N° 2207

ASSEMBLÉE NATIONALE

CONSTITUTION DU 4 OCTOBRE 1958

SEIZIÈME LÉGISLATURE

Enregistré à la Présidence de l'Assemblée nationale le 14 février 2024

RAPPORT D'INFORMATION

DÉPOSÉ
en application de l'article 145 du Règlement

PAR LA COMMISSION DES LOIS CONSTITUTIONNELLES, DE LA LÉGISLATION ET DE L'ADMINISTRATION GÉNÉRALE DE LA RÉPUBLIQUE,

en conclusion des travaux d'une mission d'information [1]

sur les **défis** *de l'***intelligence artificielle générative** *en matière de* **protection** *des* **données personnelles** *et d'***utilisation** *du* **contenu généré**

ET PRÉSENTÉ PAR

MM. PHILIPPE PRADAL ET STÉPHANE RAMBAUD,

Députés

(1) La composition de cette mission figure au verso de la présente page.

La mission d'information sur les défis de l'intelligence artificielle générative en matière de protection des données personnelles et d'utilisation du contenu généré est composée de MM. Philippe Pradal et Stéphane Rambaud, rapporteurs.

SOMMAIRE

—

PAGES

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).

*

* *

PREMIÈRE PARTIE : L'APPROCHE EUROPÉENNE, UNE RÉPONSE AXÉE SUR LA RÉGULATION PAR LES RISQUES ET SUR LA CONFIANCE

L'approche européenne n'est pas centrée exclusivement sur l'intelligence artificielle générative (IAG), qui est un sous-domaine de l'intelligence artificielle (IA). Elle appréhende l'IA comme un ensemble à réguler.

La réponse à un enjeu aussi systémique que celui de l'IA ne peut se construire uniquement au niveau national. L'harmonisation des législations au niveau européen est indispensable pour mettre en place une régulation efficace sans pénaliser l'innovation.

L'Union européenne (UE) a déjà démontré sa capacité à encadrer l'activité des plateformes et à protéger les données personnelles des citoyens européens par des textes contraignants dont la portée est extraterritoriale. Ces législations s'appliquent toutefois imparfaitement à l'IA, notamment générative, d'où l'engagement de travaux sur cette thématique qui devraient prochainement aboutir.

Cette initiative européenne va dans le bon sens, car elle protègerait le secteur de l'IA, dynamique en France, et garantirait aux citoyens l'accès à des systèmes responsables dont le fonctionnement serait conforme aux valeurs démocratiques et au respect des droits fondamentaux.

Parmi l'éventail des modes de régulation possibles, certains présentent toutefois le risque d'être trop contraignants, défavorisant l'usage de l'IA par les États membres dans le domaine régalien et ne permettant pas l'émergence de nouveaux acteurs européens. Or, ces derniers accusent aujourd'hui du retard sur leurs concurrents non européens, notamment américains.

La concurrence internationale pour fixer les grands standards devant régir l'IA risque de porter préjudice aux intérêts des citoyens européens. Certains pays essaient déjà d'assouplir leur législation en matière de protection des données personnelles en vue d'attirer les investisseurs et les infrastructures indissociables de l'IA (serveurs, stockage des données…). Le marché européen est en mesure d'imposer ses standards, car il bénéficie d'un poids démographique et financier qui incite les grands acteurs à respecter nos règles.

La législation européenne doit donc être la plus équilibrée possible, quitte à confier aux États membres une marge d'appréciation supplémentaire pour qu'ils puissent affiner leur conception de l'IAG et créer un environnement le plus favorable possible à l'innovation. Ils pourraient s'appuyer pour cela sur des autorités de régulation déjà expérimentées et sur le droit souple, dit de *compliance*.

I. DES INITIATIVES AMBITIEUSES POUR PROTÉGER LES CITOYENS ET POUR L'ÉMERGENCE D'UNE INTELLIGENCE ARTIFICIELLE (IA) DE CONFIANCE

Inspirée par l'élaboration du Règlement général sur la protection des données (RGPD) mais consciente des limites du droit en vigueur pour encadrer l'intelligence artificielle (IA), l'UE a engagé un processus législatif sur cette question qui vient de s'achever.

Ce règlement repose sur de grands principes technologiquement neutres et une série d'obligations contraignantes adaptées à différents niveaux de risque. Ces négociations se déroulent sous la pression d'autres organisations et pays souhaitant également adapter leur droit aux défis de l'IA.

A. UN CADRE JURIDIQUE DÉJÀ CONTRAIGNANT QUI S'APPLIQUE IMPARFAITEMENT À L'IA

1. Une législation numérique étoffée au niveau européen

a. Le RGPD et les autorités de protection des données

Le **Règlement européen général sur la protection des données** (RGPD) est entré en vigueur le 25 mai 2018. Il vise à renforcer la protection des données personnelles des individus et à harmoniser les lois sur la protection des données au sein de l'UE.

Le RGPD fixe un certain nombre de **principes** :

– **Traitement licite, équitable et transparent** : les données personnelles doivent être traitées de manière légale, équitable et transparente pour la personne concernée ;

– **Finalité limitée** : les données personnelles ne doivent être collectées que pour des finalités spécifiques, explicites et légitimes, et ne peuvent pas être traitées de manière incompatible avec ces finalités ;

– **Minimisation des données** : les données personnelles collectées doivent être adéquates, pertinentes et limitées à ce qui est nécessaire pour les finalités pour lesquelles elles sont traitées ;

– **Exactitude** : les données personnelles doivent être exactes et, si nécessaire, tenues à jour. Des mesures appropriées doivent être prises pour garantir leur précision ;

– **Limitation de la conservation** : les données personnelles ne doivent pas être conservées plus longtemps que nécessaire pour les finalités pour lesquelles elles sont traitées ;

– **Intégrité et confidentialité** : les données personnelles doivent être traitées de manière à garantir leur sécurité, y compris la protection contre le traitement non autorisé ou illégal et contre la perte, la destruction ou les dommages accidentels.

Par conséquent, les responsables de traitement de données sont tenus à **différentes obligations** :

– **Recueil du consentement** : lorsque le traitement des données repose sur le consentement de la personne concernée, ce consentement doit être donné de manière libre, éclairée, spécifique et révocable ;

– **Droits des personnes concernées sur leurs données** : les personnes concernées ont des droits étendus en ce qui concerne leurs données personnelles, y compris le droit d'accès, de rectification, d'effacement (« droit à l'oubli »), de portabilité des données, et le droit de s'opposer au traitement ;

– **Notification des violations de données** : les responsables du traitement des données sont tenus de prévenir les autorités de contrôle et les personnes concernées en cas de violation de données susceptible d'entraîner un risque pour les droits et libertés des personnes.

En cas de non-conformité au RGPD, **des sanctions lourdes sont prévues** et peuvent s'appliquer aux entreprises en dehors de l'UE si elles traitent des données de personnes résidant dans l'UE :

– Pour les **violations les moins graves**, par exemple des manquements aux obligations administratives, les amendes peuvent atteindre jusqu'à 10 millions d'euros ou 2 % du chiffre d'affaires mondial total de l'entreprise, selon le montant le plus élevé ;

– Pour les **violations les plus graves** (comme le non-respect des principes fondamentaux du RGPD ou des droits des personnes concernées), les amendes peuvent aller jusqu'à 20 millions d'euros ou 4 % du chiffre d'affaires mondial total de l'entreprise, selon le montant le plus élevé.

Le RGPD est mis en œuvre par les autorités nationales – la CNIL en France – et, au niveau européen, par le **Contrôleur européen de la protection des données (CEPD).** Le CEPD, créé en 2004, est ainsi chargé de superviser le respect des règles de protection des données au sein des institutions et organes de l'UE, de la conseiller sur les questions de protection des données, de gérer les plaintes des citoyens et de surveiller les évolutions technologiques liées à la vie privée.

Il existe également un **Comité européen de la protection des données** qui réunit les autorités nationales de protection des données des pays de l'Espace économique européen, ainsi que le CEPD. Ce comité veille à ce que le RGPD soit appliqué de manière cohérente et veille à la coopération européenne, notamment en matière répressive. Il adopte également des décisions contraignantes concernant les affaires transfrontalières sur lesquelles aucun consensus n'est atteint entre les autorités nationales.

b. Le Digital market act *et le* Digital services act *(DMA/DSA)*

Le règlement sur les marchés numériques (DMA) du 14 septembre 2022 vise à mettre fin à la domination des géants de l'Internet, en particulier les GAFAM (Google, Apple, Facebook, Amazon et Microsoft), en introduisant des règles strictes pour favoriser la concurrence, protéger les petites entreprises et stimuler l'innovation sur le marché numérique européen, tout en prévoyant des sanctions sévères en cas de non-respect.

Le règlement couvre dix « services de plateforme essentiels », tels que les moteurs de recherche, les réseaux sociaux, les systèmes de messagerie, les services de *cloud*, *etc*.

Les contrôleurs d'accès, c'est-à-dire les acteurs ayant une forte influence sur le marché intérieur (peu importe leur localisation), doivent se conformer à une série d'obligations et d'interdictions.

Ils doivent faciliter la désinstallation d'applications pré-installées, rendre interopérables leurs services de messagerie, autoriser les vendeurs à promouvoir leurs offres en dehors de la plateforme, partager les données de performance marketing, *etc*. Le DMA interdit aux contrôleurs d'accès d'imposer des logiciels par défaut, de favoriser leurs propres services, de réutiliser les données personnelles à des fins de publicité ciblée sans consentement, et d'imposer des services annexes aux développeurs d'applications.

La Commission européenne peut infliger des amendes pouvant aller jusqu'à 10 % du chiffre d'affaires mondial total du contrôleur d'accès en cas d'infraction – et 20 % en cas de récidive. Des astreintes allant jusqu'à 5 % du chiffre d'affaires journalier mondial total peuvent également être imposées. En cas de violations systématiques, la Commission peut prendre des mesures correctives, telles que la cession d'activités ou l'interdiction d'acquérir d'autres entreprises dans le numérique ou la collecte de données.

Le règlement européen sur les services numériques (DSA) est entré en vigueur le 25 août 2023. Il vise à surveiller et responsabiliser les plateformes en ligne et à protéger les droits des internautes ainsi qu'à soutenir les petites entreprises européennes dans ce secteur. Il a pour objectif d'appliquer les mêmes règles en ligne que hors ligne, en luttant contre la diffusion de contenus illégaux ou préjudiciables, tels que la haine en ligne, la désinformation ou la vente de produits illégaux.

Le DSA s'applique à tous les intermédiaires en ligne offrant des services sur le marché européen, qu'ils soient basés en Europe ou ailleurs dans le monde. Cela inclut les fournisseurs d'accès à Internet, les services de *cloud computing*, les réseaux sociaux, les plateformes de partage de contenu et les moteurs de recherche notamment. Les très grandes plateformes en ligne, telles que les GAFAM, sont soumises à des obligations plus strictes.

Le DSA prévoit plusieurs mesures, notamment la simplification des procédures de signalement, la transparence des décisions de modération, l'interdiction des publicités ciblées pour les mineurs et de celles fondées sur des données sensibles, telles que les opinions politiques.

Les très grandes plateformes ont l'obligation d'analyser et d'atténuer les risques systémiques, notamment liés à la désinformation et à la sécurité en ligne. Des audits indépendants et un accès aux données clés sont prévus, ainsi qu'un mécanisme de réaction aux crises pour assurer la sécurité publique.

Chaque État membre de l'UE doit désigner un « coordinateur des services numériques » pour contrôler le respect du DSA – l'Autorité de régulation de la communication audiovisuelle et numérique (ARCOM) en France. Les plateformes et moteurs de recherche les plus systémiques seront directement surveillés par la Commission européenne. Des amendes pouvant aller jusqu'à 6 % du chiffre d'affaires mondial peuvent être infligées en cas de non-respect, et les violations graves et répétées peuvent entraîner l'interdiction des activités sur le marché européen.

La mise en œuvre du DSA/DMA par le projet de loi SREN

Le projet de loi visant à sécuriser et réguler l'espace numérique adapte le droit français pour que puissent s'appliquer les règlements sur les services numériques (DSA) et sur les marchés numériques (DMA).

Il désigne notamment les autorités nationales de régulation compétentes, en l'espèce l'ARCOM pour les contenus, la CNIL pour les données et l'Autorité de la concurrence pour les questions économiques.

Le texte prévoit également une série de mesures pour protéger les enfants de la pornographie en ligne et les utilisateurs contre les arnaques, le harcèlement et la désinformation en ligne. Le projet de loi propose également une régulation renforcée des jeux numériques et des locations touristiques.

La Commission européenne a adressé un avis circonstancié à la France pour signaler la non-conformité de certaines dispositions du texte avec le DSA, notamment :

– la vérification de l'âge à l'entrée des plateformes ;

– le blocage judiciaire des comptes créés par une personne bannie d'un service en ligne ;

– l'obligation pour les plateformes de signaler les contenus pornographiques, alors même qu'elles n'ont pas nécessairement connaissance de leur existence.

Le projet de loi n'évoque l'intelligence artificielle que sous deux angles : pour l'introduire parmi les thématiques abordées dans le cadre de l'éducation au numérique à l'école et pour permettre au Pôle d'expertise et de régulation du numérique (PEREN) d'intervenir dans ce domaine.

Dans un avis circonstancié, la Commission européenne a indiqué qu'elle estimait que certaines dispositions du projet de loi en discussion ne sont pas conformes au DSA. Des échanges sont en cours avant que la navette sur ce texte ne se poursuive.

c. Le Data Act

Le règlement sur des règles harmonisées en matière d'accès et d'utilisation équitables des données (dit *Data Act*) a été définitivement adopté en juillet 2023. Il constitue un élément essentiel de la stratégie européenne en matière de données, en soutenant l'utilisation des données dans le respect du droit de l'UE.

Il vise à garantir une équité entre les acteurs économiques dans l'utilisation des données générées par les objets connectés et à permettre aux utilisateurs de tirer pleinement parti des données numériques qu'ils génèrent.

Le *Data Act* propose plusieurs mesures pour mettre davantage de données à la disposition des entreprises, des citoyens et des administrations publiques. Cela inclut des mesures visant à accroître la sécurité juridique pour les entreprises et les consommateurs, à prévenir les abus contractuels, à permettre aux organismes du secteur public d'accéder à des données privées et à établir des règles pour faciliter le passage des données entre les fournisseurs de services et les entreprises de traitement de données.

2. Des dispositifs qui s'appliquent imparfaitement à l'IA, notamment générative

Ces différents textes ne s'appliquent qu'indirectement à l'intelligence artificielle et ne couvrent ni toutes les problématiques qu'elle pose, ni l'ensemble des usages qu'elle peut avoir. Ils viennent par ailleurs s'ajouter aux règlementations sectorielles applicables.

Les DSA/DMA couvrent le cas des plateformes et pourraient donc faciliter la lutte contre la désinformation ou les discriminations si celles-ci étaient promues par une IA. Ils permettent également de contrôler certains fournisseurs de solutions d'IA.

Le RGPD encadre l'utilisation des données personnelles par les IA, qui sont un élément déterminant de leur fonctionnement et permet leur contrôle. Lors de son audition, la direction générale de la justice de la Commission européenne (DG JUST) a confirmé son interprétation selon laquelle le RGPD a vocation à s'appliquer pleinement au système d'IAG.

Quant au *Data Act,* il est en mesure de faciliter l'accès des développeurs d'IA à des données dans le but d'innover dans ce secteur, mais cela ne suffit pas à garantir la qualité des IA ainsi conçues.

La question de la protection des données ne couvrira pas l'ensemble des risques existants liés à l'IAG, le RGPD ne permettant pas d'imposer un étiquetage des contenus, ni de contrôler les biais introduits par un algorithme ou par la méthode d'entraînement de l'IA.

C'est pour ces raisons que l'Union européenne a engagé, comme d'autres pays et organisations internationales, une réflexion sur l'élaboration d'une législation spécifiquement dédiée à l'IA. Comme l'a souligné la DG JUST, il n'est pas question que ces nouvelles règles entrent en contradiction avec les dispositions précédemment présentées en matière de régulation numérique. Le règlement IA ne fera que préciser l'application des mécanismes existants et couvrir des situations qui ne font l'objet d'aucun encadrement.

B. UNE COURSE CONTRE LA MONTRE POUR IMPOSER UN MODÈLE EUROPÉEN D'IA RESPONSABLE

1. La volonté européenne de réguler l'IA

L'Europe, grâce à l'ampleur de son marché et au pouvoir d'achat de ses citoyens, dispose de la capacité d'intervenir pour réguler le secteur du numérique. Les entreprises étrangères se plieront globalement à ses exigences si elles ne veulent pas être exclues des opportunités offertes par le marché des utilisateurs européens, ni s'exposer à un risque de réputation.

Votre rapporteur Stéphane Rambaud considère que cette volonté de régulation traduit le retard de l'Union européenne en matière numérique et son incapacité à soutenir efficacement l'innovation.

Votre rapporteur Philippe Pradal estime plutôt que l'Union européenne a fait preuve de prudence, soucieuse qu'elle est de la protection des données personnelles notamment. Il commence cependant à émerger des concurrents sérieux, notamment d'origine française, qu'il est nécessaire d'accompagner par une régulation ambitieuse.

À l'occasion du déplacement de vos rapporteurs à Bruxelles, le CEPD a indiqué que l'interdiction de *ChatGPT* par la CNIL italienne en avril 2023 a démontré « *la capacité des pays européens à mettre autour de la table des acteurs globaux et à leur imposer certaines règles* » (voir encadré).

Deux défis principaux s'imposent à l'Europe :

– **réussir à protéger ses principes**, en encadrant les usages de l'intelligence artificielle et en fixant des règles équitables permettant de concilier un marché libre avec des règles de fonctionnement vertueuses ;

– permettre l'émergence de nouveaux acteurs, si possible européens, en mesure de s'imposer à l'échelle du continent, voire au niveau mondial.

En l'absence de régulation spécifique, l'Union européenne s'expose doublement. Économiquement, elle risque de faire fuir les données de ses habitants et d'offrir d'immenses revenus à des entreprises extra-européennes implantées sur son marché. Politiquement, elle ouvre la voie à de nombreuses stratégies d'influence venant des États-Unis ou de pays aux valeurs divergentes. Cela faciliterait la présence sur le marché intérieur d'acteurs peu scrupuleux entretenant une concurrence déloyale avec des entreprises européennes plus vertueuses.

La fermeture de *ChatGPT* en Italie en avril 2023

Le *Garante per la protezione dei dati personali* (GDDP), homologue italien de la CNIL, a décidé de bloquer temporairement l'accès à *ChatGPT* le 31 mars 2023. Dans un communiqué, elle a indiqué avoir constaté plusieurs manquements au RGPD de la part de l'IAG d'OpenAI : manque d'information des utilisateurs et des personnes dont les données sont collectées ; absence de base légale justifiant la collecte et le stockage massifs de données personnelles dans le but d'entraîner les algorithmes ; absence de filtre pour vérifier l'âge des utilisateurs.

OpenAI a pris en compte ces remarques pour se rapprocher d'une conformité au RGPD : en publiant sur son site une description des données personnelles traitées dans le cadre de l'entraînement de ses modèles, en rappelant que chacun a le droit de refuser un tel traitement ; en mettant à disposition des utilisateurs un formulaire pour s'opposer au traitement de leurs données personnelles ; en prévoyant un mécanisme de vérification de l'âge et du consentement des parents.

Dans un communiqué, la GDDP a indiqué qu'elle « *reconnaît les progrès réalisés par OpenAI pour concilier les avancées technologiques avec le respect des droits des personnes et espère que l'entreprise poursuivra ses efforts pour se conformer à la législation européenne en matière de protection des données* ».

ChatGPT est redevenu accessible aux internautes italiens le 28 avril 2023.

2. Une concurrence internationale en vue de réguler l'IAG

a. Plusieurs pays et organisations internationales essaient d'élaborer des règles pour encadrer l'intelligence artificielle.

Il existe deux grands modèles, deux grandes approches pour affronter les défis de l'IAG : une **approche libérale**, qui considère que le jeu naturel du marché permettra de faire le tri entre les bons et les mauvais usages de l'IAG, entre les vraies et les fausses informations, entre les acteurs de confiance et les acteurs nocifs ; et une **approche plus dirigiste**, pour laquelle l'IAG ne peut se diffuser que dans un cadre réglementaire très strict et délimité.

Ces deux approches sont souvent représentées par le monde anglo-saxon d'une part (pour l'approche libérale) et par la Chine d'autre part (pour l'approche dirigiste), même si dans les faits les réalités sont plus nuancées, comme l'a exposé Mme Asma Mhalla lors de son audition.

Ainsi, il est vrai que les Anglo-Saxons ont d'abord eu une attitude très libérale à l'égard de l'IAG.

Le **Royaume-Uni** se singularise, par exemple, pour avoir annoncé son absence d'intention de préparation d'une réglementation spécifique à l'IAG. Mais la réflexion de ce pays semble avoir évolué très récemment, celui-ci ayant été à l'initiative du premier sommet international sur les risques associés à l'IA, qui s'est tenu les 1er et 2 novembre 2023 et auquel ont participé 28 États. Le Royaume-Uni a, par ailleurs, créé un *AI Safety Institute*, qui aura pour mission de tester et évaluer les

risques des modèles des géants de l'IA, y compris l'IAG. Comme l'a indiqué le premier ministre Rishi Sunak, le Royaume-Uni poursuit le but de se présenter comme *« le meilleur endroit en Europe pour lever des capitaux, là où les géants des technologies choisissent d'installer leurs filiales européennes. Et l'IA présente d'énormes opportunités pour ceux qui sauront contrôler ses risques »* (1).

Les **États-Unis** également semblent, après des hésitations, s'être engagés dans une démarche de régulation plus prononcée. Le 30 octobre 2023, le président américain, Joe Biden, a signé un premier décret pour encadrer l'IA prévoyant, selon le communiqué de la Maison Blanche (2), plusieurs mesures justifiées par la sécurité nationale, comme un partage des informations sensibles avec le gouvernement, ou encore un programme de cybersécurité pour détecter et corriger les vulnérabilités des logiciels les plus stratégiques. Le communiqué insiste sur le fait que le plan d'actions présenté a notamment pour but de préserver le « *leadership américain dans le monde* », ce qui confirme les inquiétudes précédemment formulées sur le risque de biais et de manipulation des IAG à des fins d'influence.

Ce même texte prévoit des actions pour lutter contre la diffusion des hypertrucages (*deepfakes*) et pour prévenir les biais discriminatoires des IAG. Il énonce que le département du Commerce élaborera des directives pour l'authentification du contenu généré par l'IA. Il indique enfin que « *les agences fédérales utiliseront ces outils pour permettre aux Américains de savoir que les communications qu'ils reçoivent de leur gouvernement sont authentiques, et ainsi donner l'exemple au secteur privé et aux gouvernements du monde entier* ».

Il est frappant de constater que la **Chine** affiche la même ambition de leadership que les États-Unis, en adoptant toutefois un modèle de régulation beaucoup plus dirigiste. La Chine a d'ores et déjà adopté une réglementation abondante avec par exemple l'obligation pour les développeurs d'inscrire leur système d'IAG à un registre des algorithmes, un référentiel gouvernemental nouvellement créé qui recueille des informations sur la manière dont les algorithmes sont entraînés, tout en exigeant qu'ils passent une auto-évaluation de sécurité (3).

Si le contrôle de l'information est l'un des objectifs principaux de la réglementation chinoise, on retrouve également un volet dédié à la protection des travailleurs, ce qui illustre le fait que les craintes sur les risques sociaux de l'IAG ne sont pas limitées à la sphère occidentale.

---

(1) https://www.lemonde.fr/international/article/2023/10/30/le-royaume-uni-organise-le-premier-sommet-mondial-sur-les-risques-associes-a-l-intelligence-artificielle_6197300_3210.html

(2) https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

(3) https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117

Enfin, cette réglementation comporte aussi un volet de soutien à l'innovation, avec l'objectif explicite de créer un environnement politique propice à ce que la Chine devienne le leader mondial en matière de développement et d'applications de l'IAG.

Finalement, il apparaît que partout dans le monde, les pouvoirs publics recherchent le juste équilibre entre réglementation et soutien à l'innovation, avec une arrière-pensée géopolitique non dissimulée.

b. Des échanges ont déjà lieu à différents niveaux.

Les pays du G7 sont également en cours de discussion dans le cadre du protocole dit d'Hiroshima. Le CEPD a indiqué à vos rapporteurs que l'OCDE travaillait sur une définition de l'IAG et que l'Assemblée mondiale de la vie privée, dont il est membre et qui regroupe 150 autorités de régulations, a ouvert des discussions pour définir des principes communs applicables aux IAG.

Le 31 mai 2023, s'est également tenue la quatrième réunion ministérielle du Conseil du commerce et des technologies (CCT) entre l'Union européenne et les États-Unis pour l'application de la « feuille de route commune pour une IA digne de confiance et pour la gestion des risques » [(1)]. Cette feuille de route, qui devrait prochainement s'appliquer à l'IAG, contient des constats communs sur les opportunités et les risques liés à l'IA et prévoit une série de chantiers pour mettre en place des règles et des modalités de contrôle communes.

Il n'est, dès lors, pas envisageable de rester à l'écart des discussions internationales. La France doit s'y investir pleinement et ne pas se limiter au cadre européen.

Pour ce faire, vos rapporteurs recommandent la nomination d'un ambassadeur à l'IAG. Ce dernier pourrait s'appuyer utilement sur l'expertise du **comité de l'intelligence artificielle générative** institué par le Gouvernement au mois de septembre 2023.

Recommandation n° 1 : Nommer un ambassadeur à l'IAG pour suivre les discussions internationales autour de sa régulation.

Cette émulation peut être bénéfique si elle permet, sans pénaliser les États européens, de faire émerger un consensus, inspirant ensuite d'autres pays et conduisant à une harmonisation par le haut des législations ainsi qu'à la définition de standards communs.

Le **modèle européen**, à mi-chemin entre l'approche libérale anglo-saxonne et l'approche dirigiste de la Chine, dispose de nombreux atouts pour trouver ce juste équilibre.

(1) https://ec.europa.eu/newsroom/dae/redirection/document/92123

Votre rapporteur Philippe Pradal estime qu'il est essentiel que l'Europe soit à l'initiative de ce socle commun, afin qu'elle diffuse sa vision de l'intelligence artificielle comme ce fut le cas de la mise en œuvre du RGPD, qui est aujourd'hui un modèle repris dans de nombreux pays.

Dans l'Union européenne et en France, il n'existe pas encore de réglementation spécifique de l'IA en général, et encore moins de l'IAG en particulier. En l'état, chaque régulateur sectoriel doit contrôler la légalité du développement et des usages de l'IAG dans son champ de compétence.

Tel a été l'objet des négociations autour du règlement sur l'intelligence artificielle (*AI Act*), qui ont pris fin en décembre 2023.

3. L'état des négociations sur le règlement établissant des règles harmonisées concernant l'intelligence artificielle (*AI Act*)

L'Union européenne souhaite mettre en œuvre le premier régime de réglementation de l'IA, y compris générative, dans le cadre de sa stratégie numérique pour assurer un développement et une utilisation plus sûrs de ces technologies innovantes.

En avril 2021, la Commission européenne a proposé un cadre réglementaire qui suggère de classer les systèmes d'IA en fonction des risques qu'ils présentent pour les utilisateurs et la société, déterminant ainsi le niveau de réglementation adéquat.

Le Parlement européen a souhaité garantir que les systèmes d'IA utilisés dans l'UE soient sûrs, transparents, traçables, non discriminatoires et respectueux de l'environnement. Il souhaite, à cette fin, définir pour l'IA une norme technologique neutre, qui puisse s'adapter aux évolutions du secteur.

a. Des obligations variant selon une échelle de risque

Différentes obligations sont prévues tant pour les fournisseurs que les utilisateurs, en fonction du niveau de risque lié à chaque système.

● **Les systèmes à risque inacceptable** seraient interdits. Cela vise tous les systèmes d'IA considérés comme une menace évidente pour la sécurité, les moyens de subsistance et les droits des personnes, qui seront interdits – qu'il s'agisse par exemple de la notation sociale par les gouvernements, ou encore de jouets utilisant l'assistance vocale et qui encouragent les comportements dangereux.

● **Les systèmes à haut risque** seraient évalués avant leur mise sur le marché. Les systèmes d'IA relevant de huit domaines spécifiques devraient donc être enregistrés dans une base de données relevant de l'UE :

– les infrastructures critiques susceptibles de mettre en danger la vie et la santé des citoyens, par exemple dans le domaine des transports ;

– la formation éducative ou professionnelle, qui peut déterminer l'accès à l'éducation et au cours professionnel de la vie d'une personne, par exemple la notation des examens ;

– les composants de sécurité des produits, par exemple en matière de chirurgie assistée par robot ;

– l'emploi, la gestion des travailleurs et l'accès au travail indépendant, par exemple un logiciel de tri de CV pour les procédures de recrutement ;

– les services publics et privés essentiels, par exemple l'évaluation du client souhaitant obtenir un prêt ;

– les services répressifs susceptibles d'interférer avec les droits fondamentaux des personnes (par exemple l'évaluation de la fiabilité des preuves) ;

– la gestion des migrations, de l'asile et du contrôle aux frontières (par exemple la vérification de l'authenticité des documents de voyage)

– les processus démocratiques.

● **Les systèmes à risque limité** seraient soumis à des obligations allégées de transparence. En particulier, lors de l'utilisation de systèmes d'IA tels que les agents conversationnels (*chatbots*), les utilisateurs doivent être conscients qu'ils interagissent avec une machine afin qu'ils puissent prendre une décision éclairée de continuer ou de prendre du recul.

● **Les IA à risque minimal ou nul** seraient librement utilisables. Cela inclut des applications telles que les jeux vidéo compatibles avec l'IA ou les filtres anti-spam. La grande majorité des systèmes d'IA actuellement utilisés dans l'UE relèvent de cette catégorie.

Pour toutes les catégories de risques, une fois qu'un système d'IA est sur le marché, les autorités seraient chargées de leur surveillance en s'appuyant sur les signalements des utilisateurs et des fournisseurs. Les fournisseurs auraient également l'obligation de mettre en place des systèmes de prévention interne.

b. La place de l'IAG dans le règlement

Dans la version adoptée par le Parlement européen, les systèmes d'IAG obéiraient au cadre global et entreraient dans la catégorie des IA à risque élevé en raison « *des risques spécifiques de manipulation qu'ils présentent* » (1). Parmi les obligations incombant aux fournisseurs d'IAG, figureraient le marquage des contenus générés, la mise en place de solutions visant à entraver la production de contenus illégaux, ainsi que la réalisation d'un *reporting* sur les types de données utilisés pour l'entraînement du système et la publication, parmi elles, des données protégées par le droit d'auteur.

Dans son exposé des motifs, le règlement, dans la version issue du Parlement européen, précise : « *Les obligations de transparence s'appliqueront aux systèmes qui i) interagissent avec les humains, ii) sont utilisés pour détecter des émotions ou déterminer l'association avec des catégories (sociales) sur la base de données biométriques, ou iii) générer ou manipuler des contenus (trucages vidéo ultra-réalistes). Lorsque des personnes interagissent avec un système d'IA ou que leurs émotions ou caractéristiques sont reconnues par des moyens automatisés, elles doivent en être informées. Si un système d'IA est utilisé pour générer ou manipuler des images ou des contenus audio ou vidéo afin de produire un résultat qui ressemble sensiblement à un contenu authentique, il devrait être obligatoire de déclarer que le contenu est généré par des moyens automatisés, sauf pour certaines finalités légitimes faisant l'objet d'exceptions (domaine répressif, liberté d'expression). Cette obligation laisse la possibilité aux personnes de prendre des décisions en connaissance de cause ou de se désengager d'une situation donnée* ».

Si le Parlement européen a estimé que l'ensemble des IAG étaient à risque au regard de leur spécificité technologique, cette position n'est pas unanime. Le coordonnateur national pour l'IA en France, M. Guillaume Avrin, considère que le choix de classer l'ensemble des IAG parmi les IA à risque est discutable car le risque, et le niveau de régulation, devrait dépendre de l'usage qui est fait de l'IAG (par exemple dans le domaine de la santé) et non de la technologie utilisée.

Les négociations entre le Parlement européen et le Conseil ont permis d'aboutir en décembre 2023 à un compromis prévoyant un encadrement plus strict pour les modèles à « fort impact », c'est-à-dire ceux pouvant présenter un risque systémique au regard de la quantité de données avec lesquelles ils ont été entraînés et de leur performance de calcul. Cette solution permet de contrôler les acteurs les plus importants, déjà installés sur le marché, sans pénaliser les nouveaux entrants. Il faudra néanmoins s'assurer que cette norme, qui n'est pas technologiquement neutre, puisse s'adapter aux progrès du secteur.

---

(1) Exposé des motifs du règlement IA

C. UN ÉVENTAIL DE MODES DE RÉGULATION QUI IMPLIQUE DES PROGRÈS TECHNOLOGIQUES RAPIDES

L'UE présente à ce jour une feuille de route ambitieuse, qui mobilise l'ensemble des leviers de régulation possible, tels qu'ils avaient été identifiés par vos rapporteurs au fil de leurs auditions.

1. Le contrôle *ex ante* : la certification et le marquage des contenus

La solution la plus efficace en matière de régulation consisterait à certifier *ex ante* les modèles et les applications d'intelligence artificielle souhaitant proposer leurs services sur le marché européen. Cette certification pourrait soit conditionner l'accès au marché intérieur, soit faire office de label visant à orienter l'utilisateur vers les IA les plus vertueuses, ce qui serait moins contraignant pour les fournisseurs, mais aussi moins efficace au regard du but poursuivi.

Cette certification peut reposer sur un certain nombre de critères : nature et qualité des jeux de données utilisés pour entraîner le modèle, existence ou non de biais dans la configuration de l'algorithme et dans l'apprentissage, existence ou non de filtre dans le contenu généré, conformité de l'utilisation des données en amont (données d'entraînement) et en aval (données collectées auprès des utilisateurs) avec le RGPD, respect du droit d'auteur applicable aux données utilisées, prévention du risque d'hallucination ou de biais dans les réponses, *etc*.

L'Union européenne prévoit le recours à cette certification pour les systèmes à haut risque, dont les IAG font partie. L'évaluation reposerait sur les critères suivants :

– des systèmes adéquats d'évaluation et d'atténuation des risques ;

– la haute qualité des ensembles de données alimentant le système, afin de minimiser les risques et les résultats discriminatoires ;

– l'enregistrement des activités, afin d'assurer la traçabilité des résultats ;

– une documentation détaillée fournissant toutes les informations nécessaires sur le système et son objet, pour permettre aux autorités d'évaluer sa conformité ;

– des informations claires et adéquates pour l'utilisateur ;

– des mesures de surveillance humaines appropriées, afin de minimiser les risques ;

– un haut niveau de robustesse, de sécurité et de précision.

L'autre dimension du contrôle préventif des IA consiste à exiger le marquage des contenus générés pour les rendre identifiables. Cette solution est techniquement possible via l'introduction d'une « griffe » marquant l'intervention de l'IA sur l'image, la vidéo, le code ou le texte généré.

Elle peut être efficace dans un contexte dit « coopératif », où ni le fournisseur, ni l'utilisateur ne souhaitent dissimuler l'usage qui a été fait de l'intelligence artificielle. La charge de marquer le contenu pourrait incomber soit uniquement au fournisseur ou à l'utilisateur, soit aux deux. Dans ce cas, le marquage sert principalement à signaler au destinataire du contenu que celui-ci a été généré par tel ou tel modèle. Se posera toutefois la question de la définition d'un contenu généré par une IA.

Qu'en est-il de l'utilisation de l'IA pour résumer ou traduire son propre travail, ou bien de l'incorporation dans un travail plus vaste d'extraits réalisés à l'aide d'une IA, ou encore d'un contenu généré par IA, puis modifié de façon plus ou moins substantielle par l'utilisateur ?

Dans un contexte dit « adverse », c'est-à-dire dans lequel l'IA est utilisée de manière volontairement dissimulée ou illégale, les garanties quant à l'efficacité du marquage sont moindres. Cela pourrait néanmoins entraver certains détournements de l'IA, par exemple en alourdissant et en ralentissant la création de fausses informations ou d'arnaques, qui reposent souvent sur la production massive de contenus identiques (*phishing* ou hameçonnage). Cela pourrait également fonctionner pour lutter contre les détournements de l'IA, à condition de réussir à dissimuler le marquage pour le rendre ineffaçable et détectable seulement au moyen de certains logiciels mis à disposition par les fournisseurs d'IA eux-mêmes.

Comme l'a expliqué la Gendarmerie nationale à vos rapporteurs : « *Pour protéger les contenus de manipulation, il existe des moyens comme le tatouage qui a trait à l'audio, à l'image ou au texte. Il s'agit d'une technique qui a pour objet de défendre la propriété intellectuelle des œuvres numériques. Le tatouage est comme une signature qui est non visible dans le document d'origine. Par exemple, dans le domaine de l'audio, il est possible d'exploiter ce que l'on appelle l'étalement spectral, pour introduire dans des fréquences inaudibles, une information qui peut être du texte, une image ou un autre enregistrement sonore. Le concepteur peut choisir d'insérer un tatouage fragile ou robuste. Si le tatouage est fragile, il va se dégrader dès lors que le contenu sera modifié et ainsi attester de la manipulation. S'il est robuste, l'objectif est de pouvoir retrouver l'origine d'un fichier ou son concepteur même s'il a été manipulé.* »

En tout état de cause, certains acteurs essaieront certainement d'échapper à la régulation, tout comme leurs utilisateurs. En l'espèce, cela pourra relever davantage de poursuites judiciaires, par exemple en créant une infraction relative à la diffusion dissimulée de contenus créés par une IA [(1)].

(1) Voir deuxième partie (II. A. 2.).

2. Le contrôle *ex post* : le traitement des plaintes et un régime de sanction

Le second volet de la régulation consiste à identifier, corriger et sanctionner les dysfonctionnements des IAG selon un système de contrôle *ex post* ciblant certains acteurs en fonction des signalements reçus.

Ce contrôle pourrait notamment porter sur les jeux de données utilisés pour entraîner le modèle – en particulier l'exploitation de données personnelles ou de données soumises au droit d'auteur – et sur le fonctionnement du modèle ou de ses applications (recherche de biais, vérification des filtres, *etc.*).

Afin de faire respecter ces règles par l'ensemble des entreprises intervenant sur le marché numérique européen, un régime de sanctions pourrait être mis en place, s'inspirant là encore du mécanisme prévu par le RGPD. Ces sanctions financières pourraient être assorties, à titre provisoire ou définitif, de la fermeture ou de la suspension de l'accès des utilisateurs aux services proposés lorsque ces derniers ne respectent pas les règles prévues. Ces sanctions auraient également des répercussions sur la réputation des entreprises ciblées, et inciteraient les fournisseurs à se conformer au plus vite au droit européen.

En complément, des réflexions ont été engagées pour créer un régime de responsabilité spécifique aux IA, qui pourrait également concerner les IAG. La Commission européenne a rendu public, le 28 septembre 2022, une proposition de directive relative aux règles de responsabilité civile extracontractuelle applicables aux outils d'intelligence artificielle (*voir encadré ci-avant*). Celle-ci vise plus particulièrement à adapter les règles en matière de charge de la preuve et de lien de causalité en cas de dommage en lien avec l'utilisation d'une IA. Ceux-ci peuvent effectivement être difficiles à déterminer compte tenu de la technicité et de l'opacité des systèmes concernés. En outre, l'imputation de la faute est complexe entre le concepteur du modèle, le développeur de l'application, l'auteur du *prompt*, l'utilisateur du contenu généré, *etc*.

Une proposition de directive relative aux règles de responsabilité civile extracontractuelle applicables aux outils d'intelligence artificielle

La Commission européenne a rendu public, le 28 septembre 2022, une proposition de directive relative aux règles de responsabilité civile extracontractuelle applicables aux outils d'intelligence artificielle.

Selon l'exposé des motifs de la proposition, « *les règles nationales existant en matière de responsabilité, notamment en ce qui concerne la responsabilité pour faute, ne sont pas adaptées pour traiter les actions en responsabilité dans le cas de dommages causés par des produits et services dotés d'IA* […] *Lorsque l'IA s'interpose entre l'acte ou l'omission d'une personne et le dommage, les caractéristiques spécifiques de certains systèmes d'IA, telles que l'opacité, le comportement autonome et la complexité, peuvent rendre excessivement difficile, voire impossible, l'acquittement de cette charge de la preuve par la personne lésée* ».

La proposition de directive vise à garantir aux victimes de dommages causés par l'IA une protection équivalente à celle des victimes de dommages causés par les produits de manière générale. Elle a vocation à s'appliquer aux actions civiles fondées sur une faute extracontractuelle pour des dommages causés par un système d'IA, introduites dans le cadre de régimes de responsabilité fondés sur la faute. L'action en réparation peut être dirigée soit contre le fournisseur, soit contre l'utilisateur de l'IA soupçonné d'avoir causé le dommage.

La définition de la faute relève du droit national, ou du droit de l'Union européenne dans les secteurs qui ont fait l'objet d'une harmonisation. Ainsi la proposition de directive énonce expressément que le droit national des États membres peut établir des obligations spécifiques pour certaines applications de l'IA.

Si elle n'introduit pas de régime de responsabilité sans faute, la proposition de directive engage les États membres à adopter trois séries de mesures permettant d'alléger la charge de la preuve qui incombe aux victimes dans les actions en responsabilité :

– une juridiction pourrait ordonner au fournisseur ou à l'utilisateur d'IA à haut risque soupçonnés d'avoir causé un dommage de divulguer des éléments de preuve pertinents ;

– la juridiction pourrait fonder sa décision concernant la responsabilité sur une présomption de faute, en cas de non-respect de l'injonction judiciaire de divulgation ou de conservation des éléments de preuve.

– une présomption de causalité dispenserait les victimes d'IA de la charge de la preuve dès lors qu'elles parviendraient à démontrer qu'une personne, physique ou morale, a commis une faute en ne respectant pas une obligation liée à l'utilisation de l'IA et que l'existence d'un lien de causalité entre le dommage et la performance de l'IA est raisonnablement probable.

La définition des dommages relève du droit national et ne fait pas l'objet d'une harmonisation dans la proposition de directive. Les États membres pourraient adopter ou conserver des règles nationales plus favorables aux demandeurs.

3. Un système de régulation qui exige des compétences élevées

La régulation de l'IA exige une expertise juridique, mais aussi technique pour en comprendre les enjeux et être en capacité de dialoguer avec les acteurs du secteur et d'expertiser leurs systèmes. Le CEPD a insisté auprès de vos rapporteurs pour que les autorités en charge de la régulation de l'IA soient celles qui assurent aujourd'hui la protection des données personnelles, car elles disposent du niveau d'expertise technico-juridique le plus élevé. Les spécificités de l'IA appellent néanmoins des compétences spécifiques qu'il sera urgent de développer (1).

Comme l'a indiqué le PEREN lors de son audition, le contrôle des systèmes d'IAG est particulièrement complexe – davantage que celui des algorithmes des plateformes. Il est par exemple très difficile d'identifier un contenu écrit généré par une intelligence artificielle. Cela fonctionne plus efficacement lorsqu'il s'agit de la détection d'une fausse image ; ainsi, la gendarmerie utilise déjà des outils de reconnaissance des contenus générés par IA dans le cadre de ses enquêtes en matière d'hypertrucage (ou *deepfake*) (2). Il est également presque impossible d'identifier dans la réponse d'un agent conversationnel les sources d'information utilisées, à moins de pouvoir accéder dans le détail au fonctionnement de l'IA.

Ce contrôle implique de prévoir les modalités d'accès des régulateurs aux modèles et à leurs applications. Le renversement de la charge de la preuve – également souhaitable en ce qui concerne la responsabilité pénale et civile des IAG (3) – pourrait être nécessaire pour obliger les fournisseurs à expliquer les dysfonctionnements de leur système.

Selon la Direction générale des réseaux de communication, du contenu et des technologies (DG CNECT) de la Commission européenne, l'amélioration des capacités de contrôle devra reposer sur l'élaboration de standards clairs et contrôlables pouvant être généralisés au niveau international. Le règlement pour l'IA devrait prévoir un comité de l'IA, dont le format est encore en négociation.

Vos rapporteurs estiment qu'il serait pertinent qu'un partage de connaissances et d'expérience soit organisé au niveau européen, afin de permettre aux différentes autorités de contrôle de rapidement monter en compétence et d'ajuster les standards au fur et à mesure des progrès technologiques.

Recommandation n° 2 : Organiser un partage de connaissances et d'expérience en matière de contrôle des IA, en particulier générative, au niveau européen.

*

* *

(1) Voir deuxième partie (I. B. 2. a. et II. A. 3.).

(2) Audition de la direction générale de la gendarmerie nationale.

(3) Voir deuxième partie (II).

II. DES RISQUES D'ENTRAVE À L'INNOVATION ET DE RETARD DANS LA MISE EN ŒUVRE DE LA LÉGISLATION

Une approche trop souple n'est pas souhaitable, car elle viendrait légitimer tous les acteurs, y compris les moins vertueux, et empêcherait tout rattrapage de la part de nouveaux acteurs. Pour autant, une approche trop craintive vis-à-vis de l'IA présenterait aussi des inconvénients à ne pas sous-estimer.

Il semble donc nécessaire de trouver un juste équilibre entre contrainte et souplesse ; entre ouverture du marché à l'innovation et protection des acteurs émergents, en particulier européens. À ce titre, il est impératif que la réglementation européenne se fixe des priorités, au premier rang desquelles la protection des principes européens en matière de protection des libertés fondamentales et de l'État de droit.

A. LE RISQUE D'UNE APPROCHE TROP PRUDENTE QUI SERAIT PÉNALISANTE POUR LES ÉTATS ET POUR L'INNOVATION

1. Le risque de pénaliser l'utilisation de l'IA par les États membres

Lors de leur déplacement à Bruxelles, vos rapporteurs ont pu constater que, si la Commission avait des ambitions élevées en matière de régulation de l'IA, le Parlement est allé encore plus loin, en particulier pour garantir la protection la plus élevée possible des libertés fondamentales.

Les États-membres, dont la France, souhaitent revoir certains points afin de ne pas priver leurs administrations de l'opportunité d'utiliser l'intelligence artificielle dans le cadre de certaines politiques publiques, notamment en matière fiscale ou de sécurité.

La représentation permanente de la France auprès de l'UE a indiqué à vos rapporteurs que le texte élaboré par le Parlement pourrait conduire à un recul des possibilités offertes aujourd'hui aux États membres, par exemple en matière de détection des contenus numériques à caractère discriminatoire ou pédopornographique.

Vos rapporteurs estiment également que l'utilisation de l'IA dans certains domaines en lien avec les compétences régaliennes des États ne saurait être encadrée de la même manière selon qu'elle relève des administrations ou du secteur privé – par exemple en matière d'identification, de biométrie, de vidéoprotection, *etc*. Les potentialités de l'IA en matière de protection des populations, de sécurité et de lutte contre la délinquance seraient très élevées au regard des risques limités qu'elle présenterait, à condition que le recours à l'IA soit rigoureusement encadré dans ses usages et régulièrement contrôlé.

Recommandation n° 3 : Adapter les contraintes applicables en matière de recours à l'IA selon qu'elle est utilisée par des États membres ou par des acteurs privés.

2. Le risque de freiner l'émergence de nouveaux acteurs

Dans un contexte où l'UE accuse un retard technologique et commercial en matière d'IA, notamment générative, l'élargissement excessif de l'application du contrôle *ex ante* pourrait pénaliser l'innovation.

En effet, le défaut de la certification préalable est de permettre aux entreprises déjà implantées, qui ont développé leur modèle en dehors de ces contraintes, de préserver leur avantage concurrentiel. Il semble donc devoir être limité aux IA présentant les risques les plus graves. C'est le prisme adopté par l'Union européenne, qui établit une échelle de risque impliquant une échelle de contrôle différenciée. Pourtant, celle-ci semble vouloir placer toutes les IAG dans la même catégorie des IA à haut risque, alors même que certaines IAG pourraient être considérées comme à risque limité lorsqu'elles ont des finalités spécifiques et que leur utilisation est limitée.

Recommandation n° 4 : Prendre en compte l'audience des IAG et leur caractère systémique avant de leur assigner un niveau de risque, afin de ne pas pénaliser les nouveaux acteurs et d'encourager l'expérimentation et l'innovation.

3. Le risque d'une concurrence par le bas : vers des « paradis des données » ?

En outre, une approche trop restrictive risquerait non seulement de ne pas faire de l'Europe un territoire d'innovation, mais pourrait menacer ses intérêts et particulièrement ceux de la France, qui se distingue des autres États-membres par son potentiel en matière d'IAG.

Afin d'attirer les investisseurs et les cerveaux, certains pays pourraient utiliser leur législation pour créer un environnement plus favorable au développement de l'intelligence artificielle. Cette concurrence a déjà commencé, le Royaume-Uni ayant décidé en mars 2023 d'assouplir son régime de protection des données – jusqu'alors aligné sur le RGPD – afin de faciliter l'usage des données personnelles sans le consentement de l'utilisateur, à des fins précises.

L'existence d'une telle concurrence sur le territoire européen est d'autant plus problématique que les acteurs majeurs du secteur pourront s'en servir pour localiser leurs serveurs et les données qu'ils collectent.

Dans ce contexte, vos rapporteurs sont circonspects quant à la récente décision d'adéquation rendue par la Commission européenne le 10 juillet 2023 concernant la conformité de la législation américaine au RGPD (*voir encadré ci-après*). Cette décision reconnaît une équivalence des législations et donc la possibilité, pour les entreprises américaines, de transférer aux États-Unis les données personnelles qu'elles collectent en Europe auprès des utilisateurs européens et, par conséquent, le contentieux qui s'y rattache.

La nouvelle décision d'adéquation de la Commission européenne

Une décision d'adéquation est une décision adoptée par la Commission européenne sur la base de l'article 45 du RGPD, qui établit qu'un pays tiers ou une organisation internationale assure un niveau de protection adéquat des données personnelles, équivalent à celui de l'UE. Le RGPD prévoit une liste non exhaustive d'éléments qui, cumulés, permettent à la Commission d'évaluer le caractère adéquat du niveau de protection des données du pays tiers.

La décision d'adéquation a pour effet de permettre le transfert, sans exigences supplémentaires, de données personnelles depuis les organismes soumis au RGPD vers le pays tiers concerné. En l'absence d'une telle décision, des « garanties appropriées » doivent être mises en place.

Le 16 juillet 2020, la CJUE, dans son arrêt *Schrems II*, a invalidé la précédente décision d'adéquation de la Commission européenne à l'égard des États-Unis.

La CJUE a analysé la législation américaine alors en vigueur en matière d'accès aux données des fournisseurs de services Internet et entreprises de télécommunications par les services de renseignement américains. Elle en a conclu que les atteintes portées à la vie privée des personnes dont les données étaient traitées par les entreprises et opérateurs états-uniens soumis à cette législation étaient disproportionnées au regard des exigences de la Charte des droits fondamentaux de l'Union européenne. En particulier, la CJUE a jugé que la collecte des données par les services de renseignement n'est pas proportionnée et que les voies de recours, y compris juridictionnelles, dont disposent les personnes à l'égard du traitement de leurs données étaient insuffisantes.

En réaction à cette invalidation, le président des États-Unis, Joe Biden, a adopté le 7 octobre 2022 un décret présidentiel pour renforcer les garanties concernant la collecte et l'utilisation des données personnelles par les services de renseignement américains.

Ce nouveau cadre transatlantique a été soumis à la Commission européenne afin qu'elle évalue s'il permet d'assurer un niveau de protection adéquat des données des Européens. Avant d'adopter définitivement sa décision reconnaissant le caractère adéquat de ce nouveau dispositif, la Commission a soumis, le 13 décembre 2022, un projet de décision pour avis au Comité européen de protection des données (CEPD).

Le 28 février 2023, le CEPD a adopté et publié son avis sur ce projet d'adéquation. Le CEPD y relève les améliorations apportées par le gouvernement des États-Unis, tout en faisant part de ses préoccupations sur un certain nombre de points dont il dresse la liste.

Le 10 juillet 2023, la Commission européenne a adopté une nouvelle décision d'adéquation constatant que les États-Unis assurent un niveau de protection substantiellement équivalent à celui de l'UE, permettant ainsi, sous certaines conditions, le transfert de données personnelles vers ce pays, sans exigences supplémentaires.

Source : Site de la CNIL

Cette décision fait suite à l'invalidation de la précédente décision d'adéquation par la Cour de justice de l'Union européenne (CJUE), qui a conduit les États-Unis à prendre une série de mesures pour se rapprocher des standards européens. Elles seront complétées par les travaux en cours destinés à accélérer cette convergence et à favoriser la coopération entre l'Union européenne et les États-Unis dans le secteur de l'intelligence artificielle.

À l'occasion de son audition en vue de sa reconduction à la tête de la CNIL, Mme Marie-Laure Denis a indiqué qu'« *il est à noter que plusieurs projets de réforme de la législation états-unienne ont été présentés à la Chambre des représentants en décembre et pourraient, s'ils étaient adoptés, entraîner le réexamen de la décision d'adéquation. Les évolutions du cadre législatif américain sont donc suivies de près* » (1).

B. ÉLABORER UNE LÉGISLATION PROTECTRICE DES PRINCIPES EUROPÉENS QUI S'IMPOSE ÉGALEMENT AUX ENTREPRISES ÉTRANGÈRES

1. Trouver le juste équilibre entre droit dur et droit souple

Vos rapporteurs sont convaincus que l'IA ne peut être uniquement régulée par du droit dur, inscrit dans les textes législatifs et réglementaires européens ou nationaux. Ce droit risque de se montrer trop contraignant et trop rigide pour s'adapter aux évolutions, peu prévisibles, de ce domaine. Il ne suffira pas à répondre au besoin de sécurité juridique des entreprises, qui ne peuvent prendre le risque de s'exposer aux amendes prévues. À terme, il risque d'entraver l'innovation et la concurrence sur ce marché entre les acteurs disposant d'une expertise juridique et de moyens pour faire face au risque contentieux et les autres.

Des grands principes technologiquement neutres doivent bien être adoptés, conforme aux principes européens. Ils doivent pouvoir faire l'objet d'un contrôle *ex-ante* pour les systèmes à haut risque d'origine européenne et pour les systèmes conçus en dehors de l'Union européenne ; puis *ex-post* en cas d'incident comme cela est prévu dans le projet de règlement.

En amont de la commercialisation des IA, les règles semblent devoir être moins strictes et s'appuyer davantage sur le droit souple, coordonné au niveau européen, pour permettre les expérimentations dans un cadre juridique favorable. Les instances de dialogue entre les organes de protection de données des États membres devraient pouvoir élaborer des règles communes permettant d'adapter le droit commun à certaines exigences de l'IA. Les recommandations pourraient notamment porter sur une interprétation plus souple du RGPD, afin de permettre l'entraînement des modèles d'IA sur le sol européen.

---

(1) Réponses écrites de Mme Marie-Laure Denis en vue de son audition par la commission des Lois de l'Assemblée nationale, le 17 janvier 2024.

En France, la CNIL a commencé à le faire en accompagnant des projets pour les sécuriser juridiquement. Elle a également indiqué à vos rapporteurs qu'elle publierait prochainement un guide sur l'IA et des fiches pratiques sur la constitution des bases d'entraînement des IAG. La place accordée au droit souple peut également permettre d'étendre la marge d'appréciation accordée aux États membres [(1)].

En aval de la commercialisation, des mécanismes de conformité (ou *compliance*) pourront être mis en place pour prévenir les risques, comme cela existe aujourd'hui en matière de protection de l'environnement ou de prévention de la corruption. Ce type de mécanisme repose généralement sur une cartographie des risques, une évaluation des tiers, une procédure de contrôle et de sanction interne et un code de conduite.

Enfin, en complément des règles européennes, les États peuvent s'accorder sur des normes communes (spécification, cahier des charges, standards…) au niveau européen. L'association française de normalisation (AFNOR) a indiqué, lors de son audition, qu'il serait possible et souhaitable d'établir des standards précis en matière notamment de robustesse, de confidentialité, de cyber sécurité et d'interopérabilité. La maîtrise de ces normes peut constituer un avantage concurrentiel indéniable.

Recommandation n° 5 : Imposer le respect du RGPD et de grands principes technologiquement neutres aux services d'IAG proposés au sein de l'Union européenne.

Recommandation n° 6 : Promouvoir un droit souple et des mécanismes de conformité (*compliance*) en complément de mesures contraignantes.

Recommandation n° 7 : Laisser aux États membres de l'Union européenne une marge d'appréciation suffisante pour leur permettre de soutenir l'innovation.

2. Le droit, un outil parmi d'autres pour soutenir l'innovation au niveau européen

Le contrôle *ex post* permettrait de pénaliser les entreprises ne respectant pas les principes énoncés au niveau européen et celles bénéficiant d'une position dominante grâce au non-respect des règles édictées par l'Europe. Il obligerait en outre les acteurs présents dans le domaine de l'IA à se conformer aux exigences européennes *a posteriori* et permettrait de les sanctionner, par exemple si elles continuent d'exploiter gratuitement des données soumises au droit d'auteur. Dans le même temps, les acteurs vertueux pourraient continuer de se développer et rattraper leur retard.

Comme l'a indiqué la DG CNECT, il n'est pas possible d'interdire *a priori* les IA extra-européennes. En revanche, les IA entraînées à l'étranger, qui ne sont pas nécessairement soumises au RGPD pour l'ensemble de leurs activités, pourraient faire l'objet de contrôles très stricts quant aux données qu'elles auraient préalablement utilisées.

(1) Voir deuxième partie.

Incontestablement, la modification et les modalités d'application du droit européen seront décisives pour soutenir l'innovation en matière d'IAG. Cela ne sera toutefois pas suffisant, et il est indispensable que cet effort juridique soit accompagné d'un effort financier. Cette question ne relève pas directement de la compétence de la commission des Lois, mais vos rapporteurs sont attentifs aux propositions pouvant émaner d'autres parlementaires sur ce sujet.

L'IAG exige effectivement des investissements élevés et des infrastructures importantes (calculateurs, serveurs, jeux de données). L'UE pourrait intervenir davantage pour développer ces dernières et les mettre à disposition des acteurs économiques, afin de faciliter leur développement et de mutualiser les efforts financiers.

Vos rapporteurs plaident également en faveur d'un projet de grande ampleur, sur la base d'un accord entre les État membres qui le souhaitent et qui partagent la même vision de ce que doit être une IA européenne, par exemple via la création d'une société européenne à capitaux partiellement publics, sur le modèle d'Airbus [(1)].

Ils considèrent également que le développement de systèmes d'IAG de confiance peut aussi passer par des obligations d'investissement des grands groupes sur le modèle de celles qui existent en matière culturelle ou de logement [(2)]. Cela pourrait prendre la forme d'un « 1 % IA », mis en œuvre de manière coordonnée au niveau européen, dont seraient exonérées les entreprises qui financent déjà l'innovation dans ce domaine.

Recommandation n° 8 : Promouvoir un « Airbus de l'IAG » sur la base d'une coopération intergouvernementale entre pays européens volontaires et inciter les États membres à orienter des financements des entreprises vers l'innovation en matière d'IA.

*

* *

Les débats européens et internationaux sont riches et lourds d'enjeux. Ce sont eux qui définiront, pour les prochaines années, le cadre juridique dans lequel se développera l'IA. Les IAG, pour certaines accessibles à tous, y compris à un public non averti, font l'objet d'une attention croissante, car elles pourraient devenir les instruments d'une lutte d'influence politique ou la source de nouvelles problématiques. Plus que jamais, la cohérence du projet européen apparaît comme une priorité, et c'est l'occasion de démontrer sa capacité à défendre les principes qui sont les siens, en particulier l'État de droit.

(1) Pour mémoire, le groupe Airbus est détenu par l'État français à hauteur de 11 %, par l'État allemand à hauteur de 10,9 % et par l'État espagnol à hauteur de 4,17 %.

(2) Voir notamment les recommandations du rapport « Soutenir l'investissement dans les start-ups, PME innovantes et PME de croissance *», Paul Midy, mission pour le Gouvernement, juin 2023.*

Les pistes de régulation ont déjà fait l'objet de nombreuses discussions et un cadre réglementaire commence à se dessiner. Celui-ci doit rester attentif aux conséquences économiques d'une réglementation trop exigeante à l'égard des acteurs émergents. Afin de limiter la contrainte administrative résultant du droit de l'UE, qui menace l'innovation, il faut faire confiance aux États membres pour mettre en place des mécanismes de droit souple et affiner, au fil des avancées technologiques, leur droit national en matière civile et pénale.

*

* *

DEUXIÈME PARTIE :
LES QUESTIONS CLÉS À TRAITER AU NIVEAU NATIONAL

En parallèle de l'élaboration d'une réglementation européenne, la France peut contribuer par son droit interne au développement d'une IAG de confiance et respectueuse de l'État de droit. Par la manière dont elle organisera la gouvernance de la régulation, par le rôle moteur qu'elle peut faire jouer à la commande publique, par les outils dont elle dotera les pouvoirs publics et par les modifications législatives de son droit civil et pénal, elle peut devenir un modèle en matière de développement responsable des IAG et créer un contexte propice à l'émergence de champions économiques.

L'enjeu est tout à la fois de créer un cadre favorable à la diffusion de l'IAG et de protéger les citoyens et les libertés fondamentales.

I. CRÉER UN CADRE FAVORABLE À LA DIFFUSION DE L'IAG, NOTAMMENT DANS LE SECTEUR PUBLIC

A. ADAPTER LA GOUVERNANCE DE LA PROTECTION DES DONNÉES

1. Désigner un régulateur compétent en matière d'IAG

a. La CNIL apparaît comme la mieux à même de réguler les IAG

La réglementation européenne à venir va mettre en œuvre différents mécanismes de régulation. Cette régulation devrait être confiée, comme c'est habituellement le cas, à des autorités indépendantes.

Plusieurs autorités peuvent se considérer comme compétente au regard des enjeux concernés, en particulier la CNIL en ce qui concerne la protection des données personnelles et l'Autorité de régulation de la communication audiovisuelle et numérique (ARCOM) en ce qui concerne la régulation des plateformes.

Leurs interventions doivent être coordonnées, mais il apparaît nécessaire de désigner l'une d'entre elles pour piloter cette nouvelle politique. Vos rapporteurs estiment que la CNIL, dans un rôle redéfini, doit remplir ce rôle. La place qu'occupent les données personnelles dans les problématiques relatives à l'IAG est majeure, puisqu'elle intervient au stade de l'entraînement du modèle, de son apprentissage, de l'utilisation des données fournies par les usagers et de l'usage des données produites.

Or, seule la CNIL dispose de l'expertise et de l'expérience nécessaires pour évaluer le respect du RGPD par les modèles d'IAG. Elle a l'habitude de traiter les plaintes des usagers, de les instruire et, si nécessaire, de prononcer des sanctions. Les auditions ont mis en évidence la maturité de la réflexion de la CNIL.

Dans sa contribution écrite, la CNIL rappelle que « *l'IA n'est pas un phénomène entièrement nouveau pour la CNIL qui a eu l'occasion d'observer sa progression dans les usages quotidiens du numérique. Cela ne correspond finalement qu'aux dernières générations d'algorithmes de traitement de données, que l'informatique produit depuis plus de 40 ans. Toutefois, comme en témoigne l'importance du sujet ces dernières années, l'IA pose des questions nouvelles en matière de discriminations et de biais, mais également des questions fondamentales sur l'utilisation des données personnelles (principes de finalité, de minimisation des données, d'information et de transparence, limitation des durées de conservation, exercice des droits, impératif de sécurité notamment). La CNIL est constamment sollicitée sur la mise en œuvre de traitements de données personnelles impliquant des technologies d'IA, et cela dans de très nombreux domaines*. »

Vos rapporteurs se félicitent également que la CNIL ait déjà adopté une démarche aidante pour le développement de l'IAG en France, en apportant aux entrepreneurs les clarifications nécessaires pour l'interprétation d'une réglementation complexe en matière de protection des données.

b. La nécessaire évolution du statut et des moyens de la CNIL

Pour autant, la CNIL ne saurait parvenir à réguler efficacement l'IAG sans évoluer. Les acteurs économiques du secteur semblent en effet partager le sentiment selon lequel la CNIL ne dispose pas des moyens suffisants pour contrôler efficacement ce nouveau secteur.

Dans l'optique de répondre aux questions précédentes, la CNIL a annoncé en janvier 2023 la création d'un service dédié à l'IA. Celui-ci rassemble une équipe pluridisciplinaire de cinq personnes, composée d'analystes, juristes et ingénieurs, chargée d'apporter son expertise aux autres services et directions de la CNIL. Compte tenu de l'accélération des développements et des déploiements des systèmes d'IA (dans leur ensemble) et des questions récurrentes sur la constitution et l'utilisation de bases de données d'entraînement, la CNIL a présenté, le 16 mai dernier, un plan d'action qui s'articule autour de 4 volets :

– appréhender le fonctionnement des systèmes d'IA (et notamment d'IAG) et leurs impacts pour les personnes ;

– guider le développement d'IA respectueuses de la vie privée, en clarifiant le cadre juridique applicable aux systèmes d'IA (et notamment d'IAG) ;

– fédérer et accompagner les acteurs innovants dans le domaine en France et en Europe ;

– auditer les systèmes d'IA et protéger les personnes [(1)].

Vos Rapporteurs estiment qu'il faut aller plus loin, en redéfinissant dans leur ensemble les missions de la CNIL et son périmètre d'intervention. Ils proposent de transformer la CNIL en une Haute Autorité en charge de la protection des données et du contrôle de l'intelligence artificielle. Elle serait dotée d'un grand nombre d'experts et techniciens en mesure de contrôler des algorithmes complexes.

Recommandation n° 9 : Transformer la CNIL en une Haute autorité en charge de la protection des données et du contrôle de l'IAG.

Recommandation n° 10 : Veiller à l'articulation de cette nouvelle autorité avec l'ARCOM, compétente en matière de régulation des plateformes.

2. Favoriser le rôle d'accompagnateur du régulateur

Vos rapporteurs estiment que la CNIL, au-delà de sa mission de régulateur, doit être confortée dans son rôle d'accompagnateur du secteur. Les entreprises de l'IAG disent être confrontées à une législation complexe et parfois imparfaitement adaptée aux problématiques spécifiques qu'elles rencontrent. Elles craignent de commettre des fautes qui pourraient, à plus long terme, condamner leurs innovations.

Il est, à ce titre, indispensable que le régulateur puisse également jouer un rôle de conseil et apporter ainsi une sécurité juridique supplémentaire aux acteurs économiques.

La CNIL a indiqué déjà remplir ce rôle : « *La CNIL échange régulièrement avec les fournisseurs de solutions d'IA et évalue leurs systèmes. Elle accompagne ainsi les divers responsables de traitements de données utilisant de l'IA (start-ups, PME, grandes entreprises publiques). Le déploiement de ces outils permet à la CNIL d'expliquer l'application du RGPD à l'IA (constitution de jeux de données d'entraînement, minimisation de la collecte de données, information des personnes)* ».

La stratégie d'accompagnement repose sur quatre piliers :

– la production de « droit souple » (référentiels, lignes directrices, recommandations, etc.), qui précise l'application de la réglementation à des cas d'usage. La production de ces outils, destinés à être concrets et opérationnels, se fait généralement grâce à des concertations et consultations, que la CNIL organise en amont de sa régulation afin de mieux appréhender la réalité des activités et d'en tenir compte dans son processus décisionnel ;

(1) Contribution écrite de la CNIL.

– l'existence d'une procédure de « demande de conseils », ouverte notamment aux entreprises et fédérations professionnelles, permettant d'obtenir une réponse écrite de la CNIL sur un cas d'application incertain de la réglementation ;

– des outils de formation ;

– des « bacs à sable » permettant de tester des innovations dans un cadre juridiquement sécurisé.

– une offre d'accompagnement renforcé au cas par cas.

Comme ses représentants s'y étaient engagés lors de leur audition par vos rapporteurs, la CNIL a formulé sur son site internet des conseils et des recommandations concrètes pour le développement des systèmes d'IAG et la constitution de bases de données utilisées pour leur apprentissage, qui impliquent des données personnelles.

Une série de fiches thématiques a été élaborée et publiée. Elle permet d'accompagner utilement un large nombre de professionnels aux profils aussi bien juridique que technique, notamment s'agissant de la phase de développement d'un système d'IAG.

Il s'agit d'un guichet utile et bien identifié par les acteurs, mais ses moyens sont insuffisants. Pour remplir ces missions, des moyens humains et financiers supplémentaires sont également nécessaires. Les agents chargés de cet accompagnement doivent disposer non seulement des compétences juridiques et techniques nécessaires, mais aussi d'une bonne connaissance du secteur du numérique comme des enjeux financiers et stratégiques propres à l'IA.

Recommandation n° 11 : Promouvoir le rôle d'accompagnateur du régulateur en matière d'IAG.

B. INTRODUIRE UNE CULTURE DE L'IAG ET FAVORISER LES EXPÉRIMENTATIONS DANS LE SECTEUR PUBLIC

En août 2022, un rapport du Conseil d'État appelait à « *construire une intelligence artificielle publique de confiance* » [(1)]. Les auditions de la mission d'information ont mis en évidence une grande hétérogénéité des usages de l'IAG par les administrations. Certaines d'entre elles, comme la gendarmerie nationale, sont déjà très en avance dans leur utilisation, tandis que beaucoup d'autres n'ont encore mené aucune réflexion. Vos rapporteurs ont adressé à l'ensemble des administrations métiers un questionnaire sur leur rapport avec l'IAG, mais seules la direction des affaires criminelles et des grâces du ministère de la justice et la direction générale de la gendarmerie nationale y ont répondu.

(1) *Conseil d'État, «* Intelligence artificielle et action publique : construire la confiance, servir la performance *», août 2022.*

Il apparaît donc urgent qu'une prise de conscience ait lieu au sein de l'État pour saisir les opportunités et prévenir les risques liés à l'IAG, au moment où des expérimentations ont été – ou pourraient prochainement être – engagées dans les domaines de la sécurité, de l'accès au service public ou de la vie démocratique.

1. Introduire une culture de l'IAG dans le secteur public

a. Identifier les opportunités offertes par l'IAG grâce aux agents et aux usagers

L'utilisation de l'IAG au sein du secteur public implique, en premier lieu, d'évaluer l'apport de cette technologie aux missions de l'État et de ses agents. L'accès libre à l'agent conversationnel *ChatGPT* a permis à tout un chacun de mesurer l'intérêt et les limites de ces outils. Il est indispensable de laisser les agents et les administrations réfléchir par eux-mêmes aux usages pertinents de l'IAG dans leur domaine.

Les tâches pouvant être remplacées ou accompagnées par l'IAG pour les rendre plus efficaces ou moins pénibles doivent être identifiées et consolidées au niveau interministériel pour mutualiser ce qui peut l'être et spécialiser ensuite les usages pour chaque administration.

Les agents sont souvent les mieux placés pour identifier eux-mêmes les cas d'usage possibles de l'IAG. Il est indispensable de leur permettre de formuler des propositions en la matière et de les associer à la conception du système d'IAG. Un système d'IAG nécessite, en effet, toujours un renforcement humain pour garantir sa fiabilité. Seuls des agents qui exercent les métiers à destination desquels le système d'IAG est conçu devraient participer à ce renforcement humain. Ce renforcement humain ne devrait être délégué à des prestataires externes que dans des cas exceptionnels et justifiés.

Si les agents y sont suffisamment associés et formés, l'IAG peut être une source de gain d'efficacité permettant de dégager du temps utile pour d'autres tâches aujourd'hui délaissées (accueil du public, contrôle de légalité, accompagnement des entreprises et des collectivités locales).

Il ne s'agit pas de remplacer des agents par des robots mais, au contraire, de leur fournir des outils facilitant leur travail et leur permettant de se concentrer sur les tâches où l'intervention humaine est la plus indispensable.

Dans un second temps, les agents devront être accompagnés par des formations adaptées, afin d'éviter de mauvais usages de l'IAG et les risques qui y sont inhérents. La rédaction des prompts est certainement une nouvelle compétence à développer parmi les personnes qui seront confrontées à l'usage des IAG – y compris lors de la scolarité.

Recommandation n° 12 : Identifier dans le secteur public les tâches pouvant, à terme, être déléguées en tout ou partie à des systèmes d'IAG.

Recommandation n° 13 : Former les agents publics aux opportunités et aux risques inhérents à l'usage des systèmes d'IAG.

b. Développer des IAG au service de l'administration dans un environnement sécurisé

L'usage de l'IAG par les services publics présente en effet divers enjeux de sécurité.

● **En premier lieu**, il est indispensable d'en garantir **la sécurité informatique**.

L'Agence nationale de la sécurité des systèmes d'information (ANSSI) a souligné, lors de son audition, qu'il « *conviendra de vérifier que le déploiement de l'IAG s'effectue de manière sécurisée, c'est-à-dire qu'il n'introduit pas des* ***vulnérabilités supplémentaires dans les systèmes d'information*** *au sein desquels ils sont insérés* ».

Il faudra aussi s'assurer que l'utilisation de l'IAG par les administrations n'occasionne pas des **fuites de données sensibles vers des pays tiers**. Si les solutions utilisées par les administrations sont essentiellement des solutions dont les données sont hébergées en France, l'utilisation par les agents publics de solutions directement accessibles et hébergées dans des pays non européens n'est pas à exclure. Or, comme le rappelle l'ANSSI, « *toute utilisation d'une IAG avec des données sensibles a pour conséquence l'exfiltration de ces données vers des pays non européens. Elles peuvent ensuite être soit consultées par les entreprises qui gèrent ces IA, soit intégrées dans des résultats fournis à d'autres utilisateurs. Dans tous les cas, le risque que ces données soient divulguées à des personnes non autorisées est particulièrement élevé. Par conséquent, l'ANSSI proscrit l'usage d'IAGs pour les données soumises à une réglementation particulière (données confidentielles de défense, données marquées diffusion restreinte)* ».

Il est donc indispensable de sensibiliser les agents aux risques qu'induit l'utilisation des IAG – en complément de la formation sur ses opportunités, préalablement mentionnée.

● **En deuxième lieu**, il faut en assurer **la solidité juridique**. Les systèmes d'IAG utilisés sont susceptibles d'avoir des conséquences sur les droits et libertés des usagers. L'administration dispose de pouvoirs et de prérogatives qui confèrent à la plupart de ses décisions le caractère d'actes administratifs, dont la plupart sont désormais susceptibles de recours.

Le fait de s'appuyer sur l'IAG pour accompagner la prise des décisions implique donc une grande fiabilité et une capacité de contrôle externe par les agents.

● **En troisième lieu**, il est question de **souveraineté économique**. Pour développer ces projets, les administrations s'appuient sur des solutions clefs en main, sans que leurs données remontent vers le fournisseur de la solution. Elles peuvent ensuite entraîner spécifiquement ces IAG pour qu'elles deviennent performantes dans leur domaine d'application.

Néanmoins, ces solutions, dites « sur étagère », sont aujourd'hui proposées par des entreprises souvent extra-européennes, risquant de rendre l'administration dépendante de ces technologies et de réduire l'innovation française au perfectionnement d'IAG embarquées dont la conception et l'entraînement en amont nous échappent.

Vos rapporteurs tiennent à souligner le rôle déterminant que peut jouer la commande publique. Il convient d'inciter les acheteurs publics à s'orienter vers des systèmes d'IAG de confiance labellisés. Cette labélisation, pour laquelle l'Afnor a montré son intérêt lors de son audition, pourrait notamment porter sur la transparence du modèle et le respect des règles du RGPD lors de la phase d'apprentissage.

Recommandation n° 14 : Inciter les acheteurs publics à s'orienter vers des systèmes d'IAG de confiance labellisés.

c. Coordonner et accompagner les administrations utilisatrices d'IAG au niveau interministériel

L'**accompagnement technique des administrations** souhaitant développer l'usage des IAG en leur sein existe. Elles peuvent s'appuyer sur l'ANSSI et la CNIL en ce qui concerne la cyber sécurité et la protection des données ; et sur la direction interministérielle du numérique (DINUM) et le pôle d'expertise et de régulation du numérique (PéREN) pour le développement de leurs applications métiers.

Le développement des applications de l'IAG au sein du service public requiert en effet à la fois des compétences informatiques et une compréhension des spécificités des besoins des administrations. Le PéREN a ainsi « *pour mission de fournir aux services de l'État une assistance et une expertise techniques dans les domaines de la science des données et des procédés algorithmiques* ». Il est amené à conseiller les administrations souhaitant recourir à l'IAG et dispose en interne de capacités de développement d'applications, même si encore peu d'administrations l'ont saisi à cette fin.

Pour autant, **la coordination interministérielle apparaît insuffisante**. Plusieurs administrations disposent d'une compétence interministérielle en matière de numérique, qu'il s'agisse de l'ANSSI sous l'angle de la cybersécurité, du PEREN sous l'angle de l'expertise numérique, du ministère de la transformation publique pour la formation des agents, ou d'Étalab pour la gestion des données. En janvier 2023, un coordinateur national pour l'intelligence artificielle a été désigné par le Gouvernement en la personne de M. Guillaume Avrin, mais sa mission est principalement tournée vers l'investissement et l'accompagnement du développement des systèmes IA en France.

Malgré cette pluralité d'acteurs, les échanges interministériels sur le recours à l'IAG au sein de l'État apparaissent limités : l'ANSSI a par exemple indiqué qu'elle ne participait pour le moment à aucune instance de dialogue et de concertation entre les administrations sur la question de l'IAG, alors même que des expérimentations voient le jour (voir ci-après). La DACG évoque, quant à elle, une « *amorce de concertation* » autour de la DINUM, en lien avec le coordonnateur pour l'IA, pour « *permettre un premier dialogue entre les directions du numérique des différents ministères via notamment le réseau des administrateurs ministériels des données, des algorithmes et des codes sources* ».

Il serait utile que voit le jour une instance de dialogue interministérielle, tournée vers les technologies, mais aussi et surtout vers les usages. Au sein de l'une des entités existantes compétentes en la matière, cette instance pourrait tout à la fois permettre le partage des bonnes pratiques et la centralisation des besoins.

Recommandation n° 15 : Prévoir une stratégie pluriannuelle en matière d'IAG, comprenant un volet consacré à son usage dans les administrations.

Recommandation n° 16 : Internaliser des compétences en matière de développement d'applications d'IAG à partir de modèles d'IAG.

Recommandation n° 17 : Prévoir une instance permanente de dialogue entre les administrations sur l'usage des IAG.

2. Favoriser les expérimentations dans le domaine régalien

a. En matière de sécurité

La Gendarmerie nationale est l'administration qui apparaît la plus mature dans son utilisation de l'IAG. Cela s'explique par son exposition directe aux usages détournés de l'IAG à des fins criminelles.

La DGGN identifie en effet de nombreux risques sécuritaires directement liés à l'usage des IAG : « *à court terme, la désinformation et la fraude à l'identité ; à moyen terme*, […] *la création de nouvelles molécules, de cellules souches ou encore de vaccins comme de virus ; à long terme,* […] *des attaques cyber plus performantes comme les attaques dites adverses qui modifient l'interprétation des systèmes d'IA* ».

Pour lutter contre ces risques, la DGGN a développé ses propres outils d'IAG. En s'appuyant sur des bases de données (sons, textes, images) qui lui sont propres, elle a mené un travail d'apprentissage de systèmes d'IA, visant à détecter le détournement des IAG à des fins délictuelles ou criminelles. Elle dispose ainsi, par exemple, d'un logiciel de détection des hypertrucages (*deepfakes*).

L'IAG est également utilisée comme une technique d'enquête, par exemple en procédant au rajeunissement ou au vieillissement de portraits de personnes à des fins d'identification criminalistique ou de recherche de personnes disparues.

Le ministère de la justice a également engagé des réflexions pour recourir davantage à l'IA dans ses logiciels de rapprochement, qui ont pour finalité de faire le lien entre des affaires ou d'aider à identifier les éléments pertinents d'un dossier d'enquête [(1)].

D'autres enjeux de sécurité pourraient être traités au moyen de l'IAG, par exemple dans le champ de cybersécurité, pour prévenir ou combattre des attaques cyber. Comme l'a indiqué l'ANSSI, « *les applications de l'IAG pour la cybersécurité sont multiples. Par exemple, l'IAG pourrait assister l'ANSSI lorsqu'elle est amenée à rédiger des grandes quantités de documentation technique (spécifications, standards). On peut aussi imaginer des applications dans le domaine de l'analyse de la menace (synthèse de rapports d'incidents, de discussions sur des forums d'attaquants,* etc.*). À plus long terme, on peut espérer des progrès sur l'analyse de code informatique ou sur la production de rapports et l'aide à la décision à partir de données techniques*. »

La France a fait valoir, au niveau européen, son inquiétude – que vos rapporteurs partagent – quant à un encadrement trop strict des possibilités de recours à l'IAG dans les domaines régaliens, alors même que ces usages peuvent être utiles pour lutter contre les abus de l'IAG elle-même et qu'elle sera, en tout état de cause, encadrée très strictement par la loi et le pouvoir réglementaire [(2)].

b. En matière de service aux usagers

Le service aux usagers est certainement une voie de développement prometteuse pour l'IA dans le service public.

Une expérimentation a été lancée par le ministère de la transformation et de la fonction publique, pour une durée de six mois, et doit permettre à 200 agents des maisons France Services (MFS) de tester deux agents conversationnels pour améliorer la réponse aux usagers. Les agents des maisons France services doivent en effet accompagner les usagers dans de très nombreuses démarches, sans appartenir aux administrations qui en assurent le traitement (prestations sociales, recherche d'emploi, demandes de titres, *etc.*).

Le rapporteur pour avis de la commission des Lois sur la mission « Administration générale et territoriale de l'État (AGTE) » constatait que « *la création de MFS a pu conduire à la fermeture de certains guichets, ce qui conduit certes à une amélioration de la proximité mais aussi à une dégradation de l'efficacité de la réponse apportée. Le Défenseur des droits a réalisé des testings dans les MFS qui l'ont conduit a estimé que le taux de réponse claire et juste sur des demandes d'information était inférieur à 50 %* » [(3)].

---

(1) Voir II. A. 3.

(2) Voir première partie.

(3) Assemblée nationale, Rapport pour avis sur le projet de loi de finances pour 2024 (Mission AGTE)*, M. Ugo Bernalicis, 18 octobre 2023, n° 1778, p. 15.*

L'IAG est en mesure de répondre efficacement à des demandes directes d'usagers ou par l'intermédiaire d'agents formés à la rédaction de prompts permettant d'obtenir les bonnes réponses et capables, le cas échéant, de les vérifier. Un agent conversationnel sans intermédiaire pourrait également être envisagé à plus long terme. En tout état de cause, il faudra toutefois veiller à ce que les réponses apportées ne soient pas directement créatrices de droit car aucun système ne sera infaillible (1).

Pour encourager l'innovation en la matière, la CNIL a également mis en place un « bac à sable » sur l'usage de l'IA dans le cadre de services publics. Il s'agit d'un dispositif d'accompagnement, par lequel des organismes – publics et privés – sélectionnés sur la base d'un appel à projets thématique, bénéficient d'un accompagnement juridique et technique poussé sur leur projet. Les projets retenus concernent la qualité du service rendu aux usagers, l'accès aux services publics et le soutien de la performance des agents publics (2).

Recommandation n° 18 : Mettre à la disposition du public, à horizon 2027, un agent conversationnel fiable reposant sur l'IAG pour renforcer l'accès au droit et à l'information administrative.

c. Dans le champ de la vie démocratique

La question de la désinformation n'est pas uniquement une question de sécurité, mais également de souveraineté et de santé démocratique. Le recours à l'intelligence artificielle dans le cadre des campagnes électorales est probablement inévitable, car elle permet de démultiplier, à moindre coût, les capacités de communication et d'influence des candidats (vidéo de propagande, production d'argumentaire notamment).

Outre les obligations qui s'appliqueront de manière générale au système d'IAG, des obligations supplémentaires pourraient s'imposer aux candidats, de telle sorte qu'ils soient dans l'obligation de signaler les outils d'IAG auxquels ils ont eu recours et de marquer les contenus qu'ils ont produits grâce à ces IAG.

Il s'agirait de l'extension logique du principe selon lequel l'utilisateur d'un agent conversationnel doit savoir s'il échange avec un robot ou avec un humain – voire avec un humain assisté par un robot.

Il y va de la sincérité du scrutin, puisque les électeurs pourraient facilement être manipulés par des montages ou des trucages visant à promouvoir un candidat, ou au contraire à discréditer son adversaire.

Recommandation n° 19 : Promouvoir l'étiquetage des contenus produits par IAG en matière de propagande électorale.

Recommandation n° 20 : Étendre l'obligation d'information en cas d'interaction avec un système d'IAG aux agents conversationnels utilisés dans le cadre des campagnes électorales.

(1) Voir I. B.1.b de la présente deuxième partie.

(2) Contribution écrite de la CNIL.

D'un point de vue méthodologique, vos rapporteurs estiment qu'il faut que cette conversion de l'État à l'IAG se fasse de manière progressive, mais en se donnant les moyens d'innover y compris juridiquement. Ils suggèrent le recours à l'expérimentation et à l'évaluation, d'abord à petite échelle, avant de lancer de grands projets très coûteux et qui soulèveront nécessairement des questions éthiques lourdes.

Ces expérimentations doivent également être l'occasion de promouvoir l'innovation de manière plus large, en finançant des bourses à l'intention des chercheurs français et en mettant à disposition de petites entreprises des capacités de calcul et des jeux de données qui seront porteuses d'externalités positives pour l'ensemble de l'écosystème de l'IAG.

Recommandation n° 21 : Mettre en place le cadre juridique adéquat pour permettre des expérimentations de systèmes d'IAG dans le domaine régalien.

Recommandation n° 22 : Soutenir financièrement les chercheurs français dans le domaine de l'IAG, au moyen de bourses et de la mise à disposition de capacités de calcul.

*

* *

II. PROTÉGER LES CITOYENS ET LES LIBERTÉS FONDAMENTALES

En droit interne, l'État sera également appelé à intervenir pour réguler les usages de l'IAG, notamment lorsque les fonctionnalités offertes seront détournées pour commettre des infractions ou pour provoquer des dommages.

Le législateur français dispose d'une marge d'appréciation importante pour adapter le droit pénal et civil national à ce nouvel enjeu, en complément de la régulation européenne. Il s'agit d'apporter aux citoyens la protection nécessaire et d'encourager un usage responsable des IAG, conforme à la conception française des libertés fondamentales.

A. ADAPTER LA RÉPONSE PÉNALE AUX NOUVEAUX RISQUES

L'impact de l'IAG sur le droit pénal peut être de deux natures.

En premier lieu, l'IAG peut faciliter la commission d'infractions existantes, fournir une aide à leurs auteurs et leur permettre d'accroître les gains issus de leurs comportements frauduleux. L'IAG offre notamment une capacité à massifier les opérations délictueuses en les automatisant.

En second lieu, l'IAG peut être source de nouveaux comportements qui ne sont pas appréhendés, ou le sont mal, par les incriminations existantes.

1. Adapter la répression des infractions existantes

● Grâce à l'IAG, tout le monde peut devenir faussaire. La production de contenu synthétique vraisemblable nécessite peu de prérequis techniques.

Au titre des usages appréhendés par le droit pénal, on peut citer :

– les faux articles et fausses images à des fins de désinformation, réprimées par le délit de diffusion de fausses nouvelles prévu à l'article 27 de la loi du 29 juillet 1881 sur la liberté de la presse ;

– les usurpations d'identité par l'usage de l'hypertrucage (*deepfake*), réprimées par l'article 226-4-1 du code pénal ;

– la génération de pages réalistes pour des arnaques de type hameçonnage (« phishing ») ou la création de faux profils sur les réseaux sociaux à des fins de collecte de données personnelles ou pour inciter la victime à réaliser un paiement, réprimées par l'article 226-18 du code pénal ;

– la génération de code pour infecter des systèmes informatiques à l'aide de *malwares*, réprimée par l'article L. 163-4 du code monétaire et financier ;

– des attaques bactériologiques par la génération de nouvelles molécules, dont de nouveaux virus toxiques.

● Dans ces différentes hypothèses, l'usage de l'IAG n'est qu'un moyen d'accroître l'efficience du comportement délictueux. Elles ne nécessitent pas un changement de définition des infractions.

Toutefois, l'usage de l'IAG est susceptible de créer davantage de préjudices grâce à l'automatisation des tâches, l'amélioration de la vraisemblance des contenus ou encore par l'accroissement du nombre de contenus frauduleux produits.

Au regard du caractère massif des préjudices que peut causer l'usage d'une IAG, vos rapporteurs considèrent que la création d'une circonstance aggravante générale dans le code pénal devrait être envisagée.

Les circonstances générales sont définies aux articles 132-71 et suivants du code pénal (par exemple : la bande organisée, le guet-apens, la préméditation). L'article 132-79 prévoit une aggravation des peines d'emprisonnement encourues en cas d'usage d'un moyen de cryptologie. Cette circonstance aggravante pourrait être étendue à l'usage d'un contenu généré par un traitement algorithmique.

Recommandation n° 23 : Modifier l'article 132-79 du code pénal pour étendre la circonstance aggravante générale d'usage de moyens de cryptologie à l'usage d'un contenu généré par une IAG.

2. Adapter la définition de certaines incriminations pour appréhender de nouveaux comportements

L'apparition des IAG est susceptible de créer de nouvelles infractions jusque-là non imaginées, ou qui ne peuvent être aujourd'hui appréhendées qu'en interprétant plus largement une infraction déjà existante, ce qui présente un risque juridique non-négligeable.

Vos rapporteurs ont été surpris par l'analyse de la direction des affaires criminelles et des grâces (DACG), selon laquelle l'arsenal des infractions pénales prévues et réprimées par le code pénal permet d'appréhender l'utilisation détournée des contenus produits par les IAG. Dans ses réponses au questionnaire adressé, la DACG indique qu'il « *n'est pas identifié de carence du dispositif répressif pour appréhender l'utilisation des contenus générés par l'intelligence artificielle* ».

On peut pourtant craindre que certains usages détournés de l'IAG ne soient pas couverts par les incriminations existantes, bien qu'ils soient souhaitables de les éviter, en particulier s'agissant des plagiats et des faux.

a. Une nécessaire prise en compte des hypertrucages (« deepfake »)

La technologie de l'« hypertrucage » ou « permutation intelligente de visages » (*deepfake*) est une technique reposant sur l'intelligence artificielle et visant à fabriquer des synthèses d'images ou de vidéos réalistes.

Il s'agit d'un phénomène en pleine expansion, qui est mal appréhendé par le droit pénal. Certes, l'article 226-8 du code pénal réprime les montages réalisés sans le consentement de la personne s'il n'apparaît pas à l'évidence qu'il s'agit d'un montage. Mais il n'est pas certain qu'une représentation à l'aide d'un contenu généré par un traitement algorithmique soit toujours constitutive d'un « *montage* ».

Pour ces raisons, vos rapporteurs approuvent les amendements parlementaires, adoptés dans le cadre de l'examen du projet de loi visant à sécuriser et réguler l'espace numérique (SREN), qui ont modifié l'article 226-8 du code pénal.

L'article 4 *bis* de ce projet de loi, actuellement en navette parlementaire, ajoute à la répression du montage celle de l'usage d'un contenu généré par un traitement algorithmique.

Recommandation n° 24 : Pénaliser les hypertrucages (« *deepfake* ») réalisés sans le consentement de la personne représentée.

b. Une nécessaire adaptation de la définition du faux ou du plagiat

Il n'est pas acquis qu'un mémoire, une thèse, un rapport de stage produit par une IAG soit constitutif d'un faux ou d'un plagiat. De même, le participant à un concours de dessins, de poésie ou de nouvelles qui soumet des œuvres issues d'un système d'IAG ne pourrait probablement pas être inquiété pénalement.

Certes, au sens commun, le plagiat est l'acte qui, dans le domaine artistique ou littéraire, donne pour sien ce qui a été pris à un autre. En première analyse, on pourrait dès lors estimer que la personne qui présente comme sien le contenu produit par une IAG est un plagiaire.

Mais en droit pénal, le plagiat s'assimile à la contrefaçon et se définit comme « *toute reproduction, représentation ou diffusion, par quelque moyen que ce soit, d'une œuvre de l'esprit en violation des droits de l'auteur* » (article L 335-3 du code de la propriété intellectuelle). Or, le contenu original produit par une IAG ne heurte pas nécessairement le droit d'un auteur et l'IAG n'est pas nécessairement considérée comme l'auteur du contenu qu'elle génère.

Au surplus, il est aisé de programmer un système d'IAG pour paraphraser une œuvre préexistante en lui demandant de s'en écarter suffisamment pour rendre plus complexe la violation des droits d'auteur. L'IAG pourrait ainsi produire un contenu difficile à détecter pour les logiciels de détection des plagiats.

Vos rapporteurs proposent d'engager une réflexion sur une évolution du délit de contrefaçon pour sanctionner la pratique consistant à masquer, à l'aide d'une IAG, la reproduction de contenus préexistants.

Recommandation n° 25 : Engager une réflexion sur le délit de contrefaçon pour pénaliser la pratique consistant à masquer, à l'aide d'une IAG, la reproduction de contenus préexistants.

Dans les différentes hypothèses précitées, l'incrimination de faux n'est pas non plus établie de façon certaine. Cette infraction est en effet définie comme « *toute altération frauduleuse de la vérité, de nature à causer un préjudice et accomplie par quelque moyen que ce soit, dans un écrit ou tout autre support d'expression de la pensée qui a pour objet ou qui peut avoir pour effet d'établir la preuve d'un droit ou d'un fait ayant des conséquences juridiques* » (article 4441-1 du code pénal).

Or, le fait de s'approprier le contenu d'une IAG n'est pas à lui seul une altération frauduleuse de la vérité. La définition du faux pourrait, dès lors, être revue pour mieux appréhender certains comportements et prévoir que : « *constitue une altération frauduleuse de la vérité le fait de se prétendre auteur de l'écrit ou du support alors que celui-ci a été produit substantiellement par un système d'intelligence artificielle générative* ». Ceci permettrait d'appréhender pénalement des contenus générés par une IA en vue d'obtenir un diplôme ou d'acquérir des droits.

De même, une telle évolution permettrait de sanctionner pénalement les professionnels qui facturent des réponses « *personnalisées* » à leurs clients – écrites de façon automatisée grâce à une IAG – en leur laissant croire qu'ils ont répondu eux-mêmes.

Recommandation n° 26 : Prévoir en matière de faux que constitue une altération frauduleuse de la vérité le fait de se prétendre auteur de l'écrit ou du support alors que celui-ci a été produit substantiellement par une IAG.

c. Au-delà des usages, un renforcement de la protection des IAG pour éviter leur détournement

La DGGN et l'ANSSI s'inquiètent de la modification clandestine d'un système d'IA par « l'empoisonnement » de ses données de façon à introduire des biais dans les systèmes d'apprentissage, ou par la mise en place de restriction visant à ne fournir que certains points de vue ou informations. En effet, il a été indiqué à plusieurs reprises à vos rapporteurs que le mode d'entraînement d'une IAG ou ses réglages avaient des conséquences directes sur la nature des réponses apportées et pouvait facilement refléter une idéologie. Si une IAG est programmée pour répondre selon la position idéologique, cela va orienter ses réponses. Ce type d'acte pourrait être utilisé pour déstabiliser un pays ou simplement mener des stratégies d'influence.

L'article 323-1 du code pénal prévoit que « *le fait d'accéder ou de se maintenir, frauduleusement, dans tout ou partie d'un système de traitement automatisé de données est puni de trois ans d'emprisonnement et de 100 000 € d'amende. Lorsqu'il en est résulté soit la suppression ou la modification de données contenues dans le système, soit une altération du fonctionnement de ce système, la peine est de cinq ans d'emprisonnement et de 150 000 € d'amende* ». Cette infraction pourrait être adaptée pour couvrir explicitement les données d'entraînement des systèmes d'IAG.

Recommandation n° 27 : Adapter l'article 323-1 du code pénal pour sanctionner l'empoisonnement des données d'entraînement d'une IAG ou la modification clandestine de son algorithme.

3. Adapter les techniques d'enquête et le travail des magistrats

Comme le rappelle la DGGN, « *le droit à la protection de la vie privée est suffisamment générique sur le plan de la qualification légale de l'infraction pour sanctionner les atteintes liées à l'utilisation de l'IAG. Néanmoins, il devra permettre de mettre en évidence la matérialité des infractions afin de faire face à l'évolution de l'identité numérique* ».

Plus largement, ce sont de nombreuses techniques d'enquête qui devront être perfectionnées, souvent en s'appuyant également sur l'IAG. Comme indiqué précédemment, la gendarmerie a développé ses propres outils d'IAG : en s'appuyant sur des bases de données (sons, textes, images) qui lui sont propres, elle a mené un travail d'apprentissage de systèmes d'IA visant à détecter le détournement des IAG à des fins délictuelles ou criminelles.

Recommandation n° 28 : Développer les capacités des pouvoirs publics à identifier les contenus générés par l'intelligence artificielle et l'usage détourné de l'IAG.

Plus largement, l'IAG est en mesure d'accompagner les enquêteurs et les magistrats instructeurs dans leur analyse des pièces et des dossiers. Ces logiciels informatiques existent depuis longtemps (voir encadré) mais sont appelés à se perfectionner rapidement en s'appuyant sur l'IAG.

Enfin, l'IAG peut également être utilisée en soutien des magistrats. La DACG a ainsi indiqué que « *le principal domaine de développement pour lequel l'IAG offre des perspectives susceptibles de répondre à un besoin réel est celui de l'aide à la décision pour les magistrats. Des outils d'aide à la décision doivent en effet être proposés aux juridictions et intégrer les potentialités de l'intelligence artificielle. L'aide à la décision s'entend comme un outil facilitant, permettant d'indiquer au magistrat quelles sont les possibilités juridiques qui s'offrent à lui et non comme un outil qui rendrait la décision en lieu et place du magistrat, sans aucun élément explicatif. Il s'agirait ainsi de soutenir le magistrat dans son action, et en particulier dans son appréhension de la norme pénale et des jurisprudences* ».

En revanche, le ministère de la justice précise que l'IAG ne saurait être utilisée à des fins prédictives en matière de jugement : « *l'utilisation de l'intelligence artificielle générative reposant non seulement sur des données juridiques mais également sur les données spécifiques d'un dossier ou sur les antécédents d'une personne voire sur les habitudes d'une juridiction ou même d'un magistrat n'est pas d'actualité.* »

En tout état de cause, les IAG utilisées dans ces domaines devront être perfectionnées, car les biais et les erreurs peuvent avoir des conséquences très graves. Il serait notamment utile, selon la DACG, que les IAG utilisées à ces fins soient plus précises dans leur capacité à indiquer la source des données qu'elles mobilisent.

Les logiciels et systèmes de rapprochement judiciaires

La finalité de ce type de logiciels est d'aider l'enquêteur à gérer la complexité des informations judiciaires mises à sa disposition pour l'élucidation des faits dont il est saisi. Il s'agit de confronter l'ensemble des éléments présents en procédure, tout en établissant un dispositif progressif de levée d'anonymat des données à caractère personnel en fonction de leur utilité au regard de la manifestation de la vérité.

Ainsi, seuls les logiciels de rapprochement judiciaires aux fins d'analyse criminelle ont fait l'objet d'une déclaration auprès de la CNIL. Leur régime particulier est défini par le décret n°2012-687 du 7 mai 2012.

Il s'agit concrètement des dispositifs suivants :

– ANACRIM – ATRT est un traitement d'exploitation automatisée de relevés bancaires et de documents téléphoniques (facturation détaillée, localisation de relais, etc.), obtenus sur réquisitions judiciaires, afin de mettre en évidence les liens entre les données collectées ;

– ANACRIM-ANB est un logiciel permettant l'analyse et la représentation visuelle de données sous forme de graphiques relationnels ou événementiels ;

– MERCURE est un logiciel permettant aux services de police d'analyser et de traiter, dans tous les cadres d'enquêtes judiciaires, les données de téléphonie difficilement exploitables de façon manuelle. En outre, Mercure permet d'uniformiser les méthodes d'analyse et de traitement des données de téléphonie ;

– SALVAC (système d'analyse des liens de la violence associée aux crimes) est fichier d'analyse sérielle pour les affaires qui présentent un risque, manifeste ou seulement plausible, de relever d'une série de faits commis par un même auteur ou lorsque l'affaire paraît ne pas pouvoir être résolue par des méthodes traditionnelles d'analyse et d'investigation.

Source : Réponses de la DACG

B. ANTICIPER LES CONSÉQUENCES SUR LA RESPONSABILITÉ CIVILE

Comme toute nouveauté technologique, les systèmes d'IAG risquent, soit en raison de défauts de conception, soit en raison d'une mauvaise utilisation, de causer des dommages.

La problématique déborde le champ de l'IAG et concerne l'IA en général. Mais elle est particulièrement prégnante en matière d'IAG compte tenu de l'effet dit « *de boîte noire* », lié à l'opacité de sa phase d'apprentissage, des algorithmes et des données d'entraînement. La preuve de la responsabilité de l'éditeur du système d'IAG peut, dès lors, être difficile à apporter.

La question qui se pose est de savoir si le droit commun est suffisant et adapté pour traiter les actions en responsabilité, dans le cas de dommages causés par des produits et services dotés d'IA.

Sur le plan procédural également, le déséquilibre économique entre un éditeur puissant et un utilisateur isolé peut rendre difficile la juste réparation des préjudices subis.

Tant sur le fond du droit que sur les aspects procéduraux, il conviendra d'anticiper au plus vite les conséquences de la diffusion de l'IAG sur la responsabilité civile.

1. Un droit interne à adapter en lien avec le projet de directive européenne

À Bruxelles, la DG Justice de la Commission européenne a indiqué que différents régimes pouvaient s'appliquer aux IAG : la responsabilité sans faute, la responsabilité pour produit défectueux, ou encore la responsabilité extracontractuelle. Ces deux dernières devraient faire l'objet d'une adaptation, notamment dans le cadre d'une directive dont le projet a été rendu public en septembre 2022 (1). Lors de leur audition, les représentants du Medef ont toutefois indiqué à vos rapporteurs n'avoir aucune nouvelle de ce projet de directive.

L'objectif affiché par la Commission européenne était pourtant de garantir que les personnes lésées par les systèmes d'intelligence artificielle bénéficient du même niveau de protection que les personnes lésées par d'autres technologies.

La lenteur des institutions européennes à faire aboutir ce texte ne doit pas empêcher la France d'avancer sur l'évolution du droit de la responsabilité civile, quitte à reprendre dans sa législation les principaux aspects du projet de directive.

L'introduction d'un régime de responsabilité sans faute spécifique n'est pas une option que vos rapporteurs recommandent, dans la mesure où elle constituerait un frein évident à l'innovation. Elle est d'ailleurs écartée au niveau européen à ce stade.

En revanche, l'allègement de la charge de la preuve apparaît indispensable pour limiter l'asymétrie entre utilisateurs et fournisseurs. Il est possible d'adapter notre droit pour instaurer un partage de la preuve, en permettant au juge d'ordonner à l'éditeur de divulguer des éléments sur le fonctionnement du système d'IAG. À défaut d'éléments probants permettant d'écarter sa responsabilité, le juge pourrait retenir une présomption de faute imputable à l'éditeur.

Toujours dans une optique de soutien à l'innovation, il conviendrait d'engager une réflexion spécifique sur la responsabilité des fournisseurs des systèmes d'IAG qu'ils n'ont pas conçus eux-mêmes. Là encore, leur collaboration dans la charge de la preuve devrait être requise, mais le niveau d'exigence ne peut être le même que celui imposé aux concepteurs du système qui disposent, seuls, de certaines informations sur le fonctionnement du système.

Recommandation n° 29 : Adapter le régime de responsabilité des IAG à leurs spécificités, notamment avec un allègement de la charge de la preuve pour limiter l'asymétrie entre utilisateurs et fournisseurs.

Recommandation n° 30 : Réfléchir à la responsabilité des fournisseurs de service s'appuyant sur une IAG qu'ils n'ont pas conçue eux-mêmes.

(1) Voir encadré p. 41.

2. Une réforme du régime de l'action de groupe pour rétablir l'équilibre des forces entre utilisateurs et concepteurs des IAG

● Vos rapporteurs considèrent que, compte tenu de la taille des acteurs concernés, les préjudices causés par les systèmes d'IAG doivent pouvoir faire l'objet d'une réparation dans le cadre d'une action de groupe. L'un des intérêts de l'action de groupe est, en effet, de rééquilibrer le rapport de force entre les parties.

L'action de groupe permet à un demandeur d'agir en justice, non pas pour son propre compte, mais pour défendre les intérêts d'un groupe qui rassemble au moins deux cas individuels placés dans une situation similaire, subissant un dommage ayant pour cause commune un manquement de même nature aux obligations légales ou contractuelles d'un même défendeur.

Elle a été introduite en droit français par la loi du 17 mars 2014 relative à la consommation, dite « *loi Hamon* ». D'abord limitée au droit de la consommation, elle a été étendue en 2016 et en 2018 à plusieurs domaines, dont la protection des données personnelles (article 37 de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés).

Initialement cantonnée à une action en cessation d'un manquement, la loi du 20 juin 2018 a étendu ce type d'action de groupe à la réparation des préjudices matériels et moraux, afin de transposer le règlement général sur la protection des données (RGPD).

● L'action de groupe en matière de protection des données personnelles souffre de plusieurs inconvénients.

En premier lieu, la qualité pour agir est réservée à des syndicats, des associations de consommateurs ou des associations ayant pour objet statutaire la protection de la vie privée ou la protection des données à caractère personnel.

Autrement dit, le champ de la qualité pour agir ne couvre pas tout l'éventail des potentielles victimes d'un système d'IAG, dont par exemple les personnes morales et les entreprises. Les membres du groupe doivent être exclusivement composés de personnes physiques.

En deuxième lieu, l'action de groupe ne peut être dirigée qu'à l'encontre du responsable du traitement de données à caractère personnel ou d'un sous-traitant. Cette définition n'inclut pas de manière évidente le prestataire d'un système d'IAG.

Enfin, et surtout, elle ne vise que les litiges relatifs au traitement des données personnelles. Or, l'usage des systèmes d'IAG peut donner lieu à d'autres types de préjudice pour les utilisateurs.

● L'Assemblée nationale a adopté, à l'unanimité, le 8 mars 2023, une proposition de loi relative au régime juridique des actions de groupe. Ce texte tend à instaurer une **action de groupe universelle** en opérant un **triple élargissement** : élargissement de la **qualité pour agir**, élargissement de son **champ d'application** à tous les droits subjectifs et élargissement du **préjudice indemnisable**.

L'adoption définitive de ce texte – qui vient d'être inscrit à l'ordre du jour du Sénat – permettrait de mettre en œuvre la recommandation de vos rapporteurs.

À défaut, *a minima*, il conviendrait de procéder à une réécriture de l'article 37 de la loi « informatique et libertés », afin d'élargir le champ matériel de l'action de groupe « données personnelles » et la qualité pour agir.

Vos rapporteurs sont, enfin, favorables à la mise en place de juridictions spécialisées dans les litiges liés à l'IA, tant en matière pénale qu'en matière civile. Ce contentieux technique est appelé à croître. Il pourrait être traité par des magistrats spécialement formés, tant au siège qu'au parquet, et être rassemblé avec le reste du contentieux relatif au numérique.

Recommandation n° 31 : Réformer le régime juridique de l'action de groupe prévue en matière de protection des données personnelles, afin d'élargir la qualité pour agir et d'étendre son champ matériel à tous les préjudices en lien avec un système d'IAG.

Recommandation n° 32 : Spécialiser une ou plusieurs juridictions pour traiter du contentieux de l'IAG.

En conclusion, vos rapporteurs estiment souhaitable que le Parlement puisse réaliser un suivi permanent des questions relatives à l'intelligence artificielle, dont les implications sont multiples et évolutives. L'Office parlementaire d'évaluation des choix scientifiques et technologiques semble le mieux à même d'assumer cette responsabilité, d'autant qu'il a déjà travaillé sur ce sujet.

Recommandation n° 33 : Confier à l'OPECST un suivi permanent des questions relatives à l'intelligence artificielle.

*

* *

TRAVAUX DE LA COMMISSION

Lors de sa réunion du mercredi 14 février 2024, la Commission examine le rapport de la mission d'information sur les défis de l'intelligence artificielle générative en matière de protection des données personnelles et d'utilisation du contenu généré (MM. Philippe Pradal et Stéphane Rambaud, rapporteurs).

Ces débats ne font pas l'objet d'un compte rendu. Ils sont accessibles sur le portail vidéo du site de l'Assemblée nationale à l'adresse suivante :

Lien vidéo : https://assnat.fr/wJsSLg

À l'issue des débats, la publication du rapport d'information est autorisée.

LISTE DES RECOMMANDATIONS

Recommandation n° 1 : Nommer un ambassadeur à l'IAG pour suivre les discussions internationales autour de sa régulation.

Recommandation n° 2 : Organiser un partage de connaissances et d'expérience en matière de contrôle des IA, en particulier générative, au niveau européen.

Recommandation n° 3 : Adapter les contraintes applicables en matière de recours à l'IA selon qu'elle est utilisée par des États membres ou par des acteurs privés.

Recommandation n° 4 : Prendre en compte l'audience des IAG et leur caractère systémique avant de leur assigner un niveau de risque, afin de ne pas pénaliser les nouveaux acteurs et d'encourager l'expérimentation et l'innovation.

Recommandation n° 5 : Imposer le respect du RGPD et de grands principes technologiquement neutres aux services d'IAG proposés au sein de l'Union européenne.

Recommandation n° 6 : Promouvoir un droit souple et des mécanismes de conformité (*compliance*) en complément de mesures contraignantes.

Recommandation n° 7 : Laisser aux États membres de l'Union européenne une marge d'appréciation suffisante pour leur permettre de soutenir l'innovation.

Recommandation n° 8 : Promouvoir un « Airbus de l'IAG » sur la base d'une coopération intergouvernementale entre pays européens volontaires et inciter les États membres à orienter des financements des entreprises vers l'innovation en matière d'IA.

Recommandation n° 9 : Transformer la CNIL en une Haute autorité en charge de la protection des données et du contrôle de l'IAG.

Recommandation n° 10 : Veiller à l'articulation de cette nouvelle autorité avec l'ARCOM, compétente en matière de régulation des plateformes.

Recommandation n° 11 : Promouvoir le rôle d'accompagnateur du régulateur en matière d'IAG.

Recommandation n° 12 : Identifier dans le secteur public les tâches pouvant, à terme, être déléguées en tout ou partie à des systèmes d'IAG.

Recommandation n° 13 : Former les agents publics aux opportunités et aux risques inhérents à l'usage des systèmes d'IAG.

Recommandation n° 14 : Inciter les acheteurs publics à s'orienter vers des systèmes d'IAG de confiance labellisés.

Recommandation n° 15 : Prévoir une stratégie pluriannuelle en matière d'IAG, comprenant un volet consacré à son usage dans les administrations.

Recommandation n° 16 : Internaliser des compétences en matière de développement d'applications d'IAG à partir de modèles d'IAG.

Recommandation n° 17 : Prévoir une instance permanente de dialogue entre les administrations sur l'usage des IAG.

Recommandation n° 18 : Mettre à la disposition du public, à l'horizon 2027, un agent conversationnel fiable reposant sur l'IAG pour renforcer l'accès au droit et à l'information administrative.

Recommandation n° 19 : Promouvoir l'étiquetage des contenus produits par IAG en matière de propagande électorale.

Recommandation n° 20 : Étendre l'obligation d'information en cas d'interaction avec un système d'IAG aux agents conversationnels utilisés dans le cadre des campagnes électorales.

Recommandation n° 21 : Mettre en place le cadre juridique adéquat pour permettre des expérimentations de systèmes d'IAG dans le domaine régalien.

Recommandation n° 22 : Soutenir financièrement les chercheurs français dans le domaine de l'IAG, au moyen de bourses et de la mise à disposition de capacités de calcul.

Recommandation n° 23 : Modifier l'article 132-79 du code pénal pour étendre la circonstance aggravante générale d'usage de moyens de cryptologie à l'usage d'un contenu généré par une IAG.

Recommandation n° 24 : Pénaliser les hypertrucages (« *deepfake* ») réalisés sans le consentement de la personne représentée.

Recommandation n° 25 : Engager une réflexion sur le délit de contrefaçon pour pénaliser la pratique consistant à masquer, à l'aide d'une IAG, la reproduction de contenus préexistants.

Recommandation n° 26 : Prévoir en matière de faux que constitue une altération frauduleuse de la vérité le fait de se prétendre auteur de l'écrit ou du support alors que celui-ci a été produit substantiellement par une IAG.

Recommandation n° 27 : Adapter l'article 323-1 du code pénal pour sanctionner l'empoisonnement des données d'entraînement d'une IAG ou la modification clandestine de son algorithme.

Recommandation n° 28 : Développer les capacités des pouvoirs publics à identifier les contenus générés par l'intelligence artificielle et l'usage détourné de l'IAG.

Recommandation n° 29 : Adapter le régime de responsabilité des IAG à leurs spécificités, notamment avec un allègement de la charge de la preuve pour limiter l'asymétrie entre utilisateurs et fournisseurs.

Recommandation n° 30 : Réfléchir à la responsabilité des fournisseurs de service s'appuyant sur une IAG qu'ils n'ont pas conçue eux-mêmes.

Recommandation n° 31 : Réformer le régime juridique de l'action de groupe prévue en matière de protection des données personnelles, afin d'élargir la qualité pour agir et d'étendre son champ matériel à tous les préjudices en lien avec un système d'IAG.

Recommandation n° 32 : Spécialiser une ou plusieurs juridictions pour traiter du contentieux de l'IAG.

Recommandation n° 33 : Confier à l'OPECST un suivi permanent des questions relatives à l'intelligence artificielle.

LISTE DES PERSONNES ENTENDUES

Universitaires

— M. Hugues Bersini, professeur d'informatique à l'Université libre de Bruxelles

— M. Alain Goudey, chercheur et professeur de marketing, spécialiste des technologies disruptives

— M. Patrick Pailloux, conseiller d'État, ancien directeur technique de la direction générale de la sécurité extérieure (DGSE)

— M. Alexandre Lallet, conseiller d'État, co-rapporteur de l'étude sur l'intelligence artificielle dans le secteur public

— M. Vincent Lorphelin, coprésident de l'Institut de l'Iconomie et fondateur de Controv3rse, groupes de réflexion et d'études sur l'économie numérique

— Mme Alexandra Bensamoun, professeur de droit à l'Université Paris-Saclay/Évry, coordinatrice du manuel « Droit de l'intelligence artificielle »

— Mme Asma Mhalla, spécialiste des enjeux politiques et géopolitiques de la Tech, membre du Laboratoire d'anthropologie politique (LAP) de l'EHESS/CNRS, enseignante à Columbia GC, Sciences Po et l'École Polytechnique

— M. Charles Bouveyron, directeur, professeur des universités en mathématiques appliquées à l'Institut 3IA Côte d'Azur

— Mme Chloé Clavel, professeure associée en *affective computing* à Télécom ParisTech

— Mme Laure Soulier, maîtresse de conférences à Sorbonne Université au sein de l'équipe « *Machine Learning and Information Access* » de l'Institut des systèmes intelligents et de robotique

Autorités administratives indépendantes

- **Autorité de régulation des communications électroniques (ARCEP)**

—Mme Laure de la Raudière, présidente

- **Autorité de régulation de la communication audiovisuelle et numérique (ARCOM)**

— M. Guillaume Blanchot, directeur général

— Mme Pauline Combredet-Blassel, directrice générale adjointe

- **Commission nationale de contrôle des techniques de renseignement (CNCTR)**

  — M. Serge Lasvignes, président

  — M. Samuel Manivel, conseiller auprès du président

  — M. Guillaume Brosse, conseiller technique

- **Commission nationale de l'informatique et des libertés (CNIL)**

  — Mme Marie-Laure Denis, présidente

  — M. Thomas Dautieu, directeur de l'accompagnement juridique

  — Mme Chirine Berrichi, conseillère pour les questions parlementaires

  — M. Félicien Vallet, chef du service de l'intelligence artificielle

Administrations

- **M. Guillaume Avrin, coordinateur national pour l'intelligence artificielle**

- **Direction générale de la gendarmerie nationale**

  — Général Patrick Perrot, chef du service de la transformation

  — Mme Ysens de France, chargée de mission IA au sein du service de la transformation

- **Direction des affaires criminelles et des grâces (DACG)**

  — Mme Sophie Macquart-Moulin, adjointe au directeur des affaires criminelles et des grâces

  — Mme Élodie Buguel, cheffe de la mission transition numérique

- **Direction interministérielle du numérique (DINUM)**

  — Mme Stéphanie Schaer, directrice

  — M. Ulrich Tan, chef du pôle Datalab

- **Pôle d'expertise et de régulation numérique (PERéN)**

  — M. Nicolas Deffieux, directeur

  — M. Philéas Samir, *data scientist*

- **Agence nationale de la sécurité des systèmes d'information (ANSSI)**

  — M. Vincent Strubel, directeur général

Entreprises

- **MEDEF**

  — Mme Juliette Rouilloux-Sicre, présidente du comité régulation du numérique

  — Mme Clémentine Furigo, sherpa du comité régulation du numérique

  — Mme Fadoua Qachri, chargée de mission sénior en affaires publiques

- **France digitale**

  — Mme Marianne Tordeux Bitker, directrice des affaires publiques

  — Mme Agata Hidalgo, responsable des affaires européennes

  — M. Christophe Tricot, co-fondateur et président de La Forge, incubateur spécialisé dans l'intelligence artificielle

  — M. Guillaume Navarre, cofondateur et responsable marketing de la startup Golem.ai

- **Numeum**

  — M. Michel Combot, délégué général

  — Mme Marine Gossa, déléguée aux affaires publiques

- **Meta**

  — M. Martin Signoux, responsable des affaires publiques

  — Mme Béatrice Oeuvrard, responsable politiques publiques

- **TikTok**

  — M. Éric Garandeau, directeur des relations Institutionnelles et affaires publiques France

  — Mme Claire De Panafieu, *Public Policy Manager* France

- **X (*Twitter*)**

  — Mme Claire Dilé, directrice des affaires publiques

- **Microsoft France**

  — M. Anton Carniaux, directeur des affaires publiques et juridiques

— M. Philippe Limantour, directeur technologie et cybersécurité

- **Mistral AI**

— M. Arthur Mensch, cofondateur et président

Autres

- **Conseil national du numérique (CNN)**

— M. Adrien Basdevant, avocat et membre du Conseil

— Mme Justine Cassell, membre du Conseil

- **Association française de normalisation (AFNOR)**

— Mme Julie Latawiec, responsable développement en charge notamment de l'intelligence artificielle

— M. Thierry Geoffroy, responsable relations institutionnelles

CONTRIBUTIONS ÉCRITES

— M. Marius Bertolucci, maître de conférences en sciences de gestion

— Groupe Vivendi

— Oracle

— RELX

— Amazon Web Service

— Criteo

— Alliance française des industries du numérique (AFNUM)

— Impact AI

DÉPLACEMENT À BRUXELLES

- **Comité européen de protection des données**

  — M. Leonardo Cervera-Navas, Secrétaire Général ;

  — M. Olivier Matter, chef de la coopération internationale au sein de l'Unité Politique et Consultation ;

  — Mme Michèle Dubrocard, juriste au sein de l'Unité Politique et Consultation ;

  — M. Sebastiao De Barros Valle, juriste au sein de l'Unité Politique et Consultation ;

  — M. Mario Guglielmetti, juriste au sein de l'Unité Politique et Consultation.

- **Représentation permanente de la France auprès de l'Union européenne**

  — M. Benoît Blary, conseiller télécommunications, numérique et postes

  — M. Alain Dijan, conseiller affaires intérieures

Commission européenne

- **Cabinet du commissaire au marché intérieur**

  — M. Maurits-Jan Prinz, conseiller

- **Direction générale de la justice et des consommateurs**

  — M. Olivier Micol, unité protection des données

  — M. Srd Kisevic, unité démocratie, citoyenneté de l'Union et libre circulation

  — Mme Louisa Klingvall, unité politique des droits fondamentaux

- **Direction générale des réseaux de communication, du contenu et des technologies (CNECT)**

  — Mme Lucilla Sioli, directrice chargée de l'intelligence artificielle et de l'industrie numérique

  — M. Thibaud Kleiner, directeur chargé de la stratégie et de la diffusion des politiques