

Proposition d'un symposium lors du colloque de l'ADMEE 2025
« Paradoxes de l'innovation en évaluation au 21e siècle »

Coordination du symposium :

David Jan¹, Matthieu Hausman^{1 et 2} et Jean-Louis Berger¹

¹ Université de Fribourg, Suisse

² Université de Liège, Belgique

Discutant :

Franck Amadiou¹

¹Laboratoire CLLE, Université Toulouse Jean-Jaurès, France

Proposition d'un symposium lors du colloque de l'ADMEE 2025

**« Intelligence artificielle, évaluation et enseignement :
premiers constats, défis et opportunités »**

Le premier axe thématique du colloque interroge la manière dont les technologies peuvent être au service de l'évaluation. Cette question, en particulier concernant l'impact des intelligences artificielles (IA) sur les pratiques étudiantes, les pratiques évaluatives ainsi que la place de l'humain (Romero et al., 2024), est au cœur de la problématique de ce symposium.

L'utilisation des IA se généralise dans de multiples domaines ; servant par exemple à générer, trouver ou combiner texte, image ou encore vidéo. Le potentiel de cette « nouvelle » ressource est aujourd'hui largement débattu au sein de la communauté universitaire (Lo, 2023) et la question de savoir s'il faut considérer les IA comme une opportunité d'accroître l'attractivité de l'enseignement (Chocarro et al., 2023; Lepage et Collin, 2024; *New principles on use of AI in education*, 2023), l'implication des étudiants voire leurs apprentissages s'avère pour le moins épineuse (Romero et al., 2024).

Dès lors, le symposium interroge la manière dont les IA, qui promettent de changer les pratiques d'enseignement-apprentissage, mais aussi l'accès, la construction et l'évaluation des apprentissages, peuvent être évaluées comme un véritable outil au service des étudiants et des enseignants (Atlas, 2023; Giannini, 2023). Entre défi et opportunité, l'enjeu de ce symposium est de rassembler diverses études empiriques portant sur l'évaluation et l'utilisation de l'IA tant comme outil d'enseignement-apprentissage que comme partenaire de travail scientifique (Biswas et al., 2023).

Ce symposium court comptera trois présentations ainsi qu'une discussion générale. Les contributions seront commentées par un chercheur reconnu dans le champ de l'évaluation en éducation. La discussion mettra en évidence l'apport des trois contributions à l'évaluation de l'utilisation de l'IA en éducation, en veillant à instruire avec une spécificité accrue, la question de l'opportunité de l'implémentation de l'IA dans les dispositifs d'enseignement-apprentissage.

Références bibliographiques

- Biswas, S., Dobaria, D. et Cohen, H. L. (2023). ChatGPT and the Future of Journal Reviews: A Feasibility Study. *The Yale Journal of Biology and Medicine*, 96(3), 415-420. <https://doi.org/10.59249/SKDH9286>
- Chocarro, R., Cortiñas, M. et Marcos-Matás, G. (2023). Teachers' attitudes towards chatbots in education: a technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics. *Educational Studies*, 49(2), 295-313. <https://doi.org/10.1080/03055698.2020.1850426>
- Giannini, S. (2023). L'IA générative et le futur de l'éducation.
- Lepage, A. et Collin, S. (2024). Preserving Teacher and Student Agency: Insights from a Literature Review. Dans A. Urmeneta et M. Romero (dir.), *Creative Applications of Artificial Intelligence in Education* (p. 17-34). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-55272-4_2
- Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- New principles on use of AI in education*. (2023, 4 juillet). The Russell Group of Universities.
- Romero, M., Frosig, T., Taylor-Beswick, A. M. L., Laru, J., Bernasco, B., Urmeneta, A., Strutynska, O. et Girard, M.-A. (2024). Manifesto in Defence of Human-Centred Education in the Age of Artificial Intelligence. Dans A. Urmeneta et M. Romero (dir.), *Creative Applications of Artificial Intelligence in Education* (p. 157-178). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-55272-4_12

Contribution 1

Lionel Alvarez¹, Denise Sutter-Widmer², Giulia Ortoleva² et Mattia Fritz²

¹ CRE/ATE | Haute École pédagogique de Fribourg et CERF | Université de Fribourg, Suisse

² TECFA | Université de Genève, Suisse

Les IA génératives dans la formation à l'enseignement. Une technologie disruptive dans ses facteurs d'acceptation.

Résumé court

Les intelligences artificielles (IA) génératives progressent, créant des défis et des opportunités en éducation. Selon Zirar (2023), elles peuvent être utilisées judicieusement, mais nécessitent une évaluation critique et systématique des résultats. Les IA génératives étant récentes dans l'enseignement, surtout à l'école obligatoire (Tedre et al., 2021), la littérature est souvent théorique, discutant des avantages et menaces. Il est alors difficile de formuler des recommandations pour les usages scolaires. La formation à l'enseignement y est alors doublement complexe, car les usages concernent les étudiant·e·s et les futur·e·s enseignant·e·s. La présente étude examine l'acceptation des IA génératives par les enseignant·e·s en formation, dans leur double rôle (1) d'étudiant·e·s actuel·le·s et (2) de futurs professionnel·le·s, ainsi que les projections d'usages rapportées. Le GETAMEL (Abdullah et Ward, 2016) sert de modèle théorique pour comprendre les intentions d'usage. La recherche a consisté en une enquête par questionnaire (Sprenger et Schwaninger, 2023) rempli par un échantillon de 256 enseignant·e·s en formation initiale (primaire, secondaire 1 et 2) issus de six Hautes Écoles suisses romandes. Les analyses descriptives révèlent des déviations standards relativement élevées, ce qui indique une diversité dans les rapports aux IA. Pour les analyses par modèles d'équations structurelles, elles montrent peu de différences entre les résultats rapportés comme étudiant·e·s et comme futur·e·s enseignant·e·s. Les facteurs GETAMEL qui affectent l'acceptation technologique d'IA sont la norme subjective, l'auto-efficacité, le plaisir et l'utilité perçue. La quasi-disparition de la facilité d'usage perçue comme prédicteur de l'intention d'usage pourrait être expliquée par la nature de la technologie discutée et les qualités du public sondé.

Ces changements dans le GETAMEL pour les IA génératives dans la formation à l'enseignement soulignent la nécessité d'identifier les usages jugés utiles, afin de mieux préparer les enseignant·e·s en formation à intégrer les IA génératives dans leurs pratiques.

Mots-clés : intelligence artificielle générative, formation des enseignants, GETAMEL, acceptation de la technologie, professionnels en devenir

Résumé long

Les intelligences artificielles (IA) génératives ont connu des avancées rapides et significatives, posant des défis et ouvrant des opportunités dans le domaine de l'éducation et de la formation. Dans une récente revue de la littérature (Zirar, 2023), l'auteur conclut que des utilisations judicieuses des IA génératives sont possibles en éducation. Toutefois, pour chaque utilisatrice, qu'il s'agisse d'apprenant·e·s ou d'enseignant·e·s, une évaluation systématique et critique des résultats (output) est nécessaire, en se concentrant sur leur validité, leur fiabilité et leur précision. La recommandation propose aussi une utilisation limitée des IA génératives par les professionnel·le·s, par exemple lors de l'élaboration de matériel d'enseignement et d'évaluation, en soulignant l'importance de définir clairement le rôle de la technologie avant de lui déléguer des tâches. Étant donné que les technologies des IA génératives sont encore récentes dans l'enseignement et l'apprentissage, en particulier pour l'enseignement à l'école obligatoire (Tedre et al., 2021), la littérature concernée est souvent théorique et discute principalement les avantages et menaces potentiels de l'IA. Il est actuellement difficile de formuler des recommandations claires pour les usages scolaires, au-delà de la suggestion habituelle pour chaque nouvelle technologie : tester et développer un rapport critique à l'outil. La formation à l'enseignement devient alors défiante à double titre, car des étudiant·e·s apprennent potentiellement avec les IA génératives et de futur·e·s enseignant·e·s feront possiblement usage de ces technologies.

La présente étude vise à examiner l'acceptation des IA génératives par les enseignant·e·s en formation, compte tenu de leur double rôle (1) d'étudiant·e·s actuel·le·s et (2) de futurs professionnel·le·s, ainsi que les projections d'usages rapportées. Le GETAMEL (Abdullah et Ward, 2016), extension du *technology-acceptance model* (Davis, 1989; Venkatesh, 2022) ajoutant des facteurs externes, sert de modèle théorique pour comprendre l'intention d'usage des IA génératives par ce public. Selon ce modèle, l'intention de faire usage (*intention of use*) d'une technologie dépend de l'utilité perçue (*perceived usefulness*) et de la facilité d'usage perçue (*perceived ease of use*), et ces facteurs sont influencés par les normes subjectives (*subjective norm*), l'auto-efficacité (*self-efficacy*), le plaisir (*enjoyment*), l'angoisse face aux ordinateurs (*computer anxiety*), et l'expérience d'utilisation (*experience*). La recherche a consisté en une enquête par questionnaire emprunté à une étude au contexte similaire (Sprenger et Schwaninger, 2023), alors envoyé à un échantillon de 256 enseignant·e·s (dont 78 % de femmes) en formation initiale (pour l'enseignement au primaire, au secondaire 1 et au secondaire 2) issu de six Universités et Hautes Écoles pédagogiques situées dans la région francophone de la Suisse. À noter, les items associés à l'auto-efficacité, à l'utilité perçue et à l'intention d'usage, ont été questionnés deux fois, d'abord dans un contexte d'usage des IA génératives comme étudiant·e·s, puis dans un contexte des usages comme enseignant·e·s. Des données socio-démographiques ont aussi été collectées, telles que le genre ou le degré d'études. Les informations recueillies auprès de ces participant·e·s fournissent un cadre précieux pour comprendre l'acceptation des IA génératives par ce public, et pour interroger le GETAMEL à la lumière des caractéristiques distinctives de cette technologie.

Les analyses descriptives révèlent des déviations standards relativement élevées pour chaque facteur, ce qui indique une diversité importante dans les rapports aux IA génératives parmi les personnes sondées. Cette diversité est particulièrement marquée en ce qui concerne l'intention d'utilisation des IA génératives. L'hypothèse de l'expérience d'utilisation pour expliquer cette dispersion ne semble pas légitime face aux données collectées. Il semble que le besoin de comprendre et connaître les potentialités des IA génératives soit plus essentiel. Pour les analyses effectuées à l'aide de modèles d'équations structurelles, elles ne montrent que peu de

différences entre les résultats rapportés par les étudiant·e·s et ceux rapportés par les enseignant·e·s. Cependant, une différence notable émerge en ce qui concerne l'influence du facteur auto-efficacité sur la facilité d'usage perçue, cette influence étant distincte entre les deux groupes de répondants.

Selon nos résultats, les facteurs GETAMEL qui affectent l'acceptation technologique d'IA génératives par les futur·e·s enseignant·e·s sont la norme subjective, l'auto-efficacité, le plaisir et l'utilité perçue. L'anxiété informatique, l'expérience et la facilité d'utilisation perçue n'influencent pas l'intention d'utiliser les IA génératives. Si le niveau d'étude n'a pas d'impact sur l'acceptation des IA génératives, les disparités entre les sexes sont importantes pour la satisfaction, l'expérience et les normes subjectives, les hommes étant plus enclins à utiliser cette technologie. Les données qualitatives documentent une grande ambivalence dans les rapports que ce public entretient avec cette technologie parfois annoncée comme la solution tant attendue pour un enseignement centré sur l'élève (Devi et al., 2022).

La quasi-disparition de la facilité d'usage perçue comme prédicteur de l'intention d'usage pourrait possiblement être expliquée par la nature de la technologie discutée et les qualités du public sondé. En effet, les IA génératives fonctionnent prioritairement à l'aide d'instruction écrite (du moins, lors de la passation du questionnaire), alors que le public interrogé est universitaire et se destine à l'enseignement. Il s'agit alors de gens à priori à l'aise avec le texte.

Ces changements dans le GETAMEL pour les IA génératives dans la formation à l'enseignement soulignent la nécessité pour les programmes d'identifier les usages jugés utiles (vs. formation à la manipulation), afin de mieux préparer les enseignant·e·s en formation à intégrer les IA génératives dans leurs pratiques professionnelles. Quelques recommandations théoriques, associées aux propositions des personnes interrogées, permettent d'ores et déjà d'orienter les recommandations (par ex., à destination de l'institution, de l'enseignant·e, ou de l'élève). Au fur et à mesure que les technologies d'IA génératives évoluent, la poursuite des travaux de recherche permettra d'affiner notre compréhension de ces technologies en contexte éducatif et d'élaborer des stratégies efficaces pour clarifier l'adoption des IA en formation à l'enseignement.

Références bibliographiques

- Abdullah, F. et Ward, R. (2016). Developing a General Extended Technology Acceptance Model for E-Learning (GETAMEL) by analysing commonly used external factors. *Computers in Human Behavior*, 56, 238-256. <https://doi.org/10.1016/j.chb.2015.11.036>
- Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319-340.
- Devi, S., Sreedhar, B., Aralprakash, P., Kazi, K. et Radhakrishnan, R. (2022). A Path Towards Child-Centric Artificial Intelligence based Education. *International Journal of Early Childhood Special Education*, 14(3).
- Sprenger, D. A. et Schwaninger, A. (2023). Video demonstrations can predict the intention to use digital learning technologies. *British Journal of Educational Technology*, bjet.13298. <https://doi.org/10.1111/bjet.13298>
- Tedre, M., Toivonen, T., Kahila, J., Vartiainen, H., Valtonen, T., Jormanainen, I. et Pears, A. (2021). Teaching Machine Learning in K–12 Classroom: Pedagogical and

Proposition d'un symposium lors du colloque de l'ADMEE 2025
« Paradoxes de l'innovation en évaluation au 21e siècle »

- Technological Trajectories for Artificial Intelligence Education. *IEEE Access*, 9, 110558-110572. <https://doi.org/10.1109/ACCESS.2021.3097962>
- Venkatesh, V. (2022). Adoption and use of AI tools: a research agenda grounded in UTAUT. *Annals of Operations Research*, 308(1-2), 641-652. <https://doi.org/10.1007/s10479-020-03918-9>
- Zirar, A. (2023). Exploring the impact of language models, such as ChatGPT, on student learning and assessment. *Review of Education*, 11(3), e3433. <https://doi.org/10.1002/rev3.3433>

Contribution 2

Naomie Fournier Dubé¹ et Hébert, Marie-Hélène²

¹ Université de Montréal, Canada

² Université TÉLUQ, Canada

IA et évaluation : l'agent conversationnel ChatGPT peut-il réussir les activités d'évaluation d'un cours universitaire ?

Résumé court

Selon plusieurs études, la pression de réussir peut inciter les étudiants à tricher aux activités d'évaluation (McCabe et al., 2012), entraînant ainsi des questionnements sur la validité de leurs résultats académiques et sur la crédibilité des diplômes qui leur sont décernés au terme de leur programme d'études (Fendler et al., 2018).

Avec la montée en puissance de l'intelligence artificielle, il y a lieu de se questionner : dans quelle mesure un agent conversationnel comme ChatGPT peut-il réussir les activités d'évaluation d'un cours universitaire et, partant, être utilisé par des étudiants pour tricher ?

Professeures en évaluation des apprentissages dans deux universités québécoises, c'est dans ce contexte que nous avons testé la capacité de l'agent conversationnel ChatGPT à réussir les activités d'évaluation qui servent à évaluer les apprentissages des étudiants dans deux cours que nous offrons.

Partant d'un échantillon de 9 activités d'évaluation, de la plus simple à la plus complexe, nous avons administré chacune d'elles à ChatGPT avec l'objectif d'en apprécier ses réponses au moyen des mêmes outils de jugement (solutionnaires et grilles d'évaluation) que nous utilisons pour apprécier les performances de nos étudiants. Si ChatGPT a su réussir à la perfection les tâches simples de mémorisation et d'application, il a en revanche échoué à donner des réponses de qualité à des tâches de plus en plus complexes.

Dans le cadre de la présente communication, nous ferons voir aux participants les réponses données par l'agent conversationnel aux 9 activités d'évaluation que nous lui avons administrées, en plus de réfléchir aux impacts d'un tel outil sur les pratiques évaluatives et à son intégration de manière éthique à l'évaluation. Nous concluons la présentation en présentant quelques spécimens d'activités d'évaluation où les occasions de tricherie avec ChatGPT se révèlent limitées, voire impossibles.

Mots-clés : Plagiat et tricherie, Intelligence artificielle, Pratiques évaluatives

Résumé long

Selon plusieurs études, la pression de réussir peut inciter les étudiants à tricher aux activités d'évaluation (McCabe et al., 2012), entraînant ainsi des questionnements sur la validité de leurs résultats académiques et sur la crédibilité des diplômes qui leur sont décernés au terme de leur programme d'études (Fendler et al., 2018). Avec la montée en puissance de l'intelligence artificielle, il y a lieu pour le personnel enseignant et les établissements d'enseignement supérieur de se questionner : dans quelle mesure un agent conversationnel comme ChatGPT peut-il réussir les activités d'évaluation d'un cours universitaire et, partant, être utilisé par des étudiants pour tricher ?

Professeures en évaluation des apprentissages dans deux universités québécoises, c'est dans ce contexte que nous avons testé la capacité de l'agent conversationnel ChatGPT à réussir les activités d'évaluation qui servent à évaluer les apprentissages des étudiants dans des cours que nous offrons. Les cours en question en sont deux sur le thème de l'évaluation des apprentissages en salle de classe, dans lesquels nous formons les futurs enseignants de l'éducation préscolaire (4-5 ans), du primaire (6-12 ans) et du secondaire (13-17 ans) à l'évaluation. Ils visent à développer chez eux la compétence à évaluer les apprentissages (MEQ, 2020, p. 58-59) des élèves à qui ils enseignent, en les formant notamment à :

- « Concevoir ou choisir des instruments ou des modalités d'évaluation qui s'appuient sur les programmes d'études et permettent de vérifier les apprentissages effectués par les élèves. »
- « Repérer les forces ainsi que les défis des élèves et prévoir des interventions appropriées qui favoriseront leurs apprentissages. »
- « Utiliser l'information récoltée par l'entremise des différentes modalités d'évaluation pour avoir une vue d'ensemble de son groupe-classe et planifier son enseignement en conséquence. »
- « Donner des rétroactions fréquentes et constructives à l'élève pour lui permettre de se situer dans ses apprentissages et soutenir sa progression. »

Partant d'un échantillon de 9 activités d'évaluation tirées de nos deux cours, de la plus simple à la plus complexe, nous avons administré chacune d'elles à ChatGPT avec l'objectif d'en apprécier ses réponses au moyen des mêmes outils de jugement (solutionnaires et grilles d'évaluation) que nous utilisons pour apprécier les performances de nos étudiants. Parmi ces 9 activités d'évaluation se trouvent les tâches suivantes :

- 1) Indiquer s'il est vrai ou faux d'affirmer que, pour évaluer une compétence, il n'est pas obligatoire de recourir à plusieurs activités d'évaluation appartenant à une même famille.
- 2) Indiquer à quelle fonction de l'évaluation un énoncé donné fait référence parmi l'aide à l'apprentissage et la reconnaissance des compétences.

Proposition d'un symposium lors du colloque de l'ADMEE 2025
« Paradoxes de l'innovation en évaluation au 21e siècle »

- 3) Indiquer si une activité d'évaluation donnée est de niveau « connaissance », de niveau « habileté », de niveau « stratégie » ou de niveau « savoir-être ».
- 4) Nommer les fonctions que le ministère de l'Éducation du Québec attribue à l'évaluation des apprentissages.
- 5) Concevoir un examen et son solutionnaire.
- 6) Concevoir une tâche complexe et sa grille d'évaluation.
- 7) Réaliser une capsule vidéo d'une durée de 120 à 240 secondes s'adressant aux élèves d'une classe pour expliquer l'évaluation : à quoi sert-elle ?
- 8) Administrer auprès d'un groupe d'élèves une activité d'évaluation.
- 9) Analyser la mise en œuvre d'une activité d'évaluation vécue ou observée dans une classe, à la lumière des directives ministérielles québécoises et des considérations théoriques abordées dans le cours, afin de déterminer dans quelle mesure cette activité d'évaluation correspond ou non aux attentes ministérielles et aux considérations théoriques.

Si ChatGPT a su réussir à la perfection les tâches simples de mémorisation et d'application (tâches 1 à 4 ci-dessus), il a en revanche échoué à donner des réponses de qualité à des tâches complexes (tâches 7 à 9 ci-dessus). Pour les activités d'évaluation de niveau de complexité moyen (tâches 5 et 6 ci-dessus), il a su atteindre la note de passage du cours (60%) après plusieurs tentatives, c'est-à-dire après lui avoir demandé de perfectionner sa réponse à maintes reprises au moyen de « prompts » soigneusement rédigés.

Dans le cadre de la présente communication, nous ferons voir aux participants les réponses données par l'agent conversationnel aux 9 activités d'évaluation que nous lui avons administrées, en plus de réfléchir aux impacts d'un tel outil sur les pratiques évaluatives et à son intégration de manière éthique à l'évaluation. Nous concluons la présentation en présentant quelques spécimens d'activités d'évaluation où les occasions de tricherie avec ChatGPT se révèlent limitées, voire impossibles comme :

- Demander aux étudiants de créer un tutoriel de type « TikTok » pour informer sur une stratégie, une tâche ou une méthode efficace d'évaluation formative en classe.
- Demander aux étudiants de concevoir une tâche diagnostique d'une durée de 60 minutes, ainsi que son outil de consignation, afin de documenter les forces et les besoins des élèves, en vue de planifier et de mettre en œuvre des pratiques de flexibilité pédagogique (contenu, production, structure, processus) dans leur classe au quotidien.
- Demander aux étudiants d'enseigner à quelqu'un (ex. : nièce, ami, étudiant du cours) une connaissance devant être apprise à l'école, puis de faire une vidéo de rétroaction avec une application de leur choix (ex. : Explain Everything) basée sur les principes d'une rétroaction efficace.
- Demander aux étudiants de réaliser une entrevue d'une durée de 25 à 45 minutes avec un enseignant afin de déterminer ses besoins en matière d'évaluation, résumer l'entrevue, puis créer un module de formation continue à l'aide d'une application le permettant (ex. : Genially) pour pallier l'un de ses besoins.

Références bibliographiques

- Fendler, R. J., Yates, M. C. et Godbey, J. M. (2018). Observing and deterring social cheating on college exams.
International Journal for the Scholarship of Teaching and Learning, 12(1). doi :
10.20429/ijstl.2018.120104
- Ministère de l'Éducation du Québec (2020). *Référentiel de compétences professionnelles : profession enseignante*.
Gouvernement du Québec.
- McCabe, D. L., Butterfield, K. D. et Trevino, L. K. (2012). *Cheating in college : why students do it and what educators can do about it*. JHU Press.

Contribution 3

David Jan¹, Matthieu Hausman^{1 et 2} et Jean-Louis Berger¹

¹ Université de Fribourg, Suisse

² Université de Liège, Belgique

L'évaluation de recherche assistée par intelligence artificielle

Résumé court

L'usage des systèmes d'intelligence artificielle (IA) dans la recherche académique, notamment pour la recherche de littérature scientifique, est en pleine expansion. Ces systèmes, tels que ChatGPT et elicit.com, offrent une exploration plus large et flexible des sources scientifiques grâce à leurs capacités avancées d'analyse et de synthèse. Dans notre étude, des étudiants de master en sciences de l'éducation ont utilisé elicit.com pour rechercher des sources pertinentes afin de réfuter des mythes pédagogiques. Par exemple, ils ont exploré la question de savoir si la réussite scolaire dépend exclusivement de l'intelligence, obtenant des résumés basés sur plusieurs sources scientifiques.

Notre recherche a évalué l'utilisation de cet outil par les étudiants, leur perception de son utilité pour enrichir leur production écrite, et leur rapport aux méthodes traditionnelles de recherche documentaire. Les résultats montrent que 72,73% des étudiants recommandent l'utilisation d'Elicit, appréciant sa rapidité et la rédaction en langage naturel. Cependant, 27,27% ne le recommandent pas, critiquant la qualité et la pertinence des informations fournies, ainsi que le nombre limité de références francophones.

L'expérience des étudiants avec l'IA a été mitigée, suscitant des émotions variées comme la curiosité et la confusion, démontrant un engagement cognitif et affectif. Les tests ont révélé qu'il n'y a pas de lien significatif entre la recommandation de l'IA et les compétences auto-évaluées en informatique ou en recherche de références scientifiques. Cependant, la recommandation est significativement liée à l'évaluation de la qualité, la rapidité et la facilité d'utilisation de l'outil.

Ces résultats soulignent l'importance de la perception de l'outil plutôt que des compétences techniques pour recommander l'IA. À l'avenir, nous envisageons d'explorer si une formation spécifique à l'utilisation de l'IA pourrait améliorer l'expérience des étudiants et leurs compétences perçues.

Mots-clés : Recherche documentaire, Intelligence artificielle, Littérature scientifique, Esprit critique, Pédagogie, Enseignement supérieur

Résumé long

Selon l'Organisation de coopération et de développement économiques (OCDE, 2024), un système d'intelligence artificielle (IA) est un système automatisé générant certains résultats suite à une demande qui lui a été préalablement adressée en langage naturel. De multiples systèmes de ce type ont vu le jour ces dernières années. Qui, aujourd'hui, ignore encore l'existence de ChatGPT ? Ces outils ont pour principal atout le fait qu'ils permettent la génération rapide de multiples contenus. En contexte universitaire, une des utilisations de l'IA est la recherche de sources scientifiques en permettant *a priori* une exploration plus large, flexible et « intelligente » des sources scientifiques qu'avec des méthodes de recherche documentaire plus « classiques », soutenue par les capacités d'analyse et de synthèse plus avancées offertes par l'IA que les bases de données traditionnelles.

Dans la communauté scientifique, nous trouvons des recherches sur les méthodes (Lo, 2023; Wang, 2023), limites (Walters et Wilder, 2023), avantages (Enomoto et al., 2023) ou questions (Bridges et al., 2024) que les chercheurs ont à se poser quant à l'implémentation de l'IA dans les pratiques pédagogiques. Ces études partent généralement du postulat que ces chercheurs ont une expertise et une méthodologie en recherche de littérature scientifique, qui doit alors leur permettre de discriminer avec une certaine aisance les résultats probants et les autres, parmi les propositions qui leur sont adressées par l'IA avec laquelle ils interagissent.

Dans notre étude, des étudiants de master en sciences de l'éducation de l'Université de Fribourg, en Suisse, ont exploré et découvert la recherche de littérature scientifique, en regard d'un thème spécifique leur ayant été préalablement attribué. Ils ont eu à utiliser un moteur de recherche assisté par IA (elicit.com) afin d'obtenir des sources scientifiques pertinentes leur servant à invalider des fausses croyances ou « mythes » portant sur des pratiques pédagogiques, largement répandues. Notre recherche porte sur l'évaluation que ces étudiants ont faite de ce système d'IA, leur ayant permis d'entreprendre une recherche par langage naturel. Par exemple, l'un des groupes d'étudiants a interrogé la plateforme afin d'identifier des sources permettant de traiter la question de savoir si : « La réussite scolaire dépend exclusivement de l'intelligence ? », et offrant d'emblée un résumé sous la forme d'un texte d'une centaine de mots, fondé sur quatre sources scientifiques.

Notre étude examine (1) la manière dont les étudiants ont utilisé l'outil, (2) la façon dont ils ont jugé l'utilité du système d'IA pour alimenter et améliorer leur production écrite, ainsi que (3) leur rapport à l'IA vis-à-vis des méthodes de recherche de littérature scientifique.

En termes d'appréciation d'Elicit dans le cadre de la tâche qu'il leur a été demandé de réaliser (identifier des références pertinentes dans le but de contredire un mythe pédagogique), il ressort de cette expérience que 16 étudiants sur 22 (72,73%) recommandent l'usage d'Elicit, alors que 6 (27,27%) ne le recommandent pas. Les aspects ergonomiques ou d'utilisation d'Elicit (ex. : rapidité, rédaction en langage naturel, accès aux références proposées) semblent majoritairement expliquer l'attrait que présente cet outil chez les étudiants. À l'inverse, la qualité de l'information proposée par la plateforme (ex. : fiabilité, pertinence) est au centre des critiques que les étudiants formulent à l'encontre d'Elicit. De façon plus détaillée, nous avons dénombré 107 unités de sens parmi les réponses des 22 répondants aux questions de savoir quels étaient les avantages et inconvénients d'Elicit. Parmi celles-ci, 60 unités de sens (56,08%) portaient

sur les avantages attribués à Elicit (ex. : génération de résumés, références principalement scientifiques), tandis que 47 unités de sens (43,92%) portaient sur les inconvénients d'Elicit (ex. : crédit d'utilisation gratuite limité, peu de références francophones). Cette tendance apparaît par ailleurs plutôt congruente avec l'avis général de nos utilisateurs, qui se veut plutôt favorable à l'usage d'Elicit dans la recherche de références scientifiques, sans pour autant plébisciter cet outil.

Sur le plan expérientiel, les étudiants ont vécu l'utilisation de l'IA d'une manière que nous qualifions de mitigée. En effet, ayant interrogé ceux-ci sur leur ressenti lors de l'usage de l'outil, plusieurs ont exprimé des émotions agréables (n=20), mais presque autant d'émotions désagréables (n=17). En outre, les étudiants eux-mêmes semblent avoir vécu cette expérience de façon partagée, puisque 12 répondants sur 22 ont déclaré avec ressenti plusieurs émotions de valences différentes, de manière concomitante. Il y a donc "du pour et du contre", pour le dire très simplement. Relevons que parmi les émotions évoquées par les étudiants, des émotions épistémiques telles que la curiosité ou la confusion attestent de la réflexion suscitée par l'usage de l'IA dans l'examen de la pertinence des sources proposées, notamment. Les étudiants s'engagent donc à la fois cognitivement et affectivement dans cette tâche, selon ces modalités inédites pour eux.

Des tests non paramétriques ne révèlent aucun lien significatif entre le fait de recommander ou de ne pas recommander l'utilisation de l'IA et trois indicateurs auto-rapportés du sentiment de compétence (informatique, utilisation de l'IA et recherche de références scientifiques). Cependant, en utilisant un facteur ($\omega = .70$; 3 items) construit selon la qualité, la rapidité et la facilité d'utilisation de l'IA (Chocarro et al., 2023), nous obtenons un résultat significatif ($U = 13.00$; $p < .05$) avec le fait de recommander l'utilisation de l'IA. Nous voyons par ces résultats que les étudiants qui recommandent l'utilisation de l'IA le font sur base de l'évaluation qu'ils ont de cet outil et non selon les trois indicateurs de leurs compétences.

En termes de perspectives, la question de savoir si un entraînement spécifique à l'usage de cette IA serait susceptible d'impacter positivement à la fois la qualité de leur expérience et les indicateurs de compétence des étudiants, suscite tout particulièrement notre intérêt.

Références bibliographiques

- Bridges, L. M., McElroy, K. et Welhouse, Z. (2024). Generative Artificial Intelligence: 8 Critical Questions for Libraries. *Journal of Library Administration*, 64(1), 66-79. <https://doi.org/10.1080/01930826.2024.2292484>
- Chocarro, R., Cortiñas, M. et Marcos-Matás, G. (2023). Teachers' attitudes towards chatbots in education: a technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics. *Educational Studies*, 49(2), 295-313. <https://doi.org/10.1080/03055698.2020.1850426>
- Enomoto, M., Tseng, C.-H., Hsu, Y.-C., Thuy, L. T. T. et Nguyen, M. H. (2023). Collaborating with AI in literature search—An important frontier. *Hepatology Communications*, 7(12). <https://doi.org/10.1097/HC9.0000000000000336>

Proposition d'un symposium lors du colloque de l'ADMEE 2025
« Paradoxes de l'innovation en évaluation au 21e siècle »

- Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, 49(4), 102720. <https://doi.org/10.1016/j.acalib.2023.102720>
- OCDE. Recommandation du Conseil sur l'intelligence artificielle, OECD/LEGAL/0449 (2024).
- Walters, W. H. et Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, 13(1), 14045. <https://doi.org/10.1038/s41598-023-41032-5>
- Wang, M. (2023). Research on Text Classification Method Based on NLP. *Advances in Computer, Signals and Systems*, 7(2). <https://doi.org/10.23977/acss.2023.070213>