ABONNÉS

RÉGLEMENTATION

# IA générative et données personnelles: nouvelles pratiques, nouveaux enjeux

Publié le 13 février 2024 à 9h00
LPA-CGR | Temps de lecture 5 minutes

Les IA génératives créent du contenu sur demande, suscitant des questions sur leur fonctionnement, leurs limites et les implications juridiques, éthiques et techniques. Les experts de LPA-CGR examinent le cadre réglementaire en cours de déploiement pour ces technologies émergentes.

Eleonora Sorribes, avocat associé, et Maxime Ramos-Guerrero, avocat, LPA-CGR

Les intelligences artificielles génératives (dites IA génératives) sont des systèmes aptes à créer du texte, des images ou tout autre contenu sur instruction d'un utilisateur. Il en existe de plus en plus que ce soit dans le domaine du texte et de la conversation, via les grands modèles de langage *(Large Language Models* ou LLMs), tels que GPT-3, GPT-4, Bloom ou Megatron NLG et les agents conversationnels ou *« chatbots »* dérivés (ChatGPT ou Bard), de l'imagerie (Dall-E, Midjourney, Stable Diffusion, etc.) mais également de la parole (Vall-E). **La compréhension de leur fonctionnement, leurs limites, ainsi que les enjeux juridiques, éthiques et techniques liés à leur développement et leur usage restent encore largement en suspens.**

Ces IA sont en effet susceptibles de se tromper, et de transmettre des informations fausses, tout en exploitant, sans information préalable ni recueil de consentement éclairé, les données personnelles d'individus identifiés. À la question des dangers soulevés par les IA génératives, ChatGPT répond lui-même : *« L'IA générative présente des risques significatifs pour la protection des données personnelles. Ces modèles autonomes peuvent générer du contenu textuel, visuel ou audio, utilisant des informations personnelles réelles ou fictives. Cela pourrait être exploité pour créer des contenus trompeurs ou diffamatoires, compromettant ainsi la confidentialité des individus. De plus, la capacité des modèles à générer des informations personnelles fictives soulève des préoccupations liées à la sécurité et à la vie privée. Les technologies de l'IA générative pourraient être utilisées pour produire des données d'apparence authentique, mais fausses, aggravant ainsi les risques de fraude et d'usurpation d'identité. Les biais existants dans les ensembles de données utilisés pour l'entraînement peuvent également se traduire par des générations de contenu discriminatoire, exacerbant les problèmes de préjugés dans la société. En résumé, les dangers de l'IA générative sur les données personnelles résident dans la création de contenus manipulateurs, la falsification d'identités et le renforcement de biais discriminatoires, soulignant la nécessité d'une réglementation et de protections appropriées. »* (1)

On pourrait saluer ce qui s'apparenterait à une forme de lucidité artificielle tant la réponse créée par ChatGPT semble crédible et proche de la réalité, malgré un discours quelque peu caricatural et des arguments stéréotypés. On remarquera encore que ChatGPT n'hésite pas à prêcher contre sa propre paroisse semblant appeler lui-même au développement accéléré *« d'une réglementation et de protections appropriées ».*

En toute hypothèse, l' « analyse » de l'IA elle-même, interrogée par les auteurs des présentes lignes, semble pertinente, les questions relatives au traitement et au devenir de nos données personnelles étant citées de façon récurrente lorsqu'il s'agit d'évoquer les dangers soulevés par le développement de telles technologies. La protection des données personnelles, droit fondamental (2), revêt plusieurs aspects, en particulier leur traitement loyal, à des fins déterminées sur la base du consentement éclairé des utilisateurs. Il semble important d'identifier les risques que soulèvent les IA quant à la protection des données personnelles (I) avant d'envisager comment la réglementation tente de s'adapter à ces nouveaux enjeux (II).

## I- Les risques des IA génératives

La conception et le fonctionnement de l'IA générative posent des questions nouvelles, en particulier concernant la protection des données à caractère personnel, d'une part, transmises par les utilisateurs lorsqu'ils utilisent ces outils, mais également, d'autre part, des données personnelles contenues dans le corpus de données utilisé par la machine elle-même. Ces données peuvent être collectées, traitées ou réutilisées par des algorithmes d'apprentissage automatique sans que la personne concernée n'ait nécessairement donné son consentement ou puisse exercer ses droits au regard du RGPD.

Afin d'appréhender ces risques, il convient d'apprécier le fonctionnement des IA génératives et notamment le traitement de la masse des données qu'elles opèrent. En effet, les applications d'IA génératives modélisent des travaux à l'aide de réseaux de « neurones », lesquels s'inspirent des neurones du cerveau humain pour apprendre des schémas et des fonctionnalités à partir de données existantes. Ces modèles déterminent leurs processus d'apprentissage et « génèrent » de nouveaux contenus en se référant aux données sur lesquelles ils ont été formés.

Le modèle de fondation aussi désigné modèle de langage étendu ou *« Large Language Models »* (LLM), dont les premiers exemples sont GPT-3, BERT ou DALL-E 2, sont des modèles entraînés sur une grande quantité de données non étiquetées (3) **(généralement par apprentissage autosupervisé)** (4). Aussi, leur fonctionnement implique la collecte *(scraping)* d'une quantité massive de données qui peuvent être recueillies sur des sources disponibles publiquement, mais pouvant aussi inclure des données personnelles ou des données sensibles pour lesquelles aucun consentement n'a été donné.

À titre d'illustration, en 2017, un développeur a créé un *chatbot* nommé « AI. type » ayant pour fonction de faciliter la saisie sur les claviers virtuels pour smartphones (5). Cette technologie utilisait une base de données contenant plus de 31 millions d'enregistrements, dont des extraits de conversations privées, qui incluaient des informations personnelles telles que des numéros de téléphone, des adresses, des noms et des extraits de messages privés. Or, l'entreprise n'ayant pas correctement sécurisé cette base de données, n'importe qui a pu l'exploiter sans authentification. En conséquence, des données sensibles de millions d'utilisateurs ont été exposées publiquement. Ces inquiétudes légitimes peuvent encore concerner les utilisateurs de ces technologies eux-mêmes et mènent certaines entités à prendre des décisions drastiques.

Ainsi, depuis fin mars 2023, la mairie de Montpellier a strictement interdit l'utilisation de ChatGPT dans son administration jusqu'à nouvel ordre. La ville craint effectivement que l'IA n'ait accès à toute la base de données de la collectivité et de ses administrés, du contenu des boîtes emails aux documents stockés sur les serveurs en passant par les dossiers liés à des appels d'offres. Aussi, par mesure de sécurité, la collectivité a étendu l'interdiction à ses partenaires pour éviter toute fuite de données. Plus clairement, un élu explique que *« le sujet n'est pas l'accès au service, c'est l'utilisation des données. On ne sait pas où elles vont. Aujourd'hui, on n'a aucune visibilité sur un produit susceptible de bafouer un certain nombre de lois ».* (6)

L'Italie, elle aussi, avait bloqué temporairement l'utilisation de ChatGPT pour l'ensemble de ses citoyens, estimant que ceux-ci ne pouvaient exercer leurs droits au titre de la réglementation applicable aux données personnelles auprès d'OpenIA. Cette dernière a donné des garanties quant à l'exercice effectif de ces droits et le service a finalement été de nouveau autorisé en avril 2023.

Comme toute technologie basée sur le *Machine Learning*, les modèles linguistiques GPT-3.5 et GPT-4 d'OpenAI ont été formés sur des milliards d'échantillons de textes existants et améliorés grâce à un processus de *fine tuning* (7). Aussi, OpenAI enregistre les conversations ChatGPT et les *« prompts »* en vue d'une analyse ultérieure. Outre les *« prompts »* et les conversations, OpenAI enregistre également d'autres données telles que les détails du compte, le nom et l'adresse électronique, ainsi que l'emplacement approximatif et l'adresse IP de l'utilisateur, les informations de paiement et les informations relatives à l'appareil utilisé.

La plupart des sites web collectent ces données à des fins d'analyse, ce qui n'est donc pas propre à ChatGPT. Cependant, cela signifie qu'OpenAI serait susceptible de transmettre les conversations ChatGPT et d'autres aux autorités américaines conformément au FISA (Foreign Intelligence Surveillance Act).

Qui plus est, OpenAI n'offre aucune procédure permettant aux individus de vérifier si l'entreprise stocke leurs informations personnelles ou de demander leur suppression, alors qu'il s'agit là de droits garantis par le RGPD (8). Si, dans les mentions légales sur ses usages, l'éditeur indique la nature des données traitées, il reste imprécis sur la finalité de son service et sur la base juridique applicable. On comprend ainsi mieux pourquoi ChatGPT intéresse autant les autorités de protection de données en Europe.

## II- Impérieuse adaptation législative

Face à ces inquiétudes, la Cnil (B) et la Commission européenne (C) se sont emparées du sujet en développant de nouvelles actions et réglementations. Ceci est d'autant plus nécessaire que des actions visant des IA génératives ont déjà été engagées en France (A).

### A- Les actions menées en France contre des exploitants d'IA génératives

En France, ChatGPT est visé par plusieurs **plaintes déposées auprès de la Commission nationale de l'informatique et des libertés (Cnil).** Elles portent sur la collecte de données sans consentement ainsi que la production d'informations erronées. Parmi elles, l'une est engagée par l'avocate Zoé Vilain, présidente de l'association de sensibilisation aux enjeux du numérique Janus International, dans laquelle elle explique avoir créé un compte sur le site d'OpenAI, l'entreprise californienne à l'origine du logiciel, afin d'utiliser ChatGPT et avoir constaté l'absence de *« conditions générales d'utilisation »* à accepter ou d'*« une quelconque politique de confidentialité ».* Elle demande donc à la Cnil de l'aider à exercer son droit d'accès à ses informations personnelles collectées par OpenAI, **après une tentative infructueuse effectuée auprès de l'entreprise.**

La seconde plainte a été déposée par David Libeau, un développeur investi dans la protection des données personnelles s'étant rendu compte en interrogeant ChatGPT à son sujet, que l'IA lui transmettait des informations erronées. En effet, *« l'algorithme a commencé à affabuler et à m'attribuer la création de sites web ou l'organisation de manifestations en ligne, ce qui est totalement faux »,* écrit-il.

Une autre plainte a été déposée auprès de la Cnil par le député Éric Bothorel contre OpenAI pour les mêmes raisons, mentionnant avoir repéré de fausses informations personnelles le concernant ou diffusées sans son consentement en interrogeant ChatGPT sur son profil et ce en violation des articles 5 et 6 du RGPD.

### B- Le plan d'action et les lignes directrices de la Cnil

Devant les récentes actualités sur l'intelligence artificielle, la Cnil a publié le 16 mai 2023 un plan d'action pour un déploiement de systèmes d'IA respectueux de la vie privée des individus. Cette régulation se structure autour de quatre objectifs :

- appréhender le fonctionnement des systèmes d'IA et leurs impacts pour les personnes ;
- permettre et encadrer le développement d'IA respectueuses des données personnelles ;
- fédérer et accompagner les acteurs innovants de l'écosystème IA en France et en Europe ;
- auditer et contrôler les systèmes d'IA et protéger les personnes.

Marie-Laure Denis, présidente de la Cnil, auditionnée le 11 septembre 2023 devant la Commission des lois de l'Assemblée nationale (9), a déclaré, conformément au premier objectif du plan d'action qu'*« on ne peut bien réguler un objet que si on le comprend ».* Selon elle, il faut se *« concentrer sur l'émergence d'un cadre légal et opérationnel permettant à l'innovation de se déployer dans le respect des libertés publiques et individuelles ».* Elle appelle ainsi au développement d'un « outillage » pour auditer les systèmes d'IA afin de s'assurer du respect de la vie privée, les investigations devant se dérouler à trois niveaux, au niveau de l'application, de la base de données et du modèle sous-jacent. Et afin de s'assurer que les utilisateurs soient informés de la façon dont leurs données sont traitées pour pouvoir exercer opérationnellement leur droit.

En outre, en ce qui concerne le droit à la rectification, techniquement impossible à mettre en œuvre, la présidente de la Cnil suggère de combler cette incapacité par des modules permettant de corriger les erreurs ou inexactitudes du modèle.

Dans le même sens, le 11 octobre 2023, la Cnil, pour clarifier les règles applicables, a élaboré une première série de fiches pratiques (10) qui visent à assurer la prise en compte du respect des données personnelles dès la phase de développement des IA (lors de la conception du système, la constitution de la base de données et l'apprentissage). La Cnil a confirmé, à cette occasion, la compatibilité des recherches et développements en IA avec le RGPD *« à condition de ne pas franchir certaines lignes rouges et de respecter certaines conditions ».* Parmi ces garde-fous, la Cnil évoque le principe de finalité (en vertu duquel les données personnelles doivent être utilisées dans un objectif précis uniquement), le principe de minimisation (imposant une sélection des données pour optimiser l'entraînement de l'algorithme), le principe de transparence et le principe de conservation pour une durée limitée des données.

En parallèle, l'État s'entoure d'experts pour forger sa doctrine, Élisabeth Borne a réuni le 19 septembre 2023 le premier comité interministériel sur l'intelligence artificielle générative, composé d'une quinzaine de spécialistes afin d'aider à forger la position de l'État français sur les différentes problématiques que soulève cette technologie.

### C- Règlement européen sur l'IA et RGPD

En avril 2021, la Commission européenne a proposé l'IA Act, le premier règlement mondial encadrant les IA. Il y est mentionné que les systèmes d'IA qui peuvent être utilisés dans différents cas de figure sont analysés et classifiés d'après les risques que leur utilisation peut poser à leurs utilisateurs. En fonction de ces différents niveaux de risque, il faudra adapter une législation adéquate :

- les systèmes d'IA interdits car considérés comme particulièrement dangereux pour les droits fondamentaux, tels que les systèmes de surveillance sociale utilisés par les autorités publiques pour surveiller ou évaluer le comportement des personnes ;
- les systèmes d'IA à haut risque tels que les dispositifs médicaux ou les technologies de transport autonomes, qui nécessitent des exigences strictes de transparence, de traçabilité, d'évaluation de la conformité, de documentation technique, de surveillance et de contrôle ;
- les systèmes d'IA à risque limité soumis à des exigences moins strictes que les systèmes à haut risque, mais qui doivent néanmoins respecter certaines normes.

Le règlement européen prévoit également des mécanismes de certification par des organismes tiers pour les systèmes d'IA à haut risque afin de garantir leur conformité aux règles établies et l'établissement de normes auxquelles les fournisseurs de systèmes d'IA devront se conformer pour se déployer sur le marché européen. Le 14 juin 2023, les députés ont adopté le projet d'acte. Un accord provisoire a été conclu le 8 décembre 2023 entre la présidence du Conseil et les négociateurs du Parlement européen pour faire avancer le projet.

Cet accord fait un cas particulier des IA génératives et d'une manière générale de tout IA interagissant directement avec les utilisateurs. Le règlement obligera les entreprises et les organisations publiques à indiquer à l'utilisateur qu'il est en relation avec une machine et tout contenu généré par une IA devra être identifié comme tel.

En outre, le non-respect des règles pourra entraîner des amendes allant de 35 M€ ou 7 % du chiffre d'affaires mondial pour les violations des applications d'IA interdites, à 15 M€ ou 3 % pour les violations des obligations prévues par la loi sur l'IA et à 7,5 M€ ou 1,5 % pour la fourniture d'informations inexactes. Toutefois, l'accord provisoire prévoit des plafonds d'amendes administratives plus proportionnés pour les PME et les jeunes entreprises en cas d'infraction aux dispositions de la loi sur l'IA.

Néanmoins, certains estiment que ce règlement est déjà obsolète. Plutôt que de proposer des dispositions juridiques spécifiques sur l'« explicabilité » des systèmes d'IA, le projet de règlement se contente de mesures de transparence et de traçabilité. Hormis les critiques liées à la transparence, les mesures spécifiques visant à garantir la résilience des systèmes d'IA face aux cyberattaques sont inexistantes et l'absence de consensus sur l'interdiction de l'utilisation de l'IA à des fins répressives est également reprochée (11).

## Conclusion

Du point de vue de l'Association française des correspondants à la protection des données à caractère personnel (AFCDP), *« ChatGPT et ses usages restent à encadrer, plutôt qu'à interdire ».* Patrick Blum, délégué général de l'AFCDP, déclare que *« sans attendre le futur règlement européen sur l'intelligence artificielle, qui risque d'ailleurs d'être dépassé par les outils d'IA conversationnelle, il convient certainement d'imposer aux éditeurs concernés la plus grande transparence sur les sources exactes de données de l'IA générative, en particulier celles contenant des données à caractère personnel, et le respect des obligations de base prévues par le RGPD, en particulier le droit d'accès, le droit de rectification et le droit d'opposition ».* (12)

S'il est nécessaire et rassurant à certains égards que la Cnil, la Commission européenne, le G7 des autorités de protection des données (13), et, plus globalement la communauté internationale (14) s'emparent du sujet, il semble surtout primordial que les acteurs économiques, qui voient dans les IA génératives de nouvelles opportunités de croissance, adoptent systématiquement une approche protectrice et sécuritaire des données personnelles afin de garantir un bon équilibre entre innovation et préservation des droits.

(1) Texte généré par ChatGPT

(2) Consacré à l'article 8 de la Charte des droits fondamentaux

(3) L'étiquetage des données consiste à identifier des données brutes (image, texte, vidéo, son).

(4) Les modèles peuvent être supervisés (lorsque les données d'entrées sont étiquetées) ou non supervisés (lorsque les données d'entrées ne sont pas étiquetées). Dans le deuxième cas, les algorithmes découvrent des modèles cachés ou des groupements de données sans nécessiter d'intervention humaine. Sa capacité à découvrir les similitudes et les différences d'informations permet l'analyse d'exploration des données. Les données étiquetées sont moins chronophages et coûteuses à acquérir et stocker mais sont plus limitées dans leur utilité.

(5) l'entreprise de recherche en sécurité Kromtech Security

(6) Manu Raynaud, adjoint au maire chargé des questions numériques

(7) Le *fine-tuning* consiste à entraîner à nouveau le modèle sur un petit ensemble de données (comme les conversations des utilisateurs).

(8) Le droit d'accès est protégé par l'article 15 du RGPD.

(9) Audition devant la Commission des lois de l'Assemblée nationale, Mission d'information de l'Assemblée nationale, Intelligence artificielle et protection des données personnelles, propos liminaire de Marie-Laure Denis, présidente de la Cnil, 11 septembre 2023

(10) Fiches pratiques sur l'IA de la Cnil publiées le 11 octobre 2023

(11) P. Alix, avocat DPO, post LinkedIn du 12 décembre 2023

(12) Interview de Patrick Blum, délégué général de l'Association française des correspondants à la protection des données à caractère personnel (AFCDP) du 6 juin 2023 par *Les Echos* concernant la protection des données personnelles dans le tourbillon, algorithmique de l'IA générative

(13) Déclaration commune du G7 du 21 juin 2023 et article « IA générative : adoption par le G7 des autorités de protection des données d'une déclaration commune », Revue *Lamy Droit de l'Immatériel,* n°205, 1er juillet 2023

(14) Le premier sommet sur la sécurité de l'IA aura lieu les 1er et 2 novembre à Bletchley Park

© 2024 Option Finance Tous droits réservés