La CNIL publie un outil pour la traçabilité des modèles d'IA publiés en source ouverte

18 décembre 2025

La CNIL met à disposition un démonstrateur pour naviguer à travers la généalogie des modèles d'IA publiés en source ouverte et étudier la traçabilité de cet écosystème, notamment pour faciliter l'exercice de droits d'opposition, d'accès ou d'effacement.

La traçabilité des modèles d'IA en source ouverte

La mise à disposition des modèles d'IA en source ouverte permet de proposer cette technologie au plus grand nombre. Chercheurs, entreprises ou même particuliers peuvent désormais accéder à de nombreux modèles pour différents usages tels que la génération de texte ou d'images, la traduction, ou encore la transcription audio.

De nombreux utilisateurs téléchargent également ces modèles pour les modifier ou les spécialiser sur une tâche spécifique à l'aide de nouvelles données. Souvent, ces nouveaux modèles sont alors à nouveau mis à disposition en source ouverte.

Ainsi, chaque modèle disponible en source ouverte fait partie d'une généalogie, constituée de l'ensemble des modèles :

- dont il provient directement ou après plusieurs modifications (ses ascendants) ;
- auxquels il a contribué à la constitution (ses descendants).

Pouvoir décrire et rechercher dans une généalogie de modèle d'IA en source ouverte est donc une étape indispensable pour comprendre comment un modèle a été constitué.

La mémorisation des modèles d'IA et le RGPD

La communauté académique a établi de longue date qu'il est souvent possible d'extraire des informations sur la base d'entraînement d'un modèle d'IA, simplement à travers un accès à ce dernier. Ce phénomène se manifeste par la régurgitation des modèles génératifs, lorsque ceux-ci génèrent des données qui sont très similaires à des éléments de la base d'entraînement, mais ne s'y limite pas (voir par exemple l'article « Petite taxonomie des attaques des systèmes d'IA »).

Lorsqu'un modèle a été entraîné en partie sur des données personnelles (ce qui est généralement le cas pour l'IA générative), le Comité européen de la protection des données a énoncé dans son avis qu'il faudrait considérer dans la plupart des cas que celui-ci est soumis au RGPD. Le **responsable de traitement** pourra néanmoins démontrer, notamment à l'aide de tests, qu'il n'est pas possible d'extraire ou déduire de données personnelles à partir du modèle et que le RGPD ne viendrait pas à s'appliquer.

Une expérimentation pour étudier les d'IA en source ouverte

Dans ce contexte, la CNIL a mené une expérimentation visant à explorer des scénarios d'exercice des droits d'opposition, d'accès ou d'effacement, pour des personnes qui seraient concernées par la mémorisation de leurs données dans un modèle d'IA en source ouverte. La première étape pour cela vise à identifier, partant de la connaissance qu'un modèle a mémorisé les données d'une personne, les autres modèles de sa généalogie qui seraient susceptibles d'avoir également mémorisé ces données.

Dans ce but, le service IA de la CNIL a développé, en collaboration avec le Laboratoire d'Innovation Numérique de la CNIL (LINC), un outil de démonstration qui permet d'explorer la généalogie d'un modèle d'IA présent sur la plateforme *HuggingFace*.

L'outil

- Expérimenter l'outil sur la plateforme HuggingFace
- Lire l'article de présentation de l'expérimentation