DAILY SCIENCE
DÉCOUVREZ LA SCIENCE, LA RECHERCHE ET L'INNOVATION "MADE IN BELGIUM"

# MESURER L'ANONYMAT DE NOS DONNÉES SUR LE WEB

*Publié le 6 mars 2025*

par Daily Science

L'émergence de l'intelligence artificielle a, notamment, pour conséquence la quasi-impossibilité de garantir l'anonymat des personnes sur le web. Des scientifiques belges ont développé un modèle mathématique révolutionnaire pour évaluer les risques posés par l'IA et ainsi permettre la protection de la vie privée des individus. Cet outil devrait permettre de mieux réguler les codes publicitaires ou trackers invisibles qui permettent d'identifier les utilisateurs en ligne.

**Un modèle mathématique**

L'anonymat est essentiel pour protéger la liberté d'expression et les droits numériques dans nos démocraties. Il repose sur l'absence d'identification, de surveillance ou de traçabilité des individus. Cependant, avec les avancées de l'intelligence artificielle, garantir cet anonymat devient de plus en plus difficile. Effet, dans une étude précédente, réalisée en 2019, les scientifiques étaient parvenus à démontrer la facilité à réidentifier les personnes prétendument anonymisées sur le web, sur base de quelques informations partielles (âge, code postal, genre). Ce travail avait révélé l'ampleur des risques liés à la diffusion de données sensibles, même après anonymisation.

Ces mêmes chercheurs – Julien Hendrickx, professeur à l'Ecole polytechnique de l'UCLouvain, Yves-Alexandre de Montjoye, ingénieur UCLouvain et professeur associé à l'Imperial College London et Luc Rocher, ex-doctorant UCLouvain, professeur à la Oxford University – ont mis au point un nouveau modèle mathématique pour mieux comprendre les risques posés par l'IA.

## Statistiques bayésiennes

Dans cette nouvelle étude, ils proposent un modèle innovant (baptisé modèle de correction Pitman-Yor (PYC)) qui évalue les performances des techniques d'identification à grande échelle, dans différents contextes d'application et de comportement.

Julien Hendrickx explique : « notre nouvel outil s'appuie sur les statistiques dites bayésiennes pour apprendre à quel point les individus sont similaires, et extrapoler la précision de l'identification à des populations plus importantes, avec une performance jusqu'à 10 fois supérieure aux règles précédentes. Ce travail fournit, pour la première fois, un cadre scientifique robuste permettant d'évaluer les techniques d'identification, pour les données à grande échelle. »

## Lutter contre l'empreinte digitale de l'appareil

L'objectif de cette recherche ? « Outre aider à mieux comprendre les risques posés par l'IA, il s'agit de permettre aux régulateurs de mieux protéger la vie privée des personnes. Même si des réglementations telles que le RGPD encadrent strictement l'utilisation et le partage des données personnelles, les données anonymisées échappent à ces restrictions. Il était donc essentiel de déterminer si des données sont réellement anonymes ou peuvent être réidentifiées, afin de contribuer au respect à la vie privée. »

Par exemple, dans des études médicales, l'outil peut aider à déterminer si des informations sur des patients peuvent être utilisées pour retrouver leur identité, et ainsi contribuer à empêcher cette identification. Dans la vie quotidienne, il permet également de surveiller (et donc contrer) la précision des codes publicitaires et des trackers invisibles qui identifient les utilisateurs en ligne à partir de détails, comme le fuseau horaire ou les paramètres de navigateur, selon une technique appelée "empreinte digitale de l'appareil".

« Nous pensons que ce travail constitue une étape cruciale vers le développement de méthodes rigoureuses pour évaluer les risques posés par les techniques d'IA de plus en plus avancées et la nature de l'identification des traces humaines en ligne. Nous espérons que ce travail sera d'une grande aide aux scientifiques, responsables de la protection des données, membres de comités d'éthique et autres praticiens qui cherchent à équilibrer le partage de données pour la recherche et la protection de la vie privée des citoyens. »