

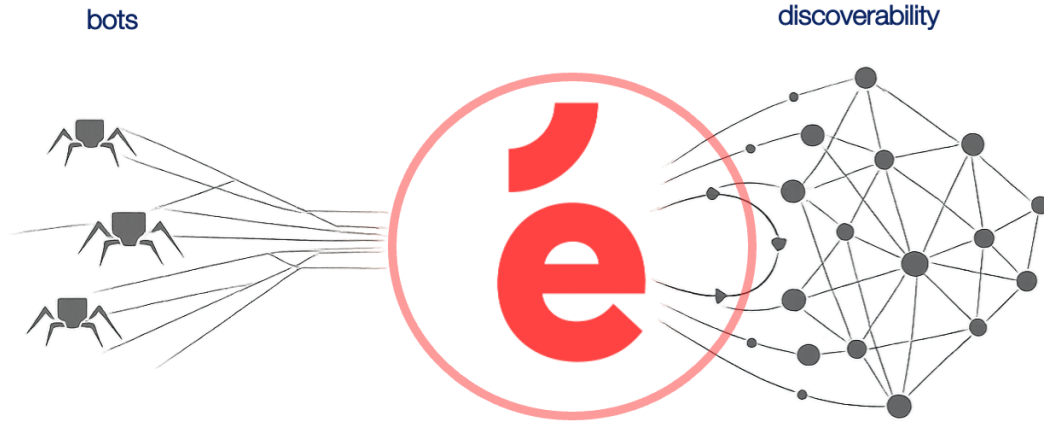
Le défi des bots et les opportunités offertes par l'IA chez Érudit

David Cormier (responsable de l'équipe informatique) & Yves Terrat (spécialiste en IA)

érudit



Aperçu



1. Le défi de l'entraînement des modèles d'IA et de l'extraction massive de données
2. La solution initiale (et actuelle) d'Érudit
3. Prochaines étapes

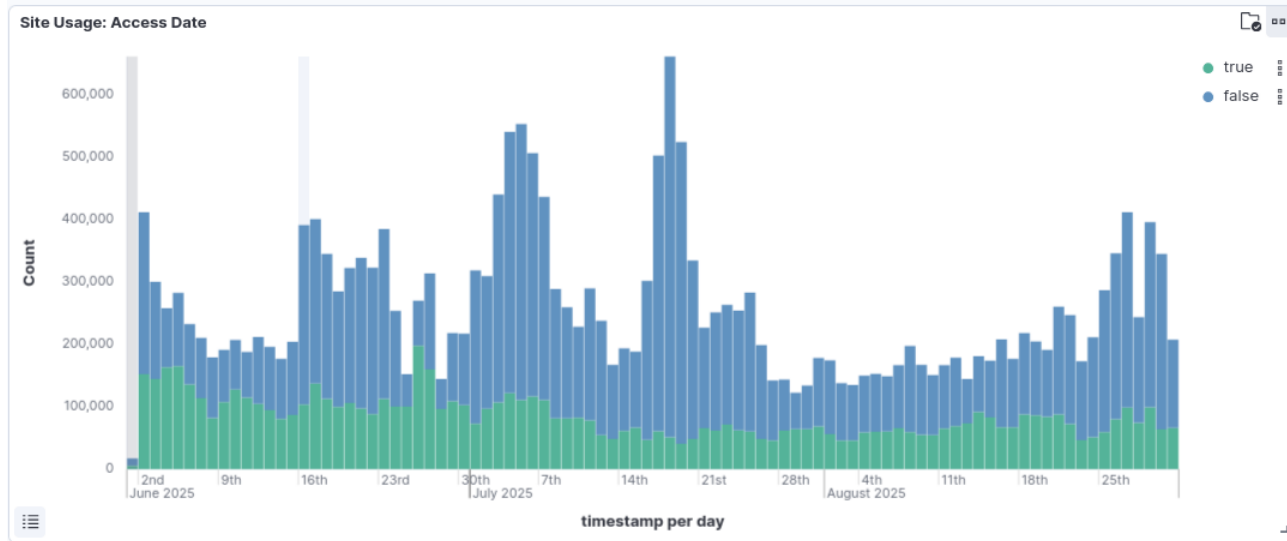
Érudit fait face à une augmentation du trafic généré par les bots

- Les modèles linguistiques reposent sur un contenu textuel de haute qualité
- Les opérateurs de bots utilisent des méthodes sophistiquées pour échapper à la détection
- Le trafic généré par les bots devient la norme plutôt que l'exception et doit être détecté automatiquement (à l'instar des spams par e-mail)

Conséquences de l'augmentation du trafic généré par les bots

- Le code de conduite COUNTER exige que les accès par des robots soient identifiés comme tels
- Cela met à rude épreuve l'infrastructure d'Érudit

Le trafic généré par les bots sur Érudit a augmenté en 2025



Rapport d'utilisation pour l'été 2025 (du 1er juin au 1er septembre)

- bleu : trafic des bots exclu – plus de 50 % du trafic
- Cela correspond à nos observations : par exemple, l'université Duke fait état de 90 %

Filtrage automatique des accès par des bots avec Anubis chez Érudit

- Utilise un mécanisme de preuve de travail pour détecter les accès automatisés
- Logiciel open source, développé au Canada
- **Auto-hébergé : l'utilisation du site n'est pas suivie par un tiers**
- Érudit a configuré Anubis pour minimiser les effets secondaires (les adresses IP des institutions canadiennes ne sont pas filtrées)
- A contribué à réduire les accès par des robots et à maintenir à un niveau gérable l'effort nécessaire à la gestion des rapports d'utilisation COUNTER



Vers une solution technique nuancée

- La mise en œuvre d'Anubis représente une « posture défensive »
- Cependant, elle bloque également de nouvelles utilisations légitimes
- Le défi d'Érudit consiste à prendre en compte ces nouvelles utilisations

L'IA chez Érudit

- La charte sur l'IA, pilier des développements futurs
- Méthodes de surveillance pour détecter les utilisations légitimes
- Projet Studia (CFI Innovation) et logiciels de recherche intégrant l'IA (Digital Research Alliance)

L'IA dans l'édition et la bibliothéconomie

Mike  , CRKN Virtual | 2026

Il est important de préciser d'emblée que le terme « IA » est un concept marketing qui ne nous rend pas vraiment service dans ce contexte. Le traitement du langage naturel et l'apprentissage automatique existaient bien avant que Sam Altman et OpenAI ne sortent de leur chrysalide et ne se mettent à se nourrir de la culture.

La spécificité est l'âme du récit (Hodgman).

IA discriminatoire

L'IA discriminatoire comprend les logiciels qui exploitent les réseaux neuronaux et l'apprentissage automatique pour trier ou traiter des données. Ces modèles *établissent une distinction* entre un type d'élément et un autre. La reconnaissance faciale est un exemple d'IA discriminatoire. Les logiciels capables de détecter des cellules cancéreuses dans l'imagerie médicale et d'établir des diagnostics relèvent également de l'IA discriminatoire.

Ils analysent les modèles et évaluent les probabilités.

[Discours de clôture : Gérer l'IA générative, ses risques et les techniques d'atténuation. Ben Zhao. Open Repositories 2025.](#)

IA générative

L'IA générative trouve son origine dans une technologie que l'on appelait autrefois les « chaînes de Markov ». En termes très simples, l'IA générative *crée le texte le plus probable* en fonction d'une invite et du corpus sur lequel le modèle a été entraîné.

Dans ce cas, des modèles comme Gemini ont indiqué qu'il fallait utiliser de la colle pour maintenir le fromage sur une pizza, car leur corpus couvrait *l'ensemble d'Internet, y compris Reddit, et l'IA générative ne sait pas ce qu'est le sarcasme.*

En fait, [l'IA générative ne **sait** rien.](#)

Ces concepts sont souvent regroupés sous une même appellation générale, ce qui pose toutes sortes de problèmes lorsqu'il s'agit de présumer de la bonne foi. Outre une adoption aveugle au niveau institutionnel, c'est peut-être là le plus grand obstacle dans mon travail.

Dans le cadre de notre discussion d'aujourd'hui, je parle principalement de *l'IA générative*.

Connaître les différents aspects

les implications d'une utilisation plus large ou « meilleure » de l'IA dans l'édition sont difficiles à cerner. Mais nous allons essayer d'aborder brièvement les points suivants :

- L'IA dans la création littéraire et l'évaluation par les pairs
- comment les éditeurs utilisent l'IA aujourd'hui
- comment les revues tentent de se protéger contre les abus de l'IA, et en quoi cela nécessite de la bonne foi
- quelques exemples de bonnes pratiques, et où trouver des discussions et des exemples

L'IA dans l'évaluation par les pairs et l'édition

Nous savons que [certaines personnes font appel à des agents pour rédiger des articles de recherche dans leur intégralité.](#)

Nous savons également qu'un nombre important de [personnes ont recours à des agents pour effectuer l'évaluation par les pairs.](#)

Nous savons également que certains éditeurs utilisent l'IA pour [les rejets initiaux et la présélection.](#)

Nous savons que des entreprises vendent des solutions qui exploitent leurs propres outils basés sur des agents, dans une sorte de course aux armements « guerre des machines » extrêmement coûteuse et perverse.

Nous savons que l'argument « ça fait gagner du temps » est utilisé comme une sorte de chèque en blanc éthique.

Au final, **tout cela représente davantage de travail pour les personnes qui effectuent réellement le travail de rédaction et d'évaluation par les pairs.**

Et, [ce qui est](#) peut-être le plus alarmant, [il s'avère que cela dégrade notre capacité à accomplir ces tâches nous-mêmes dès lors que nous comptons sur l'IA pour les faire à notre place.](#)

Actuellement, dans le meilleur des cas, nous **ne pouvons vraiment compter que sur la bonne foi, la transparence et la divulgation au sein de la communauté.**

Les rédacteurs en chef s'inquiètent autant des abus de l'IA dans leurs revues que de ceux commis par leurs étudiants. Ils doivent donc se fier à leur intuition, aux indices révélateurs et aux signes de bonne foi.

Et sur la bonne foi.

Heureusement pour eux, je m'abreuve à la source.

politiques, outils et vœux pieux

L'une des façons dont les revues tentent de tracer une ligne claire consiste à mettre en place des politiques sans ambiguïté en matière de « divulgation ».

Ceux qui utilisent l'IA dans leur travail doivent le dire ouvertement et être disposés/capables de révéler comment ces outils ont été utilisés dans leur rédaction.

Les revues s'efforcent de préciser ce point aux auteurs.

Nous attendons de tous les manuscrits soumis qu'ils soient rédigés par un ou plusieurs auteurs humains. Nous encourageons les auteurs à nous contacter s'ils ont des inquiétudes ou besoin d'aide.

Nous n'acceptons pas les contenus produits ou édités par une IA générative. Cela inclut (sans s'y limiter) l'analyse de données, les textes générés, les images ou les traductions. Nous n'autorisons pas non plus l'utilisation de l'IA générative pendant le processus d'évaluation par les pairs.

Question de divulgation qui vous sera posée lors du processus de soumission :

Avez-vous utilisé un outil d'IA générative pour rédiger ou éditer ce manuscrit, analyser des données ou créer des images ? (Si vous ne l'avez utilisé que pendant les phases de brainstorming ou d'ébauche, sélectionnez « non ». Si votre utilisation s'est poursuivie au-delà de la phase de pré-rédaction, nous vous encourageons à soumettre votre manuscrit ailleurs ; nous ne l'examinerons pas.)

- [In the Library with the Lead Pipe, Politique relative à l'IA.](#)

politiques, outils et vœux pieux

L'une des façons dont les revues tentent de tracer une ligne claire consiste à mettre en place des politiques sans ambiguïté en matière de « divulgation ».

Ceux qui utilisent l'IA dans leur travail doivent le dire ouvertement et être disposés/capables de révéler comment ces outils ont été utilisés dans leur rédaction.

Les revues s'efforcent de préciser ce point à l'intention des auteurs.

Nous constatons une différence fondamentale entre les outils d'aide intégrés aux logiciels de bureau ou basés sur le cloud couramment utilisés, qui affinent le travail linguistique ou intellectuel déjà effectué par un auteur (par exemple, les suggestions grammaticales dans MS Word ou les suggestions de formules dans Google Sheets), et les outils génératifs qui produisent un nouveau texte ou une nouvelle image et/ou développent un argument, tirent des conclusions, synthétisent des concepts ou analysent des informations pour l'auteur (par exemple, ChatGPT, Claude). Nous constatons une distinction significative entre le cas où un auteur connaît la réponse ou a déjà mûri ses idées, et celui où un auteur utilise des outils externes pour réfléchir à sa place.

- [In the Library with the Lead Pipe, Politique en matière d'IA.](#)

politiques, outils et vœux pieux

L'une des façons dont les revues tentent de tracer une ligne claire consiste à mettre en place des politiques sans ambiguïté en matière de « divulgation ».

Ceux qui utilisent l'IA dans leur travail doivent le dire ouvertement et être disposés/capables de révéler comment ces outils ont été utilisés dans leur rédaction.

Les revues s'efforcent de préciser ce point aux auteurs.

D'après notre expérience de la lecture de contenus générés par ces outils, ceux-ci sont souvent truffés d'erreurs. De plus, à moins d'être utilisés avec beaucoup de prudence, notre expérience montre que les outils d'IA générative à usage général ne sont pas adaptés à la révision de textes et ont tendance à introduire autant de problèmes de fond (concernant les faits et la formulation) qu'ils en éliminent au niveau des phrases (tels que l'accord sujet-verbe ou l'utilisation répétitive de mots). Pour ces raisons, nous n'acceptons actuellement pas les articles ayant été traités par des outils GenAI. Nous encourageons vivement les auteurs à consulter les centres d'aide à la rédaction, les relecteurs ou les guides de style académiques à leur disposition, car ceux-ci permettront généralement de préserver l'intention de l'auteur tout en l'aidant à affiner ses propres compétences en matière de rédaction et de révision.

- [In the Library with the Lead Pipe, Politique relative à l'IA.](#)

J'apprécie beaucoup cette politique, car je pense qu'elle met en lumière les problèmes liés à l'IA générative dans le domaine universitaire *tout en* précisant clairement les attentes de la revue.

Mais je sais aussi que certaines personnes l'ignoreront complètement et tenteront quand même leur chance. On ne peut pas faire grand-chose contre les personnes de mauvaise foi.

revues partenaires

D'après mon expérience, le degré d'inquiétude *actuel* des rédacteurs en chef face aux articles rédigés par l'IA parmi les soumissions qu'ils reçoivent est probablement le fruit de la confiance qu'ils accordent à leur communauté de chercheurs.

Je les renvoie, comme je le fais si souvent, au [Comité sur l'éthique de la publication](#) (COPE).

Le COPE fait partie d'un collectif d'organisations travaillant à l'élaboration d'une [norme mondiale de déclaration de l'utilisation de l'IA dans la recherche](#), qui s'est réuni très récemment à Vancouver.

Parmi les autres recommandations, on peut citer :

EIFL. (2025). [L'IA générative dans l'édition en libre accès Diamond](#). Zenodo.

[Webinaire LPC : Élaboration de politiques en matière d'IA](#) pour les publications (2025)

[Politiques éditoriales des revues sur l'IA : liste collaborative \(LPC\)](#)

[Recommandations du DOAJ sur la formulation des politiques](#)

C'est un peu comme la réaction des gens pendant la pandémie de Covid : ils attendent que les dirigeants donnent l'exemple ou qu'on leur fournisse des modèles à suivre... des « bonnes pratiques » dans un contexte où, il y a quelques années encore, il n'y en avait pratiquement pas.

En tant que bibliothécaire (et, plus particulièrement, en tant que bibliothécaire sceptique vis-à-vis de l'IA), je dois reconnaître qu'à l'heure actuelle, le mieux que la plupart d'entre nous puissent faire dans ce domaine est de formuler des demandes avec courtoisie, d'établir des lignes directrices, de se tenir informés et vigilants, et de contribuer collectivement, en toute bonne foi, à la communauté plus large des chercheurs.

Tout cela signifie que mon travail consiste à essayer de fournir aux éditeurs et aux auteurs autant d'informations que possible afin de leur permettre de réfléchir à :

- **leur rapport à la propriété de leurs travaux ou des fruits de leur travail**
- **leur rapport aux entreprises technologiques, aux outils, aux éditeurs et au concept plus large d'« intégrité académique », ainsi que la confiance qu'ils leur accordent**
- **leur intérêt à déconstruire le discours de l'« inévitabilité »**

En fin de compte, je suppose qu'on aurait dû en parler depuis le début. C'est une bonne chose qu'on le fasse maintenant et qu'on essaie de résoudre ces problèmes de manière collective et culturelle, au lieu de se contenter d'ajouter encore plus de robots (et de réservoirs d'eau communautaires) au bûcher.

La question de l'évaluation par les pairs est peut-être plus inquiétante...