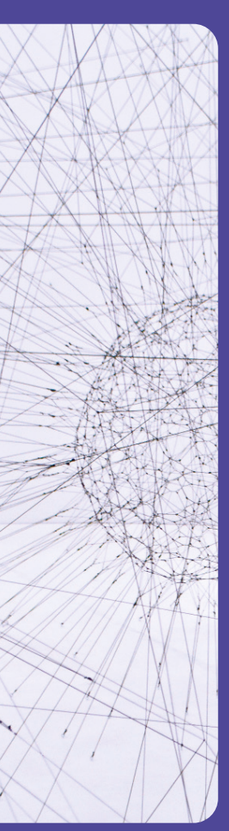
MICROPOLE

# L'IA et le RGPD : le cas de ChatGPT

# L'IA et le RGPD : le cas de ChatGPT

S'il est porteur d'une innovation certaine, l'essor impressionnant, ces derniers mois, des IA génératives comme ChatGPT soulève aussi de plus en plus de questions concernant la protection et la sécurisation des données personnelles traitées par ces outils.



Ismaël Koné, Consultant Senior Data Compliance chez Micropole

Éliott Mourier, Data Compliance & Data Privacy Manager chez Micropole

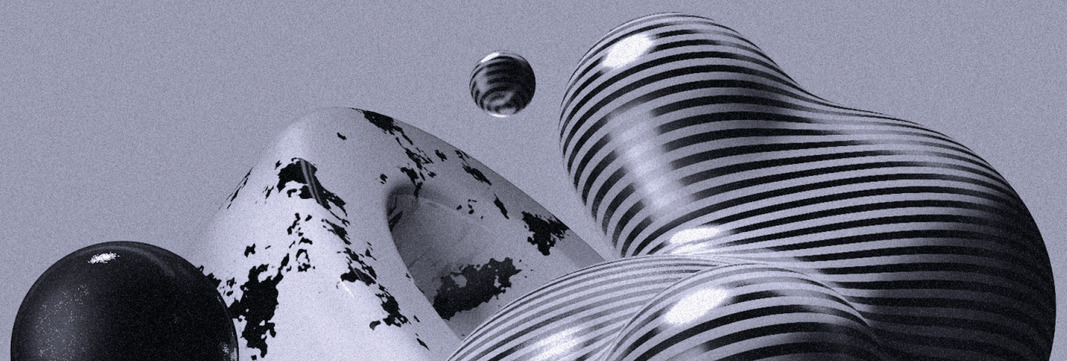
# La protection des données en question

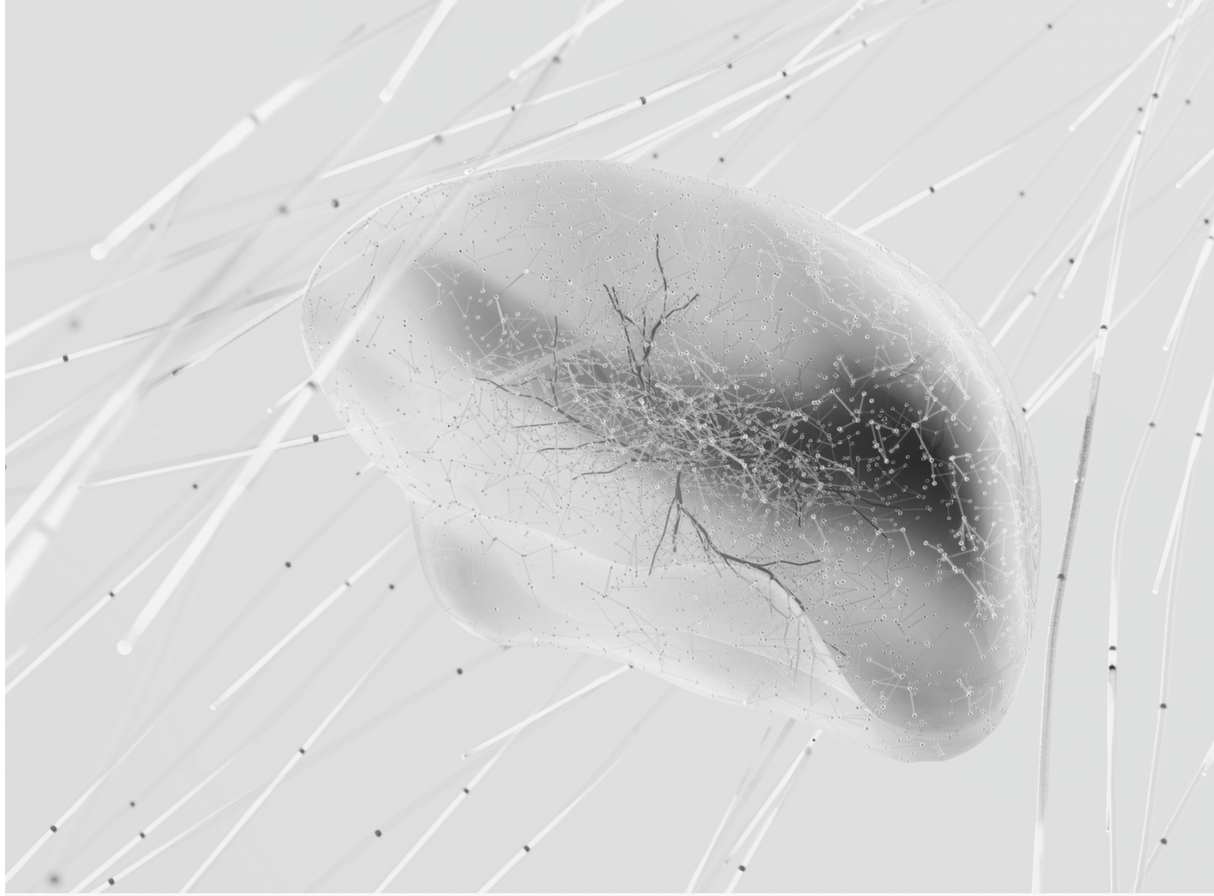
L'interdiction de ChatGPT par le régulateur national italien de la protection des données le 31 mars 2023 l'illustre bien. En effet, le régulateur italien a jugé bon de faire suspendre l'IA après avoir soulevé des inquiétudes concernant, entre autres, les récentes violations de données qu'a subies OpenAI, et quant à la base juridique de l'utilisation de données personnelles pour « entraîner » le chatbot.

OpenAI, a en effet confirmé une violation de données le 20 mars 2023 causée par un bug dans une bibliothèque open source, alors qu'une société de cybersécurité avait remarqué qu'un composant récemment introduit a été affecté par une vulnérabilité activement exploitée.

**Selon l'enquête d'OpenAI, les titres de l'historique des conversations des utilisateurs actifs et le premier message d'une conversation nouvellement créée ont été exposés lors de cette violation de données.**

Le bug a également révélé des informations relatives au paiement appartenant à 1,2 % des abonnés de ChatGPT ainsi que le nom et le prénom, l'adresse électronique, l'adresse de paiement, la date d'expiration de la carte de paiement et les quatre derniers chiffres du numéro de la carte du client.





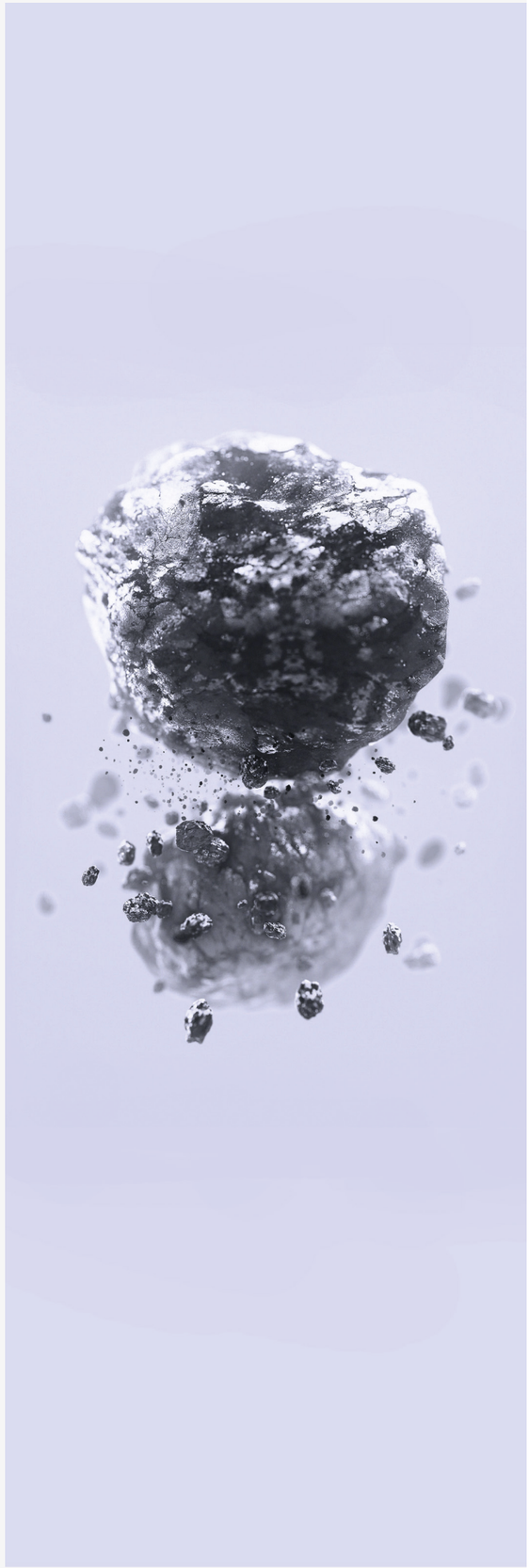
Aussi, en France, comme le révèle l'AFP, deux plaintes ont été déposées contre ChatGPT auprès de la CNIL. Elles portent sur la collecte des données personnelles sans consentement ainsi que la production d'informations erronées. Sur les questions relatives à l'utilisation et la conservation des données personnelles, OpenAI enregistre bien les conversations ChatGPT et les "prompts" en vue d'une analyse ultérieure. Selon une page de FAQ publiée par l'entreprise, ses employés peuvent examiner les conversations de manière sélective pour des raisons de sécurité. En d'autres termes, on ne peut pas supposer que tout ce que l'on communique à ChatGPT reste privé ou confidentiel.

Outre les prompts et les conversations, OpenAI enregistre également d'autres données tels 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 données aux tribunaux ou aux forces de l'ordre américaines conformément au FISA (Foreign Intelligence Surveillance Act).

# «Fine-tuning»

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. Cependant, ces modèles peuvent également être améliorés grâce à un processus connu sous le nom de «fine-tuning», qui consiste à entraîner à nouveau le modèle sur un petit ensemble de données (comme les conversations des utilisateurs). Il est de notoriété publique qu'OpenAI procède à ce type de mise au point des modèles depuis que la société a admis avoir embauché des humains pour simuler des conversations de chat idéales.



Toutefois, même lorsque les données sont accessibles au public, leur utilisation peut porter atteinte à ce que l'on appelle l'intégrité contextuelle. Il s'agit d'un principe fondamental dans les discussions juridiques sur la vie privée. Il exige que les informations relatives aux personnes ne soient pas révélées en dehors du contexte dans lequel elles ont été produites à l'origine.

# Plans d'actions

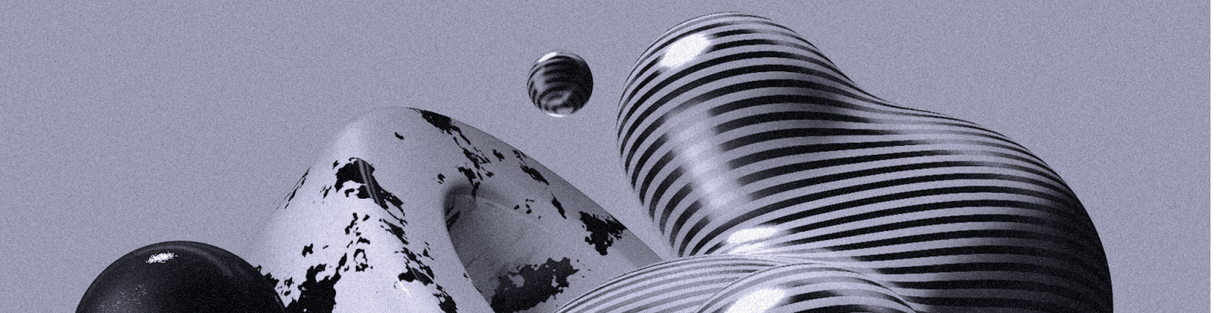
Par ailleurs, OpenAI n'offre aucune procédure permettant aux individus de vérifier si l'entreprise stocke leurs informations personnelles ou de demander leur suppression, alors qu'il s'agit là de droits garantis par le règlement général européen sur la protection des données (RGPD). On comprend ainsi mieux pourquoi ChatGPT intéresse autant les autorités de protection de données en Europe.

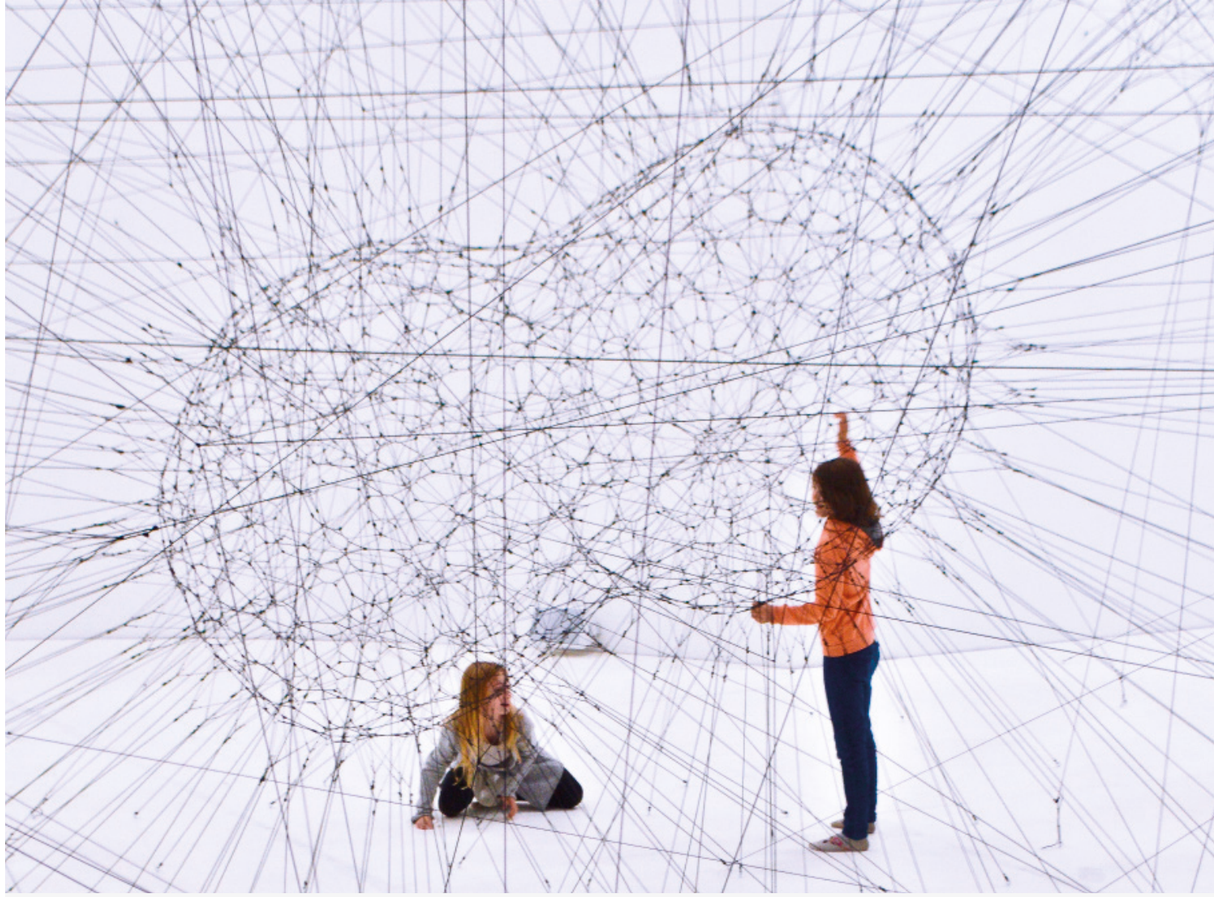
Fait intéressant, la CNIL a publié le 16 mai 2023 un "Plan d'action IA" afin de mieux appréhender et mieux répondre aux nombreux enjeux suscités par le développement de l'IA et tout particulièrement des IA génératives, telles que ChatGPT ou Midjourney (pour la génération d'images).

# Ce plan repose sur 4 volets :

- Comprendre comment fonctionnent les systèmes d'IA et comment ils affectent les humains ;
- Promouvoir et superviser le développement d'une intelligence artificielle respectueuse de la vie privée ;
- Fédérer et accompagner les acteurs innovants de l'écosystème français et européen de l'IA ;
- Auditer et contrôler les systèmes d'IA et protéger les personnes.

Ces travaux contribueront certainement également aux discussions en cours autour de l'IA Act européen, dont les dernières moutures intègrent davantage ces nouvelles réalités technologiques.





Toutefois, s'il est rassurant à certains égards que la CNIL ou la Commission européenne s'emparent du sujet, il est surtout primordial que les acteurs économiques, qui voient dans les IA génératives de possibles nouvelles opportunités de gain, adoptent systématiquement sur ces sujets une approche "privacy & security by design" ; que dès la réalisation de POC ou de prototypes, les considérations sur la protection et la sécurité des données personnelles soient intégrées par défaut afin de garantir un bon équilibre entre innovation et éthique.

Trouver ce juste milieu n'est jamais simple, mais c'est probablement l'un des principaux défis que posera le déploiement de l'IA dans les semaines et les mois à venir. C'est précisément ce dont témoigne cet appel de la Future of Life Institute le 22 mars 2023 à "suspendre immédiatement et pour au moins 6 mois l'entraînement des IA génératives plus puissantes que GPT-4», avec cette justification : "Les systèmes d'IA puissants ne devraient être développés qu'une fois que nous sommes confiants dans le fait que ses effets soient bénéfiques et dans notre capacité à en gérer les risques". Celui de la protection de la vie privée n'est pas le moindre d'entre eux.

Ismaël Koné,
**Consultant Senior Data Compliance**

Éliott Mourier,
**Data Compliance & Data Privacy Manager**