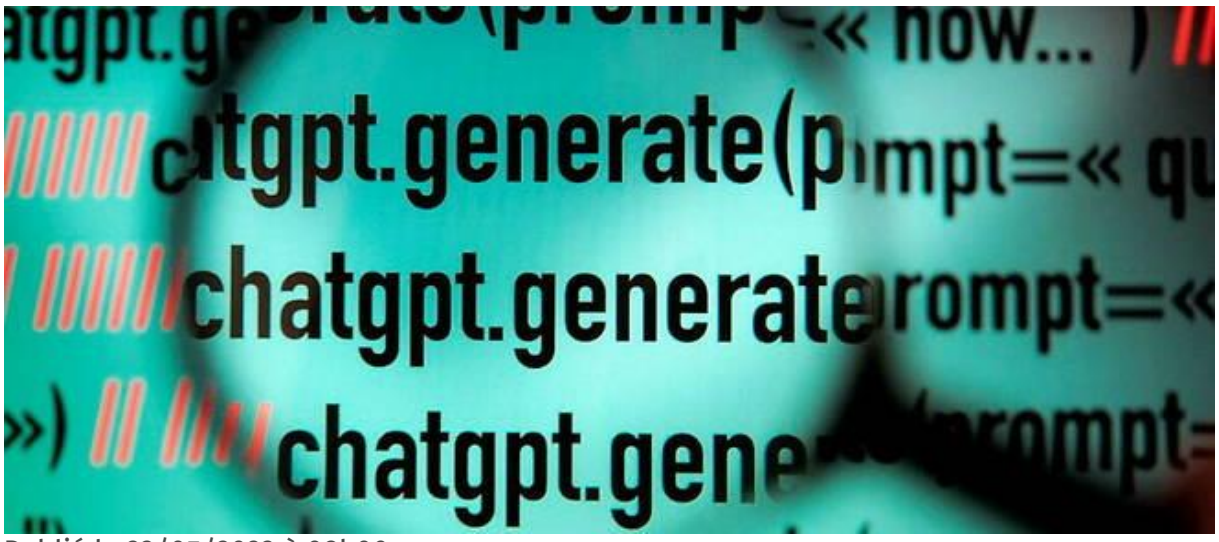
ChatGPT : quels dangers pour nos vies privées ?

Les agents conversationnels lancent de nouveaux défis à la protection et la sécurité des données personnelles. Décryptage avec Constantin Pavléas, avocat spécialisé en droit des technologies.

Par Laurence Neuer



Publié le 23/05/2023 à 08h00

À peine lancé, déjà accusé ! Le robot conversationnel (chatbot) ChatGPT est dans le collimateur des autorités de protection des données. Ce qui coince ? Entre autres griefs, le service ne garantit pas l'exactitude des informations qu'il délivre.

Plusieurs plaintes ont été déposées auprès de la Commission nationale de l'informatique et des libertés (Cnil) contre cette intelligence artificielle (IA) générative capable de converser en langage naturel et de générer des réponses semblables à celles d'un être humain.

David Libeau, développeur investi dans la protection des données personnelles, lui reproche d'avoir sorti des informations fausses sur lui. L'avocate Zoé Vilain, présidente de l'association de sensibilisation aux enjeux du numérique Janus International, interroge, de son côté, la transparence de l'outil.

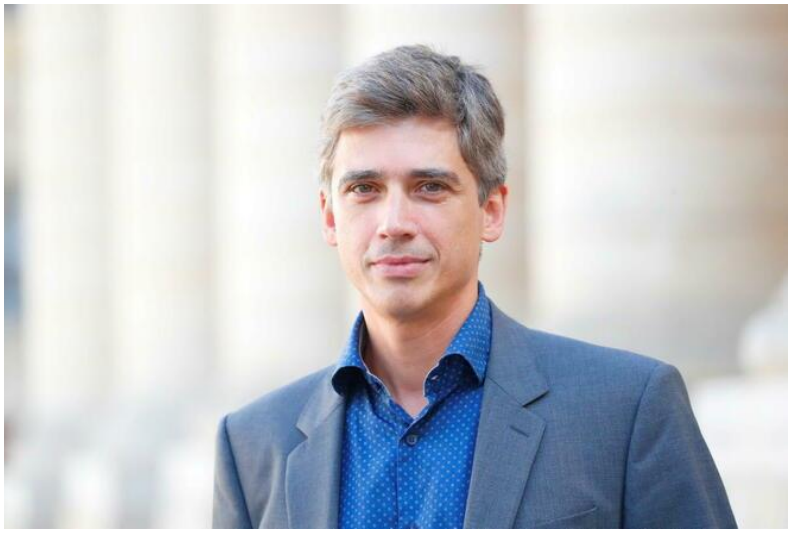
« Je demande à la Cnil d'enjoindre à OpenAI [le développeur de ChatGPT, NDLR] de me permettre l'accès à ces données, conformément à l'article 15 du RGPD [Règlement général sur la protection des données, NDLR] et d'indiquer si mes requêtes ont été communiquées à d'autres utilisateurs ou à des partenaires commerciaux et si je fais l'objet d'un traitement automatisé. En dehors de mon activité sur ChatGPT, j'aimerais savoir si OpenAI a collecté des données personnelles me concernant sur des applications tierces ou sur Internet, et notamment sur le blog de mon cabinet. En clair, je souhaite comprendre comment fonctionne l'outil », détaille-t-elle.

« Appréhender le fonctionnement des systèmes d'IA génératives et leurs impacts pour les personnes » est justement l'une des priorités de la Cnil. Partant du constat que « la protection des données personnelles est un enjeu majeur pour la conception et l'utilisation de ce type d'outil », l'autorité indépendante « souhaiteinstaurer des règles claires, protectrices des données personnelles des citoyens européens afin de contribuer au développement de systèmes d'IA respectueux de la vie privée », précise-t-elle dans son plan d'action publié le16 mai 2023.

Quels risques ChatGPT fait-il peser sur nos données ? Que prévoient ses conditions d'utilisation ? Le futur règlement européen sur l'IA est-il suffisamment armé pour protéger les utilisateurs face aux dérives de l'outil ? Les réponses de Constantin Pavléas, avocat spécialisé en droit du numérique.

Le Point : Chat GPT est accusé de tous les maux : violer les règles européennes sur les données personnelles des utilisateurs, inventer des coupables, notamment, des professeurs ayant faussement harcelé des étudiantes, utiliser des œuvres protégées pour entraîner l'algorithme… Qu'est-ce que cela vous inspire ?

Constantin Pavléas : Les IA génératives, qui permettent à une machine de comprendre et de générer du texte humain ou des images, créent de formidables opportunités pour l'humanité, avec toutefois de nombreux risques. Par exemple, le détournement de ces outils par des personnes ou des organisations malveillantes ou des États non démocratiques, pour tromper, désinformer, manipuler l'opinion, pour faciliter les actes de cybercriminalité



On peut également craindre que les IA génératives, accessibles au grand public, amplifient les effets des réseaux sociaux sur nos comportements par la propagation de biais et discriminations à grande échelle ou l'enfermement dans des bulles cognitives. La protection de nos données personnelles est aussi un enjeu important : il faut savoir que ces outils ont été entraînés en utilisant de manière indifférenciée toutes les données disponibles sur le Web !

C'est pour cette raison que l'autorité de protection des données italienne a suspendu pendant un mois l'utilisation du chatbot, ce qui est une première !

L'autorité de contrôle italienne, la GPDP [Garante per la Protezione dei dati personali, NDLR] a pris une mesure radicale pour suspendre ChatGPT en Italie. Le 31 mars 2023, son président a en effet ordonné de limiter temporairement les traitements de données personnelles opérés par OpenAI, l'éditeur de l'outil.

La décision de l'autorité italienne concernait les personnes établies en Italie et était fondée sur une série de signalements constituant des violations du RGPD : absence d'information fournie aux utilisateurs et aux personnes concernées dont les données sont collectées par OpenAI, absence de base juridique justifiant la collecte et le traitement massifs de données à caractère personnel afin de « former » les algorithmes sur lesquels la plateforme s'appuie, traitement de données inexactes par l'IA, absence de tout mécanisme de vérification de l'âge des utilisateurs qui expose les enfants à recevoir des réponses inappropriées, etc. L'autorité de

contrôle italienne a, par la suite, ordonné à OpenAI de corriger ces dysfonctionnements avant le 30 avril afin de se conformer au RGPD.

Et OpenAI a obtempéré !

Oui, OpenAI a publié des mises à jour de sa politique de confidentialité et de protection de données – les *privacy policy*. L'entreprise américaine a ainsi tenu compte des critiques de l'autorité italienne en informant les utilisateurs de ChatGPT des données personnelles collectées, de la finalité des traitements, de leur possibilité de s'opposer à la collecte de leurs données personnelles, de l'interdiction faite aux mineurs de 13 ans d'utiliser ChatGPT et de la nécessité de l'accord parental pour les mineurs de 13 à 18 ans....

OpenAI a aussi mis en place des formulaires par lesquels toute personne, qu'elle utilise ou non le service, peut demander que ses données personnelles soient effacées. L'utilisateur peut aussi s'opposer à ce que ses interactions avec l'outil entraînent les algorithmes sous-jacents. C'est une avancée notable d'OpenAI, qui lui a valu la levée de la suspension.

OpenAI s'est-elle conformée à l'ensemble des exigences de l'autorité italienne ?

L'autorité italienne continue à investiguer la conformité d'OpenAI au RGPD, y compris dans le cadre du groupe de travail créé en avril par les autres autorités de contrôle de l'Union européenne (UE), au sein de l'EDPB [European Data Protection Board, Comité européen de la protection des données, NDLR].

Mais une mesure demandée par l'autorité italienne attire tout spécialement l'attention : le fait d'organiser une campagne d'information sur tous les médias grand public italiens, y compris radio, télévision, journaux et Internet, afin d'informer le public que les données personnelles sont susceptibles d'être collectées pour former les algorithmes d'OpenAI, qu'une information spécifique sur ce point a été publiée sur le site Web de l'entreprise et qu'un outil est disponible sur ce site pour permettre aux personnes concernées de demander la suppression de leurs données personnelles. Le contenu de cette campagne d'information devra au préalable être validé par l'autorité de contrôle italienne.

Du jamais-vu !

Tout à fait ! Non seulement les Italiens ont créé un électrochoc avec la mesure de suspension de ChatGPT, mais la campagne d'information demandée est sans précédent dans son ampleur, tant en termes de contenu que de médias visés. Une telle réaction me semble tout à fait justifiée au regard des enjeux suscités par l'IA générative pour la protection des données personnelles.

La Cnil française a adopté une approche différente et a dévoilé son « plan d'action » le 16 mai dernier. Que prévoit-il ?

La Cnil décline un plan d'action en quatre axes : comprendre comment fonctionnent les systèmes d'IA ; accompagner les IA respectueuses des données personnelles ; contrôler le fonctionnement des IA par des audits et protéger ainsi les personnes ; favoriser l'écosystème IA en France et en Europe. Cette approche révèle la vision sphérique et pragmatique de la Cnil pour prendre en compte les enjeux économiques et concurrentiels de l'IA tout en protégeant les personnes physiques et leurs données.

La Commission instruit actuellement plusieurs plaintes d'utilisateurs « abusés » par l'outil… Peut-on s'attendre à des sanctions ?

La Cnil a été saisie de cinq plaintes à ce jour. En cause, le droit à l'information des personnes et l'exercice de leurs droits au titre du RGPD, tels que le droit d'accès ou le droit de rectification. Mais il y a également le problème des erreurs dans les réponses concernant les personnes : c'est notamment l'objet de la plainte du député Éric Bothorel.

Interrogé sur la biographie du parlementaire, l'outil a affabulé, ce qui contrevient à l'article 5 du RGPD selon lequel les informations sur les personnes doivent être « exactes et si nécessaires tenues à jour ». En cas de manquement constaté, les textes prévoient des sanctions pécuniaires allant jusqu'à 20 millions d'euros ou 4 % du chiffre d'affaires annuel mondial de l'entreprise.

Que disent les conditions d'utilisation du service sur nos données et la possibilité d'en demander la suppression ?

Dans la dernière version mise à jour le 27 avril 2023, OpenAI met à disposition des utilisateurs une adresse e-mail pour demander la rectification de données personnelles inexactes. Cependant, OpenAI avertit qu'au vu de la complexité du fonctionnement de ses modèles, il peut être difficile pour eux de rectifier les données incorrectes.

Dans ce cas, un formulaire est mis à disposition de toute personne, qu'elle soit utilisatrice du service ou non, pour demander la suppression de ses données. Autre nouveauté mise en place en avril, il est possible d'utiliser ChatGPT en demandant que l'outil n'utilise pas l'historique des conversations pour entraîner ses algorithmes. Pour répondre aux plaintes concernant le droit d'accès, OpenAI a ajouté un e-mail de contact si l'utilisateur ne peut faire valoir ses droits via son compte en ligne.

Ces efforts de mise en conformité avec le RGPD sont encore insuffisants car l'exactitude des données personnelles traitées est un des principes essentiels présidant à tout traitement de données. Autre lacune, on notera que la notice sur la politique de confidentialité n'est disponible qu'en anglais sur le site d'OpenAI. Or, au titre du RGPD, l'information ne peut être censée valablement donnée dans une langue non comprise par l'utilisateur.

Enfin, OpenAI fait état de mesures « commercialement raisonnables » pour assurer la sécurité des données traitées. Or, le règlement institue une obligation absolue dans ce domaine. Les autorités de contrôle apprécieront… Et il subsiste d'autres zones d'ombre.

Lesquelles ?

La définition du « contenu » des informations susceptibles d'être traitées par l'IA n'est pas claire. Selon les conditions d'utilisation, celui-ci comprend l'« input » et l'« output », à savoir la requête de l'utilisateur et la réponse du robot. Mais dans la dernière version de la politique de confidentialité, le « contenu » ne comprend que l'input, la requête, dont les données peuvent être analysées par la société. Quid de la réponse ? Est-elle également analysée par la société, comme le laissent penser les conditions d'utilisation ?

Après une hésitation entre opt-in et opt-out concernant l'utilisation du « contenu » par OpenAI pour entraîner ses algorithmes, l'entreprise a

institué un opt-out. Ce sera donc à l'utilisateur de s'opposer à l'utilisation de ses données personnelles pour l'entraînement des algorithmes.
À défaut, ses données seront ainsi traitées par l'outil.

Un projet de règlement européen sur l'IA, dont un texte amendé a été adopté le 11 mai, va être soumis au vote du Parlement européen : quel est son objet ? Pourrait-il intégrer ces problématiques et apporter les garanties nécessaires pour obliger ChatGPT à se conformer au RGPD ?

En effet, le projet de règlement européen sur l'IA – *AI Act* – présenté en avril 2021 est toujours en cours de discussion au sein des institutions européennes [Conseil européen et Parlement européen, NDLR]. L'objet du texte est de réguler le développement, la commercialisation et l'usage des systèmes d'IA, en adoptant une approche basée sur les risques.

Ainsi seraient interdits, parce que comportant des risques « inacceptables », les systèmes de « notation sociale », les « techniques subliminales » visant à manipuler les citoyens, les systèmes « qui exploitent les vulnérabilités dues à l'âge, au handicap ou à la situation sociale », une assistance vocale incitant les enfants à avoir un comportement dangereux. À l'opposé, les systèmes comportant des risques minimes ne seraient pas réglementés.

Entre les deux, le texte imposerait des obligations de sécurité strictes sur les éditeurs d'IA « à haut risque », celles qui comportent un effet négatif sur la santé ou la sécurité des personnes ou les droits fondamentaux. Par souci de sécurité juridique, l'approche du législateur européen est de lister les types d'IA « à haut risque » dans une annexe III au règlement.

La question est de savoir si les IA génératives, tel ChatGpt, seront listées dans cette annexe et seront soumises *ipso facto* aux règles contraignantes de l'*AI Act*, imposant par exemple de vérifier la « qualité » des données utilisées pour entraîner les logiciels, de « minimiser les risques et les résultats discriminatoires » ou d'assurer l'exactitude des résultats.

Le 11 mai, les eurodéputés réunis en commission du marché intérieur et des libertés civiles ont amendé le texte de la Commission, en ajoutant de nouvelles catégories d'IA interdites, notamment l'interdiction de la surveillance biométrique, de la reconnaissance des émotions, des systèmes

d'IA de police prédictive. Ils ont aussi élargi le champ des IA « à haut risque » aux atteintes à l'environnement.

Les IA génératives ne sont pas listées en tant que telles comme des IA « à haut risque ». Cependant, des règles spécifiques ont été ajoutées pour obliger les fournisseurs de modèles à se conformer à des exigences de transparence supplémentaires, par exemple l'indication que le contenu a été généré par une IA.

En outre, l'entraînement, la conception et le développement du modèle doivent être réalisés de manière à assurer que le contenu généré n'enfreigne pas les lois de l'UE. Les fournisseurs d'IA générative devraient également rendre publique la liste des œuvres protégées par le droit d'auteur qui ont été utilisées pour la formation des algorithmes. Le texte doit encore être approuvé par le Parlement européen en juin. Après ce vote, le projet d'*AI Act* sera alors négocié avec le Conseil européen, avant qu'un texte définitif ne soit adopté.

Le futur *Digital Services Act* s'appliquera-t-il aux IA génératives ?

En effet, le *Digital Services Act* (DSA), qui entrera pleinement en vigueur en 2024, impose plusieurs choses aux très grandes plateformes et aux très grands moteurs de recherche, les « TGP » [des plateformes qui comptent plus de 45 millions d'utilisateurs actifs chaque mois, soit 10 % de la population de l'UE, un chiffre qui évolue avec la population européenne. Tous les six mois, le coordinateur des services numériques vérifiera le franchissement de ces seuils par les intermédiaires placés sous sa juridiction, NDLR].

ll impose notamment des obligations d'analyse et de traitement des risques systémiques, la diffusion de contenus illicites tels que les contenus pédopornographiques ou haineux ou encore la manière dont les services sont utilisés pour propager ou amplifier des contenus fallacieux ou trompeurs. Ces risques peuvent également être liés à des campagnes coordonnées de désinformation relatives à la santé publique ou la conception d'interfaces qui stimulent des addictions comportementales des individus.

L'enjeu sera de savoir si le DSA s'appliquera aux plateformes d'IA générative. La question de leur statut d'éditeur de contenu ou d'hébergeur

de contenus se pose aussi. Selon la qualification retenue, l'éditeur de l'IA pourrait être considéré comme responsable du contenu généré ou pas. Il y a là de grandes batailles juridiques en perspective.