L'IA générative affecte-t-elle la consultation sur Érudit ? Une première analyse d'internautes et de robots

Simon van Bellen

Conseiller principal en recherche chez Érudit-Coalition Publica et membre de la Chaire UNESCO sur la science ouverte, École de bibliothéconomie et des sciences de l'information, Université de Montréal, Québec, Canada

simon.van.bellen@erudit.org - orcid.org/0000-0002-1698-8530

Introduction

Les applications de l'intelligence artificielle (IA) générative bouleversent le monde de la publication. Depuis le développement d'outils comme ChatGPT et Google AI Overviews, les internautes cherchent de plus en plus de l'information à travers ces outils, plutôt qu'auprès des sources originales, comme le démontrent les statistiques récentes de consultation de Wikipédia[1]. De plus, les grands modèles de langue associés à ces outils sont basés sur de très grandes quantités de données textuelles, ce qui fait en sorte qu'un nombre croissant de robots moissonnent massivement les pages web qui leur sont accessibles (Kwon, 2025). Ainsi, ils contaminent les statistiques de téléchargement, les rendant inutiles pour l'analyse et l'interprétation, et par des accès excessifs à haute fréquence ils menacent la stabilité et la disponibilité des infrastructures.

Érudit, en tant qu'infrastructure ouverte diffusant plus de 170 000 documents scientifiques, n'y fait pas exception. À l'automne 2025 Érudit a déployé Anubis[2], un logiciel libre qui permet de bloquer des tentatives d'accès illégitimes à la plateforme. Sur les pages d'Érudit, la légitimité d'un usage est basée sur deux aspects : la présence humaine et le type de comportement. Premièrement, nous considérons comme légitime tout accès direct par un être humain à travers un navigateur. Les robots peuvent être considérés comme légitimes, à condition qu'ils respectent nos règles d'accès, telles que précisées dans le fichier robots.txt sur nos serveurs. Les accès semi-automatisés des grands modèles de langue en réponse à une requête d'un·e internaute sont également autorisés et

[1] https://diff.wikimedia.org/2025/10/17/new-user-trends-on-wikipedia/

[2] Le déploiement d'Anubis sur Érudit : https://apropos.erudit.org/deploiement-anubis/

CC BY érudit COALITION PUBLICA

comptabilisés parmi les consultations légitimes dans la mesure où ce modèle fournit des plages d'adresses IP publiques qui nous permettent d'identifier clairement la provenance des requêtes. Cependant, les accès automatisés qui ne respectent pas les règles d'accès, qui cachent leur origine, qui utilisent de faux agents utilisateurs, ou qui moissonnent à très haute fréquence nos contenus sont bloqués par Anubis ou enlevés des statistiques a posteriori, par une routine de « nettoyage » manuel.

Sur Érudit, les tendances de consultation sont d'un grand intérêt pour les revues diffusées, puisque les redevances qu'elles reçoivent en dépendent partiellement. Elles sont importantes pour les bibliothèques partenaires et abonnées, qui peuvent en tirer de précieuses informations pour mieux connaître les besoins de leurs communautés - et y trouver des arguments supplémentaires pour solidifier leurs liens avec Érudit. Bien sûr, nos bailleurs de fonds suivent le nombre de visites enregistrées sur la plateforme, dont les tendances documentent la circulation des savoirs à partir du Canada. En soi, les statistiques de consultation sont d'intérêt pour la recherche sur une multitude de sujets potentiels, par exemple l'effet du libre accès sur la consultation[3]. L'objectif de ce texte est de présenter notre compréhension des tendances de consultation du contenu savant sur Érudit.

Un an après le déploiement d'Anubis, nous avons remarqué que le nombre de consultations classées comme étant légitimes a baissé de façon importante, et ce, depuis l'automne 2024 (Figure 1). Une explication potentielle est liée à un usage croissant des résumés générés par des outils de l'IA générative. L'accessibilité de ces réponses condensées fait en sorte que les internautes s'informent moins souvent à partir des sources originales. Ce phénomène, que l'on appelle « zero-click search » en anglais, est bien documenté dans différents contextes. Par exemple, lors d'une expérience, Khosravi et Yoganarasimhan (2026) ont démontré que l'intégration de l'aperçu IA de Google a mené à une baisse significative des consultations des pages de Wikipédia à l'origine des résumés générés.

De l'autre côté, une baisse apparente des accès à la plateforme pourrait être liée à un nettoyage accru lors de la dernière année, ou, formulé autrement, à une surestimation des consultations légitimes lors de la période précédant le déploiement d'Anubis, alors que le nettoyage ne se faisait que manuellement. Peu familiers avec le phénomène croissant des robots moissonnant l'intégralité de nos collections il y a à peine un an, nous avons possiblement classé comme légitimes des accès qui auraient dû être classés illégitimes. Il nous importe

[3] https://apropos.erudit.org/effet-libre-acces-revues/

d’obtenir une idée, ne serait-ce que globale, de l‘importance relative des deux explications.

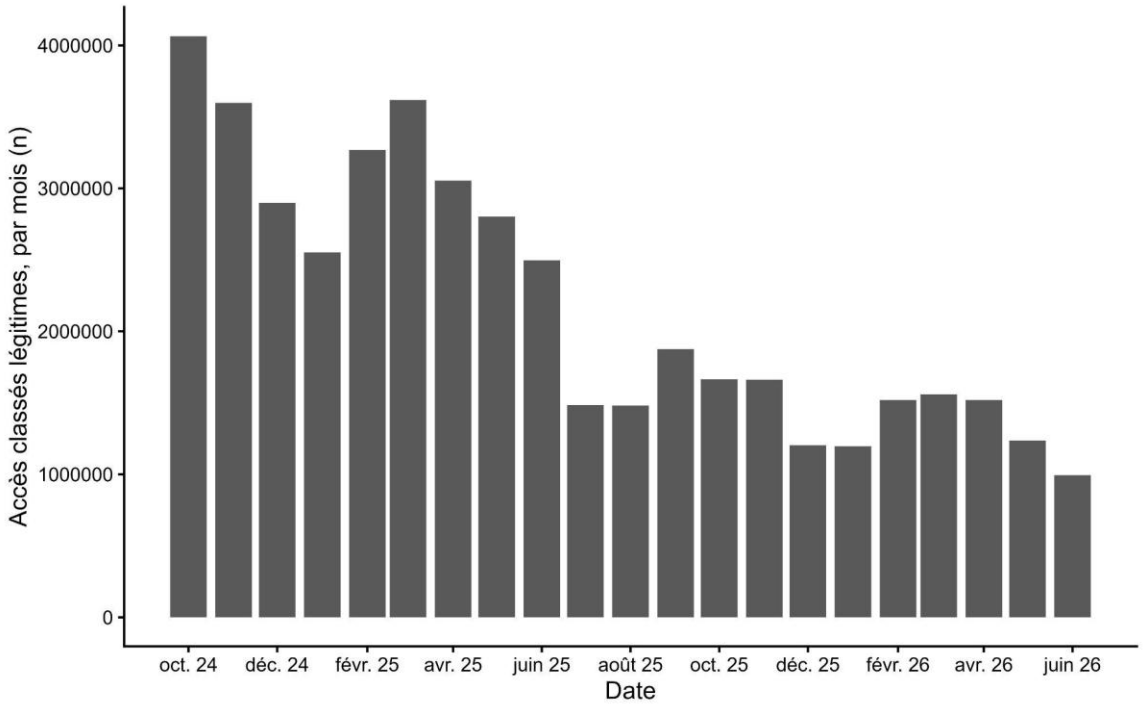


Figure 1 : Accès classés légitimes du contenu savant sur Érudit, 1er octobre 2024 au 30 juin 2026.

Méthodologie

Les accès sur Érudit sont enregistrés suivant la norme internationale COUNTER R5. Puisque la norme COUNTER se limite à l’exclusion des robots selon une liste préétablie par l’organisation, la classification des accès par l’équipe d’Érudit représente une étape supplémentaire facultative, permettant de garantir la qualité des données de consultation retenues.

Sur Érudit les accès peuvent également être classifiés selon l’affiliation des internautes. D’abord, Érudit a environ 700 institutions et bibliothèques partenaires et abonnées, qui contribuent au développement des revues ainsi qu’à la plateforme, en échange d’un accès sans restriction. Lorsque leurs communautés (chercheur·e·s, étudiant·e·s, enseignant·e·s, personnel de soutien) se connectent à travers le proxy de leur institution, ces accès sont enregistrés comme étant des accès « par abonnement ». Ensuite, la grande majorité des articles diffusés sont accessibles même si l’internaute n’est pas affilié·e à une institution ou bibliothèque partenaire ou abonnée; la seule

exigence est de disposer d'une connexion web. Ces accès sont classés comme en provenance de « non abonnés ».

Pour l'analyse, nous avons retenu toutes les consultations, légitimes et illégitimes, du contenu savant sur Érudit. Nous avons comparé deux périodes, du 1er octobre 2024 au 30 juin 2025, et du 1er octobre 2025 au 30 juin 2026. La première période précède le déploiement d'Anubis en septembre 2025. Dans ce billet, nous référons généralement à « 2024-2025 » et « 2025-2026 » pour identifier ces périodes.

Résultats

La part d'un nettoyage accru lors de la dernière année

Un effet de nettoyage accru lors de la dernière année pourrait expliquer une partie de la baisse enregistrée. Autrement dit, le nettoyage des accès aurait été insuffisant lors de la période 2024-2025. Cette situation pourrait être attribuée au fait que ce phénomène est très récent, et que peu de documentation pouvant guider les plateformes est disponible. Puisque les accès par abonnement sont légitimes par définition (les proxys institutionnels hébergent très peu de robots), nous n'avons considéré que les accès par non-abonnés. Afin d'estimer si certains accès ont pu passer entre les mailles du filet, il est d'intérêt de quantifier les accès en fonction des pays à l'origine de ceux-ci. Cette approche s'appuie sur le fait que certains pays se trouvent à l'origine de grands nombres de visites par robots, alors que d'autres n'hébergent pas ou très peu de robots[4].

Nous avons comparé le profil géographique du lectorat de 2024-2025 avec celui de 2025-2026. En présumant que le nettoyage y était sans faille, cette dernière période agit comme étalon pour l'évaluation de la période précédente. En comparant 2024-2025 et 2025-2026, on aperçoit des différences marquées pour ce qui est de l'origine géographique des accès jugés légitimes (Figure 2). Lors de la période 2024-2025, antérieure à l'intégration d'Anubis et donc au nettoyage accru, des pays comme Hong Kong, le Singapour, la Russie et le Brésil enregistraient des nombres relativement élevés d'accès classés légitimes en comparaison avec 2025-2026. À l'inverse, le Canada, la France, le Maroc et la Belgique étaient nettement plus présents en 2025-2026 qu'en 2024-2025; ces pays seraient caractérisés par relativement peu d'accès illégitimes.

[4] Pour un aperçu actuel des pays hébergeant de nombres élevés de robots, consultez https://gs.statcounter.com/bots-by-country

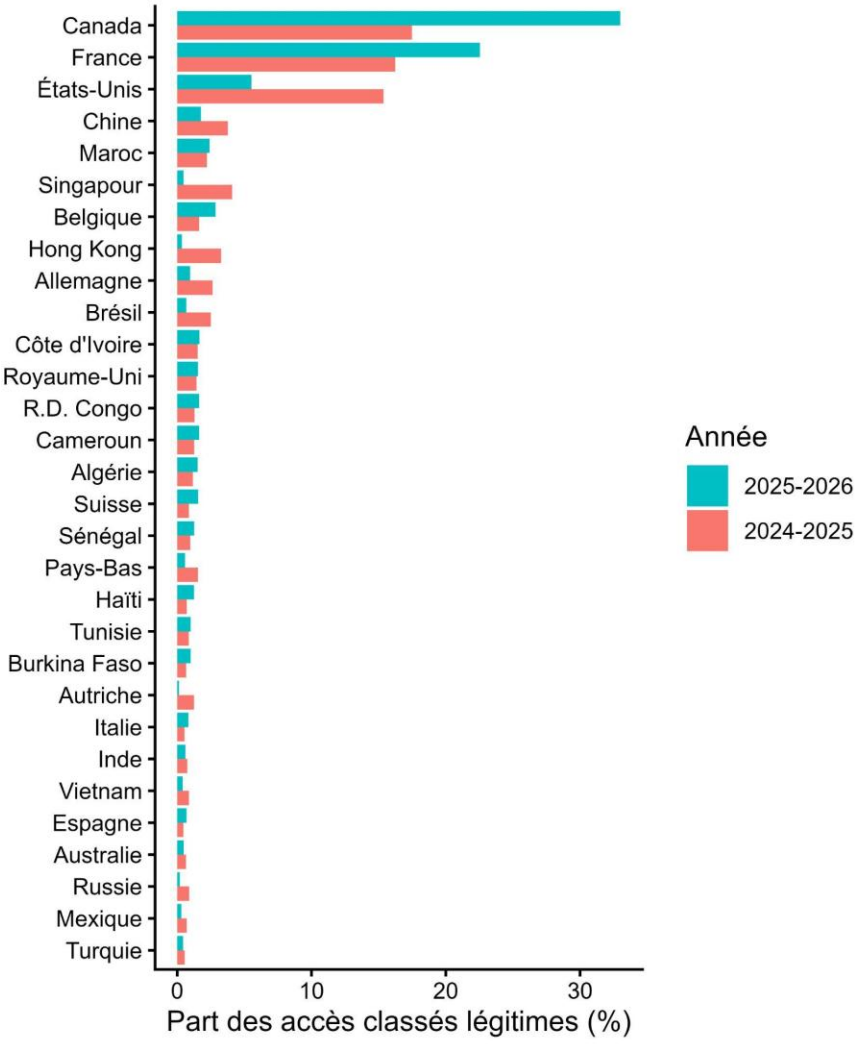


Figure 2 : Part des accès classés légitimes par rapport à la somme de l'ensemble des pays, pour les périodes 2024-2025 et 2025-2026.

De plus, il est d'intérêt de quantifier, pour chaque pays, la proportion des accès classés illégitimes par rapport à l'ensemble de leurs accès. Ainsi, on remarque que les pays ayant de fortes proportions d'accès classés illégitimes (Figure 3) correspondent aux pays montrant de fortes baisses dans les accès classés légitimes entre 2024-2025 et 2025-2026 (Figure 2), soit Singapour, la Chine et le Brésil.

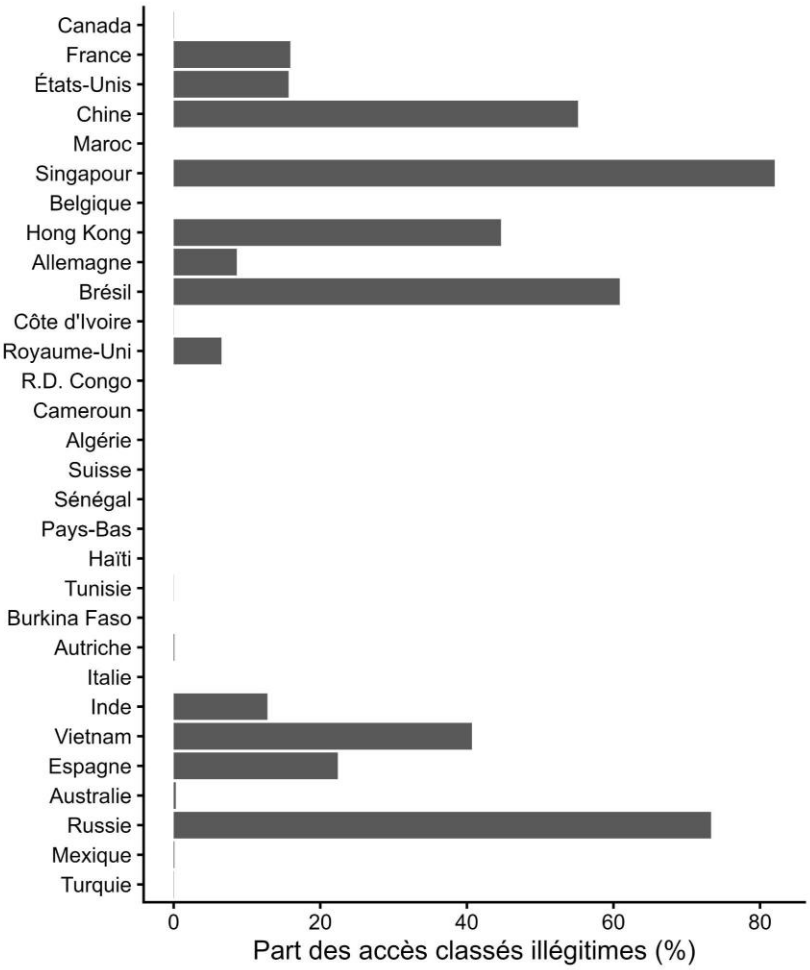


Figure 3 : Part des accès classés illégitimes par rapport au total par pays, en 2024-2025.

La part des accès illégitimes par pays n'est pas le seul indicateur de la nature des accès. En affichant les consultations par pays à travers le temps, on remarque des structures particulières. Alors que certains pays affichent une stabilité relative dans leurs accès, parfois avec une cyclicité hebdomadaire, d'autres montrent des niveaux d'accès très variables, en alternance avec des pics marqués (Figure 4).

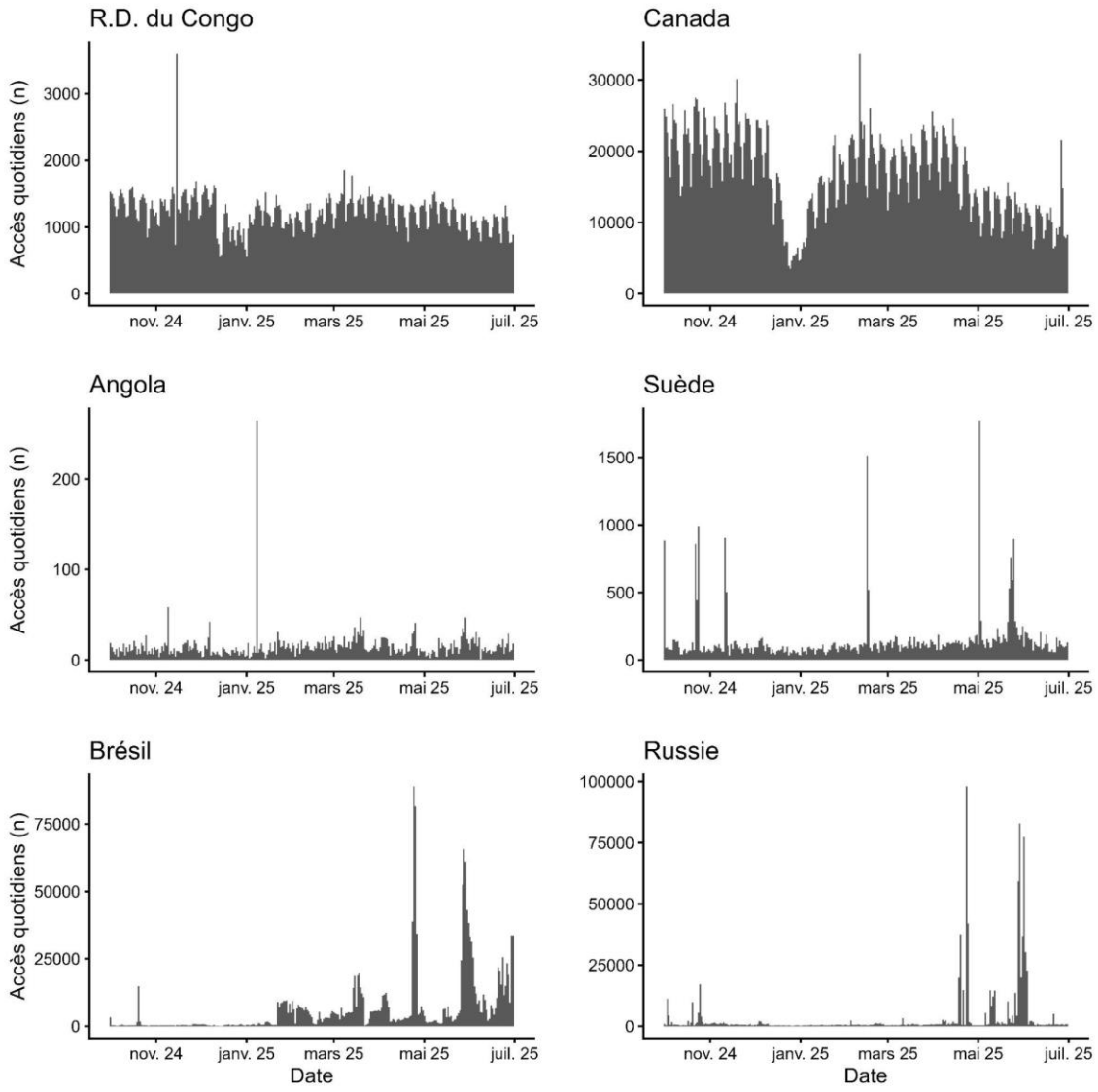


Figure 4 : Nombres d'accès quotidiens enregistrés pour six pays, 2024-2025. Les valeurs incluent les accès classés légitimes et illégitimes.

En postulant qu'un usage humain exclusif génère une série temporelle relativement stable – avec des baisses associées aux périodes de faible activité scolaire, par exemple, lors de l'été, ou à la fin de décembre – il devient possible de préciser les pays ayant une faible activité de robots. Ce profil est visible, par exemple, auprès du R.D. du Congo et du Canada (Figure 4). L'Angola et la Suède montrent une certaine stabilité des accès au cours de l'année, quoiqu'on y aperçoive une présence de journées ponctuelles ayant des valeurs nettement plus élevées. Finalement, le Brésil et la Russie enregistrent des valeurs très variables au cours de l'année, incluant des pics très marqués.

Évolution différenciée selon les pays

L'intégration de l'information du profil géographique des consultations permet d'identifier un regroupement de pays ayant très peu ou pas d'accès illégitimes à la fois pour 2024-2025 et pour 2025-2026. En les regroupant, il est possible de préciser la baisse réelle des accès légitimes. L'identification de ces pays est basée sur deux indicateurs : la proportion d'accès classés illégitimes enregistrés en 2024-2025 et la structure de la série temporelle des accès. Une heuristique mathématique a permis de classifier les pays. Cette heuristique intègre plusieurs paramètres, calculés pour chaque pays. Ensemble, ces paramètres devaient assurer que chaque série présente un volume d'accès significatif, n'a pas une valeur maximale (pic) irréaliste et n'a pas un trop grand nombre de valeurs maximales aberrantes. Spécifiquement, les paramètres étaient 1) le nombre total d'accès (>1000 pour inclusion); 2) le ratio de la valeur maximale de la série et la somme de la série (<0.017 pour inclusion) et 3) le nombre de valeurs aberrantes de la série (<12 pour inclusion). De plus, pour qu'un pays soit classé parmi les pays de référence, la proportion d'accès classés illégitimes par rapport à l'ensemble des accès devait être en dessous de 2%. Dès qu'un paramètre dépasse la valeur référence, le pays est exclu pour la suite de l'analyse.

Ainsi, 34 pays ont été identifiés comme ayant des séries d'accès très peu ou pas affectées par des accès illégitimes. Les pays les plus importants en termes de volume d'accès sont le Canada, le Maroc, la Belgique et la Côte d'Ivoire. Fait intéressant, à quelques exceptions près (la Grèce, le Nigéria, le Sri Lanka, la Tanzanie et le Zimbabwe), tous les pays sont francophones ou hébergent de grandes communautés francophones. Ensemble, ils ont accumulé 8,7 millions d'accès classés légitimes en 2024-2025.

D'une uniformité assez remarquable, tous ces pays ont enregistré une baisse des accès entre 2024-2025 et 2025-2026, allant de -25% pour le Canada à -63% pour le Nigéria; la valeur médiane est de -49%. Lorsque pondérée par le nombre d'accès, la baisse de l'ensemble de ces 34 pays est de -36%. Comme la construction des groupes le laissait attendre, la baisse des 201 pays ayant des profils caractérisés par une activité robotique est nettement plus élevée, à -73%.

Puisque les pays ayant une faible présence d'accès illégitimes semblent présenter des profils d'usage exclusivement humain, nos données rendent peu plausible que la baisse d'accès de -36% soit associée à un nettoyage déficient en 2024-2025; elle devrait plutôt refléter une baisse réelle des accès. Autrement dit, l'écart entre le -36% des pays n'étant pas associés aux accès illégitimes et le -73% des pays associés aux accès illégitimes serait attribuable au nettoyage incomplet des accès pour ces derniers.

En appliquant la baisse de -36% aux pays associés aux accès classés illégitimes, nous pouvons obtenir une nouvelle estimation des accès légitimes en 2024-2025 pour l'ensemble des pays. Ainsi, les 17,3 millions d'accès classés légitimes pour les pays associés à des niveaux d'accès illégitimes sont corrigés à la baisse pour atteindre 7,3 millions. En y ajoutant les 8,7 millions d'accès des autres pays, la nouvelle estimation s'élève à 16,0 millions.

Les institutions abonnées et partenaires

Sur Érudit, 18% du nombre total d'accès légitimes est associé à une institution abonnée ou partenaire, qui est identifiée en fonction d'un répertoire d'adresses IP. Ce pourcentage est sans doute une sous-estimation du total, parce qu'un usager ou une usagère d'une institution abonnée pourrait naviguer sur Érudit, par exemple à partir de son domicile, en omettant le proxy de son institution. Puisque près de 99% du contenu sur la plateforme est librement accessible, les chances sont minces de rencontrer une barrière.

Bien qu'il s'agisse d'une minorité des consultations, les accès par les abonnés sont d'intérêt pour évaluer un possible changement de comportement des internautes. D'abord, il s'agit d'un groupe d'institutions dont la population est relativement stable et dont l'identification est fiable, et ses accès sont par définition légitimes. Un tour d'horizon rapide démontre que peu d'universités canadiennes ont à ce jour acquis des outils spécifiques à la recherche documentaire basés sur l'IA; cependant, l'usage de l'IA générative est bien établi parmi les communautés de recherche (Crescenzi et al., 2026). Un réel changement du comportement en matière de recherche documentaire devrait se refléter dans les accès sur Érudit. De plus, les outils de l'IA générative étant très accessibles, on s'attendrait à observer des baisses de consultation pour l'ensemble des institutions abonnées.

Or, le nombre d'accès par abonnement a diminué assez modestement, à -4,4% par rapport à l'année précédente, soit d'environ 2,4 millions à 2,3 millions. De plus, les valeurs varient selon l'institution : elle est -8,7% pour l'Université de Montréal, -8,1% pour l'Université de Moncton, alors que d'autres universités montrent une hausse, comme l'UQAM, à +30,9% et l'Université Laval, à +14,3%.

Ces chiffres suggèrent qu'un éventuel effet d'usage accru de l'IA générative reste limité lorsque la personne consulte au sein de son institution. La variabilité observée entre institutions abonnées implique que cet effet n'est pas dominant parmi d'autres effets potentiels spécifiques aux institutions individuelles. Par exemple, un ajustement des paramètres dans l'outil de recherche offert par une bibliothèque peut affecter le nombre d'internautes référés aux pages d'Érudit. Ou encore, un projet de recherche ponctuel au sein d'une institution impliquant la

fouille textuelle, faisant un moissonnage complet des contenus sur Érudit – ce qui est un usage légitime selon notre définition. Ces développements ou évènements peuvent générer une croissance soudaine des accès, propre à une institution.

Discussion et conclusion

Conséquence potentielle d'un usage croissant des outils de l'IA générative, il est probable que nous observons une baisse des consultations légitimes du contenu savant sur Érudit. En cumulant les accès par institutions abonnées et partenaires et par internautes accédant sans abonnement, on obtient un total estimé de 18,4 millions accès légitimes en 2024-2025, la même période l'année suivante a enregistré un total de 12,6 millions, soit une baisse de -32% (Figure 5). Pour comprendre le comportement des internautes, il est pertinent de distinguer les accès par abonnement et ceux sans abonnement. Pour les premiers, la baisse est modeste, à environ -4%, alors que pour les derniers, il serait assez important, estimé à environ -36%. Bien que cette distinction donne l'impression d'une plus grande stabilité en matière des accès du côté des internautes ayant accès par abonnement, il est possible que les accès « sans abonnement » intègrent un certain nombre d'usagers et usagères ayant en fait une appartenance à une institution abonnée.

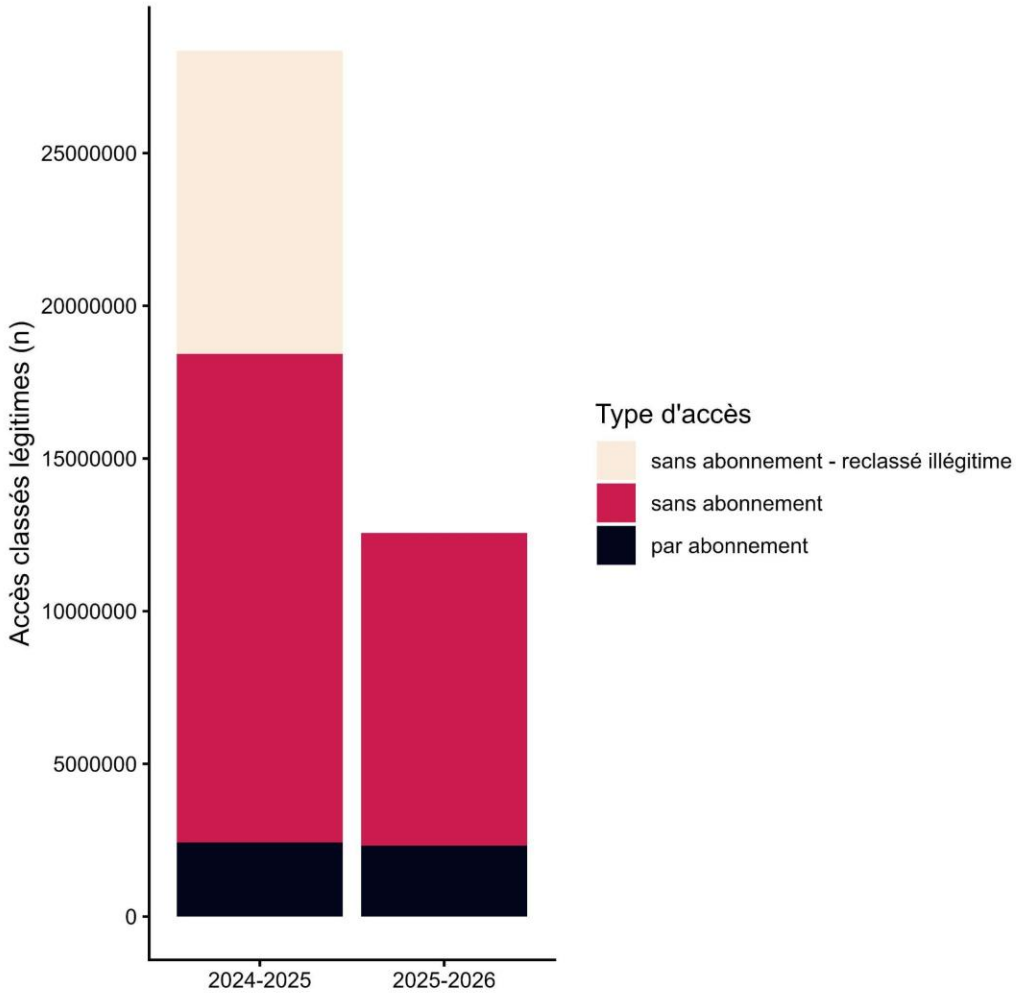


Figure 5 : Comparaison des accès classés légitimes à la suite des analyses, pour les périodes 2024-2025 et 2025-2026.

Ces estimations ont été réalisées après vérification d'une possible surestimation des accès classés légitimes effectués en 2024-2025. Leur analyse suggère qu'environ 9,9 millions d'accès illégitimes auraient été classés légitimes pendant la période 2024-2025 (Figure 5).

Cette étude présente une première estimation de la baisse réelle des accès légitimes à la plateforme. Nous soulignons que sa méthodologie a des limites. Elle est basée sur la présomption que, pour la période 2025-2026, l'identification des accès illégitimes ne présente aucun faux négatif ou faux positif. De plus, elle implique que, pour chaque pays, les parts des accès illégitimes par rapport à l'ensemble de ses accès n'ont pas changé entre 2024-2025 et 2025-2026. Aussi, l'approche employée présume que les pays associés aux bots présentent le même niveau de changement des accès humains, entre 2024-2025 et 2025-

2026, que les pays sans présence apparente de robots. Finalement, le regroupement des pays en fonction de leurs proportions d'accès illégitimes et leurs profils de séries temporelles est une approximation basée sur un jugement humain, formalisé en heuristique mathématique. Par conséquent, les valeurs présentées devraient être considérées comme une première tentative d'analyse, permettant d'établir un ordre de grandeur du phénomène.

Afin de mieux comprendre les tendances d'accès illégitimes, nous prévoyons d'analyser en détail les accès bloqués en amont par Anubis. Ces données détaillées ne sont pas disponibles actuellement. Ces futures analyses auront comme objectif d'assurer que Anubis ne bloque pas d'accès légitimes et de préciser les origines géographiques des entités responsables des accès illégitimes.

Le déploiement d'Anubis s'est avéré une grande réussite, car ce logiciel a filtré de grands nombres d'accès illégitimes depuis octobre 2025 – un travail qui aurait été impossible manuellement. Cependant, le nettoyage manuel se poursuit depuis, puisque certains accès identifiés comme étant illégitimes ont réussi à contourner Anubis dans les derniers mois et ont dû être enlevés des statistiques manuellement. L'équipe d'Érudit reste à l'affût des tentatives de contournement, et continue à surveiller et analyser les consultations de la plateforme.

Remerciements

Cette étude et le texte ont bénéficié des contributions d'Éric Araujo, Suzanne Beth, Lucía Céspedes, David Cormier, Bérenger Enselme, Gwendal Henry, Vincent Larivière, Grégory Moullec et Yves Terrat.

Bibliographie

Crescenzi, A., Choi, B., Huang, P.-P., Gautier, E., Little, R., & Pandya, S. (2026). A Multi-Year Survey of Use and Perceptions of Generative AI in Higher Education. In *Proceedings of the 2026 Conference on Human Information Interaction and Retrieval* (pp. 140–150). Association for Computing Machinery. https://doi.org/10.1145/3786304.3787867

Khosravi, M., Yoganarasimhan, H. (2026). Impact of AI Search Summaries on Website Traffic: Evidence from Google AI Overviews and Wikipedia. arXiv https://doi.org/10.48550/arXiv.2602.18455

Kwon, D. (2025). Web-scraping AI bots cause disruption for scientific databases and journals. *Nature*, *642*(8067), 281–282. https://doi.org/10.1038/d41586-025-01661-4