

INTELLIGENCE ARTIFICIELLE: LES LLM NE REPRODUISENT PAS DE DONNÉES

Posted on juillet 27, 2024 by Philippe Gilliéron



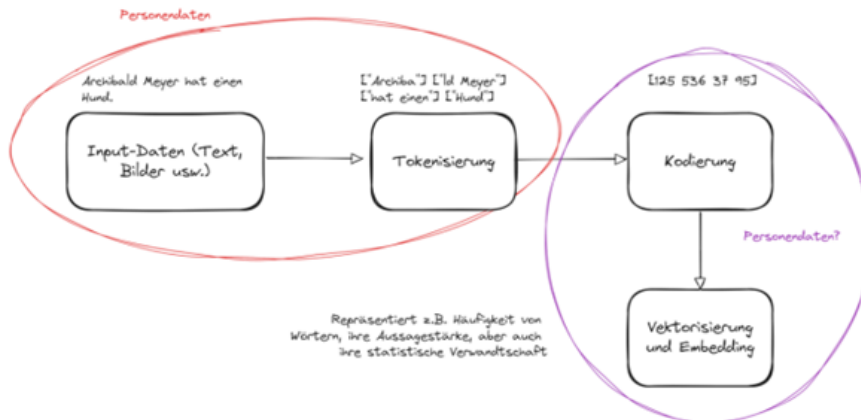
Les questions entourant l'exploitation de données personnelles ou de contenus protégés par des droits d'auteur sont essentielles pour permettre aux développeurs et utilisateurs d'apprécier leurs risques et de prendre les mesures propres à les minimiser.

Le [préposé à la protection des données de Hambourg a publié le 15 juillet 2024 un document](#) dans lequel il émet l'opinion suivant laquelle les LLM ne traiteraient pas de données personnelles. Le 23 juillet 2024, sans mentionner l'opinion du préposé dont il n'avait pas connaissance, Flemming Moos a publié dans *Computer und Recht* (CR 2024, 442) une contribution intitulée "[Personenbezug von Large Language Modles - Eine datenschutzrechtliche Grundsatzfrage bei der Nutzung generativer KI-Modelle](#)" dans laquelle il émet le même avis. [David Rosenthal](#) et [David Vasella](#) ont quelque peu nuancé cet avis.

Que faut-il en retenir?

1. Il convient de distinguer le niveau auquel on se place, la réponse à la question pouvant être différente suivant que l'on se trouve au niveau du modèle lui-même (GPT 4.0), de l'application qui en est dérivée (ChatGPT), de l'entreprise qui pourrait en faire une application dérivée en recourant à diverses techniques comme le RAG (*Retrieval Augmented Generation*) ou encore de l'utilisateur qui se contente de formuler des prompts pour générer une suggestion.
2. Suivant l'approche relative désormais suivie tant au niveau européen qu'en Suisse ensuite des arrêts Breyer ([C-582/14](#)) et VIN ([C-319/22](#)) respectivement rendus les 19 octobre 2016 et 9 novembre 2023 par la CJUE, ce dans la droite ligne du considérant 26 du RGPD, les réglementations en la matière ne peuvent s'appliquer que si des données permettant d'identifier une personne ou la rendre à tout le moins identifiable sont traitées. Or, pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre, soit par le responsable du traitement, soit par une autre personne, pour identifier ladite personne. Les principes de la protection ne s'appliquent pas aux données rendues anonymes d'une manière telle que la personne concernée n'est plus identifiable.
3. Les LLM ne reproduisent pas à proprement parler de données personnelles, puisque seuls des parties de mots sont utilisées (ou techniquement parlant "tokenisées") pour entraîner le modèle et permettre de calculer au travers d'innombrables paramètres la plus grande probabilité quant à la suite à donner au token concerné. Ce n'est en réalité que lors de modèles améliorés

(*fine-tuning*) ou recourant à des techniques particulières sur des données internes (RAG, *Retrieval Augmented Generation*) que le risque de voir certains mots, et donc des données personnelles, reproduits tels quels sous forme de token peut exister. Pour reprendre le schéma simple mais parlant dessiné par Vasella dans la contribution susmentionnée:



4.

5. Il ne peut en aller différemment que dans l'hypothèse, certes possible mais rare, où la probabilité que le token suivant corresponde à une donnée personnelle a pour conséquence que la donnée personnelle elle-même est érigée en token.
6. Si certaines attaques permettent de remonter de la suggestion aux données utilisées pour aboutir audit résultat (*membership inference attack, model inversion attack* ou encore les *training-data extraction attacks*), et donc à une telle donnée, le recours à de telles méthodes présuppose à l'aune de l'approche relative susmentionnée un niveau de connaissance qui en font des moyens a priori peu susceptibles d'être mis en oeuvre. A l'aune de l'approche relative susmentionnée, il faudrait en conclure qu'aucune donnée personnelle susceptible d'intéresser le RPGD ou la LPD n'entrerait en ligne de compte.

Au final, il découle de ces intéressantes contributions que les LLM ne contiennent en principe pas de données personnelles, réserve étant faite de cas particuliers. Même en ces hypothèses, encore faudrait-il que l'utilisateur ait un intérêt à recourir à des moyens lui permettant de remonter auxdites données, ce qui devrait être encore plus exceptionnel.

Même si la discussion n'est pas close, ces conclusions liminaires sont évidemment significatives. Transposées au droit d'auteur, elles signifient que, là encore sous réserve de cas exceptionnel, aucune oeuvre n'est en réalité reproduite dans un LLM, seuls des morceaux de prime abord

manquant d'individualité (point à vérifier) étant susceptibles de l'être sous forme de token. Inutile de souligner les conséquences qui pourraient en résulter sur les litiges en cours et l'impact sur les titulaires de droit, puisqu'aucune violation de droits d'auteur ne résulterait alors de la mise sur pied de tels modèles. Affaire à suivre.

