MESDAMES, MESSIEURS

Depuis une année, l'**intelligence artificielle générative (IAG)** a fait une irruption spectaculaire dans le débat public.

La mise à disposition gratuite au grand public par la société californienne OpenAI de son robot conversationnel (« *chatbot* ») ChatGPT a fait prendre conscience aux citoyens, aux médias et aux décideurs que la démocratisation de l'IAG en langage naturel allait devenir une **réalité incontournable**.

La diffusion de l'IAG semble inéluctable et ses usages possibles sont sources de curiosité, d'intérêt et d'espoir pour les uns ; d'inquiétude voire d'angoisse pour les autres.

Ces réactions s'expliquent par le fait que l'IAG – qui est un sous-domaine de l'intelligence artificielle (IA) – a pour caractéristique de produire rapidement des **contenus originaux, visuels, sonores ou écrits**, parfois grâce à une interface simple n'exigeant pas de compétences informatiques particulières. Parmi eux, les modèles de langage (*large language model* ou *LLM* en anglais) permettent de converser dans une langue humaine avec l'IAG.

L'ensemble de ces systèmes reposent sur des modèles mathématiques statistiques, qui identifient la réponse la plus probable à l'injonction donnée : « *Un LLM "prédit" le résultat (en l'occurrence le mot suivant) le plus vraisemblable au vu de la distribution statistique des données d'entraînement* » (1).

Ces modèles ne sont donc pas infaillibles, en particulier lorsque le nombre de données dont ils disposent sur une question est limité. Ils tendent alors à donner une réponse probable, voire plausible, mais qui peut être factuellement erronée. On parle alors « d'hallucination ».

Ces systèmes peuvent aussi être utilisés de manière détournée, ce qui conduit les concepteurs d'IAG à « brider » certaines de leurs fonctionnalités afin d'empêcher qu'elles puissent produire des contenus offensant ou dangereux.

L'IAG soulève, pour une grande partie, les mêmes enjeux que l'IA dans son ensemble, tout en renouvelant certaines problématiques et en en posant de nouvelles.

---

(1) Pôle d'expertise et de régulation du numérique, « ChatGPT ou la percée des modèles d'IA conversationnels *», avril 2023.*

D'emblée, des interrogations sont apparues sur la conformité de ChatGPT au règlement européen sur la protection des données (RGPD). L'homologue italien de la commission nationale de l'informatique et des libertés (CNIL) a ainsi bloqué, le 31 mars 2023, l'accès au robot conversationnel, avant de l'autoriser à nouveau le 28 avril de la même année sous réserve que la société OpenAI poursuive ses efforts pour appliquer la législation européenne sur la **protection des données**.

Ce sujet relève directement de la compétence de la commission des Lois. C'est la raison pour laquelle, dès le 3 mai 2023, le bureau de notre commission a souhaité créer une **mission d'information sur les défis de l'intelligence artificielle générative** en matière de protection des données personnelles et d'utilisation du contenu généré.

L'**utilisation du contenu généré**, au même titre que la protection des données, suscite également de nombreuses questions, notamment en matière de responsabilité civile, voire pénale – deux domaines qui relèvent également de la commission des Lois.

Il en est de même du respect des libertés fondamentales, un sujet de préoccupation important pour l'IA en général et l'IAG en particulier, au regard de son utilisation potentielle en matière de manipulation de l'information ou dans le cadre de campagnes électorales, par exemple.

Pour autant, certains sujets n'ont pu être éludés compte tenu de leur **caractère transversal** (droit d'auteurs, manipulation de l'information, utilisation de l'IAG dans le domaine de la santé ou de l'éducation). Pour ces raisons, vos rapporteurs ont été conduits ponctuellement à formuler des appréciations plus générales.

Vos rapporteurs ont conscience que les discussions sur l'IAG ne font que débuter et que leurs travaux n'épuiseront pas le sujet. Leur rapport constitue une première contribution parlementaire à ce débat. Les travaux complémentaires d'autres commissions permanentes seront précieux pour approfondir et nourrir la réflexion du Parlement.

La tentation est grande de réduire la discussion à un **clivage entre technophiles et technophobes**. Vos rapporteurs ont tenu à ne pas y céder. Pour cette raison, ils ont organisé leurs auditions en deux phases.

• **En premier lieu, ils ont débuté leurs travaux par une phase conceptuelle**, donnant lieu à plusieurs auditions d'experts, de professeurs et de chercheurs afin d'appréhender l'état de la technique.

Pour résumer, beaucoup d'entre eux considèrent que l'IAG, pour spectaculaire qu'elle soit aux yeux du grand public, se place dans la continuité du développement de l'IA et n'est pas une révolution sur le plan technique, même si elle peut conduire à une révolution de certains usages.

Travaux parlementaires antérieurs ayant pour thème l'intelligence artificielle

Il n'y a pas eu, à ce jour, de rapport parlementaire spécifique à l'intelligence artificielle générative, bien qu'un débat se soit tenu sur ce thème au Sénat lors de la séance du 12 avril 2023 (débat sur les « *impacts économique, social et politique de l'intelligence artificielle générative* »).

L'intelligence artificielle, en revanche, a fait l'objet d'une étude approfondie par l'office parlementaire d'évaluation des choix scientifiques et technologiques (OPECST), dans un rapport du 15 mars 2017 intitulé « *Pour une intelligence artificielle maîtrisée, utile et démystifiée* ».

Sous la XV^{ème} législature, le député Cédric Villani s'était également vu confier par le Premier ministre une mission qui a donné lieu à un rapport, publié le 28 mars 2018 et intitulé « *Donner du sens à l'intelligence artificielle – Pour une stratégie nationale et européenne* ».

On peut également citer deux rapports d'information établis au nom de la commission des Affaires européennes du Sénat :

– celui de MM. André Gattolin, Claude Kern, Cyril Pellevat et Pierre Ouzoullas, publié le 31 janvier 2019 et portant sur la stratégie européenne pour l'intelligence artificielle ;

– et, plus récemment, celui de M. André Gattolin, Mme Catherine Morin-Desailly, M. Cyril Pellevat et Mme Elsa Schalck, publié le 30 mars 2023, relatif à la proposition de législation européenne sur l'intelligence artificielle.

L'appréhension des défis posés par l'IAG nécessite donc au préalable de cerner et délimiter ce qu'est l'IA.

S'il existe plusieurs définitions de l'IA, vos rapporteurs ont constaté un consensus pour considérer que celle-ci est un procédé qui permet à une machine de produire un **résultat intellectuel** reproduisant ou simulant l'intelligence humaine. De ce point de vue, l'IA est présente depuis longtemps dans la vie quotidienne des citoyens. L'expression « intelligence artificielle » a été créée dans les années 1950, après l'apparition des premiers programmes informatiques, et une simple calculatrice peut être considérée une forme d'IA basique.

Dans certains cas, le résultat produit peut aboutir à une prise de **décision** par la machine elle-même (par exemple, le freinage d'un véhicule autonome en cas de détection d'un danger).

L'IA se distingue toutefois d'un algorithme de base, dans la mesure où elle est capable de répondre à une situation nouvelle à partir de situations antérieures, simulant ainsi l'apprentissage humain, au point qu'un même *stimulus* peut parfois entraîner des réponses distinctes.

Lors de sa conception, l'IA suppose au minimum trois ingrédients de base pour fonctionner : des **capacités de calculs,** des **algorithmes** et des **données**. Elle est ensuite « entraînée » avec ses données, qui lui permettent de connaître un grand nombre de cas de figure et de rapprocher une nouvelle situation (images, *prompt*, comportement) et des situations déjà existantes afin, éventuellement, de réagir en conséquence.

Le choix des données et la manière dont le système est programmé et entraîné ont un effet direct sur les réponses fournies par les IA, d'où l'intérêt de réfléchir à la manière dont la conception des IA doit être encadrée.

Les capacités de calculs de l'IA doivent permettre de dépasser la performance d'une intelligence humaine (tout comme les outils et l'industrie ont permis, au cours de l'histoire, d'améliorer les performances physiques humaines et d'accroître les capacités de production). **L'algorithme** permet de remplacer l'intelligence humaine dans le processus d'élaboration du résultat de l'IA.

Par exemple, dans le domaine médical, une IA est capable d'analyser des centaines de milliers de radiographies et d'observer des corrélations entre certains symptômes et certaines maladies. Il revient ensuite à l'humain d'apporter un degré d'analyse supplémentaire.

L'IA ne semble pas pouvoir rendre l'humain inutile, mais au contraire lui permettre de se concentrer sur les tâches pour lesquelles son cerveau est plus performant que l'IA.

Les données sont fournies à l'IA par les concepteurs (modèle fermé) et parfois par l'utilisateur de l'IA (modèle ouvert). Elles sont traitées par l'IA grâce à l'algorithme. Les modèles ouverts sont plus particulièrement exposés au risque dit « *d'hallucination* », puisqu'ils réintègrent des données qu'ils ont eux-mêmes produites, parfois par erreur.

Les développements récents de l'IA sont dus, pour l'essentiel, à l'amélioration des capacités de calcul des microprocesseurs et à la sophistication croissante des algorithmes. En outre, la diffusion d'internet et la disponibilité de données en nombre considérable (les mégadonnées ou *big data* en anglais) offrent un **cadre renouvelé favorable aux progrès exponentiels de l'IA.**

Ces progrès, pour partie déjà visibles, s'apparentent pour beaucoup à une « *révolution* » (en ce sens, voir le rapport du 31 janvier 2019 [(1)] de la commission des Affaires européennes du Sénat, qui recommande de « *préparer la quatrième révolution industrielle* »).

Vos rapporteurs ont cependant été alertés sur certaines limites à ce développement, l'IA étant très gourmande en énergie. Cela est particulièrement vrai pour l'IAG, qui nécessite des millions de calculs pour produire un contenu vraisemblable. Le coût et l'impact environnemental de l'IAG peuvent dès lors, à terme, être considérables. Les progrès de l'IAG, bien qu'exponentiels, ne seront pas infinis.

---

(1) Rapport d'information du Sénat n° 279 de MM. André Gattolin, Claude Kern, Cyril Pellevat et Pierre Ouzoulias au nom de la commission des Affaires européennes, sur la stratégie européenne pour l'intelligence artificielle, déposé le 31 janvier 2019.

• **En second lieu, vos rapporteurs ont été particulièrement attentifs à identifier les défis spécifiques** posés par la diffusion inéluctable de l'IAG dans la société. Pour cela, ils ont auditionné des administrations centrales, des autorités administratives et des représentants du monde de l'entreprise. Au-delà des technologies, c'est la question de la régulation des usages qui se posera rapidement, l'accès à l'IAG étant appelé à devenir de plus en plus simple et bon marché.

Ils ont noté un contraste marqué dans les approches entre les représentants d'entreprises, d'une part, et les administrations, d'autre part. Alors que les premiers semblent déjà envisager un avenir où l'IAG transformera radicalement l'économie, les administrations adoptent une posture plus attentiste. Globalement, les entités administratives consultées, à l'exception de la CNIL, ont manifesté peu d'enthousiasme et d'initiative pour contribuer à la mission d'information. Leurs réponses aux questionnaires étaient plutôt convenues et leur réflexion sur les impacts de l'IAG en est encore à ses prémices.

S'agissant des entreprises, vos rapporteurs sont convaincus que la France dispose de nombreux atouts pour réussir dans le secteur de l'IAG. Elle dispose des ressources humaines et d'un écosystème de grande qualité.

Vos rapporteurs se sont également rendus à Bruxelles pour étudier le règlement européen en préparation, avant que celui-ci soit finalement adopté au début du mois de décembre 2023. Ils en sont revenus convaincus que les défis de l'IAG sont mondiaux et devaient donc être traités tant à l'échelle européenne qu'à l'échelle nationale.

Le défi de la régulation à venir sera de parvenir à encadrer les usages de l'IAG et à s'assurer que son développement demeure compatible avec les principes européens, sans entraver l'innovation et l'émergence de nouveaux acteurs, français ou européen.

Le présent rapport est construit en deux parties, **l'une consacrée aux enjeux de l'élaboration d'une régulation au niveau européen** et **l'autre aux questions clés à traiter au niveau national**.

Au préalable, **une introduction générale, plus transversale, présente les défis généraux** auxquels sont confrontés, partout dans le monde, les pouvoirs publics face aux opportunités et aux risques de l'IAG.

*

* *

GLOSSAIRE

Algorithme : ensemble de règles opératoires propres à un calcul. Il constitue une suite d'étapes permettant d'obtenir un résultat à partir d'éléments fournis en entrée.

***Big data* ou mégadonnées** : ensemble très volumineux de données, ingrédient indispensable pour le développement des systèmes d'IAG.

***Chatbot* ou agent conversationnel** : modèle d'IAG permettant de converser dans un langage humain.

***Deepfake* ou hypertrucage** : création de fausses images ou vidéos, souvent à partir de la fusion d'images existantes, et visant à tromper celui qui les regarde en les faisant passer pour vrai.

***Data mining* ou fouille de données** : extraction d'un savoir ou d'une connaissance à partir de grandes quantités de données, par des méthodes automatiques ou semi-automatiques.

***Opt out* ou droit d'opposition** : possibilité offerte aux propriétaires de leurs données personnelles ou de données protégées par le droit d'auteur de s'opposer à ce qu'elles soient utilisées pour l'entraînement de systèmes d'IAG.

Entraînement : phase de la création d'une IAG consistant à lui fournir un grand nombre d'informations pour qu'elle soit en mesure de fournir la meilleure réponse possible.

***Fine tuning* ou réglage fin** : approche d'apprentissage dans laquelle un modèle pré-entraîné est formé sur de nouvelles données pour en spécifier l'usage.

Hallucination : situation dans laquelle une IAG fournit des réponses erronées ou inadaptées, en raison de biais dans ses calculs ou dans ses bases de données.

***Machine learning* ou algorithme apprenant :** algorithme conçu de telle sorte qu'il peut découvrir lui-même les opérations à suivre et progresser dans le temps à partir de son expérience.

Modèle de fondation ou *foundation model* : modèle d'IA entraîné sur une grande quantité de données et pouvant être adapté à un large éventail de tâches en aval.

Modèle de langage étendu ou *Large language model* : type de programme d'IA capable de reconnaître et de générer du texte.

Open data : ouverture et mise à disposition des données produites et collectées, notamment par les services publics.

Open source : mise à disposition de logiciels ou programme pouvant être librement redistribué, dont le code source est accessible et qui peut être réutilisé pour travaux dérivés.

Phishing **ou hameçonnage** : forme d'escroquerie qui se déroule sur internet, consistant à récupérer des données personnelles par la tromperie, puis à les utiliser de manière malveillante

Prompt **ou commande** : phrase ou texte court entré par un utilisateur pour initier un échange avec une IAG et servant à guider la réponse en fonction de l'objectif poursuivi.

Renforcement humain : intervention humaine dans l'interaction entre l'usager et l'IAG pour en améliorer la fiabilité.

Robustesse : capacité d'une IAG à fournir des réponses fiables dans le temps et à résister aux tentatives de détournement.

INTRODUCTION GÉNÉRALE : LES POUVOIRS PUBLICS FACE AUX OPPORTUNITÉS ET AUX RISQUES DE L'INTELLIGENCE ARTIFICIELLE GÉNÉRATIVE (IAG)

Une constante fondamentale ressort des nombreuses auditions menées par vos rapporteurs : **l'intelligence artificielle générative (IAG) est intrinsèquement neutre, ni bénéfique ni préjudiciable**. Son impact sur la société dépend entièrement de l'application qui en sera faite.

Ce nouvel outil est source à la fois de dangers et de bénéfices pour les citoyens individuellement et pour la société dans son ensemble.

Les pouvoirs publics doivent en mesurer les opportunités pour ne pas freiner cette nouvelle étape de la révolution numérique. Ils doivent, dans le même temps, appréhender l'éventail des risques inhérents à la diffusion inéluctable de l'IAG pour préparer la société à les maîtriser.

Vos rapporteurs sont convaincus qu'**un constat partagé sur ces opportunités et ces risques est un préalable indispensable pour que les pouvoirs publics surmontent le dilemme auxquels ils font face aujourd'hui et qui consiste à accompagner l'essor de l'IAG tout en protégeant les citoyens**.

À l'heure où, partout dans le monde, émergent des voies concurrentes pour réguler l'IA dont l'IAG, l'Europe et la France doivent rechercher le **juste équilibre entre soutien à l'innovation et réglementation**.

I. LES OPPORTUNITÉS DE CETTE NOUVELLE ÉTAPE DE LA RÉVOLUTION NUMÉRIQUE

La diffusion des systèmes d'IAG ouvre un champ vaste d'opportunités dans divers secteurs et domaines.

A. UN ACCROISSEMENT DE LA PRODUCTIVITÉ

Tout d'abord, l'un des avantages attendus les plus manifestes de l'IAG réside dans **l'accroissement de la productivité, y compris dans le fonctionnement des administrations**.

Les gains de productivité pourraient résulter de **l'automatisation des tâches cognitives routinières, mais aussi de la possibilité d'analyser rapidement de grands volumes de données**. Les promoteurs des IAG alimentent l'espoir, peut-être présomptueux, que les employés pourraient à l'avenir se focaliser principalement sur des missions plus stratégiques et sophistiquées. Plus modestement, ces évolutions technologiques pourront permettre de simplifier certaines tâches ou de les rendre moins pénibles, par exemple avec des solutions bureautiques de meilleure qualité ou des moteurs de recherche plus précis.

Le secteur tertiaire connaîtrait alors des gains de productivité, en particulier pour les emplois de bureau, comparables à ceux connus autrefois dans les secteurs agricoles et industriels.

Certes, une telle prédiction avait déjà été formulée avec l'arrivée de l'informatique, puis moquée par l'économiste Robert Solow qui avait constaté en 1987 qu'on pouvait « *voir les ordinateurs partout, sauf dans les statistiques de la productivité* » (1). Les gains de productivité de l'informatique n'ont été constatés qu'une à deux décennies après la diffusion des ordinateurs. Il n'est pas écrit que les gains de productivité attendus des IAG soient plus rapides.

Qu'elles soient à court terme ou à plus long terme, les perspectives en matière de gains de productivité suscitent des inquiétudes quant à la suppression éventuelle d'emplois tant dans le secteur public que le secteur privé.

En ce qui concerne le secteur privé, les experts auditionnés par la mission ont formulé deux types de réponses aux craintes exprimées concernant les emplois.

En premier lieu, ils ont unanimement objecté que, selon le principe schumpetérien de destruction-créatrice, de nouveaux emplois et nouveaux secteurs d'activité devraient émerger. Ils ont fait observer, qu'alors même que l'IAG est peu diffusée, des profils tels que « *Data Scientist* », « *Ingénieur en IA* » ou même des spécialistes en éthique de l'IA sont de plus en plus recherchés pour pourvoir de nouveaux postes. Ces métiers qui viennent compléter les missions des informaticiens et programmeurs témoignent de l'attention portée désormais par les entreprises à leur responsabilité quant aux usages de leurs produits et à la protection des données personnelles qu'elles collectent et exploitent.

En second lieu, comme l'a souligné par exemple M. Alain Goudey lors de la première audition de la mission, l'IAG est moins susceptible de supprimer des emplois que de redéfinir les compétences nécessaires pour les occuper. Le risque est moins celui de la destruction d'un emploi que celui de la perte d'employabilité des personnes qui n'auront pas été formées aux usages de l'IAG. Autrement dit, l'IAG va nécessiter de nouvelles compétences au même titre que l'arrivée de l'informatique, ce qui ouvre encore un nouveau gisement d'emplois dans le domaine de la formation.

Si les gains de productivité attendus se concrétisent, **le secteur public devrait en être à terme l'un des premiers bénéficiaires**. Le secteur public relève, en effet, pour l'essentiel du secteur tertiaire et produit un grand nombre de décisions et de réglementations.

À condition que les processus de décision restent exclusivement à la main et sous le contrôle des agents, l'IAG peut grandement améliorer l'efficacité de l'action administrative. Elle peut automatiser le tri et l'analyse de documents, optimiser la gestion des ressources et même aider au travail de rédaction et de mise en forme des documents administratifs. Un bon usage des IAG doit permettre aux usagers d'obtenir des réponses plus rapides et plus fiables à leurs sollicitations.

---

(1) « You can see the computer age everywhere, but in the productivity statistics *», New York Times Book Review.*

Vos rapporteurs ont observé que l'usage de l'IAG dans l'administration en est encore à ses balbutiements et ne fait l'objet que de rares expérimentations (1). À l'avenir, on peut imaginer que certaines administrations, telles que l'administration fiscale, exploitent davantage ces technologies pour réaliser de nouveaux gains de productivité. D'autres administrations pourraient, en revanche, utiliser l'IAG pour renforcer leur expertise stratégique et rediriger les effectifs vers des missions plus importantes, notamment dans les domaines de la santé et de l'éducation.

Pour l'instant, ces scénarios appartiennent encore au domaine de la science-fiction. Mais il est important de souligner que **l'utilisation des gains de productivité espérés relèvera exclusivement de choix politiques**. L'IAG pourrait, selon ces choix, soit être source de destructions d'emplois, soit constituer une opportunité pour allouer de nouveaux moyens à des tâches plus essentielles.

À court terme, ces inquiétudes ne sont toutefois pas infondées et elles impliquent une vigilance de la part des pouvoirs publics qui doivent, dans la mesure du possible, anticiper les besoins en formation et en reconversion. Ce domaine dépasse le champ de la mission.

B. UN ENRICHISSEMENT DE LA CRÉATIVITÉ HUMAINE

Ensuite, il ne faut pas négliger le potentiel de l'IAG à enrichir la créativité humaine plutôt qu'à la supplanter.

Dans le domaine de la recherche et du développement, l'IA peut, grâce à la quantité de données qu'elle est capable de traiter, proposer des solutions novatrices à des problèmes complexes, accélérant ainsi le cycle d'innovation.

Dans le domaine de l'art, les algorithmes peuvent générer des esquisses ou des compositions musicales, servant de point de départ pour de nouvelles créations. De même que la photographie a changé la manière de peindre, il est prévisible que de nouveaux courants artistiques vont se déployer, stimulés par les IAG. On en perçoit les premières manifestations dans le domaine de la production vidéo. L'IAG peut offrir de nouveaux décors, de nouveaux personnages ou de nouveaux effets spéciaux à moindre coût.

Enfin, plus largement, l'IAG démocratise la capacité à créer dans des domaines qui supposent actuellement des prérequis techniques importants en matière de codage informatique. L'IAG offrira certainement à chacun la possibilité de devenir créateur de jeux vidéo ou de sites internet.

Pour atteindre cet objectif, les pouvoirs publics doivent se montrer particulièrement vigilants quant à l'origine des IAG utilisées et à la qualité des données utilisées et des conditions de leur entraînement.

---

(1) Voir deuxième partie (I. B.).

C. DE NOUVELLES RESSOURCES POUR COMMUNIQUER, FORMER ET ÉDUQUER

● L'IAG peut apporter de **nouvelles ressources pour communiquer**, que cela soit à titre privé ou professionnel. L'aide apportée à la mise en forme, à l'amélioration des formules stylistiques ou à la correction syntaxique et lexicale dépasse largement les outils actuellement présents dans les suites bureautiques disponibles. Certains éditeurs ont d'ailleurs annoncé que les prochaines versions de leurs suites bureautiques intégreront de l'IAG.

Les progrès en matière de traduction ont été impressionnants grâce à l'IAG et son mode de fonctionnement probabiliste qui consiste à calculer la probabilité des mots les plus pertinents à retenir. Comme l'un des experts l'a énoncé de manière un peu provocatrice lors de son audition, on observe que la traduction automatique n'a jamais autant progressé depuis que « *la linguistique a été remplacée par le calcul mathématique* ». Le calcul mathématique semble plus performant que le codage minutieux des règles de grammaire et des définitions.

Les nouvelles opportunités offertes en matière de traduction vont permettre aux citoyens, aux étudiants, aux chercheurs, aux journalistes d'accéder plus facilement à des ressources en langue étrangère.

L'IAG sera aussi particulièrement utile aux étrangers ou d'une manière générale aux personnes ayant des difficultés avec le maniement de la langue écrite. À titre d'exemple, un restaurateur étranger peut aisément s'appuyer sur l'IAG pour traduire ses menus, répondre aux mails et commentaires de ses clients, ou encore pour élaborer une publicité sans passer par un intermédiaire, sous réserve d'en contrôler le résultat.

● L'IAG peut également jouer un rôle important **en matière d'éducation et de formation**.

Les IAG disponibles permettent d'ores et déjà, et facilement, à leurs utilisateurs de progresser dans l'apprentissage de certaines disciplines. Les IAG peuvent en effet converser avec les utilisateurs, corriger leurs erreurs et leur proposer des exercices personnalisés.

L'IAG dispose encore d'un fort potentiel inexploité pour transformer l'éducation, en mettant de nouvelles ressources personnalisées à disposition des éducateurs et en apportant une aide complémentaire, voire ludique, aux élèves et aux parents, par exemple au moyen de tutorats adaptés.

Au final, l'IAG présente une multitude d'opportunités qui ont le potentiel de transformer notre manière de travailler, d'apprendre, de créer et d'administrer. Pour autant, ce panorama optimiste ne doit pas occulter les dangers de l'IAG.

*

* *

II. L'ÉVENTAIL DES RISQUES INHÉRENTS À LA DIFFUSION INÉLUCTABLE DE L'IAG

Les discours les plus alarmistes décrivent les futures générations d'IAG comme des outils qui faciliteront les cyberattaques massives, les manipulations d'opinions publiques à grande échelle ou encore des attentats terroristes, sans oublier les millions de destructions d'emplois et un appauvrissement culturel généralisé.

Sans tomber dans ce catastrophisme, de nombreux risques sont d'ores et déjà apparents avec les générations actuelles de l'IAG et peuvent être décrits.

Ces risques sont connus, identifiés et bien documentés. Ils concernent tant l'utilisateur de l'IAG que le destinataire du contenu produit qui n'est pas toujours la même personne que l'utilisateur (risques individuels). Certains risques peuvent présenter un effet systémique et concerner la société dans son ensemble (risques collectifs).

A. LES RISQUES INDIVIDUELS

1. Les risques pour la vie privée

Les risques pour la vie privée sont liés à l'usage important de données qui est nécessaire au développement, puis au fonctionnement d'un système d'IAG.

En amont, **de nombreuses données personnelles peuvent être utilisées pour l'entraînement des systèmes d'IAG**. En aval, l'utilisateur du système d'IAG peut également fournir de nombreuses données personnelles pour obtenir la production d'un contenu qui lui convient (par exemple, pour une personne qui souhaite éditer un CV ou un patient qui sollicite une interprétation par l'IAG d'analyses médicales).

Les IAG vont donc accroître les risques qui pèsent sur le respect de la vie privée à l'ère du numérique dès lors que l'utilisation de ces données reste méconnue. Certains experts considèrent qu'il s'agit simplement de données utilisées de manière anonyme à des fins statistiques, tandis que d'autres estiment que les données, notamment fournies par l'intermédiaire des demandes formulées (*prompts* [1]), peuvent être analysées et interprétées par l'IAG pour être réutilisées à d'autres fins (commerciales, renseignement…).

2. Les risques de non confidentialité ou de fuite de données

Au-delà de ses données personnelles, l'utilisateur est incité à fournir de nombreuses données aux systèmes d'IAG afin d'améliorer le résultat généré. Le risque, pour ce dernier, est de divulguer des données sensibles le concernant ou bien concernant son employeur ou son client s'il agit dans un cadre professionnel. Il pourrait alors mettre en péril son organisation (entreprise ou administration) et sa carrière professionnelle.

---

(1) Un prompt est un mot anglais qui désigne toute commande écrite envoyée à une « intelligence artificielle » spécialisée dans la génération de contenu, comme du texte ou des images.

3. Le risque de biais et d'influences extérieures

L'utilisateur risque aussi d'obtenir un contenu biaisé ou influencé, volontairement ou involontairement, par les concepteurs de l'IAG. Deux facteurs au stade de la conception du système d'IAG expliquent ces risques de biais ou d'influences extérieures : le renforcement humain et les données d'entraînement.

Pour améliorer la fiabilité du contenu produit, les développeurs procèdent en effet à un « *renforcement humain* » lors de la phase d'apprentissage. Par exemple, des individus vont attribuer une note, sur une échelle de 1 à 10, évaluant la qualité du contenu produit selon différents critères subjectifs. Sur cette base, l'IA va pouvoir être « *entraînée* » à produire un contenu plus fiable ou de meilleure qualité selon des critères définis par l'homme.

Les experts auditionnés par la mission ont tous souligné que **ce renforcement humain est susceptible d'introduire de nombreux biais**, y compris de manière parfaitement involontaire de la part des concepteurs d'une IAG.

Les données d'entraînement peuvent aussi être biaisées. Le contenu généré est nécessairement influencé par le corpus qui alimente en données l'IAG. Ainsi, un corpus essentiellement anglophone peut conduire à la génération de contenus qui accordent une prévalence aux points de vue anglo-saxons.

Par exemple, lors d'un test réalisé par la mission le 25 mai 2023, à la question « *Saddam Hussein disposait-il d'armes de destructions massives ?* », ChatGPT concluait après un long développement qu'il n'y a pas de consensus clair sur la question alors même qu'il n'y a désormais plus de doute sur le caractère mensonger de cette affirmation [(1)].

À n'en pas douter, un robot entraîné sur la base de textes francophones produirait une réponse plus catégorique et négative à la même question. D'ailleurs, dans la version française de Wikipédia, la thèse selon laquelle Saddam Hussein aurait détruit ses ADM juste avant l'invasion était, à la même date, classée dans une section intitulée « *Théories du complot* » [(2)] (section qui n'existait pas dans la version anglaise).

---

(1) Contenu généré par Chatgpt le 25 mai 2023 : « Il est important de noter que cette question a été l'objet de débats politiques intenses et de controverses, et différentes opinions existent quant à la nature et à l'étendue des ADM en Irak à l'époque de Saddam Hussein. Certains soutiennent que Saddam Hussein avait effectivement des ADM, mais qu'elles ont été dissimulées ou détruites avant l'invasion, tandis que d'autres estiment qu'il n'y avait aucune base solide pour justifier l'intervention militaire. En définitive, il n'y a pas de consensus clair sur la question de savoir si Saddam Hussein avait effectivement des ADM au moment de l'invasion de l'Irak en 2003. »

(2) Extrait de la page Wikipédia « Armes de destruction massive en Irak » au 25 mai 2023 : « En 2005, l'administration Bush reconnaît que les armes de destruction massive irakiennes n'existaient pas *[…]* Colin Powell exprimera deux ans plus tard son « amertume » : interrogé sur ABC, il explique que cette présentation, en grande partie fausse, fait « tache » dans sa carrière *[…]* Le 27 novembre 2009, devant la commission Chilcot, William Ehrman, haut responsable au ministère britannique des Affaires étrangères entre 2000 et 2002, déclare que Tony Blair savait que l'Irak n'avait plus d'ADM avant d'envoyer ses troupes dans le pays ».

Les spécificités et singularités du modèle français justifient une attention toute particulière eu égard au risque de biais ou d'influences extérieures résultant du choix des données et de la méthode d'entraînement. Les IAG pourraient à terme fragiliser des principes spécifiques à la France, par exemple la notion de laïcité, au profit de positions intellectuelles davantage partagées dans le monde.

4. Les risques d'erreur et d'hallucination

Malgré le renforcement humain, il existe un risque important que le contenu produit par les IAG contienne des erreurs.

Cela s'explique par le fait que **l'IAG n'est pas en mesure de distinguer le vrai et le faux compte tenu de son mode de fonctionnement qui repose sur des calculs probabilistes**. L'IAG va produire un contenu probable, un contenu qui ne dépareille pas avec son corpus d'entraînement.

En première analyse, on pourrait penser que ce risque n'est pas plus important que le risque d'erreur relatif aux contenus librement accessibles et modifiables sur internet. Tel n'est pourtant pas le cas. Le risque d'être induit en erreur est bien plus important avec les IAG notamment dans les domaines où les données initiales disponibles sont réduites.

En effet, l'IAG étant programmée pour produire « *coûte que coûte* » un contenu, celle-ci peut générer un résultat en partie faux pour répondre à la requête de l'utilisateur. À partir d'un résultat partiellement faux, le contenu peut progressivement « dériver » pour devenir complètement erroné. Ceci s'explique par le mode de fonctionnement probabiliste de l'IAG.

L'exemple le plus typique est celui relatif à la production de biographies de personnes qui, sans disposer d'une grande notoriété, sont présentes dans les données disponibles de l'IAG. À partir d'une information, **l'IAG peut imaginer une biographie, certes probable, mais très loin de la réalité**.

Les experts désignent ce risque comme un « *risque d'hallucination* ». Ce **risque d'hallucination existe dans tous les domaines où le champ du savoir est encore réduit**. Il peut donner l'impression qu'il existe des réponses documentées à toutes les problématiques et requêtes. Le risque d'hallucination est dangereux pour l'utilisateur notamment s'il utilise l'IAG comme un moteur de recherche.

5. Le risque de tromperie

Ce risque individuel concerne plus particulièrement le destinataire du contenu lorsqu'il n'est pas lui-même l'utilisateur du système d'IAG. Le destinataire peut être trompé quant à la question de savoir quel est le véritable auteur du contenu qu'il reçoit.

Si l'on peut admettre que le spectateur d'un film accepte, par convention, de visualiser des plans produits par une IAG, la réponse est moins évidente pour l'usager d'un service public ou le client d'un avocat qui reçoit un mail en réponse à sa question.

La problématique est la même dans toutes les conventions *intuitu personae*. Tout dépend du fait de savoir si le destinataire considère que l'auteur du contenu est une qualité substantielle de son consentement. En droit civil, la qualité substantielle est, en effet, celle qui détermine le consentement du contractant. Il n'est pas évident, *a priori*, de savoir si le destinataire d'un contenu fait du mode de production de ce contenu une qualité substantielle.

Le risque est aussi, pour le destinataire, de ne plus savoir s'il interagit avec une personne humaine ou non. Ce risque est particulièrement présent dans le domaine de la consommation lors d'interactions avec un service clientèle.

Enfin, cette incertitude peut avoir des répercussions sur l'imputation de la responsabilité civile en cas de dommage en lien avec l'utilisation du système d'IAG ou du contenu produit par celui-ci.

6. Le risque de fraude à l'IAG

Le risque d'être victime d'une fraude à l'IAG est moins connu, mais déjà avéré. Il s'agit, en quelque sorte, de la vente frauduleuse du développement d'une fausse IAG. Ce procédé consiste à tromper le destinataire et à lui faire croire que le contenu produit est le fruit d'une IAG en développement, alors qu'il a été conçu en grande partie par un travail humain. Le but de la fraude est d'inciter le destinataire à investir dans le développement d'un système d'IAG prétendument prometteur.

Cette fraude n'est pas sans rappeler le Turc mécanique, une machine que son prétendu concepteur présentait à la fin du XVIII^{ème} siècle comme un automate capable de jouer des parties d'échecs (en réalité, un véritable joueur d'échecs dissimulé à l'intérieur du meuble décidait des coups à jouer).

B. LES RISQUES COLLECTIFS

1. Les risques d'usages détournés

La démocratisation de l'IAG peut faire craindre une multiplication de comportements délictueux, rendus plus aisés en raison de la production de contenu synthétique vraisemblable sans prérequis techniques de la part de l'utilisateur.

Les usurpations d'identité, la création de faux sites internet ou courriers électroniques en vue de réaliser des opérations de type hameçonnage (« phishing ») ou d'une manière générale divers types d'escroquerie seront plus faciles à mettre en œuvre et plus difficiles à détecter pour les victimes. L'IAG est en mesure de reproduire fidèlement un courrier de l'administration, voire de s'approprier la voix d'une personne pour ensuite appeler sa banque ou ses proches en se faisant passer pour elle.

Au-delà de comportements aujourd'hui appréhendés par le droit pénal, l'IAG peut aussi faire l'objet d'usages non sanctionnés mais socialement largement réprouvés comme le fait, par exemple, de se prétendre auteur du contenu généré sans en avertir les destinataires (fausse lettre de motivation, faux rapport de stage, fausse consultation juridique ou tout autre document produit par une IAG que s'approprie l'utilisateur sans avertir le destinataire).

2. Les risques sociaux

● La crainte principale, en matière sociale, est liée au risque de suppression d'emplois consécutive aux gains de productivité permis par l'IAG.

En droit du travail, des **mutations technologiques** peuvent justifier des licenciements pour motifs économiques (2° de l'article L. 1233-3 du code du travail). Les professionnels indépendants qui effectuent des tâches intellectuelles sont aussi particulièrement menacés dans certains secteurs, notamment celui du divertissement.

La grève des scénaristes à Hollywood en mai 2023, qui n'est pas sans rappeler la révolte au XIX^{ème} siècle des canuts lyonnais, qui protestaient contre l'arrivée de nouvelles machines à tisser, est l'une des premières à avoir été motivée principalement par le sujet de l'IAG. Après cinq mois de conflits, les scénaristes – qui sont pour l'essentiel des travailleurs indépendants – ont obtenu un accord prévoyant des mesures de protection contre l'arrivée des robots écrivains, notamment le maintien d'un nombre minimal d'humains dans la conception des spectacles de cinéma et de télévision.

En France, la Sacem (société des auteurs, compositeurs et éditeurs de musique) a mis en œuvre en octobre 2023 son droit d'opposition (*opt out*) [(1)]. Désormais, les activités de fouilles de données (*data-mining*) sur les œuvres du répertoire de la Sacem par les entités développant des outils d'intelligence artificielle devront faire l'objet de son autorisation préalable.

Les craintes sur les emplois se sont également concrétisées dans le secteur de la veille médiatique, ou encore pour certaines tâches de traduction.

La question se pose de savoir si la vitesse des progrès attendus de l'IAG est de nature à jouer sur l'acceptation sociale de cette nouvelle technologie.

● Au titre des risques sociaux, il faut également citer les risques liés à l'exploitation à l'étranger de travailleurs faiblement qualifiés.

---

(1) Le droit d'opposition est reconnu par l'article L. 122-5-3 du code de la propriété intellectuelle et permet de rendre inopérante l'exception, prévue par ce même article, autorisant la fouille de données dans le cadre des techniques d'analyse automatisée de données inhérente aux outils d'intelligence artificielle.

Le développement des IAG nécessite, en effet, un important travail humain lors de la phase d'apprentissage. Or, l'entraînement des IAG peut souvent être sous-traité sous forme de micro-tâches à des personnes peu qualifiées et peu rémunérées. Par exemple, des milliers d'heures de travail sont nécessaires pour apprendre à un système d'IAG à reconnaître un chat parmi des images d'animaux.

Il existe dès lors un risque que ces tâches soient en grande partie localisées dans les pays où les droits des travailleurs et la protection sociale sont les plus faibles.

3. Le risque d'amplification des discriminations

● De manière générale, l'intelligence artificielle présente des risques de reproduction, voire d'amplification des discriminations. Fondé sur l'apprentissage de régularités statistiques, un système d'IA va incorporer des biais présents dans les données d'entraînement.

Cela n'est pas problématique en soi selon l'usage qui est fait de la « *discrimination algorithmique* » produite par l'IA. M. Hugues Bersini a expliqué, lors de son audition, qu'il était vain d'interdire à une IA de produire un résultat qui serait qualifié, au regard de la loi, de discriminatoire. Prenant l'exemple des candidatures à une formation universitaire, il a expliqué qu'une IA qui évaluerait la probabilité de réussite ou d'échec d'un étudiant peut être utile même si ses prédictions font apparaître des discriminations (par exemple, en calculant qu'un étudiant aurait moins de chance de réussir qu'une étudiante dans le domaine concerné). L'objectivation de discriminations par une IA peut, en effet, permettre aux pouvoirs publics de prendre les mesures correctives pour mieux les prévenir.

À l'inverse, **une discrimination algorithmique qui n'est pas identifiée et perçue par l'utilisateur du système d'IAG peut être dangereuse et contribuer à accroître les discriminations dans la vie réelle.**

● Les risques sont également très forts, et plus difficiles à réglementer, pour l'IAG. Avant renforcement humain, une IAG qui produit des contenus visuels a tendance à reproduire, les discriminations ou les stéréotypes présents dans les données d'entraînement, par exemple les biais de genre. Le renforcement humain ne permet pas d'éliminer totalement les biais issus des données d'entraînement, dans la mesure où les personnes qui procéderont à ce renforcement peuvent elles-mêmes avoir des biais et des préjugés.

4. Les risques sur l'information

Les risques sur l'information sont de plusieurs ordres.

Tout d'abord, la réalisation aisée d'hypertrucages de type « deepfake » pourrait multiplier les risques de diffusion de fausses nouvelles.

Ensuite, les avancées de l'IAG pourraient conduire à une diminution de la visibilité et de l'audience des médias en ligne. Il est envisageable que les moteurs de recherche se servent de l'IA afin de fournir des réponses aux questions des

utilisateurs, ce qui aurait pour effet de restreindre la mise en avant des sources d'information tierces telles que la presse ou les médias. La rétribution en ligne des contenus de magazines et de journaux pourrait se retrouver menacée dans son ensemble.

Enfin, la montée en puissance des contenus générés par des systèmes d'IAG pourrait ébranler la confiance du public et affecter la valeur de l'écosystème économique des médias en ligne.

5. Le risque d'attrition et d'appauvrissement culturel

Sur le long terme, la diffusion de l'IAG peut être à l'origine d'un appauvrissement culturel à raison d'une attrition progressive des contenus d'origine humaine.

En effet, les prochaines générations d'IAG peuvent être entraînées sur des données elles-mêmes issues de l'IAG. Il s'ensuit que les contenus produits par IAG pourraient nourrir les contenus produits par d'autres IAG. Sans davantage d'encadrement, elles tendent à s'approprier des contenus protégés par le droit d'auteur sans les rémunérer à leur juste valeur.

Le risque, à raison du mode de fonctionnement probabiliste de ces systèmes, est d'aboutir à une relative uniformisation des types de contenus. Cela augmentera aussi le risque d'hallucination qui pourrait brouiller davantage encore la frontière entre le vrai et le faux.

6. Les risques environnementaux

Les IAG nécessitent des capacités de calculs importantes et donc une abondante consommation électrique. L'impact énergétique et environnemental de ces technologies n'est donc pas à négliger et constitue même une limite à leur développement prétendument exponentiel.

Pour mémoire, selon les sources, le numérique représente d'ores et déjà 3 à 4 % des émissions de gaz à effet de serre (GES).

*

* *

En conclusion, il s'agit de surmonter les dilemmes auxquels sont confrontés les pouvoirs publics

Nul ne peut ignorer les opportunités offertes par l'IAG. Dans le même temps, l'éventail considérable des risques suscite des préoccupations légitimes tant chez les décideurs publics que les citoyens.

Les pouvoirs publics sont dès lors confrontés à un **double dilemme**.

Le premier est de nature **économique, voire géopolitique** dans un contexte de compétition internationale. Il s'agit de savoir comment réglementer ou réguler l'IAG sans freiner l'innovation. Autrement dit, il convient de rechercher le juste équilibre entre encadrement de l'IAG et stimulation de l'innovation.

Le second a trait à la **protection des personnes et de la société**. Comment protéger au mieux les citoyens tout en accompagnant l'essor de l'IAG ?

Les enjeux de l'IAG étant mondiaux, les réponses à apporter à ces dilemmes devront tenir compte du contexte international et s'inscrire pour une large part dans un cadre européen (première partie).

Concomitamment, au plan national, il convient dès à présent de préparer la société aux implications de l'IAG (seconde partie).

*

* *