

Ouvrir la science à l'heure de l'intelligence artificielle : pourquoi et comment ?

Eléments de réflexion à destination de nos communautés de recherche

Argumentaire élaboré par la commission recherche et documentation de l'ADBU en décembre 2025.

Depuis un quart de siècle, le développement de la science ouverte a eu de nombreux effets positifs sur la diffusion de la connaissance scientifique : elle a facilité l'accès aux publications et aux données, mis l'accent sur la reproductibilité de la recherche et rendu possible l'exploitation en masse de données scientifiques ouvertes.

Le développement plus récent des intelligences artificielles vient changer en profondeur les pratiques de la recherche. Les données ouvertes par les communautés de recherche deviennent une source privilégiée pour l'entraînement des grands modèles de langage développés par des entreprises commerciales, en particulier les GAFAM.

Il devient dès lors essentiel d'interroger les valeurs et les pratiques de la science ouverte au prisme de cette nouvelle donne. Pourquoi et comment ouvrir la science à l'ère des intelligences artificielles génératives ?

1. La fiabilité des intelligences artificielles génératives est aujourd'hui un enjeu pour la qualité de l'information et la diffusion des connaissances scientifiques

Valeurs fortes pour la recherche scientifique, la transparence et la reproductibilité deviennent essentielles dans une époque où la science est questionnée et remise en cause par des pratiques de désinformation. L'accès ouvert et libre aux sources primaires d'une information doit ainsi être préservé, afin qu'il soit toujours possible d'appuyer son discours ou sa recherche sur des contenus validés scientifiquement.

Parce qu'ils sont très largement utilisés, en particulier par les jeunes générations, les systèmes d'IAG constituent aujourd'hui des médias essentiels dans la diffusion de l'information. Les intelligences artificielles génératives se nourrissent en priorité des données librement disponibles sur internet. La qualité de l'information synthétisée par les IAG dépend donc en grande partie de la qualité des corpus disponibles. La mise à disposition ouverte de résultats scientifiques fiables est ainsi de nature à améliorer la qualité des informations accessibles depuis une IAG.

2. Publier des travaux dans des revues en accès fermé ne garantit pas à l'auteur l'absence d'exploitation par une intelligence artificielle générative.

Dès lors qu'un auteur a cédé ses droits sur une publication, l'éditeur peut choisir d'utiliser cette publication pour entraîner une IAG : entraînement de modèles ou mise en place d'assistants de recherche, vente de corpus à un tiers moyennant un mécanisme de licence. Dans un monde en constante évolution, les données scientifiques possédées par un éditeur représentent un capital d'une valeur importante et un atout pour le développement de nouveaux modèles économiques.

Ainsi, la question de l'exploitation des travaux scientifiques par les IAG dépasse-t-elle le débat sur l'ouverture et la fermeture. Elle concerne de fait l'ensemble des publications et travaux universitaires. Elle incite à la vigilance sur les contrats de cession des droits signés au moment de la publication.

Le retour à la publication en accès fermé ne constitue donc pas une solution pour résoudre les difficultés posées par l'exploitation des publications scientifiques par les intelligences artificielles génératives.

3. Ouvrir les contenus contribue à la diversité des solutions d'intelligence artificielle générative et leur adaptation aux spécificités de la recherche

L'accès à des contenus scientifiques librement disponibles permet à une diversité d'acteurs de se positionner sur le développement de solutions d'intelligence artificielle générative. Une pluralité d'acteurs garantit l'existence d'une pluralité de solutions : intelligences artificielles génératives plus transparentes, fonctionnant sur un périmètre de contenus validés scientifiquement, solutions moins gourmandes énergétiquement...

La recherche scientifique réclame des IAG exigeantes dans leur fonctionnement, capables de décrire leur processus de recherche et le périmètre de corpus fouillés. Le développement de technologies fiables et précises s'appuie sur la disponibilité de publications et de données ouvertes et de qualité.

4. Les intelligences artificielles génératives enferment les processus de recherche dans une boîte noire : la recherche doit les en libérer !

Les intelligences artificielles génératives nuisent à la reproductibilité de la recherche. D'une part, les systèmes d'IAG et les modèles de langage demeurent en grande majorité opaque. Les fournisseurs de modèles de langage rendent généralement accessibles une partie de leurs paramètres (leurs poids ou weights en anglais) mais restent par ailleurs propriétaires de leurs composants. Il ne s'agit donc pas de logiciels libres.

Leur nature probabiliste explique également la difficulté à reproduire un processus de recherche mobilisant un modèle de langage. Par défaut, le même prompt fournit des résultats différents sur deux ordinateurs différents, ou sur le même ordinateur à quelques minutes d'intervalle. Cet obstacle à la reproductibilité est encore aggravé par les changements de version très réguliers des modèles de langage.

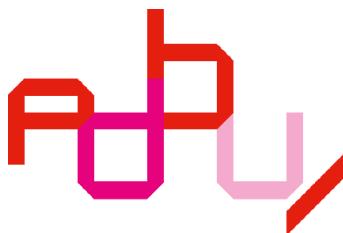
La transparence de l'ensemble des sources de la recherche (publications, données, logiciels) demeure une clé essentielle pour la reproductibilité de la recherche scientifique. Les communautés de la recherche peuvent et doivent produire de la recherche ouverte sur les intelligences artificielles génératives, afin d'en limiter les effets négatifs sur la reproductibilité.

Ouvrez les intelligences artificielles génératives ! Ne refermez pas la science !

Les intelligences artificielles génératives représentent des opportunités et un défi important pour nos communautés de recherche et pour la société. Elles posent des questions majeures dans le champ de l'écologie, de l'information et de la transparence. Face à ces technologies, la tentation est grande de protéger la science par de nouvelles barrières pour éviter son exploitation non maîtrisée. Mais c'est bien le mouvement inverse, la poursuite de l'ouverture de la science, qui offre pour demain les meilleures garanties pour la transparence et la reproductibilité de la recherche, pour un partage des connaissances scientifiques avec le plus grand nombre. Comme nous contribuons au quotidien à une science ouverte et partagée, participons à la mise en œuvre d'intelligences artificielles ouvertes et de qualité.

Pour aller plus loin :

Cours en libre-accès de l'URFIST de Toulouse : *Intelligence Artificielle et Science Ouverte : tension ou complémentarité ?* : <https://callisto-formation.fr/course/view.php?id=912>
(septembre 2025)



L'ADBU est l'Association française des directeurs et personnels de direction des bibliothèques universitaires et de la documentation. Dans l'écosystème en transformation de la connaissance et de l'information, l'ADBU porte la vision d'une bibliothèque universitaire (BU) au cœur des enjeux sociaux. L'ADBU s'intéresse aux questions de stratégie documentaire, de diffusion de l'information scientifique et technique et ne cesse de réaffirmer le rôle, les missions, l'ambition et les valeurs des bibliothèques universitaires dans la construction d'une société démocratique, ouverte à toutes et à tous.