2018). De ces spécificités et des usages potentiels vus précédemment, il est possible d'envisager les bénéfices de l'IAG pour les administrations publiques.

Réduction des coûts

En premier lieu, la généralisation de l'assistance éditoriale dans les administrations et les organisations gouvernementales internationales entraînerait l'internalisation d'une partie des travaux qu'elles font réaliser à l'extérieur, en raison des coûts moindres de production via l'IAG. Sont directement concernées l'édition, la traduction (Jiao et al. 2023) et la création de contenus aux fins de communication et formation. De même, les systèmes de traduction automatique deviennent très performants avec des technologies proches de celle de l'IAG et seraient complétés par l'assistance stylistique de l'IAG pour réduire les coûts²¹. Les gains de productivité et financiers semblent évidents pour les organisations internationales dont les productions textuelles sont multilingues, écrites par des experts non natifs des langues de l'organisation.

S'agissant de l'analyse de données, le coût de GPT4 se situe entre 0.45% et 0.70% de celui d'un analyste de données suivant son expérience (Cheng et al. 2023). Pour tirer bénéfice de l'IAG dans ce domaine, les administrations devront compter sur des fonctionnaires qui disposent de suffisamment de connaissance statistique pour formuler des requêtes ayant un sens mathématique et valider les réponses de l'IAG. L'IAG constituerait ainsi une interface de métalangage pour l'analyse de données. Cela suppose une évolution des administrations dans leurs stratégies data. Jusqu'à présent, leur souci était de recruter – et garder – des scientifiques des données. La rentabilité de l'IAG devrait inciter les administrations à recruter ou former des fonctionnaires disposant d'une culture quantitative, mathématique, statistique de base. A très court terme, au vu des limitations de l'IAG sur l'analyse de données dues à sa faiblesse calculatoire²², ces fonctionnaires profiteraient déjà de l'IAG comme « tuteur » à la création de code informatique pour l'analyse de données²³.

Uniformité et correction de la langue administrative

Le deuxième bénéfice est un gain en qualité et uniformité de style, deux paramètres des langues administratives très formatées. Nous pourrions reprocher à l'IAG, d'être du niveau d'un fonctionnaire "moyen", d'adopter une langue standardisée moins diversifiée

²¹ La traduction du français vers l'anglais pour un document de 10 000 mots est d'environ 1 500 €. Ceci est à comparer avec un abonnement de 7,49€/mois pour DeepL et de 20€/mois pour ChatGPT, sans limitation de taille de texte.

²² Cheng et al. 2023 ont testé l'analyse de données sur GPT3, les résultats de GPT4 aux examens mathématiques et statistiques fournis par OpenAI mais non confirmés de façon indépendante rendent probablement déjà obsolète cette limite (OpenAI, 2023).

²³ Ceci s'insère déjà particulièrement dans la politique de l'OMD qui fournit des formations en ligne et en présentiel à des douaniers non spécialistes pour leur faire acquérir les bases du métier d'analyse de données. Le programme OMD-Corée BACUDA a délivré plus de 2000 certificats de formation en ligne aux outils de l'analyse de données et forme en présentiel une douzaine de douaniers par an sur des cursus longs de science des données. Un programme similaire, OMD-France, DATAFID, est en cours pour les administrations africaines francophones. Les douaniers certifiés dans ces programmes bénéficieraient d'ores et déjà de l'appui de l'IAG en matière de code pour la science des données.

que celle utilisée par les humains (Pu et Demberg 2023), voire avec des tics de langage comme la façon qu'a ChatGPT de résumer sa réponse à la fin de celle-ci. Nous pourrions même craindre que l'utilisation généralisée d'agents conversationnels comme assistants de rédaction appauvrisse le langage voire standardise la réflexion et l'écriture, à la façon dont nous critiquons la "pensée powerpoint"²⁴. Si ces craintes sont légitimes, il faut les ramener à l'état actuel de la production intellectuelle dans les administrations dans le monde. Dire que l'agent conversationnel produit un résultat « moyen » signifie aussi qu'il est déjà meilleur que de nombreux agents humains. (Herbold et al. 2023). Cela a été (malheureusement) expérimenté par l'auteur (francophone) de ce papier, en langue anglaise. Une note interne a été traduite intégralement du français vers l'anglais par un traducteur automatique performant, donc avec une qualité de style très correcte et surtout continue sur tout le document, puis une partie de la note a été améliorée, linguistiquement, avec ChatGPT. Le défi a été présenté aux lecteurs de trouver laquelle. Une collègue de l'auteur a rapidement deviné quelle partie n'avait pas bénéficié de l'amélioration par ChatGPT, du fait d'un style de moindre qualité et de quelques erreurs grammaticales. Au sein des organisations internationales techniques, ou des directions générales des administrations, dont les activités principales sont la production textuelle (conventions, guides, rapports, notes, etc...), l'usage des agents conversationnels hisserait tous les fonctionnaires à un standard d'écriture de qualité minimale commune, sans faute de langue, ni faute logique.

Accès à une connaissance large

Le troisième bénéfice est un gain en substance. Les capacités d'assistant de recherche de ChatGPT sont plutôt bonnes. Elles le sont d'autant plus lorsque l'IAG est interrogée sur des sujets où le fonctionnaire n'est pas spécialiste et ne dispose que de quelques intuitions préliminaires, comme cela est souvent le cas lorsque les administrations sont confrontées à de nouveaux sujets. Là où l'on attend du fonctionnaire qu'il fasse une synthèse des idées ou situations existantes, l'IAG sera plus rapide et plus complète dans les connaissances collectées, donc plus performante.

De ce point de vue, nous n'avons pas été la seule victime de ChatGPT. Il a été demandé à ChatGPT de rédiger une quinzaine de lignes pour la plaquette d'un master dans une discipline scientifique appliquée à un milieu professionnel et de développer un cursus. La réponse a été jugée très satisfaisante par l'auteur et un autre expert et envoyée à un spécialiste de la discipline dont la tâche était de l'améliorer. Malgré les qualités scientifiques reconnues de ce spécialiste dans sa discipline, sa présentation finale était d'une qualité moindre que celle de ChatGPT : le style de la langue était inférieur, et son texte comportait des inexactitudes simples à relever pour les professionnels destinataires de la plaquette du master.

En disposant d'un accès à une connaissance large et d'une capacité à produire des synthèses, les agents d'IAG sont particulièrement performants pour assister les

²⁴ <https://www.theguardian.com/commentisfree/2015/sep/23/powerpoint-thought-students-bullet-points-information>

fonctionnaires en poste dans les directions générales ou les cabinets ministériels en charge de la rédaction de notes politiques.

Dépasser l'IA numérique en matière de lutte contre la fraude ?

L'IA pourrait bien renverser les paradigmes actuels dans les cas où l'humain doit analyser du texte pour détecter de la fraude. Ce renversement est le suivant : alors que l'IA spécialisée détecte de la fraude à partir d'un entraînement sur des informations du passé, qui comportent fraudes détectées et des fraudes non détectées, l'IA pourrait être entraînée sur une information exclusivement « vraie », au sens administratif du terme, constituée de textes ayant force de loi et aider le fonctionnaire à détecter la fraude en lui fournissant l'information juste.

Pour illustrer ce renversement possible, nous développons le cas simple du classement tarifaire en douane évoqué précédemment : le douanier classe les marchandises dans une nomenclature internationale, le système harmonisé, en confrontant la description textuelle des marchandises énoncées sur les déclarations en douane à des textes réglementaires²⁵. Nous avons testé la version publique de ChatGPT contre un algorithme spécialisé de recommandation de classement tarifaire BACUDA²⁶. L'algorithme BACUDA est une aide au douanier vérificateur, il se fonde sur les réseaux neuronaux et les technologies de traitement du langage naturel pour fournir des probabilités de classement associées à une description de marchandise. La différence fondamentale entre cet algorithme et ChatGPT est qu'il est construit pour cet objectif et entraîné sur un corpus de déclarations en douane fourni par l'administration. La version publique de ChatGPT a été testée sur 4 produits, choisis parce qu'ils n'avaient pas été correctement classés par l'algorithme BACUDA²⁷. ChatGPT a réussi à en classer trois sur les quatre²⁸.

Quelles leçons tirer de cette (brève) expérience pour conduire une exploration plus approfondie de cette fonctionnalité ? L'IA semble plus prometteuse car non sensible aux erreurs de l'administration. En effet, l'algorithme BACUDA repose sur la qualité des données fournies, les déclarations en douane, qui ne fait que refléter la qualité des classements tarifaires de l'administration. Si les douaniers se trompent régulièrement, alors l'algorithme renverra des réponses fausses. C'est un point commun à la plupart sinon tous les algorithmes actuellement utilisés pour la lutte contre la fraude : ils n'ajoutent pas d'information exogène à l'univers d'information dans lequel ils sont déployés, ils sont donc influencés par de l'information incorrecte, qu'il s'agisse d'un mauvais classement tarifaire accepté par l'administration ou d'une déclaration frauduleuse non détectée (Mikuriya et Cantens 2020). En revanche, dans notre cas, ChatGPT présente l'avantage d'être entraînée sur de l'information exogène à l'administration, dans notre cas un corpus, certes inconnu

²⁵ L'idée d'utiliser ChatGPT pour faire de l'*identity matching* - reconnaître que deux descriptions textuelles renvoient au même objet - constitue un domaine d'application à part entière où l'IA est confrontée à d'autres types d'IA (Peeters et Bizer, 2023).

²⁶ <https://www.wcoomd.org/en/media/newsroom/2022/march/wco-bacuda-experts-develop-a-neural-network-model-to-assist-classification-of-goods-in-hs.aspx>

²⁷ Échanges internes à l'OMD (Avril 2023).

²⁸ Le classement du chocolat blanc est demeuré erroné. En effet, par définition un modèle de langage s'appuie sur des mots et le chocolat blanc n'est pas, au sens du système harmonisé, du chocolat.

mais certainement pas de déclarations en douane. L'IAG pourrait ainsi en terminer avec un reproche régulier fait aux algorithmes de lutte contre la fraude, à savoir qu'ils vont vite, qu'ils font gagner du temps mais qu'à eux seuls, ils ne peuvent pas détecter plus de fraude que ce que des fonctionnaires ont déjà détecté dans le passé (*idem*). Paradoxalement, l'IA spécialisée en détection de la fraude, comme l'algorithme de BACUDA, devrait alors être utilisée à contre-emploi et en complément de l'IAG : si l'IAG fournit les réponses justes, alors les erreurs de l'algorithme spécialisé, entraînée de l'information passée, pointent sur les fraudes et les erreurs habituelles de l'administration.

L'enjeu des corpus d'entraînement

Ces bénéfices peuvent être maximisés au niveau national, régional ou global. Plus que la technologie de l'agent conversationnel, le corpus de textes d'entraînement est le choix stratégique du moment pour les administrations. Comment l'administration peut-elle construire un corpus d'entraînement qui lui soit propre, réponde à ses questionnements de telle sorte que les réponses des agents d'IAG qui sont entraînés sur ce corpus soient les plus précises possibles ? La possibilité de construire ce type de corpus en fonction des besoins est déjà une réalité. Il s'agit du processus de *finetuning*²⁹ qui est proposé par les compagnies d'IAG pour adapter le modèle aux besoins de l'utilisateur. Des développements commerciaux plus complets sont déjà disponibles dans le domaine juridique³⁰, mais aussi dans des compagnies privées qui mettent des assistants d'écriture à disposition de leur personnel³¹. Dans notre dernier cas de classification tarifaire, un corpus d'entraînement compilerait tous les textes réglementaires propres au système harmonisé, la classification elle-même, ses notes explicatives, mais aussi les décisions précédentes prises par l'administration, les délibérations sur le SH à l'OMD. Dans une vision plus générale d'assistance à l'analyse politique, le corpus d'entraînement inclurait les documents sur les thèmes stratégiques ou innovants du moment pour l'administration et les textes qui font le cœur de l'activité de l'administration (les textes officiels publics, les notes techniques, et les textes non publics produits par l'administration elle-même ou l'État en général).

Les bénéfices devraient également être envisagés globalement. L'usage généralisé de l'IAG dans les administrations, en particulier dans les pays à faible revenu, devrait contribuer à plus d'égalité et à lisser les disparités de productions intellectuelles entre les états. Tous les fonctionnaires, y compris non anglophones, ont accès à un même corpus de connaissances, indépendamment des langues de production de celles-ci, à une même capacité de les synthétiser sur des questions qui leur sont communes et à une même

²⁹ OpenAI a déjà cette offre en ligne de raffiner son propre agent conversationnel, ce qui a bien entendu un coût <https://platform.openai.com/docs/guides/fine-tuning>

³⁰ Voir <https://www.reuters.com/technology/bar-exam-score-shows-ai-can-keep-up-with-human-lawyers-researchers-say-2023-03-15/> Voir Harvey pour les avocats <https://www.sequoiacap.com/article/partnering-with-harvey-putting-llms-to-work/> et des projets en cours <https://www.pwc.com/gx/en/news-room/press-releases/2023/pwc-announces-strategic-alliance-with-harvey-positioning-pwcs-legal-business-solutions-at-the-forefront-of-legal-generative-ai.html>

³¹ <https://asia.nikkei.com/Business/Companies/Panasonic-unit-deploys-ChatGPT-style-AI-to-improve-productivity>

capacité de produire des analyses intelligibles par tous. Ceci devrait faciliter la diffusion globale des idées et analyses produites dans les administrations, en particulier, celles aux capacités financières plus limitées.

Enfin, pour les fonctionnaires non anglophones, l'IAG leur permet de retrouver l'agilité de penser et écrire dans leur langue maternelle et de diffuser leurs analyses au niveau global dans une langue anglaise plus riche, contribuant ainsi à la fois à la préservation des langues et, souhaitons-le, à l'affaiblissement du *globish* et l'appauvrissement de la réflexion fondée sur la circulation de mots valises qu'il véhicule.

S'il est difficile de prévoir l'avenir du modèle commercial de l'IAG, nous pouvons d'ores et déjà supposer que l'engouement actuel l'orientera vers des usages gratuits ou à bas coût, comme les moteurs de recherche. Il est également tout à fait probable que l'IAG suive des développements open source, comme cela est le cas des outils de la science des données. Même si les disparités demeurent en termes de formation ou de développements spécifiques, les administrations des pays à faible revenu progressent par *leapfrog*³², absorbant rapidement les innovations technologiques qui leur sont utiles. C'est de ce point de vue aussi qu'il faut envisager la diffusion de l'IAG dans les administrations.

Dans ce contexte, les organisations internationales pourraient se positionner sur l'IAG à des fins de développement économique et de dialogue politique global : construire des corpus d'entraînement «de confiance» constitués de rapports d'institutions internationales, d'articles universitaires, de conventions internationales, de normes et standards adoptés, de médias de référence ; mettre à disposition des agents conversationnels pour les acteurs du développement, incluant notamment les agences bilatérales d'aide, les administrations des pays bénéficiaires d'aide publique au développement, les centres de recherche ; développer des projets régionaux ou nationaux d'IAG pour les états. Certaines organisations à vocation plus technique, telles que l'OMD ou INTERPOL, pourraient mettre à disposition de leurs états membres des agents conversationnels entraînés sur des usages spécifiques de lutte contre la fraude par exemple. Jusqu'à présent, les organisations internationales ont favorisé le dialogue politique global et la dissémination de normes en mettant en ligne leurs documents, puis leurs données, il s'agit maintenant de mettre à disposition les outils pour les exploiter plus simplement. Outre la sanctuarisation de la connaissance établie dans le domaine du développement, de telles initiatives préviendraient une nouvelle fracture digitale qui ne manquera pas de s'établir lorsqu'émergera le marché d'IAG pour les services publics.

Au niveau national, l'enjeu des corpus d'entraînement est stratégique pour les administrations, en termes d'investissements, plus que les algorithmes. Construire son corpus d'entraînement pour ses besoins va requérir des ressources humaines et financières

³² Le projet DATAFID de sciences des données pour les douanes africaines, mené conjointement par l'IHEDD, Expertise France et l'OMD, a mobilisé les mêmes outils que ceux utilisés par les douanes des pays à haut revenu, et formé des experts dont un poursuit une formation hautement sélective et spécialisée dans une école de statistiques française. Par ailleurs, le projet a introduit des formations en geodata et système d'information géographique qui ont suscité un vif intérêt dans les services d'une douane européenne.

non négligeables. Cela est nouveau : jusqu'à présent le corpus d'entraînement était plus ou moins « gratuit » car constitué de données extraites de bases de données et l'effort portait plutôt sur l'algorithme. Au niveau global, les corpus d'entraînement sont tout aussi stratégiques, moins au niveau financier qu'au niveau politique voire idéologique, car leur construction imposerait de choisir les contenus à y incorporer, donc le choix des connaissances considérées comme « vraies ». Dans une perspective purement géopolitique, il pourrait même être envisagé que ces corpus d'entraînement constituent de nouveaux espaces d'influence politique.

Surmonter la défiance : distinguer les risques et les limites

Il nous semble que s'installe un malentendu entre ce que peut l'IAG actuelle, les risques qu'elle fait peser et les risques sur lesquels on réfléchit dans les administrations publiques. De nombreux risques sont peu pertinents du fait des limites actuelles de l'IAG et des usages que nous en ferions.

Quelques malentendus sur les risques de l'IAG dans les administrations

En juin 2023, le secteur de l'IAG est dans un « brouillard de guerre ». Quelques compagnies mettent leurs agents d'IAG en ligne mais demeurent peu transparentes sur leurs détails techniques et plutôt obscures sur leurs stratégies commerciales³³. Quant aux gouvernements, leur action est défensive. Ils semblent avoir été surpris et commencent seulement à orienter leurs réflexions stratégiques et réglementaires sur l'IA vers l'IAG, à la poursuite d'une réglementation idéale et craignant de décourager les grands opérateurs privés³⁴, sans pouvoir imaginer tous les usages de l'IAG. L'Union Européenne (UE) débute ses travaux pour un *AI Act* en 2021, mais les spécificités de l'IAG ne sont prises en compte qu'in extremis dans les amendements du Parlement européen en mai 2023 avant une adoption définitive de l'Acte prévue fin 2023³⁵. En mai 2023, le gouvernement américain lance un groupe de réflexion sur les limites et les risques associés à l'IAG³⁶. Toujours en mai 2023, le gouvernement chinois publie une version provisoire des mesures pour réguler l'IAG, après

³³ Ces compagnies, dont OpenAI, ne diffusent que très peu de détails techniques et financiers sur la taille de leur corpus d'entraînement, le coût de l'entraînement et des opérations des agents d'IAG en ligne, ou sur leur stratégie à moyen terme. OpenAI propose une nouvelle version de GPT4 dont des tests sont publiés (OpenAI, 2023), une version payante de ChatGPT pour un accès privilégié, mais les perspectives techniques et commerciales demeurent floues. Le secteur se trouve dans un « brouillard de guerre » entretenu par les grandes compagnies, ce qui dénote peut-être l'importance du tournant technologique qui est en train de se jouer avec l'IAG.

³⁴ Par exemple l'IAG de Google n'est pas accessible en Europe à la date de cet article. (<https://support.google.com/bard/answer/13575153?hl=en>) and discussions are ongoing (<https://www.euronews.com/my-europe/2023/05/25/we-cant-afford-to-wait-brussels-and-google-pitch-voluntary-ai-pact-to-fill-legislative-gap>).

³⁵ Voir <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence> D'une manière générale, les amendements ont consisté à accoler l'IAG ("foundation models") aux IA "high risks" (sans classer l'IAG dans cette catégorie) sur de nombreux articles et de demander des mesures spécifiques de transparence sur l'identification des productions de l'IAG, les techniques et tailles des algorithmes utilisés et les corpus d'entraînement, notamment lorsqu'ils comprennent des textes à droits d'auteur (Parlement européen 2023).

³⁶ <https://www.whitehouse.gov/pcast/briefing-room/2023/05/13/pcast-working-group-on-generative-ai-invites-public-input/>

que Ali Baba et Huawei ont mis en ligne leurs agents IAG³⁷. Le gouvernement islandais coopère avec OpenAI pour mieux intégrer la langue islandaise dans ChatGPT, constatant l'inefficacité de l'agent dans cette langue³⁸. Quant au gouvernement britannique, sa position est peut-être la plus originale pour l'instant puisqu'il a décidé de n'imposer aucune nouvelle règle spécifique, mais plutôt de mettre à la disposition du secteur privé un jeu de données "bac-à-sable" pour tester la conformité des modèles d'IA, y compris d'IAG, aux exigences réglementaires (UK government 2023, p. 51 et suivantes).

La plupart des risques identifiés pour l'IAG par les gouvernements et dans les administrations publiques (Conseil de l'Europe, 2023) sont en fait ceux traditionnellement associés à l'IA. Transparence, égalité de traitement, devoir de redevabilité, impartialité sont des valeurs bureaucratiques spécifiques aux administrations des états modernes. Comment assurer la transparence des décisions ? Comment générer la confiance des citoyens à l'égard de ce qui serait produit par l'IA ? Comment assurer la capacité de l'administration à expliquer au nom de quoi une décision est prise ? Comment rendre la prise de décision opposable ? Comment garantir l'égalité de traitement et la reproductibilité de la réponse de l'IA ? Comment détecter et éviter les biais³⁹ ? Ces questions sont au cœur de la littérature sur l'IA (Desouza et al. 2020, Mikuriya et Cantens 2020, Conseil de l'Europe 2023). Les biais de l'IA dans le secteur public ont déjà donné lieu à de multiples scandales : la notation automatique des étudiants au Royaume-Uni en 2020⁴⁰, ou les enquêtes discriminatoires sur la fraude aux avantages sociaux menées par le gouvernement des Pays-Bas en 2012 et qui ont saccagé la vie de plus de 20000 familles⁴¹.

Toutefois, ces risques valent essentiellement pour une IA décisionnaire. Il est vrai que l'IA est de plus en plus souvent intégrée à des systèmes de décision dans les administrations. Même si la "décision" de l'IA est comprise comme une "aide à la décision humaine", Alon-Barkat et Busuioc (2023) ont montré que l'interaction de l'IA avec les fonctionnaires est souvent empreinte de multiples biais qui font de la décision finale du fonctionnaire une décision fortement influencée par celle de la machine. Nous pouvons même ajouter que dans certains cas, la machine est perçue comme la réponse aux mauvaises décisions humaines et devient pleinement décisionnaire. Par exemple, dans certaines administrations douanières et dans un objectif de lutter contre la corruption, les douaniers ont l'obligation de suivre les indications de la machine.

Or, de tous les usages vus précédemment, il n'est pas de cas où l'IAG prend une décision engageant l'administration, et, compte tenu de la nature technique de l'IAG pour l'instant, cela ne devrait pas changer. En effet, le critère décisif pour l'IA dans les services publics est l'explicabilité (Lasmar Amlada et al. 2023) : l'IA et sa mise en œuvre doivent laisser

³⁷ http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm et <https://www.globalpolicywatch.com/2023/04/china-proposes-draft-measures-to-regulate-generative-ai/>

³⁸ Le gouvernement islandais coopère avec OpenAI pour une meilleure intégration de la langue islandaise dans les nouveaux modèles (<https://openai.com/customer-stories/government-of-iceland>).

³⁹ idem.

⁴⁰ <https://blogs.lse.ac.uk/impactofsocialsciences/2020/08/26/fk-the-algorithm-what-the-world-can-learn-from-the-uks-a-level-grading-fiasco/>

⁴¹ Voir Hadwick et Lan (2021) et, pour un résumé du cas, <https://www.uantwerpen.be/en/projects/aitax/publications/toeslagen/>

la possibilité d'expliquer comment l'IA a rendu sa décision. Ce critère est une nécessité absolue pour respecter les autres valeurs, notamment comprendre pourquoi elles ne sont pas respectées par la machine. En douane, l'explicabilité d'un algorithme de ciblage est l'un des critères de choix d'algorithmes pour les administrations qui ont pour contrainte légale de devoir expliquer leurs décisions, ce qui fait la popularité des arbres de décisions par exemple (Mikuriya et Cantens 2020).

Or l'explicabilité des agents d'IAG actuelle est pour l'instant inaccessible puisqu'ils reposent sur des réseaux neuronaux (Jovanovic et Campbell 2022). Il est impossible, même pour ses concepteurs, d'expliquer les résultats de l'IAG. Ce point constitue une limite majeure au déploiement de l'IAG au-delà du champ du conseil et de l'assistance, en même temps qu'il invalide la pertinence d'appliquer à l'IAG les risques généralement associés à l'IA dans les administrations publiques.

Un risque spécifique à l'IAG est souvent évoqué, y compris dans la future réglementation européenne, et il nous semble également contestable : le fait pour le citoyen, de ne pas savoir qu'il échange avec une IAG de l'administration plutôt qu'avec l'un de ses fonctionnaires. Tout d'abord, contrairement au manque d'explicabilité, ce risque est traité techniquement. Aujourd'hui, des réponses légales sont apportées, pour imposer aux IAG qu'elles incluent des formes de marquage qui permettent à chacun de vérifier la source humaine ou machine d'un discours (lettre, texte, etc.)⁴². Ce point est inclus dans l'*AI Act* de l'UE et des solutions techniques sont explorées. OpenAI a récemment mis en ligne pour le public un instrument qui renvoie le caractère probable (et non une probabilité numérique) que le texte soit écrit par une IAG⁴³, dans le cadre de la lutte anti-plagiat dans les universités et avec de nombreuses restrictions, notamment linguistiques.

Toutefois, nous pourrions aussi nous questionner sur la pertinence d'un tel risque dans le cadre des rapports avec l'administration. L'humain *doit-il* être informé qu'il converse avec une IAG dans le cadre de ses relations avec l'État ? Dans un état webérien moderne bureaucratique, la figure idéale du fonctionnaire est quasi "machinique", il doit fonctionner de façon impartiale, objective et rationnelle. En outre, au final, l'utilisateur attend une réponse qui engage l'administration, et qu'elle vienne d'un humain ou d'une machine devrait peu lui importer, si l'administration la crédite d'une valeur identique. C'est à l'administration elle-même de savoir jusqu'où elle veut rendre la réponse de la machine engageante et quels filtres humains elle doit s'imposer entre la machine et l'utilisateur. Si la réponse n'a pas valeur légale, alors la question pourrait même ne pas se poser, c'est le cas des *chatbot* qui sont autonomes. L'État est bien une personne morale contre laquelle les citoyens se retournent en cas de litige, il n'est pas besoin de connaître le fonctionnaire qui a pris une mauvaise décision pour la contester.

Il est compréhensible que les autorités publiques cherchent à examiner tous les "risques" possibles afin que les sociétés aient confiance en l'IA. Toutefois, il ne faudrait pas transposer les craintes de l'imaginaire social dans les environnements légaux et techniques

⁴² Voir Kirchenbauer et al. (2023) et <https://www.nytimes.com/interactive/2023/02/17/business/ai-text-detection.html>

⁴³ <https://platform.openai.com/ai-text-classifier>

de l'administration. Savoir que la décision d'enquêter sur elles a été prise par la machine ou pas n'a rien changé à la détresse des familles néerlandaises. Quant à l'IAG, que sa nature langagière et conversationnelle la fasse passer pour un humain ne devrait pas poser de problème a priori dans les rapports entre usagers et administrations, du fait de l'autonomie légale et de la nature impersonnelle de la parole de l'administration.

Les limites techniques surmontables

La confiance dans les réponses de l'IA est affectée par quatre facteurs spécifiques à l'IAG : l'agent conversationnel ne cite pas ses sources, il répond à partir d'un corpus de textes limités dans le temps, jusqu'en septembre 2021 pour ChatGPT (encore que parfois des faits récents soient inclus sans qu'il soit possible de comprendre pourquoi), il ne répond pas toujours de la même façon aux mêmes requêtes, et il commet des erreurs.

S'agissant des sources d'information non citées, il est probable que ce problème soit résolu techniquement, la question étant plutôt légale à ce stade d'imposer aux agents conversationnels qu'ils respectent l'attribution des sources et des droits de reproduction, surtout que l'accès à ces agents devient payant. Au stade actuel d'un usage comme assistant de recherche, il est possible de continuer la conversation en posant des questions sur les sources des idées relevées par l'IAG. En revanche, ses réponses sont décevantes car souvent erronées⁴⁴. La meilleure solution actuelle consiste à chercher les sources sur les moteurs de recherche universitaires.

S'agissant de la limite temporelle de la connaissance de l'IAG, elle est en passe de ne plus être technique. Il est possible de connecter l'IAG à Internet⁴⁵. Toutefois, les compagnies d'IAG vont se montrer réticentes à inclure ce type de fonctionnalités comme service car Internet n'est pas un environnement contrôlé, et l'agent conversationnel pourrait relayer des fausses nouvelles et des textes malveillants. Il est possible que le modèle économique pour des usages professionnels évoluera vers des services de mise à jour du corpus de textes dans certains domaines d'intérêt.

Quant à la versatilité de l'IAG dans ses réponses à une même question, c'est une limite pour l'instant inhérente à la technique elle-même. Cette versatilité peut être due à l'imprécision de nos usages de la langue. Par exemple, ChatGPT ajoute des idées lorsqu'on lui demande d'améliorer le texte si nous formulons la requête de façon imprécise "améliore le texte" plutôt que "améliore uniquement le style du texte". L'agent conversationnel manque d'intuition sur notre demande. Le problème se pose de façon plus complexe dans l'exemple du classement tarifaire où toutes les formulations de la requête sont quasi-identiques et non ambiguës linguistiquement et pourtant les réponses varient. Il est difficile, à ce stade, d'anticiper les raisons, ne serait-ce que parce que, comme cela a été dit précédemment, il est impossible de reconstituer la logique suivie par une IA fondée sur les réseaux neuronaux. Toutefois, le modèle demeurant probabiliste, il est raisonnablement optimiste de penser que des corpus dédiés de textes devraient améliorer la stabilité du sens des réponses.

⁴⁴ Notre expérience et Hueber et Kleyer (2023).

⁴⁵ <https://openai.com/blog/chatgpt-plugins>

L'erreur est un sujet plus sensible car il renvoie à notre conception mécaniste de ce que devrait être une machine qui se trompe de façon systématique, ce qui n'est pas le cas de l'IAG. L'IAG commet des erreurs, des « hallucinations », ce problème est inhérent aux modèles de génération automatique de textes (*natural language generation*) (Ji et al. 2023). Le terme d'hallucination permet de distinguer ce qui relève de l'erreur de ce qui relève du biais. Les biais sont des tendances à accentuer un phénomène, ils ne sont pas à proprement parler des « erreurs » commises par la machine mais des tendances humaines sous-jacentes présentes dans les données d'entraînement, que l'automatisation du traitement par la machine reproduit à grande échelle et plus grande vitesse. Il n'est pas reprochable à la machine de se "tromper" mais de faire aussi mal que les humains plus vite et avec plus de constance. Les biais demeurent compréhensibles au sens où les tendances sont présentes donc découvrables ⁴⁶[OBB]. Ce caractère est d'autant plus inexplicable de par la nature des réseaux de neurones dont il semble impossible de reconstituer la logique de production d'un résultat.

S'agissant des erreurs de l'IAG, il faut rappeler que certains agents conversationnels ont déjà réussi des examens d'entrée, avec des scores moyens mais suffisants pour l'admission en écoles de commerce, de gestion ou au barreau⁴⁷.

Plus que l'erreur en soi, la perception de l'erreur donc la confiance en l'IAG est un obstacle à son déploiement rapide dans les administrations. Longoni et al. (2022) ont demandé à des lecteurs de titres d'articles de presse si ceux-ci leur semblaient vrais ou faux, selon qu'ils étaient rédigés par une IAG ou un humain. Leurs résultats démontrent un manque de confiance : les lecteurs jugent de façon erronée que l'IAG se trompe lorsque ce n'est pas le cas, alors qu'ils accordent plus de confiance aux titres rédigés par des humains, même erronés. Longoni et al. (2023) attribue cela au fait que les algorithmes sont perçus comme plus homogènes que les humains, donc les erreurs d'un groupe d'algorithmes sont généralisées à tous les algorithmes. Ces études, sur la perception humaine de l'erreur face à l'IA, illustrent à quel point il serait risqué de déployer des IAG dans les administrations sans s'assurer que les erreurs commises soient réduites et traitées, au risque d'ancrer profondément une méfiance des usagers envers l'IA.

Il semble impossible que l'IAG ne commette pas d'erreur, mais il est très probable qu'elle en commettra de moins en moins, surtout pour des usages où le langage est très normé comme les usages administratifs. Les dimensions culturelles, l'humour, les expressions propres à certaines langues devraient continuer de demeurer des sources d'erreurs pour l'IAG, d'autant plus que les langues évoluent. L'émergence de l'IAG nous indique seulement que nous ne devons plus considérer l'erreur comme le propre de

⁴⁶ Un avocat américain a utilisé ChatGPT pour établir sa plaidoirie. ChatGPT lui a renvoyé des références très précises de cas qui n'existaient tout simplement pas. <https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html>

⁴⁷ Voir <https://www.reuters.com/legal/legal-ai-race-draws-more-investors-law-firms-line-up-2023-04-26/> and how some AI models passed the bar <https://www.reuters.com/technology/bar-exam-score-shows-ai-can-keep-up-with-human-lawyers-researchers-say-2023-03-15/> et plus récemment les performances plutôt stupéfiantes de GPT4 publiées par OpenAI (OpenAI 2023), sans toutefois avoir été contre-validées par une étude indépendante.

l'humain et qu'il n'y a pas moins de précaution à prendre en lisant le texte produit par une IA qu'en lisant celui écrit par un humain.

La confidentialité, un problème spécifique à l'IA

Si la confiance en l'IA et la perception de ses limites relèvent autant de la technique que de la culture actuelle, la préservation de la confidentialité des informations et du fonctionnement interne de l'administration demeure une contrainte forte, spécifiquement affectée par la nature conversationnelle de l'IA.

En effet, les compagnies produisant et gérant les agents conversationnels actuels ont tout intérêt à recevoir un maximum d'informations en retour de la part des usagers humains pour évaluer et corriger les résultats de l'IA. Entre l'entraînement de l'IA et son ouverture au public, il se déroule une phase de tests réalisés par des humains sur des questions morales, afin d'éviter le risque de toxicité, à savoir la tendance de l'IA à reproduire, donc à répandre des discours qui seraient racistes, sexistes et immoraux (Jo 2023) Évaluer l'étendue de ce risque au sein des administrations publiques est difficile. La toxicité de l'IA requiert une attention de la part des fonctionnaires, en particulier lorsqu'ils sont en charge d'un sujet à dimension discriminatoire : la police des étrangers, le contrôle des entreprises étrangères, la lutte contre la corruption, le traitement de trafics illégaux spécifiques en provenance de l'étranger⁴⁸, tout domaine où la machine est incapable, par apprentissage automatique, de distinguer le moral de l'immoral dans le contexte culturel de son déploiement. La part humaine de jugement sur les réponses apportées par l'IA est donc éthiquement critique⁴⁹, mais elle est tout aussi essentielle pour améliorer la précision des réponses. C'est ainsi que les utilisateurs de ChatGPT sont invités à évaluer les réponses fournies et que les conversations sont stockées, pour être utilisées par la suite dans un nouvel entraînement.

Cela pose un problème aux administrations, lorsque les fonctionnaires posent des questions à l'IA sur des domaines sensibles voire confidentiels. Les administrations n'ont pas l'assurance que les sociétés privées gérant les agents conversationnels ne stockent pas leurs conversations, même si de nouvelles fonctionnalités pour effacer ces conversations sont maintenant disponibles. Il n'est pas exclu que ces conversations, dévoilant potentiellement les réflexions stratégiques d'un gouvernement, soient stockées malgré tout, et accessibles à une compagnie privée voire à un autre état.

La solution pour les administrations serait l'internalisation de l'IA et le *finetuning* vu précédemment. C'est la décision du gouvernement de Singapour qui envisage de mettre à la disposition de tous ses fonctionnaires un assistant, centralisant de nombreuses applications à l'échelle de toutes les administrations, en partenariat avec Google⁵⁰. Si la faible confidentialité des conversations est un risque nouveau pour les administrations lié à

⁴⁸ Bien que les administrations fiscalo-douanières aient des pratiques discriminatoires par la création d'unités spécialisées dans les « grands contribuables » ou par des statuts spéciaux « d'opérateurs économiques agréés », le risque semble minime car les discriminations positives sont appliquées suivant un seuil – objectif - de contribution fiscale.

⁴⁹ OpenAI n'est pas transparente sur ses processus, toutefois un récent scandale a révélé une partie de ses méthodes de construction du corpus de ChatGPT. <https://time.com/6247678/openai-chatgpt-kenya-workers/>

⁵⁰ <https://www.smartnation.gov.sg/media-hub/press-releases/31052023/>

l'IA, la confidentialité des corpus n'est pas nouvelle. Ces corpus, gérés en interne, peuvent être stockés par des tiers de confiance. Les administrations font déjà face à la confidentialité des données d'entraînement, lorsqu'elles coopèrent avec le secteur privé pour élaborer leurs algorithmes. Cet aspect de la confidentialité concernerait une part moindre des données d'entraînement dans le cas de l'IA, puisqu'une large partie des textes de l'administration est publique.

Toutefois, à son tour, cette mutualisation des moyens de l'État, et de ses données, pose de nouveaux problèmes. Certains états interdisent la fusion des données des administrations, a fortiori automatique, pour garder les consultations traçables, justifiables et empêcher la création d'un système unique et centralisé de surveillance de la société. Jusqu'où mutualiser, notamment si des données personnelles de fichiers sont mises en commun ? Dans quelle mesure l'IA ferait des recoupements de données personnelles, quand bien même cela ne lui serait pas demandé ?

Le caractère langagier de l'IA : l'appropriation, pas l'asservissement

Les discussions arides sur les avantages, les limites et les possibilités ont occupé la sphère publique depuis la fin de l'année 2022 et le lancement de ChatGPT. Revenons à la stupeur enthousiaste que nous avons tous éprouvée à notre première requête lancée à un agent conversationnel gratuit sur Internet, ChatGPT⁵¹, ou BARD ou un autre qui aura émergé entre l'écriture de cet article et sa publication. À ceux qui ne l'auraient pas encore fait, une quasi-injonction - ne continuez pas la lecture de ce papier sans avoir essayé par vous-même - pour lire ce qui va suivre avec en tête la sensation quasi magique d'échanger avec une machine et la joie primale de la découverte d'une forme extra-humaine qui nous fait l'effet d'un égal encore un peu naïf. Même après leurs premiers pas avec la machine, certains d'entre nous continuent de saluer ChatGPT et de lui demander poliment d'exécuter leurs requêtes, persuadés qu'il y a dans l'intimité avec la machine quelque chose qui se joue de l'ordre de l'efficacité.

Floridi (2023) relève que les humains ont jusqu'à présent interagi avec des agents animaux, qu'ils ont cru interagir avec des agents spirituels, et qu'ils devraient maintenant interagir avec des agents artificiels qui sont capables de s'améliorer sans comprendre. Il propose ironiquement un nouvel acronyme, pour Artificial Intelligence, *Agere sine Intelligere, Acting without Intelligence* (AI). Toutefois, nous nous accordons avec Coeckelbergh et Gunkel (2023) pour dire que l'IA ne doit pas être considérée comme simplement instrumentale et que les liens entre homme, machine et langage sont tellement intriqués dans la production de sens, que travailler avec une IA qui produit du langage aura inévitablement des effets sur notre manière de penser.

⁵¹ ChatGPT est actuellement le plus connu de tous les modèles génératifs d'IA, il est développé par OpenAI et son accès est désormais gratuit (<https://chat.openai.com>). Le vocabulaire technique utilisé peut varier : "chatbot" a été utilisé jusqu'à présent pour l'IA limitée, mais il fait toujours référence aux outils de type ChatGPT, les "modèles de base" ou "modèles à usage général" sont explicites et incluent les "modèles de langage large". Dans le présent document, nous parlons d'"intelligence artificielle générative" pour insister sur la capacité à créer du contenu.

Nous ajoutons trois points sur le caractère fondamentalement nouveau de ce qu'il se passe avec l'IAG dans les administrations. Premièrement, l'IA spécialisée est en relation avec le fonctionnaire spécialiste : l'IA en charge de l'analyse des risques en douane est en contact avec les douaniers qui contrôlent les déclarations. Le caractère général de l'IAG et l'étendue des fonctionnaires concernés font de son appropriation par les administrations une réforme radicale. Le deuxième point est que travailler avec l'IAG ne fera probablement pas émerger de nouvelles idées - encore qu'il soit possible que ses "hallucinations" soient inspirantes⁵². – mais de nouvelles façons de penser "avec" l'IAG, de formuler les questions à l'IAG, d'analyser des réponses et de construire notre propre discours. Ce "penser avec" nous semble essentiel pour les fonctionnaires, il est lié à notre dernier point. L'administration produit du langage et ce qu'elle énonce est un acte d'autorité et de pouvoir. C'est bien avec le langage que les états décrivent et norment les espaces et les populations, qu'ils produisent leur autorité par des lois écrites, que les fonctionnaires contraignent, surveillent et sanctionnent, que les réformes se conduisent et propagent des mots et concepts clefs, que les fonctionnaires se saisissent de nouvelles idées et formulent de nouvelles problématiques.

Le caractère langagier de l'IAG

Le langage est maintenant au cœur du rapport entre humaines et machines, et c'est un point de première importance pour les administrations. Même si sa conception relève toujours de l'apprentissage machine donc fondée sur une approche probabiliste, l'IAG est *énonciatrice*, elle produit des énoncés, à la différence des algorithmes déjà utilisés qui sont des calculateurs, produisant des nombres, des scores, des probabilités de risque. Dans le cas de l'IAG, le résultat est bien du langage, non pas un effet de langage. Que l'IAG ne comprenne pas ce qu'elle produit, qu'elle se trompe ou qu'elle nous semble farfelue, cela n'enlève rien à la validité sémiotique et sémantique de ce qu'elle nous renvoie. Ce que l'IAG produit est du discours qui a son autonomie en soi, au même titre que n'importe quel texte. En revanche, et à la différence de l'IA numérique, l'IAG ne quantifie pas la probabilité qu'elle se trompe. Elle produit des énoncés avec un effet de certitude qui n'existait pas auparavant et qui rend plus difficile la détection de ses erreurs⁵³.

En deuxième lieu, l'IAG est *éclectique* car elle est conçue sans objectif, et cela lui procure un avantage essentiel par rapport à l'humain dans le contexte moderne de la division du travail. Comme nous l'avons vu, l'IAG peut aussi être entraînée sur un corpus précis tout comme les autres IA, mais il n'empêche que, même dans ce cas, l'IAG intègre toute l'information présente et non celle qui lui sert à un objectif précis. C'est en ce sens général qu'elle est éclectique en nature, indépendamment du corpus sur lequel elle est entraînée.

⁵² Un exemple assez similaire pourrait être trouvé dans le domaine des mathématiques où l'IA (et non l'IAG) est utilisée pour alimenter les intuitions des mathématiciens, voir Davies, A., Veličković, P., Buesing, L. et al. (2021). "Advancing mathematics by guiding human intuition with AI". *Nature* 600, 70–74. <https://doi.org/10.1038/s41586-021-04086-x>

⁵³ ChatGPT tente d'atténuer ce problème en rappelant toujours que les résultats peuvent être faux ou doivent être vérifiés, mais il ne peut pas quantifier la probabilité de commettre une erreur.

Dernier élément de langage, l'IAG est *interlocutrice*. Fondée sur une interaction directe par le langage humain et non via du code informatique, l'IAG ne demande pas d'intermédiaire humain expert pour interagir avec l'utilisateur final. Nous demandons à ces agents conversationnels, par exemple, de développer des aspects précis de leurs réponses ; nous pouvons aussi les amener à reconnaître leurs erreurs logiques, même si c'est l'embryon d'un raisonnement sitôt oublié puisqu'il suffit de poser à nouveau la même question à la machine pour s'apercevoir qu'elle produit une réponse tout aussi erronée que la réponse initiale. Il n'empêche que nous entrons en interaction directe avec un agent, qui sans être intelligent, produit bien du langage, un langage cohérent, efficace en connaissance.

Dépasser la crainte du remplacement

La crainte du remplacement du fonctionnaire est légitime ; pour certains usages, comme nous l'avons vu, la machine peut être plus efficace et moins coûteuse. L'IAG aura un effet économique sur la structure des emplois de production intellectuelle (Eloundou et al. 2023), en particulier dans les fonctions d'écriture. Cette tendance au remplacement s'insère dans la vision wéberienne à l'origine des conceptions contemporaines de l'État : remplacer l'humain par la machine pour faire fonctionner des états selon une rationalité bureaucratique idéale semble *a priori* plus simple à garantir *via* une machine « froide » que *via* un humain animé par les passions et la subjectivité. L'argument du biais a peu de prise politique contre cette tendance. La machine n'est biaisée que parce qu'elle reproduit les biais humains, elle ne fait donc pas « pire », en nature, que l'humain, elle le fait juste avec plus de célérité et de discipline, toute chose qui par ailleurs domine aujourd'hui les discours sur l'efficacité des administrations.

Comme pour l'IA en général, les administrations et les experts veulent rassurer et affirment que les fonctionnaires ne seront pas remplacés par l'IAG, mais qu'ils seront déchargés de certaines tâches. Toute la difficulté, non encore résolue, réside dans la délimitation du périmètre de ce qui serait délégué à l'IAG. Les coûts de cette délégation seront un paramètre essentiel des choix stratégiques dans un contexte néolibéral global où les états tendent depuis longtemps à réduire leur périmètre d'action⁵⁴. L'internalisation de l'IAG par une administration ou un état coûtera plus cher qu'une utilisation d'agents conversationnels publics, mais ces coûts demeureront très probablement en-deçà du coût d'un fonctionnaire et rendront l'IAG très rentable. L'IAG entre donc aujourd'hui en compétition avec des emplois qualifiés, à la différence des machines au 19^e siècle qui entraient en compétition - et la gagnaient - avec les ouvriers.

Il n'y a pas à proprement parler de remplacement du fonctionnaire. Le fonctionnaire devra plutôt intégrer dans ses pratiques quotidiennes des tâches qui étaient dévolues auparavant à des acteurs extérieurs. Nous ne connaissons pas d'administration qui emploie des « assistants de recherche », et beaucoup de tâches qui ne sont pas au cœur de l'action administrative sont externalisées. Le changement sera pour le fonctionnaire de faire un peu

⁵⁴ Le même argument financier s'applique à d'autres secteurs, comme le journalisme, qui sont fortement touchés par la concurrence et la réduction des coûts et où l'IAG est considérée comme essentielle pour compléter le travail des journalistes (Ophal et al. 2023).

plus par lui-même, faire ses propres traductions et relectures avec l'assistance de l'IAG et porter ses efforts sur certaines parties du processus de réflexion.

Intégrer de nouvelles tâches dans son quotidien est une constante du progrès technologique. Plus la machine automatise des tâches basiques, plus l'utilisateur final doit les faire exécuter lui-même par la machine, en lieu et place de ceux qui les exécutaient avant l'introduction de la machine. La micro-informatique et le traitement de texte en ont terminé avec des fonctionnaires écrivant leurs notes manuscrites et les transmettant à des équipes de secrétaires qui les tapaient à la machine.

Il nous faudra donc développer nos qualités d'écriture. Les fonctionnaires et les experts devront *mieux écrire*, au sens où ils sauront interroger l'agent conversationnel et mobiliser ses réponses pour écrire en fonction d'objectifs ou de formats auxquels l'IA ne s'adapterait pas compte tenu de la finesse des nuances⁵⁵. Il y a une différence, par exemple, entre rédiger une note politique sur le positionnement de la douane dans la lutte anti-corruption, une note conceptuelle pour un projet de lutte contre la corruption et un rapport d'expertise sur la politique anti-corruption. Même si l'IAG améliore linguistiquement le texte en dernière instance, il est peu probable, dans un court terme, qu'elle rédige en fonction d'objectifs aussi précis que ce qui est demandé dans les administrations.

Il nous faudra également *renforcer notre vigilance critique*, face aux erreurs de l'IA et face aux erreurs humaines dans l'analyse des données en général. Cette faculté de vigilance critique relève de la logique, de la cognition et de la communication, comme capacité à détecter les biais autant que les paralogismes d'un discours en général.

Éviter le risque d'asservissement

Il y aurait néanmoins un risque d'asservissement à l'IAG, si nous nous bornions à un rôle de « vérificateur-ajusteur » des productions de l'IAG. Noam Chomsky⁵⁶ identifie trois différences fondamentales entre l'IAG et l'humain : l'IAG ne distingue pas le possible de l'impossible car elle peut tout apprendre, y compris l'impossible ; l'IAG décrit et prédit mais n'explique pas ; enfin l'IAG n'est pas morale, elle propose un ensemble d'avis sans prendre parti. Il faut ajouter à Chomsky que l'IAG ne prend pas parti même si on le lui demande, ce qui est une restriction de ses concepteurs qui préfèrent une IA amoral, qui ne répond pas aux questions sur les choix moraux, plutôt qu'une IA qui serait immorale en y répondant.

C'est dans ces écarts qu'il faut chercher la valeur ajoutée humaine, dans la volonté de faire des choix politiques et moraux soutenue par une faculté de les expliquer. Ces principes généraux s'appliquent particulièrement aux fonctionnaires utilisant l'IAG et se résument dans des capacités de nuance, de distinction et de critique.

Les fonctionnaires devront *développer leur pensée critique*. Face à l'IA et sa capacité synthétique et normative, le fonctionnaire a une valeur ajoutée par sa créativité, sa capacité à penser hors du cadre, capacité facilitée par l'usage de l'IA qui construit une connaissance basique mais suivant un spectre de perspectives assez vaste. Ceci est valable pour les administrations comme pour les organisations internationales. Leur rôle est de plus en plus

⁵⁵ À ce stade, l'IAG peut écrire « à la manière de » mais dans le champ littéraire uniquement ; les essais ne sont pas concluants et la fonctionnalité est finalement plutôt inutile.

⁵⁶ Chomsky, N., Roberts, I., & Watumull, J. (2023). *The False Promise of ChatGPT*. The New York Times, 8.

« politique » et « analytique », de conseiller leurs gouvernements pour prendre part à la gouvernance globale. Ceci est particulièrement vrai dans le monde du développement et de la coopération internationale. Le rôle des organisations internationales s'équilibre entre fonctions « normatives », l'édiction de règles, de normes et de standards au sein d'instruments internationaux, de guides de meilleures pratiques, et une fonction « prospective » où les états membres des organisations attendent qu'elles mobilisent une connaissance et une analyse globale sur des questions pour lesquelles ils éprouvent des difficultés à élaborer leur propre vision stratégique. De nombreuses organisations internationales disposent d'unités « politiques, de « recherche » ou de « foresight ». Par exemple, il n'est pas anodin que les thèmes principaux d'une organisation à vocation technique comme l'OMD soient « technologies/données », « environnement » et « fragilité », fortement empreints d'une connotation prospective.

Conclusion

Nous sommes loin d'avoir épuisé le sujet de l'IAG dans les administrations. Celles-ci seront soumises aux mêmes questions qui animent tous ceux qui réfléchissent à l'avenir de l'IAG d'un point de vue non purement technique, à savoir l'exploitation commerciale d'une connaissance humaine libre et gratuite par les compagnies produisant les agents conversationnels, l'impact environnemental de l'IAG, l'usage de l'IAG pour favoriser la vie démocratique et intensifier les rapports entre citoyens et autorités publiques⁵⁷, ou encore la politisation des corpus d'entraînement des agents conversationnels⁵⁸.

Malgré ces questions auxquelles s'ajoutent les limites techniques de l'IAG et l'avenir flou de son modèle économique, nous pouvons avoir quelques certitudes. La première est que l'IAG marque une rupture fondamentale en ce qu'elle produit du langage et non des chiffres. Les sociétés dans leur intégralité vont interagir avec l'IA directement. Un nouveau rapport à l'IA s'installe, et le langage est au cœur de celui-ci.

La deuxième certitude est qu'il n'y a pas d'IAG émancipée de l'humain. Outre le fait que nous la façonnons, lorsque l'IAG produit une réponse elle le fait par calcul de probabilités sur du contenu que nous avons déjà produit. Toutefois, la relation va dans les deux sens. Alors que nous façonnons nos outils, nos outils nous façonnent aussi. Dans le cas de l'IAG, nos pratiques d'écrire, donc de penser, vont changer dans le trouble du discours humain produit par un agent non-humain. Seront renouvelés des moments dans la chaîne de production intellectuelle : formuler une problématique, penser nos questions, lire les réponses de l'IAG, les situer dans le contexte national ou culturel, les critiquer aussi. La valeur ajoutée de l'humain ne se situe donc peut-être pas dans les limites perçues de l'IAG mais

⁵⁷ Voir Vrabie (2023) pour un exemple particulièrement intéressant sur l'usage de l'IAG pour le dépeuplement des pétitions.

⁵⁸ ChatGPT est interdit par la Russie, la Chine, Cuba, l'Iran et la Syrie. L'Italie a fait part de ses préoccupations en matière de protection de la vie privée, mais OpenAI, la société qui gère ChatGPT, a réagi et l'Italie a levé l'interdiction. (<https://apnews.com/article/chatgpt-openai-data-privacy-italy-b9ab3d12f2b2cfe493237fd2b9675e21>). Alors que tous les pays mettront en œuvre et autoriseront des outils de type ChatGPT, certains pays peuvent concevoir le corpus de textes de formation de manière à limiter l'accès des personnes à certaines informations, comme c'est déjà le cas avec l'interdiction de certains moteurs de recherche dans certains pays...

dans la différence entre l'IAG et l'humain. Alors que le débat public porte surtout sur l'écart qui se resserre entre la machine et l'humain, il faudrait plus s'intéresser à ce que cet écart a d'irréductible pour que le fonctionnaire trouve sa place dans la production textuelle, intellectuelle, de l'État.

Une troisième certitude est que les capacités d'écriture et de critique des fonctionnaires travaillant avec l'IAG s'appliquent tout autant à l'analyse des données. Dans un environnement de gouvernance où il est demandé aux chefs d'administration de décider à partir d'analyses de faits et de données, ceux qui produisent ces analyses ont besoin d'être compris au-delà de leur champ d'expertise. L'émergence de l'IAG force la réflexion sur l'appropriation de l'IA à s'élargir aux non-spécialistes, au-delà des questions de recrutement et de formation de spécialistes de données.

Nous pouvons donc supposer que l'engouement pour l'IAG suscité par sa simplicité d'utilisation et son large spectre d'applications dépassera rapidement les réserves pour l'instant observées dans les administrations, réserves purement formelles puisque les fonctionnaires utilisent probablement déjà ChatGPT. L'IAG devrait devenir un outil de travail routinier, intégré dans les pratiques professionnelles en général, au même titre qu'Internet et les moteurs de recherche. En outre, les administrations n'auront pas le choix d'adopter ou non l'IAG, car elles y seront quasiment contraintes par l'évolution de la société. Pour les administrations en charge de contrôles, comme les administrations fiscales et douanières par exemple, l'IAG posera le même problème que toute technologie, elle sera utilisée par les entreprises qui fraudent, pour se défendre devant les commissions de recours et les tribunaux, mais aussi probablement pour construire de nouveaux schémas de fraude, sur tous les segments du processus administratif qui sont essentiellement textuels.

Il y aurait un nouveau risque finalement que nous nous privions formellement de nombreux usages de l'IAG possibles maintenant, au motif que nous ne savons pas comment traiter les risques qu'elle porte pour des usages que nous n'avons de toutes les façons pas encore adoptés ni peut-être même imaginés. Il nous semble plus raisonnable d'utiliser une IAG dont la technique ne remplit pas tous les critères éthiques et légaux et d'en limiter l'usage pour les pratiques administratives où ces critères sont peu ou pas pertinents.

L'adoption de l'IAG dans le monde du travail et dans les administrations en particulier marque une « révolution intellectuelle » au sens où l'adoption des machines dans les usines a caractérisé la « révolution industrielle » aux 19^e et 20^e siècles : une machinisation de fonctionnalités basiques de la pensée, au sein d'un processus de production intellectuelle, introduisant des gains d'efficacité mais nécessitant aussi une adaptation et une spécialisation des humains.

Plus que le dépassement ou le remplacement, c'est le risque d'asservissement de l'humain à la machine. Ce point semble tellement intolérable a priori qu'il est quasi absent du débat public. À ce titre, généraliser et renforcer notre appétit, notre faculté et notre habileté critique nous permettrait de mieux garder la main sur la machine, au moment où une machine éclectique nous aide, nous fonctionnaires-experts élevés mais aussi affaiblis par cette extrême division du travail qui caractérise notre modernité. En 1945, dans son pamphlet *La France contre les robots*, mettant dos-à-dos les économistes anglais et Karl Marx

dans une même critique de la réduction de l'homme à ses quantités, Georges Bernanos caractérisait cette modernité machinique en déplorant que « le progrès n'est plus dans l'homme, il est dans la technique, dans le perfectionnement des méthodes capables de permettre une utilisation chaque jour plus efficace du matériel humain ». L'idée serait plutôt que la machine nous aide à devenir de meilleurs experts, critiques de la connaissance que nous produisons pour nous gouverner.

Bibliographie

Alon-Barkat, S., & Busuioc, M. (2023). Human–AI interactions in public sector decision making: “automation bias” and “selective adherence” to algorithmic advice. *Journal of Public Administration Research and Theory*, 33(1), 153-169.

Babl, F. E., & Babl, M. P. (2023). Generative artificial intelligence: Can ChatGPT write a quality abstract?. *Emergency Medicine Australasia*.

Beheshti, A., Yang, J., Sheng, Q. Z., Benatallah, B., Casati, F., Dustdar, S., ... & Xue, S. (2023). ProcessGPT: Transforming Business Process Management with Generative Artificial Intelligence. *arXiv preprint arXiv:2306.01771*.

Chen, Z., Zheng, L. N., Lu, C., Yuan, J., & Zhu, D. (2023). ChatGPT Informed Graph Neural Network for Stock Movement Prediction. *arXiv preprint arXiv:2306.03763*.

Cheng, L., Li, X., & Bing, L. (2023). Is GPT-4 a Good Data Analyst?. *arXiv preprint arXiv:2305.15038*. <https://arxiv.org/pdf/2305.15038.pdf>

Chomsky, N., Roberts, I., & Watumull, J. (2023). Noam Chomsky: The False Promise of ChatGPT. *The New York Times*, 8.

Coeckelbergh, M., & Gunkel, D. J. (2023). ChatGPT: deconstructing the debate and moving it forward. *AI & SOCIETY*, 1-11.

Council of European Union, 2023. ChatGPT in the Public Sector – overhyped or overlooked? https://www.consilium.europa.eu/media/63818/art-paper-chatgpt-in-the-public-sector-overhyped-or-overlooked-24-april-2023_ext.pdf

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard business review*, 96(1), 108-116.

Desouza, K. C., Dawson, G. S., & Chenok, D. (2020). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. *Business Horizons*, 63(2), 205-213.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.

European Parliament, 2023. Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. URL: <https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf>

Europol (2023). ChatGPT. The Impact of Large Language Models on Law Enforcement. URL : <https://www.europol.europa.eu/cms/sites/default/files/documents/Public%20Version%20-%20ChatGPT%20report.pdf>

Floridi, L. (2023). AI as Agency without Intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology*, 36(1), 15.

Graham, S., Yates, D., & El-Roby, A. (2023). Investigating antiquities trafficking with generative pre-trained transformer (GPT)-3 enabled knowledge graphs: A case study. *Open Research Europe*, 3, 100.

Hadwick, D., & Lan, S. (2021). Lessons to be learned from the dutch childcare allowance scandal: a comparative review of algorithmic governance by tax administrations in the Netherlands, France and Germany. *World tax journal.-Amsterdam*, 13(4), 609-645.

Henseler, H., & van Beek, H. (2023). ChatGPT as a Copilot for Investigating Digital Evidence. <https://ceur-ws.org/Vol-3423/paper6.pdf>

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). AI, write an essay for me: A large-scale comparison of human-written versus ChatGPT-generated essays. *arXiv preprint arXiv:2304.14276*.

Huang, J., & Tan, M. (2023). The role of ChatGPT in scientific communication: writing better scientific review articles. *American Journal of Cancer Research*, 13(4), 1148.

Hueber, A. J., & Kleyer, A. (2023). Quality of citation data using the natural language processing tool ChatGPT in rheumatology: creation of false references. *RMD open*, 9(2), e003248.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1-38.

Jiao, W., Wang, W., Huang, J. T., Wang, X., & Tu, Z. P. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv preprint arXiv:2301.08745*.

Jovanovic, M., & Campbell, M. (2022). Generative Artificial Intelligence: Trends and Prospects. *Computer*, 55(10), 107-112.

Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., & Goldstein, T. (2023). A watermark for large language models. *arXiv preprint arXiv:2301.10226*.

LASMAR ALMADA, M. A., GÓRSKI, Ł., KUŹNIACKI, B., TYLIŃSKI, K., Winogradská, B., & Zeldenrust, R. (2022). Towards eXplainable Artificial Intelligence (XAI) in tax law: the need for a minimum legal standard. *World tax journal*, 14.

Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022, June). News from generative artificial intelligence is believed less. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 97-106).

Longoni, C., Cian, L., & Kyung, E. J. (2023). Algorithmic Transference: People Overgeneralize Failures of AI in the Government. *Journal of Marketing Research*, 60(1), 170-188.

Mikuriya, K., & Cantens, T. (2020). "If algorithms dream of Customs, do customs officials dream of algorithms? A manifesto for data mobilisation in Customs". *World Customs Journal*, 14(2).

Minelle, F., & Stolfi, F. (2023). AI to support PM: a ChatGPT quality assessment (β test). URL: <https://pmworldlibrary.net/wp-content/uploads/2023/05/pmwj129-May2023-Minelle-Stolfi-AI-to-support-PM-a-ChatGPT-quality-assessment-3-1.pdf>

Mökander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2023). Auditing large language models: a three-layered approach. *AI and Ethics*, 1-31.

Opdahl, A. L., Tessem, B., Dang-Nguyen, D. T., Motta, E., Setty, V., Throndsen, E., ... & Trattner, C. (2023). Trustworthy journalism through AI. *Data & Knowledge Engineering*, 146, 102182.

OpenAI. 2023. Gpt-4 technical report.

Peeters, R., & Bizer, C. (2023). Using ChatGPT for Entity Matching. *arXiv preprint arXiv:2305.03423*.

Pierce, N., & Goutos, S. (2023). Why Law Firms Must Responsibly Embrace Generative AI. Available at SSRN 4477704.

Pu, D., & Demberg, V. (2023). ChatGPT vs Human-authored Text: Insights into Controllable Text Summarization and Sentence Style Transfer. *arXiv preprint arXiv:2306.07799*.

Rodrigues, F. B., Giozza, W. F., de Oliveira Albuquerque, R., & Villalba, L. J. G. (2022). Natural language processing applied to forensics information extraction with transformers and graph visualization. *IEEE Transactions on Computational Social Systems*.

United Kingdom Government, Secretary of State for Science, Innovation and Technology (2023). A pro-innovation approach to AI regulation, URL : https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1146950/a_pro-innovation_approach_to_AI_regulation_print_ready_version.pdf

Vrabie, C. (2023). E-Government 3.0: An AI Model to Use for Enhanced Local Democracies. *Sustainability*, 15(12), 9572.

Wagner, M. W., & Ertl-Wagner, B. B. (2023). Accuracy of information and references using ChatGPT-3 for retrieval of clinical radiological information. *Canadian Association of Radiologists Journal*, 08465371231171125.

Xames, M. D., & Shefa, J. (2023). ChatGPT for research and publication: Opportunities and challenges. Available at SSRN 4381803.

Yang, X., Li, Y., Zhang, X., Chen, H., & Cheng, W. (2023). Exploring the limits of chatgpt for query or aspect-based text summarization. *arXiv preprint arXiv:2302.08081*.

“Sur quoi la fondera-t-il l'économie du monde qu'il veut gouverner? Sera-ce sur le caprice de chaque particulier? Quelle confusion! Sera-ce sur la justice? Il l'ignore.”

Pascal

FERDi

Créée en 2003, la **Fondation pour les études et recherches sur le développement international** vise à favoriser la compréhension du développement économique international et des politiques qui l'influencent.



Contact

www.ferdi.fr

contact@ferdi.fr

+33 (0)4 43 97 64 60