

IA : la CNIL met à jour son outil de traçabilité des modèles publiés en source ouverte

26.8.2026 - | Commission Nationale de l'Informatique et des Libertés

La CNIL publie une nouvelle version de son démonstrateur permettant d'explorer la généalogie des modèles d'IA publiés en source ouverte. Cette mise à jour améliore notamment les performances de l'outil, son ergonomie et automatise l'actualisation des données. Une version en anglais est désormais disponible.

Explorer la généalogie des modèles d'IA

Les modèles d'IA publiés en source ouverte peuvent être téléchargés, modifiés, spécialisés à l'aide de nouvelles données ou combinés avec d'autres modèles, avant d'être à nouveau mis à disposition. Un même modèle (par exemple, Kimi K3, Mistral Medium, LLaMa etc.) peut ainsi être à l'origine de nombreux modèles dérivés.

Le démonstrateur Genmod développé par le service IA de la CNIL en collaboration avec le Laboratoire d'innovation numérique de la CNIL (LINC), initialement publié en novembre 2025, permet d'explorer ces liens et de retrouver les **ascendants** d'un modèle (les modèles dont il provient) ainsi que ses **descendants** (les modèles auxquels il a contribué).

Cette traçabilité est notamment utile pour étudier les conséquences de la **mémorisation de données d'entraînement** par les modèles d'IA. Il s'agit d'identifier, à partir d'un modèle ayant mémorisé des données personnelles, les autres modèles de sa « généalogie » susceptibles d'avoir également conservé ces informations. Cela permet notamment d'étudier les conditions d'exercice des droits prévus par le RGPD.

Une nouvelle version plus rapide et plus simple à utiliser

Cette nouvelle version apporte plusieurs améliorations au démonstrateur.

Une interface disponible en français et en anglais

L'ensemble de l'application est désormais utilisable dans les deux langues, via un bouton en haut à droite de la fenêtre.

Des recherches plus rapides

Le moteur d'exploration du graphe a été optimisé afin de fortement réduire le temps nécessaire aux recherches les plus étendues. Ainsi, une recherche sans limite de profondeur prend une vingtaine de secondes en moyenne.

Une meilleure visibilité sur l'avancement des recherches

L'interface affiche désormais la progression d'une recherche et, lorsqu'il peut être estimé, le temps restant avant son achèvement. Plusieurs recherches peuvent également être traitées simultanément.

En cas d'affluence, l'utilisateur peut connaître sa position dans la file d'attente ainsi que le nombre de recherches en cours.

Des données actualisées régulièrement

Un nouveau procédé permet de reconstruire la base utilisée par le démonstrateur à partir de données publiques relatives aux modèles et jeux de données disponibles sur HuggingFace. Il permet notamment d'automatiser l'actualisation hebdomadaire du graphe afin de mieux suivre l'évolution rapide de l'écosystème des modèles publiés en source ouverte. La date de mise à jour des données est désormais affichée dans l'application.

Une navigation et une consultation des résultats améliorées

Les modèles les plus téléchargés sont désormais proposés en priorité dans les suggestions de recherche et les résultats détaillés sont, par défaut, classés selon leur nombre de téléchargements. Un accès direct à la page d'accueil a également été ajouté aux différentes pages de l'application.

Enfin, la mise en page a été revue afin d'assurer un affichage plus homogène sur les principaux navigateurs modernes (Firefox, Chrome, Edge et Safari) ainsi que sur différentes tailles d'écran.

Cette nouvelle version poursuit ainsi l'expérimentation menée par la CNIL autour de la traçabilité des modèles d'IA en source ouverte, tout en rendant l'exploration de leurs généalogies plus rapide et plus facilement actualisable.

Découvrir l'outil

<https://www.cnil.fr/fr/ia-la-cnil-met-jour-son-outil-de-tracabilite-des-modeles-publies-en-source-ouverte>