

Outils d'intelligence artificielle dans le périmètre des bibliothèques des universités de la FWB — étude comparative

Rapport réalisé à la demande du CA de la BICfB (février 2026)

Auteur-es : Vincent BRUNIN (UMONS), François FRÉDÉRIC (ULB), Élodie MERCY (UNamur), François RENAVILLE (ULiège), Bérénice VANHUFFLEN (UCLouvain)

Certaines formulations de ce rapport ont été assistées par des IA génératives (janvier 2026).

Ce document est sous licence [Creative Commons CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/).

Table des matières

Introduction	2
Analyse	4
Méthodologie	4
Synthèse par outils	9
Consensus	9
Elicit	10
Scholar GPT	11
SciSpace	12
Scite	13
Undermind	13
Pour conclure	15
D'autres modalités de choix d'outils	15
Risque de détournement de l'Open Access	16
Enjeux internationaux	16
Annexes	17
Consensus	17
Elicit	21
Scholar GPT	25
SciSpace	32
Scite	37
Undermind	41



Introduction

À la demande du Conseil d'administration de la BICfB, un groupe de travail interuniversitaire intitulé « Outils d'intelligence artificielle dans le périmètre des bibliothèques des universités de la Fédération Wallonie-Bruxelles » a été constitué au printemps 2025 afin d'examiner le potentiel d'outils d'IA appliqués à la recherche documentaire. La lettre de mission confiée au groupe précisait un double objectif : d'une part, identifier des solutions pertinentes au regard des besoins des communautés académiques ; d'autre part, formuler des recommandations destinées à éclairer les bibliothèques universitaires dans leurs choix futurs, tant en matière d'accompagnement que, le cas échéant, d'abonnements. Il était également attendu que le groupe adopte une méthode de test partagée, coordonne les évaluations entre institutions et produise un rapport synthétique, fondé sur des constats factuels plutôt que sur une analyse exhaustive du marché, eu égard au rythme de proposition de nouveaux outils et d'évolution des outils existants.

La durée initialement prévue pour la mission était de neuf mois à compter de son lancement. La réunion de démarrage organisée par la BICfB s'est tenue le 11 avril 2025. Le groupe de travail s'est ensuite réuni à treize reprises entre avril 2025 et février 2026, aux dates suivantes : 23 avril, 12 mai, 6 juin, 19 juin, 9 juillet, 20 août, 24 septembre, 20 octobre, 5 novembre, 2 décembre 2025, ainsi que le 19 janvier et les 2 et 6 février 2026.

Le groupe de travail a procédé à une sélection raisonnée de quelques outils d'IA centrés sur l'assistance à la recherche documentaire et à l'exploration de la littérature scientifique. Le choix s'est porté sur six solutions : **Consensus**, **Elicit**, **ScholarGPT**, **SciSpace**, **Scite** et **Undermind**. Ces outils ont en commun de proposer des fonctionnalités telles que la formulation assistée de questions de recherche, l'identification et la synthèse d'articles scientifiques, l'analyse des citations ou encore la contextualisation des résultats.

Signalons que **Perplexity AI** a un temps été envisagé comme solution à évaluer par le groupe. Toutefois, au regard de diverses critiques et polémiques concernant ses pratiques de collecte et de traitement des données¹ (notamment l'absence de paramètres forts d'anonymisation ou de chiffrement de bout en bout, l'utilisation de données de navigation et d'interactions pour améliorer les services), le groupe a finalement décidé de ne pas poursuivre le test de cette solution compte tenu du contexte académique.

Scopus AI constituait également une solution que le groupe de travail souhaitait analyser de manière approfondie. Bien que des contacts aient été établis avec Elsevier et qu'une démonstration, particulièrement prometteuse, ait pu être organisée, l'obtention d'un accès de test pour l'ensemble des membres du groupe s'est révélée techniquement complexe. Le fournisseur indiquait que l'accès à Scopus AI était effectif, alors que celui-ci s'avérait inexistant ou limité à certains utilisateurs, pour des durées difficiles à déterminer. Face à ces contraintes persistantes, le groupe de travail a finalement décidé d'abandonner cette piste au début du mois de décembre.

¹ Molden, J. (6 mai 2025). The Dark Side of Perplexity AI: Privacy Risks You Should Know. PC Matic <https://www.pcmatic.com/blog/the-dark-side-of-perplexity-ai-privacy-risks-you-should-know/>

Néanmoins, sur base des quelques utilisations de l'outil, les éléments prometteurs suivants peuvent être soulevés :

- Accompagnement qualitatif dans l'aide à l'amélioration de la requête (reformulation, suggestions de concepts connexes et de mots-clés disciplinaires) ;
- Transparence du raisonnement avec une visibilité sur la reformulation de la requête en équation booléenne ;
- Fiabilité de l'outil : la technologie du RAG travaillant sur un corpus de données fermé et qualitatif, les risques d'hallucinations et d'erreurs sont extrêmement faibles.

Le Conseil a souligné la nécessité d'objectiver les analyses en s'appuyant sur des critères clairement définis et, le cas échéant, sur des scénarios de test communs. Le groupe a pris en compte ces recommandations en ajustant sa grille d'évaluation. Si l'instauration d'une cotation et d'une pondération par grands critères a été envisagée et partiellement amorcée, la mise en œuvre d'une évaluation chiffrée s'est révélée plus complexe et plus chronophage que prévu, en particulier dans la perspective de limiter au maximum toute part de subjectivité. Dans ce contexte, le groupe a décidé de renoncer à une évaluation chiffrée des caractéristiques et fonctionnalités des solutions.

Bien que les outils proposent des conditions d'accès différenciées (versions gratuites et payantes), les écarts de performance observés relevaient le plus souvent d'un ordre quantitatif plutôt que qualitatif : les accès payants permettent généralement un nombre plus élevé de requêtes et un volume accru de références en retour. Seules certaines fonctionnalités spécifiques, notamment liées aux mécanismes de rebond dans quelques solutions, étaient réservées aux versions payantes. Les tests ont été réalisés exclusivement sur des versions payantes d'entrée de gamme ; les offres individuelles plus avancées ainsi que les versions destinées aux entreprises n'ont pas été évaluées.

Le présent rapport rend compte de ces travaux et propose une synthèse des enseignements et constats tirés des tests menés, dans le respect du cadre fixé par la mission initiale.

Analyse

Méthodologie

Afin de garantir une analyse comparable des différents outils, le groupe de travail a élaboré et appliqué une méthodologie commune à l'ensemble des tests.

Chaque outil a été testé par 3 personnes du GT-IA afin de réduire les biais individuels, d'identifier d'éventuelles divergences d'usage et d'assurer une représentativité des pratiques. Des périodes de tests communes ont été choisies afin de tester la même version de l'outil.

Une exception concerne Scite, pour laquelle l'analyse s'est basée sur deux évaluations institutionnelles antérieures, l'une menée à l'ULB et l'autre à l'UCLouvain (printemps 2024 et 2025).

Les tests ont été menés en deux phases. La première s'est concentrée sur les versions gratuites des IA et s'est déroulée de juillet à septembre 2025. La deuxième phase concernait les versions payantes et a été réalisée du 15 octobre 2025 au 15 novembre 2025.

Le groupe a fait le choix de ne pas standardiser les questions posées aux outils (prompts). Toutefois, les testeurs ont veillé à ce que les questions soient variées en termes de thématiques et de requêtes, avec une prédominance des sciences humaines et sociales, en raison des domaines de compétences des testeurs.

Afin de structurer l'évaluation, une grille d'analyse a été définie et appliquée à chacun des outils. Elle comprend six indicateurs, eux-mêmes déclinés en sous-critères :

1. Informations générales
(éditeur, public cible, couverture, mise à jour du corpus, fonctionnalités, etc.)
2. Tarification & modèle économique
(modèle économique, tarification, limite d'usages, etc.)
3. Performance
(qualité et transparence des sources, présence de sources hallucinées, reproductibilité, etc.)
4. Ergonomie et usage
(accessibilité linguistique, option de personnalisation, aide et support, etc.)
5. Interopérabilité & compatibilité
(formats import/export, connecteurs disponibles)
6. Aspects éthiques & juridiques
(Respect des données personnelles et biais algorithmiques connus)

1. Informations générales

L'enjeu fondamental réside dans l'étendue et la qualité des bases de connaissance indexées. Plus la couverture est large, multidisciplinaire, à jour et transparente, plus l'intérêt pour nos communautés scientifiques est susceptible de croître. Or, l'ensemble des outils testés communiquent de manière très partielle sur l'origine et la composition exacte de leur indexation ou sur la fréquence de mise à jour.

À l'exception de Scite, qui bénéficie de partenariats spécifiques avec plusieurs éditeurs, les outils reposent sur des corpus relativement similaires : Semantic Scholar, OpenAlex et certaines

archives institutionnelles ou thématiques. Ces limites s'inscrivent dans un contexte plus large de gouvernance de l'information scientifique. Les grands éditeurs scientifiques n'ouvrent pas tous l'accès à leurs corpus. Certains choisissent même de développer leur propre « couche IA », de façon plus ou moins satisfaisante.

Un point récurrent observé dans l'ensemble des outils évalués (Scite inclus) concerne les disparités disciplinaires, linguistiques et géographiques, qui affectent directement la représentativité et la pertinence des résultats obtenus. Dans la majorité des cas, la couverture documentaire apparaît nettement plus faible en sciences humaines et sociales (SHS), tant en volume qu'en profondeur, entraînant un appauvrissement des réponses pour certains champs disciplinaires et des angles morts importants sur des thématiques pourtant bien documentées dans les bases traditionnelles. À cette limite s'ajoute une forte dominance de la littérature anglophone, issue d'éditeurs ou d'archives principalement nord-américains, tandis que la littérature européenne — en particulier francophone, germanophone ou issue de revues régionales — demeure sous-représentée. Cette concentration linguistique induit des biais notables dans les analyses et synthèses générées, en orientant les résultats vers des perspectives épistémologiques et méthodologiques majoritairement anglo-saxonnes. De tels déséquilibres rendent les outils moins adaptés pour des recherches nécessitant une couverture internationale équilibrée ou une attention aux spécificités des SHS et imposent une intégration prudente de leurs résultats, systématiquement complétée par des recherches dans des bases disciplinaires et multilingues plus établies.

Soulignons que les tests menés ont relevés plusieurs fonctionnalités intéressantes. Globalement, ces outils permettent un rapide cadrage d'un sujet de recherche avec la génération d'un résumé synthétique mobilisant les références trouvées (le nombre de références pris en compte varie d'un outil à l'autre). L'évaluation de la qualité de ces résumés n'a pas été effectuée de manière systématique lors de cette analyse comparative cependant, la plupart de ces outils se basent exclusivement sur les abstracts pour cette génération, ce qui limite la profondeur et la qualité attendue de ces résumés. Certains peuvent intégrer les contenus des full-text quand ceux-ci sont disponibles, c'est le cas pour Consensus, Elicit et Scite.

L'outil Consensus se distingue avec son *Consensus meter*, qui permet d'appréhender rapidement les débats scientifiques autour d'un sujet par la mise en avant des articles en accord ou en désaccord. Elicit et SciSpace apportent une valeur ajoutée avec leurs synthèses sous forme de tableaux personnalisables (extraction des données pour chaque référence, par exemple : méthode, échantillon, résultats, limites) qui favorisent la comparaison systématique, tandis que les fonctionnalités de screening d'Elicit assistent la sélection des références pertinentes selon des critères explicites.

L'outil Scite propose quant à lui une analyse qualitative des citations, qui permet de distinguer les usages soutenant ou contrastifs d'une référence.

Undermind se démarque par ses fonctionnalités de visualisation des données. Les représentations graphiques et les *timelines* facilitent la compréhension de l'évolution d'un sujet de recherche. De plus, l'outil propose systématiquement des interactions préalables à sa recherche afin de préciser la requête.

Par ailleurs, notons que l'ensemble de ces outils prennent en charge la recherche sémantique, c'est-à-dire l'interprétation d'une requête documentaire formulée en langage naturel. Bien que les degrés de satisfaction de cette recherche varient d'un outil à l'autre, il faut souligner le caractère ergonomique et la rapide prise en main de ce type de recherche.

2. Tarification et modèle économique

Le groupe de travail a contacté les différents fournisseurs des outils testés, à l'exception de Scite, afin d'obtenir la tarification appliquée aux universités. Aucune information n'a toutefois pu être obtenue, ce qui ne permet pas de réaliser des comparaisons entre outils. Seul Consensus indique la tarification appliquée de manière transparente sur son site web. Celle-ci se calcule sur le nombre de FTE et varie de 2 500 \$ par an (moins de 5 000 FTE) à 10 000 \$ (plus de 25 000 FTE).

Les tests des versions gratuites ont mis en évidence les limites d'utilisation qui pour certains outils (Elicit, SciSpace, Undermind et Scite) sont tellement contraignantes qu'elles rendent l'outil inutilisable en version gratuite, l'outil le moins contraignant étant Scholar GPT.

3. Performances et qualité des sources

De manière transversale, la qualité des sources constitue un critère déterminant pour l'évaluation des outils étudiés. Bien qu'ils s'appuient tous sur Semantic Scholar comme corpus principal — parfois complété par d'autres bases reconnues telles qu'OpenAlex — la fiabilité apparente de ces sources montre rapidement ses limites lorsque l'on examine plus finement la couverture disciplinaire, souvent inégale, ainsi que le manque de transparence quant aux données réellement indexées.

Les outils reposant largement sur des modèles génératifs (Elicit, Scholar GPT) présentent des fragilités plus marquées : erreurs dans les métadonnées, résumés imprécis, voire références inventées dans le cas de Scholar GPT.

Certains outils comme Consensus ou Undermind offrent toutefois une meilleure traçabilité en indiquant clairement s'ils utilisent le texte intégral ou seulement l'abstract. Néanmoins, aucun ne garantit un niveau de robustesse suffisant pour être utilisé comme source unique dans un processus de recherche exigeant.

Un élément supplémentaire relevé lors de l'évaluation concerne la présence de revues prédatrices dans certains résultats. Ce constat peut être nuancé. D'une part, les outils testés, comme mentionnés précédemment, s'appuient sur des corpus ouverts tels que Semantic Scholar ou Open Alex. Ceux-ci ne disposent pas d'un filtrage comparable à celui mis en place par les indexations sélectives traditionnelles que l'on retrouve dans les bases de données académiques. D'autre part, cette situation n'est pas propre aux outils étudiés dans le cadre de ce rapport : certaines plateformes largement utilisées (comme Google Scholar) indexent aussi régulièrement des revues prédatrices et/ou de qualité incertaine. La présence de ces sources ne résulte donc pas d'un défaut des outils étudiés, mais plutôt de la nature des bases ouvertes avec lesquelles ils sont alimentés.

La question de la reproductibilité des résultats s'est révélée particulièrement difficile à évaluer de manière rigoureuse. Des tests ont été menés en soumettant une même requête à plusieurs reprises. Ils ont mis en évidence que la répétition d'une requête identique ne garantissait pas l'obtention de résultats strictement similaires dans la majorité des outils étudiés. Lorsqu'une même requête est répétée à différents moments, les résultats peuvent varier de manière significative, mais il n'est pas toujours possible d'identifier précisément pourquoi ces variations se produisent.

Plusieurs éléments peuvent expliquer ces écarts.

- Premièrement, les corpus utilisés font l'objet de mises à jour régulières (ajout de nouvelles références, modifications de métadonnées, retrait de publications) sans que ces évolutions soient systématiquement documentées. Il devient dès lors impossible de savoir si une variation dans les résultats provient d'un enrichissement du corpus ou d'un autre facteur.
- Deuxièmement, les modèles d'intelligence artificielle sont régulièrement réentraînés ou ajustés ce qui peut modifier la manière dont les requêtes sont interprétées et les résultats classés.
- Troisièmement, les algorithmes de classement et de pertinence ne sont pas détaillés par les fournisseurs, rendant opaque le processus qui détermine l'ordre d'apparition des résultats.
- Enfin, pour certains outils qui utilisent des modèles génératifs, la nature même de ces technologies implique qu'ils ne garantissent pas des résultats identiques d'une interrogation à l'autre, même lorsque la requête est formulée de manière strictement identique.

Certains outils semblent toutefois plus stables que d'autres. Par exemple, Consensus affiche explicitement son raisonnement et la reformulation de la requête en équation booléenne, ce qui offre une certaine traçabilité et donne une impression de constance dans l'approche, même si les résultats finaux peuvent malgré tout varier. D'autres outils, notamment ceux reposant davantage sur des modèles génératifs, présentent des variations plus marquées sans qu'il soit possible d'en déterminer la cause exacte.

4. Ergonomie et facilité de prise en main

De manière générale les interfaces des différents outils ont été jugées intuitives, même lorsque certaines fonctionnalités avancées requièrent un temps d'apprentissage (ex. : Elicit, SciSpace). Seul Scite a été jugé avec une ergonomie assez complexe empêchant une prise en main efficace dès la première utilisation. Scholar GPT bénéficie, quant à lui, de l'environnement familier de ChatGPT, facilitant son utilisation et son appropriation.

En revanche, les possibilités de personnalisation restent peu développées de manière générale. Elles sont inexistantes pour ScholarGPT, limitées pour Elicit, SciSpace, Consensus et Scite ou bien présentes pour Undermind. Les options de personnalisation concernent principalement la langue de l'interface, la création de dossier/alerte et l'intégration LibKey.

Tous les outils proposent des supports d'accompagnement que ce soit via des FAQ, blogs ou des tutoriels (vidéos).

5. Interopérabilité & compatibilité

Les six solutions analysées présentent des niveaux d'interopérabilité assez contrastés en contexte académique. Scite et Undermind se distinguent par une bonne intégration aux workflows de la recherche car ils combinent une compatibilité avec des sources et bases de données majeures (PubMed, ArXiv ont été considérés) et des points de connexion explicites avec les outils de gestion bibliographique. Elicit occupe une position intermédiaire : si l'outil prend bien en charge plusieurs formats standards (PDF, BibTeX, RIS) et propose des mécanismes d'intégration concrets via extensions (Chrome, Zotero) et d'export bibliographique, nous n'avons pu y trouver aucune compatibilité avec les trois grandes bases de données. Scholar GPT offre la couverture la plus large en

termes de formats d'entrée et de sortie, mais son absence de connecteurs impose des échanges essentiellement manuels via fichiers, ce qui est loin d'être l'idéal. Lors des tests, il est apparu que Consensus se positionnait comme un outil interopérable via l'écosystème (export bibliographique, LibKey), dont on peut souligner une toute récente possibilité de connexion à Zotero (inexistante lors de nos tests)². SciSpace apparaît comme le moins compatible en raison de formats limités et d'intégrations faibles avec les bases et outils bibliographiques, hormis via LibKey.

Globalement, il nous semble que les potentiels de connectivité avec les applications bibliographiques restent assez faibles dans l'ensemble des outils analysés.

6. Aspects éthiques & juridiques

Les pratiques liées à l'usage, au traitement et au stockage des données présentent plusieurs zones d'incertitude. L'utilisation des données personnelles est rarement clairement expliquée et semble souvent non alignée avec les exigences du RGPD. Les informations collectées — telles que les adresses e-mail, cookies et journaux — servent à communiquer avec les utilisateurs ou répondre à leurs questions ; mais ils peuvent être partagées avec des tiers « si nécessaire ». De plus, les données sont stockées sur des serveurs situés aux États-Unis, ce qui pose un problème de conformité additionnel, en particulier concernant les transferts de données hors UE.

La politique de confidentialité apparaît, dans la plupart des outils, incomplète, peu transparente et ne fournit pas d'informations détaillées sur la protection des données, ni sur les garanties offertes. Les données privées ne sont pas utilisées pour entraîner le modèle mais demeurent accessibles aux sociétés, qui peuvent les exploiter à des « fins commerciales légitimes ». Enfin, plusieurs zones d'ombre persistent quant aux droits des utilisateurs sur les contenus générés ou traités.

En conséquence, des mesures de prudence s'imposent dans le choix éventuel d'une licence institutionnelle à l'un ou l'autre des outils analysés : éviter d'utiliser la plateforme pour des données sensibles, confidentielles ou liées à la recherche ; limiter au strict minimum la fourniture de données personnelles (e-mail, informations d'identification) ; vérifier les exigences de conformité au RGPD avant tout usage institutionnel.

² Turner, S. (29 janvier 2026). Zotero + Consensus AI. <https://blog.stephenturner.us/p/zotero-consensus-ai>

Synthèse par outils

NB : Les tests ont été réalisés fin 2025, les outils ont pu évoluer depuis et intégrer de nouvelles fonctionnalités.

Consensus

<https://consensus.app/>

Consensus est un moteur de recherche IA développé par la société américaine du même nom : Consensus NPL. L'outil s'appuie sur le corpus provenant de Semantic Scholar, OpenAlex et sa propre indexation du « Scholarly web » pour un total annoncé de 220 millions d'articles évalués par les pairs sans toutefois qu'on sache la proportion d'articles référencés en texte intégral ou uniquement avec abstract.

L'accès à Consensus nécessite la création d'un compte personnel avec trois tarifications possibles : gratuit, pro (10€/mois) et *deep* (45€/mois). Chaque tarification donnant accès à un nombre limité de requêtes. Des offres pour les universités existent avec une tarification transparente basée sur le nombre de FTE.

Consensus propose trois modes de recherche : *quick*, pro et *deep*. Chaque mode permet de générer un rapport en se référant à un ensemble limité d'articles : 10 articles (*quick*), 20 articles (pro), 50 articles (*deep*). Les articles analysés sont les premiers résultats de la requête. Il n'est dès lors pas possible de sélectionner manuellement les articles qu'on souhaiterait utiliser pour l'analyse. Il est cependant possible de charger jusqu'à 20 articles pour travailler sur un corpus spécifique. Le nombre de requête *quick* est illimité quelle que soit la tarification choisie ce qui permet une utilisation régulière pour des recherches exploratoires même avec un compte gratuit. À noter que la limite du nombre d'articles porte sur le nombre d'articles utilisés pour la création du rapport mais pas sur le nombre d'articles trouvés. L'interface peut donc être utilisée comme un outil de recherche documentaire traditionnel. De plus, l'outil se distingue au niveau des filtres disponibles (année, méthodologie, thématique, pays, etc.) Un autre point de distinction sont les citations dans le rapport permettant de visualiser rapidement si l'outil a utilisé le texte intégral ou l'abstract de l'article cité.

Au niveau des performances, la qualité des sources est jugée bonne mais dépend des domaines couverts par les sources utilisées par l'outil. Des limites sont ainsi apparues dans le domaine des sciences humaines et sociales avec un nombre de réponses plus réduit que Google Scholar. Aucune source hallucinée n'a été détectée mais des erreurs dans les références provenant des sources utilisées (Semantic Scholar) ont été repérées.

Le raisonnement de l'IA ainsi que la reformulation de la question en équation booléenne sont visibles. Ce qui permet de constater qu'une même question posée à différents moments produira le même raisonnement avec cependant des variations dans les résultats obtenus, sans qu'on ne puisse dire si cela est dû à l'évolution du corpus ou au tri par pertinence des résultats.

Nos tests indiquent que l'outil est particulièrement efficace pour des questions fermées qu'il traite différemment des questions ouvertes en signalant le niveau d'accord ou de désaccord des sources analysées.

En ce qui concerne l'interface, celle-ci est simple et intuitive mais elle peut surprendre pas sa lenteur en mode de recherche *deep*. Bien que son affichage soit uniquement possible en anglais, l'IA interagit en français. Consensus propose en outre une documentation et des supports clairs (tutoriel et chaîne YouTube).

Un atout de l'outil est l'intégration avec LibKey qui, pour les institutions ayant un abonnement LibKey, permet un rebond direct vers les ressources documentaires de l'institution.

Sur le plan économique et en comparaison à d'autres outils, le modèle *freemium* est intéressant en permettant un volume acceptable de requêtes avancées et en donnant accès à toutes les fonctionnalités de l'interface (export des résultats en CSV, import d'articles, etc.)

Un point faible important réside dans le fait que la plateforme ne respecte pas les exigences du RGPD.

En définitive, Consensus est un outil attractif avec son modèle économique permettant un usage fonctionnel (et proportionnel à l'abonnement choisi) pour obtenir des synthèses rapides, structurées et référencées avec des sources traçables et non hallucinées. Toutefois, son manque de transparence sur son corpus, ses limites d'analyse à 50 articles, ses lacunes en SHS et son non-respect du RGPD imposent une utilisation prudente. Il peut constituer un complément efficace aux bases de données traditionnelles, mais ne saurait s'y substituer dans une démarche de revue systématique ou d'exhaustivité.

Elicit

<https://elicit.com/>

Elicit est un outil d'aide à la recherche documentaire, à l'exploration et à la synthèse de la littérature scientifique. L'interface est relativement ergonomique, bien que la prise en main de toutes les fonctionnalités demande d'y consacrer un peu de temps.

La couverture documentaire de la fonctionnalité recherche se base en grande majorité sur Semantic Scholar mais un manque de transparence sur la couverture réelle de l'outil est à prendre en compte. L'approche par recherche sémantique, permettant de poser des questions en langage naturel, facilite globalement le processus de recherche. Cependant, l'outil n'offre pas un accompagnement pour améliorer la requête lancée (questions de clarification par exemple). Les résultats obtenus dépendront donc de la qualité de la question (mobilisation de concepts scientifiques, vocabulaire adéquat, délimitation...). Cette limite restreint l'utilisation de l'outil à un public plus expert pour en tirer un maximum de potentiel.

De plus, les résultats obtenus varieront sensiblement en fonction du domaine de recherche et du sujet d'étude, avec une moins bonne représentation de la littérature en SHS et de tous les sujets de « niche ». Notons que l'interface est uniquement disponible en anglais et que les requêtes doivent également être introduites dans cette langue. La qualité de la recherche ne varie pas entre l'option gratuite et payante sauf pour les filtres (filtre *Full-text* en version payante). En revanche, l'exportation des références est uniquement disponible en version payante (formats : PDF, BibTeX, RIS).

La synthèse des références trouvées est proposée sous forme de tableau qui est personnalisable en fonction des données souhaitées. Ce format est appréciable car il permet d'accélérer le processus de sélection des références. Cependant, des erreurs ont été relevées lors des tests au

niveau des données bibliographiques ainsi que la présence d'inexactitudes dans la génération des résumés et des autres données. Une vérification manuelle est donc nécessaire. En complément, un chat permet de poser des questions sur le contenu des références trouvées.

Le screening des références effectué lors de la génération d'un rapport est qualitatif et peut être personnalisable en fonction des critères souhaités. Il est de plus reproductible d'une demande de rapport à l'autre. Cependant, cette fonctionnalité est très lente. Le contenu du rapport reste très généraliste, se basant presque exclusivement sur les résumés des références. Aucune citation n'est présente dans le corps du texte, seule une liste des références utilisées est indiquée en fin de document.

Finalement, la possibilité d'importer des documents (BIB, RIS, PDF) pour en extraire des informations et lancer un chat avec le contenu enrichit encore les usages possibles.

L'outil Elicit est donc globalement intéressant au niveau des fonctionnalités proposées pour sélectionner et exploiter la littérature mais reste perfectible au niveau de la qualité de la couverture et de la recherche ainsi que dans sa rigueur dans l'extraction des données.

Scholar GPT

<https://chatgpt.com/g/g-kZ0eYXUe-scholar-gpt>

Scholar GPT se présente comme une extension spécialisée pour la recherche académique, développée par Sider.ai et reposant sur GPT (OpenAI), avec un positionnement multidisciplinaire et orienté vers des publics allant des étudiants aux doctorants et chercheurs-enseignants. L'accès à Scholar GPT se fait dans le même environnement que ChatGPT : il est possible de passer d'un GPT à un autre sans changer d'interface ni de contexte de travail, ce qui simplifie l'adoption et réduit le coût « cognitif » d'entrée pour les usagers. C'est un élément opérationnel qui mérite d'être souligné.

Sur le plan ergonomique, l'interface a été jugée simple, intuitive et globalement rapide, avec une prise en charge fluide des principaux formats d'entrée et des possibilités d'export assez larges.

En revanche, plusieurs limites affectent directement l'usage documentaire et la confiance que l'on peut accorder à l'outil. D'abord, le périmètre exact du corpus interrogé reste insuffisamment transparent. Cette opacité se combine à une intégration faible avec les outils et connecteurs couramment utilisés (absence d'API, pas de connecteurs type LibKey et nécessité de passer par des échanges de fichiers BIB/RIS pour la bibliographie), ce qui limite l'intégration dans les écosystèmes institutionnels de nos universités. À cela s'ajoute un point particulièrement critique : la fiabilité des références. Des sources hallucinées ont été observées à plusieurs reprises lors des tests et des réponses ont parfois mobilisé de la documentation non scientifique. La traçabilité des références est perfectible, ce qui impose, en pratique, une vérification systématique des citations et des liens entre affirmations et sources. Impraticable.

Une autre faiblesse significative concerne la pertinence de la sélection d'articles lors des recherches prospectives à partir d'un document uploadé. Par exemple, un test a montré que, sur 10 études annoncées comme pertinentes par Scholar GPT, seules 2 prolongeaient réellement le sujet de fond du texte initial. Ce cas de figure a pu être reproduit avec d'autres fichiers. Cette variabilité d'une recherche à l'autre et cette pertinence globalement insatisfaisante conduisent à

relativiser fortement l'intérêt de Scholar GPT comme outil de recherche bibliographique à conseiller.

Dans nos tests, certains usages se sont distingués positivement, notamment la génération de cartes mentales à partir d'articles fournis ainsi que des analyses comparatives capables de mettre en regard un cas d'étude uploadé avec d'autres approches, en soulignant similitudes, différences et enseignements. Si l'outil peut apporter une plus-value quand la tâche attendue consiste à structurer, reformuler et comparer des contenus déjà apportés par l'utilisateur, il ne nous paraît pas recommandé pour découvrir de la littérature scientifique de manière fiable.

En synthèse, l'outil est robuste pour accélérer des tâches de compréhension, de synthèse et de comparaison sur un corpus fourni, mais il présente des faiblesses importantes dès qu'il s'agit de fiabilité des références et de recommandation documentaire, avec un corpus insuffisamment explicite et une intégration limitée à nos environnements académiques.

SciSpace

<https://scispace.com/>

Outil polyvalent mais à relativiser. Il se présente comme une plateforme « tout-en-un » pour la recherche et la rédaction scientifique, avec des fonctionnalités intéressantes (revue de littérature, chat avec PDF, génération de citations, intégration LibKey). Il s'adresse à tous les publics académiques, du bachelier au chercheur confirmé. L'outil bénéficie d'une interface ergonomique facilitant la prise en main (avec des options de personnalisation comme les filtres et colonnes personnalisables, l'organisation en collections thématiques), d'une recherche sémantique performante pour découvrir des articles connexes, pas d'hallucinations constatées dans les résultats, et d'une base de données volumineuse (Semantic Scholar, OpenAlex) utile pour les explorations préliminaires.

L'outil propose également un tableau de synthèse comparative permettant d'extraire et de structurer les données clés de plusieurs références simultanément (méthodologie, résultats principaux, limites), facilitant ainsi l'analyse comparative systématique dans le cadre d'une revue de littérature. Cette fonctionnalité s'avère particulièrement pertinente pour initier les étudiants à la lecture analytique croisée de corpus scientifiques, bien qu'elle reste tributaire du système de crédits de la version gratuite.

Le support multilingue accepte les requêtes en français, bien que les réponses restent en anglais par défaut (paramétrage possible pour certaines fonctionnalités).

Toutefois, son efficacité demeure fortement tributaire des sources indexées, induisant des biais disciplinaires et linguistiques significatifs. La version gratuite est très limitée (analyse restreinte à 5 articles par requête, système de crédits opaque), et les fonctionnalités avancées sont verrouillées derrière un abonnement premium.

Si la recherche sémantique est performante, la transparence algorithmique est quasi inexistante, rendant les résultats non reproductibles. L'outil ne propose par ailleurs aucune fonctionnalité d'analyse critique (détection de biais, contre-arguments, question de clarification). Il n'offre pas d'intégration fluide avec les workflows bibliographiques standards. Enfin, des zones d'ombre subsistent sur la conformité RGPD et les droits liés aux contenus générés. En résumé : intéressant

pour une exploration rapide ou un travail préliminaire, mais insuffisant pour un usage académique rigoureux ou institutionnel.

Scite

<https://scite.ai/home>

Scite.ai est un outil dont les fonctionnalités permettent d'évaluer rapidement la crédibilité et l'impact des publications scientifiques. Il mobilise l'intelligence artificielle pour analyser les citations de manière quantitative mais aussi qualitative, en les classant selon qu'elles soutiennent, contredisent ou mentionnent un travail donné. L'outil propose également un Assistant IA qui génère des réponses argumentées et référencées.

L'analyse de cet outil n'a pas été menée par le GT-IA de la BICfB mais s'est appuyée sur deux études menées, l'une par l'ULB (en 2024), l'autre par l'UCLouvain (en 2025), dont les rapports n'ont pas été publiés mais communiqués aux membres du GT. Ces études ont associé chacune des bibliothécaires et des chercheurs. Elles concluent toutes deux que cet outil ne peut suffire comme seul outil de recherche ; il pourrait être cependant un bon complément à des outils bibliographiques plus classiques.

Parmi les avantages relevés, la catégorisation des citations est largement appréciée pour l'éclairage qu'elle apporte sur les débats scientifiques. Plusieurs chercheurs ont également souligné l'intérêt de Scite.ai en phase exploratoire notamment pour dégrossir un sujet ou découvrir de nouveaux axes de recherche.

Les limites de l'outil sont toutefois nombreuses. Alors que la couverture annoncée de Scite est multidisciplinaire avec contrairement aux autres outils concurrents analysés dans ce rapport, des accords avec de nombreux éditeurs, elle a été jugée insuffisante, en particulier en sciences humaines. Des références essentielles n'apparaissent pas parce la couverture de la base de données ne s'étend pas à certains éditeurs incontournables. Par ailleurs, la couverture géographique des revues est inégale (surtout anglo-saxonne ; moins bonne pour d'autres régions/langues). L'Assistant, quant à lui, produit des réponses parfois incomplètes, peu pertinentes ou erronées, avec des phénomènes d'hallucination, imposant une vérification systématique des informations et des sources citées. L'interface est également perçue comme peu intuitive.

De plus, le prix de licence institutionnelle est un frein à l'adoption de cet outil. Les deux universités qui ont testé l'outil ont choisi de ne pas recommander l'acquisition d'une licence institutionnelle.

Undermind

<https://www.undermind.ai/>

Undermind.ai, fondé en 2023 par Joshua Ramette et Thomas Hartke, tous deux docteurs en physique (MIT/Princeton), est un assistant de recherche fondé sur l'IA qui mène des recherches itératives et adaptatives dans la littérature scientifique (via Semantic Scholar et ses corpus agrégés comme PubMed, ArXiv, IEEE, ACM ; pas toujours d'accès au texte intégral), puis synthétise les

résultats dans un rapport structuré (articles classés par pertinence, tableaux récapitulatifs, chronologie et réseau de citations).

L'outil vise des questions complexes, en mimant le processus de découverte d'un expert (affinage séquentiel, suivi des chaînes de citations, relectures ciblées) plutôt qu'une seule requête statique.

Undermind lance un dialogue initial et une précision de la requête par son interface de type agent conversationnel, qui clarifie la question pour générer un *proposed search topic* (explicité) avant le lancement d'une *Deep Search*.

Il propose une classification des articles trouvés et une mesure de leur pertinence. Les citations en ligne dans les synthèses renvoient aux articles sources ; le rapport indique des mises en garde et des éléments de type « ce qui manque ». Il produit un rapport interactif (et exportable) avec classement par pertinence, *timeline*, réseau de citations et tableaux synthétiques ; une possibilité d'alertes existe pour les nouvelles publications.

Le moteur s'appuie principalement sur Semantic Scholar, il fournit des résultats solides mais non exhaustifs au regard d'autres bases disciplinaires (Scopus, Web of Science, etc.)

Le public-cible est celui des chercheurs, surtout en STM (Sciences technologie, médecine, informatique, ingénierie, biotechnologie, mathématiques) mais les SHS sont plutôt bien couvertes, notamment quant à la profondeur chronologique des références trouvées et la prise en compte de la littérature non anglophone.

U.AI est accessible gratuitement pour un usage limité (ex. 5 recherches/mois) ; l'analyse peut cependant se limiter aux résumés plutôt qu'au plein texte (références masquées). Il existe des formules d'abonnement : *Academic/Pro* individuel (20 \$/mois ; 16 \$ si annuel) ; existence de formules *Team* et *Enterprise* (non documentées dans ce rapport).

En bref, nous pouvons dire que les points forts de Undermind sont :

- Une meilleure couverture des SHS par rapport à d'autres outils ;
- il retrouve des références anciennes (utiles en SHS) ;
- la visualisation des données : *timeline* et réseau de citations (même si l'ergonomie est perfectible) ;
- impressionnant de prime abord : les interactions préalables à l'interrogation sont intéressantes et montrent en général une bonne compréhension du sujet et de son contexte ; les premières références proposées sont souvent très pertinentes.

Ses points faibles sont :

- Une légère lenteur à produire le rapport ;
- dans la version *Free* : des publicités couvrent certaines références ;
- il perd de sa force dans l'offre documentaire en SHS, les références s'écartent vite du sujet précis mais restent utiles pour le contexte en fonction dont le sujet s'inscrit dans celui-ci (ce qu'il admet en signalant le degré de pertinence de chaque référence proposée) ; il signale nombre de citations par références proposées mais ne pointe pas vers les articles citants.

Pour conclure

À l'issue des tests et des analyses effectués sur l'ensemble des outils précités, notre observation générale est qu'aucun de ces outils ne sort du lot, si ce n'est Scholar GPT dont les références hallucinées en font l'outil le moins fiable du test. Cette observation découle d'une analyse comparative montrant que chaque outil présente des forces significatives compensées par des faiblesses importantes. Aucune solution ne parvient à conjuguer simultanément, dans notre contexte académique, exhaustivité documentaire satisfaisante, robustesse méthodologique, transparence algorithmique, ergonomie optimale, interopérabilité avec les workflows bibliographiques existants, conformité réglementaire claire et modèle économique.

Malgré les faiblesses énoncées, ces outils sont utilisés par nos chercheurs et étudiants. Cet usage effectif souligne une attente forte. L'enjeu ne réside pas tant dans la sélection d'un outil « parfait », qui n'existe pas à ce stade, que dans la proposition d'un outil, comme l'ont fait d'autres universités. Même imparfait, l'outil sélectionné permettrait de communiquer, former nos usagers, etc. dans le cadre d'une stratégie (multi)institutionnelle.

Nous souhaitons, par ailleurs, présenter comment d'autres institutions académiques ont choisi de s'abonner à un ou plusieurs outils d'IA documentaire ; et, pour conclure, relever certaines questions posées par l'essor des outils d'intelligence artificielle documentaire.

D'autres modalités de choix d'outils

L'Université Laval (Québec) nous a transmis les critères qui ont amené au choix de l'outil Elicit : « Nous n'avons pas utilisé de grille d'analyse, mais nous avons appliqué certains critères de façon informelle :

- Évolution rapide des fonctionnalités de l'outil
- Facilité d'utilisation
- Fonctionnalité d'extraction de l'information
- Référencement par d'autres institutions
- Conservation de l'historique des recherches
- Option de filtre sur la qualité des revues
- Possibilité de téléverser ses propres articles (PDF)
- Conversation avec les publications (*Chat with papers*)
- Politiques et culture de l'entreprise

Dans les outils envisagés au moment de la prise de décision (printemps 2024), nous avons principalement évalué Consensus et Elicit qui étaient considérés comme les meilleures options pour la recherche d'articles scientifiques. »³

Les pays nordiques privilégient l'outil développé localement Keenious⁴ : « Initialement conçu comme un outil de découverte, Keenious est devenu une plateforme complète qui aide les utilisateurs à trouver des recherches pertinentes et à comprendre la littérature académique complexe grâce à une analyse guidée par l'IA » (site Keenious/About). Cet outil, conçu à l'Université arctique de Norvège (UiT) a été choisi comme outil d'IA documentaire par des institutions

³ Echange avec Audrey Lefebvre, bibliothécaire à l'Université de Laval. <https://www.bibl.ulaval.ca/services/soutien-a-lapprentissage/outils-dia-en-recherche-documentaire/utiliser-elicit>

⁴Keenious. <https://keenious.com/landing>

académiques et de recherche des pays nordiques, notamment pour sa proximité géographique, son adéquation avec les règles locales de protection des données et sa moindre dépendance aux outils nord-américains ; mais aussi pour certaines de ses fonctionnalités, dont la bonne intégration aux ressources des bibliothèques (voir par exemple : Norwegian University of Life Sciences/NMBU⁵).

Risque de détournement de l'Open Access

L'intérêt des outils analysés est qu'ils explorent la littérature scientifique ; leur limite est qu'ils ne peuvent accéder à la totalité de celle-ci en raison de la politique de restriction d'accès à leurs publications de certains éditeurs (certains outils, tel Scite, négocient cependant avec des éditeurs, l'accès à leurs publications mais tous, à commencer par Elsevier, ne l'ouvrent pas). Dès lors, les outils IA moissonnent essentiellement les ressources en accès libre (notamment les dépôts institutionnels⁶ et autres portails de publications scientifiques, archives ouvertes, etc.). Dans la mesure où la plupart de ces outils proposent des accès payants aux fonctionnalités plus étendues ou pour l'obtention d'un plus grand nombre de références, ne risque-t-on pas, s'ils décident de rendre payant également l'accès aux documents cités (comme le fait Undermind), un détournement complet des principes de l'accès libre ? Et de retomber dans le carrousel de l'édition commerciale scientifique (nous payons de multiples fois ce que nous avons produit) ?

Enjeux internationaux

L'exemple norvégien illustre également d'autres enjeux liés au déploiement d'outils d'intelligence artificielle documentaire : captation du savoir scientifique librement accessible soit à des fins mercantiles, comme évoqué ci-dessus, soit en vue d'alimenter des outils aux performances et à la fiabilité critiquables ; risque de concentration quasi monopolistique (que l'on observe par ailleurs chez d'autres acteurs de l'édition et la diffusion académiques) qui, dans le contexte d'instabilité internationale actuelle, pourrait amener à la limitation, voire l'interdiction d'accès à certaines ressources ; risque de manipulation et de dévoiement des données de la recherche ; dépendance accrue à quelques acteurs aux visées politiques et économiques discutables ; moindre visibilité, voire invisibilisation de certains domaines de recherche pointus ou trop localisés, et de langues... Tous ces risques devraient amener les institutions scientifiques et académiques européennes à s'interroger sur le choix de leurs fournisseurs documentaires et la nécessité de préserver leur savoir, ses données et l'accès à ceux-ci.

⁵Norwegian University of Life Sciences. (23 janvier 2025). Keenious: An intelligent tool for finding literature. <https://www.nmbu.no/en/library/keenious-intelligent-tool-finding-literature>

⁶ Il faut aussi signaler les inconvénients provoqués aux serveurs qui hébergent les DI, soumis à des interrogations massives d'outils IA, qui peuvent provoquer le blocage de ceux-ci.

Annexes

Grilles d'analyse

Consensus

CRITÈRES	VALEURS	Commentaires
Informations générales		
Editeur/fournisseur	Consensus NPL Inc	Société américaine
Descriptif / Fonctionnalités	• Recherche documentaire • Résume et synthétise les meilleurs articles sur un sujet • Conversation interactive avec un article • Indique l'accord ou le désaccord dans la littérature en réponse à des questions fermées	Les questions peuvent être soumises en 3 modes : • Quick : recherche et résumé à partir des 10 premiers articles • Pro : analyse à partir des 20 premiers articles • Deep : revue de la littérature sur base des 50 premiers articles
Public ciblé	☒ Étudiants (BA) ☒ Ma-Doc ☒ Chercheurs-Enseignants	Consensus semble cibler les universités et les instituts de recherches en proposant des licences institutionnelles et en faisant de la communication ciblée.
Domaines couverts	☒ Multidisciplinaire ☐ Sciences & Techniques ☐ Sciences humaines et sociales ☐ Sciences de la santé	A côté du corpus général (20 M papers), Consensus permet une limite au domaine médical (8M papers)
Couverture (sources)	• Semantic Scholar • OpenAlex • "Scholarly web"	220 millions d'articles revus par les pairs. Pas de transparence sur le nombre d'articles avec abstract ou texte intégral.

Actualisation des données	☒ Mise à jour ☐ Corpus figé	Hebdomadaire
Tarification & modèle économique		
Modèle économique	☒ Gratuit ☒ Freemium ☒ Payant personnel ☒ Payant institutionnel	
Tarif indicatif	• Free – 0\$ • Pro-10\$/mois • Deep- 45\$/mois • Teams-20\$/personne/mois • Entreprise et université-personnalisé ○ <5,000 FTE : \$2,500/year ○ 5,000-10,000 FTE : \$5,000/year ○ 10,000-25,000 FTE : \$7,500/year ○ >25,000 FTE : \$10,000/year	
Limites d’usage (du modèle testé)	☒ Nombre de requêtes limité : 25 requêtes Pro, 3 Deep/mois ☐ Volume limité : ☐ Durée limitée : ☐ Fonctionnalités verrouillées : ☒ Autres limites : Scholar agent ☐ Aucune limite	Consensus a été testé avec un compte freemium et compte Pro Modèle freemium Pas de limite pour le nombre de requêtes “quick” Scholar agent = amélioration du raisonnement avec l’intégration de filtres automatiquement, les résultats sont classés par sous thèmes
Limites d’usage (du modèle testé)	☒ Nombre de requêtes limité : 15 requêtes Deep/mois ☐ Volume limité : ☐ Durée limitée : ☐ Fonctionnalités verrouillées : ☐ Autres limites : ☐ Aucune limite	Modèle Premium Pro

Performance		
Qualité des sources (pertinence de la source)	☐ Très bonne ☒ Bonne ☐ Moyenne ☐ Faible	La qualité des sources dépend du domaine des sources mobilisées par l'outil. Le tri par pertinence est relativement bon en comparaison à d'autres outils.
Transparence des sources	☐ 100 % des références sont complètes et traçables ☒ Au moins 80 % des références sont complètes et traçables ☐ Moins 80 % des références sont complètes et traçables	18/20 Certaines sources sont mal identifiées (type de document erroné, DOI inexistant). Rien d'inventé en soi, mais des erreurs dans les données des corpus utilisés
Présence de sources hallucinées	☐ Oui ☒ Non	
Transparence du raisonnement	☒ Communiquée ☐ Non communiquée	Reformulation de la question • Quick = pas de reformulation • Pro = reformulation en plusieurs questions • Deep = reformulation en plusieurs questions pour chaque concept identifié
Reproductibilité (si transparence du raisonnement accessible)	☒ Recherche reproductible ☐ Non reproductible	La conversion de la même question formulée en langage naturel en une équation de recherche booléenne est similaire à plusieurs temps-T. En revanche, les résultats peuvent être très différents d'un moment T à un autre.
Capacité auto-critique	☐ Détecte biais ☐ Suggère contre-arguments ☐ Propose des questions de clarification ☒ Non pertinent	L'outil ne présente pas de fonctionnalités d'analyse critique intégrée.

Ergonomie et usage		
Accessibilité linguistique	☒ Français ☒ Anglais	Pour les questions/réponses avec l'IA. L'interface est uniquement disponible en anglais.
Interface utilisateur	☒ Intuitive ☐ Assez complexe ☐ Complexe	
Option de personnalisation	☐ Oui ☒ Limitée ☐ Non	La langue de l'IA est personnalisable et il est possible de renseigner son institution afin de bénéficier d'une intégration automatique avec LibKey
Aide et support en ligne	☒ Disponible ☐ Non disponible	Chaîne YouTube + tutoriel
Interopérabilité & compatibilité		
Formats acceptés (input)	☒ PDF ☐ DOCX ☐ BibTeX ☐ RIS ☐ DOI ☐ URL	
Formats acceptés (export)	☒ PDF ☐ DOCX ☒ EXCEL/CSV ☐ BibTeX ☒ RIS	
Connecteurs disponibles	☐ API ☐ Extensions ☒ Export vers bibliographie ☒ LibKey ☐ Autre :	
Compatibilité écosystème académique	☐ PubMed ☐ ArXiv ☐ Scopus ☐ Autre :	Couvrir intégralement Pubmed
Aspects éthiques & juridiques		
Respect des données personnelles	☐ RGPD respecté ☐ Anonymisation ☒ Non respecté	

	☐ Aucune information trouvée	
Biais algorithmiques connus	☐ Explicités ☒ Non explicités	Sources sans rapport avec le sujet

Elicit

CRITÈRES	VALEURS	Commentaires
Informations générales		
Editeur/fournisseur	Elicit Research / United States	
Descriptif / Fonctionnalités	• Recherche documentaire • Synthèse des résultats • Rapport de recherche • Conversation interactive avec un article / résumé • Extractions de données des références ou de documents uploadés • Revue systématique	
Public ciblé	☐ Étudiants (BA) ☒ Ma-Doc ☒ Chercheurs-Enseignants	Produit orienté vers l'exploration approfondie de la documentation/revue systématique – niveau master minimum
Domaines couverts	☒ Multidisciplinaire ☐ Sciences & Techniques ☐ Sciences humaines et sociales ☐ Sciences de la santé	

Couverture (sources)	• Semantic Scholar corpus • OpenAlex • PubMed • ClinicalTrials.gov	
Actualisation des données	☒ Mise à jour ☐ Corpus figé	Hebdomadaire pour Semantic Scholar et PubMed ; quotidienne pour OpenAlex
Tarification & modèle économique		
Modèle économique	☐ Gratuit ☒ Freemium ☒ Payant personnel ☒ Payant institutionnel	
Tarif indicatif	\$12 / mois (plus) \$49 / mois (pro)	
Limites d'usage (du modèle testé)	☐ Nombre de requêtes limité : ☒ Volume limité : 20 articles analysés /mois (2 rapports), chat avec 4 documents simultanément ☐ Durée limitée : ☒ Fonctionnalités verrouillées : revue systématique, exportation, collaboration, alertes ☒ Autres limites : tableau d'analyse des références (2 colonnes) ☐ Aucune limite	Freemium
Limites d'usage (du modèle testé)	☐ Nombre de requêtes limité : ☒ Volume limité : 40 articles analysés /mois (4 rapports) ☐ Durée limitée : ☒ Fonctionnalités verrouillées : revue systématique, collaboration, alertes ☒ Autres limites : tableau d'analyse des références (5 colonnes) ☐ Aucune limite	Plus

Performance		
Qualité des sources (pertinence de la source)	☐ Très bonne ☒ Bonne ☒ Moyenne ☐ Faible	
Transparence des sources	☐ 100 % des références sont complètes et traçables ☒ Au moins 80 % des références sont complètes et traçables ☐ Moins 80 % des références sont complètes et traçables	Des références sont incomplètes, des informations bibliographiques sont erronées. Le lien vers la source (Semantic Scholar) est toujours présent mais pas vers le FT.
Présence de sources hallucinées	☐ Oui ☒ Non	
Transparence du raisonnement	☐ Communiquée ☒ Non communiquée	Rien n'est communiqué sur le raisonnement de recherche documentaire
Reproductibilité (si transparence du raisonnement accessible)	☐ Recherche reproductible ☒ Non reproductible	Les résultats peuvent être drastiquement différents d'une recherche à l'autre avec la même question initiale. A titre d'exemple, la recherche lancée en version gratuite et version payante avec la même question initiale, n'ont qu'un résultat commun !
Capacité auto-critique	☐ Détecte biais ☐ Suggère contre-arguments ☒ Propose des questions de clarification ☐ Non pertinent	Il s'agit plutôt de suggestions d'amélioration de la question de recherche. C'est option n'est disponible que dans la génération de rapport (pas de différence entre les suggestions du mode « gratuit » et du mode « payant »).
Ergonomie et usage		
Accessibilité linguistique	☐ Français ☒ Anglais	Préciser pour question et réponse
Interface utilisateur	☐ Intuitive ☒ Assez complexe ☐ Complexe	
Option de personnalisation	☐ Oui ☐ Limitée ☒ Non	
Aide et support en ligne	☒ Disponible	YouTube + Blog

	☐ Non disponible	
Interopérabilité & compatibilité		
Formats acceptés (input)	☒ PDF ☐ DOCX ☒ BibTeX ☒ RIS ☐ DOI ☐ URL	
Formats acceptés (export)	☒ PDF ☐ DOCX ☒ EXCEL/CSV ☒ BibTeX ☒ RIS	Modèle « Plus »
Connecteurs disponibles	☐ API ☒ Extensions ☒ Export vers bibliographie ☐ LibKey ☐ Autre :	Extension Chrome et connexion à Zotero web
Compatibilité écosystème académique	☐ PubMed ☐ ArXiv ☐ Scopus ☐ Autre :	
Aspects éthiques & juridiques		
Respect des données personnelles	☒ RGPD respecté (indiqué comme tel mais à relativiser) ☐ Anonymisation ☐ Non respecté ☐ Aucune information trouvée	Attention les données collectées sont cependant conservées sur des serveurs américains. Les données privées ne sont pas collectées pour entrainer le modèle mais sont utilisées par la société.
Biais algorithmiques connus	☐ Explicités ☒ Non explicités	Aucune information trouvée à ce sujet
Conditions d'utilisation	☐ Open access ☐ Usages restreints ☐ Autre :	

Scholar GPT

CRITÈRES	VALEURS	Commentaires
Informations générales		
Editeur/fournisseur	Extension spécialisée pour la recherche académique, basée sur GPT (Open Ai). Développé par l'équipe Sider.ai	
Descriptif / Fonctionnalités	Extension spécialisée pour la recherche académique • Recherche d'articles scientifiques • Résumés académiques • Aide à la rédaction scientifique • Analyse documentaire avancée • Aide à la synthèse de l'information académique • Ecriture et exécution de code Python • Visualisation de données • Traitement du langage naturel • Résolution de problèmes complexes	
Public ciblé	☒ Étudiants (BA) ☒ Ma-Doc ☒ Chercheurs-Enseignants	
Domaines couverts	☒ Multidisciplinaire ☐ Sciences & Techniques ☐ Sciences humaines et sociales ☐ Sciences de la santé	
Couverture (sources)	La liste exacte des sources utilisées n'est pas communiquée. Le système annonce « Access Google Scholar, PubMed,	

	bioRxiv, arXiv, and more », mais cette information n'est pas transparente.	
Actualisation des données	☒ Mise à jour ☐ Corpus figé	
Tarification & modèle économique		
Modèle économique	☒ Gratuit ☐ Freemium ☒ Payant personnel ☒ Payant institutionnel	
Tarif indicatif		Plus : \$20 USD /mois Pro : \$200 USD /mois ChatGPT Business : Devis demandé en novembre (pour le périmètre d'une des 3 grandes universités), mais aucun suivi n'a été accordé à notre demande.
Limites d'usage (du modèle testé)	☒ Nombre de requêtes limité : <u>Gratuit</u> : jusqu'à 10 messages avec GPT-5.2 toutes les 5 heures. Après, ChatGPT bascule sur une version « mini » jusqu'au reset. <u>Payant</u> (compte Plus) : GPT-5.2 (défaut) : jusqu'à 160 messages toutes les 3 heures. Une fois le plafond atteint, la conversation bascule vers une version « mini » jusqu'au reset. GPT-5.2 Thinking (après sélection manuelle) : limite jusqu'à 3.000 messages par semaine. Après ce seuil, Thinking n'est plus sélectionnable manuellement.	Scholar GPT a été testé avec un compte gratuit et un compte Plus. Les réponses concernent la version 5.2 (actuellement la plus récente) de ChatGPT, mais sont différentes d'un modèle à l'autre. Il est donc très difficile de fournir une réponse unique. Le dernier modèle sorti n'étant pas forcément le plus approprié pour tous les usages. En outre, Scholar GPT tourne sur le modèle 4o. La pertinence des limites sur Scholar GPT 4o n'est donc pas garantie. La plupart des informations ont été trouvées sur des pages d'aide OpenAI comme https://help.openai.com/en/articles/11909943-gpt-52-in-chatgpt et

	Important : les limites GPT-5 sont partagées entre ChatGPT et les différents GPTs par défaut (donc Scholar GPT) et créés. ☒ Volume limité : <u>Pour tous les comptes, gratuit et Plus :</u> Limites publiées (qui s'appliquent aussi dans un GPT comme Scholar GPT) : Taille max par fichier : 512 MB. Textes/documents : 2M tokens par fichier (pas pour les feuilles de calcul). Images : 20 MB par image. CSV/ feuilles de calcul : env. 50 MB (variable). Rythme d'upload : jusqu'à 80 fichiers / 3 heures (OpenAI précise qu'ils peuvent baisser ces limites en période de pointe). <u>Spécifique aux comptes Plus :</u> Projets : jusqu'à 20 fichiers par projet. Par GPT: jusqu'à 20 fichiers par GPT « pour la durée de vie de ce GPT » (selon la formulation d'OpenAI). ☐ Durée limitée : pas de limite dans le temps ☒ Fonctionnalités verrouillées : <u>Gratuit</u> : Les usagers peuvent utiliser des GPTs du store (donc Scholar GPT), mais ne peuvent pas créer leur propre GPT via l'outil GPT Builder (réservé aux profils payants Plus/Team/Enterprise). <u>Payant</u> :	https://help.openai.com/en/articles/8555545-file-uploads-faq
--	--	--

	Rien de particulier à révéler, si ce n'est que certaines fonctionnalités sont réservées aux comptes Pro et Business ☒ Autres limites : Les comptes gratuits sont dépriorisés : la génération des réponses peut parfois être plus lentes et des interruptions sont toujours possibles, ou plus risquées, lors de forte charge et d'utilisation plus intensive. ☐ Aucune limite	
Performance		
Qualité des sources (pertinence de la source)	☐ Très bonne ☐ Bonne ☐ Moyenne ☒ Faible	La qualité des résultats peut fortement varier en fonction de la question posée à Scholar GPT. Lors de tests, il n'a pas été rare de trouver des réponses basées sur de la documentation non scientifique et d'obtenir des hallucinations en sources. Ainsi, suite à un pdf uploadé où l'on demandait à Scholar GPT de réaliser une recherche prospective et de proposer davantage d'articles scientifiques en lien avec le sujet de l'article chargé, la qualité des articles proposés s'est avérée insatisfaisante : le sujet de fond de l'article uploadé ne se retrouvait que de manière anecdotique dans les résultats proposés. Ainsi, sur 10 études et publications entre 2021 et 2025 jugées pertinentes par Scholar GPT, seules 2 se sont avérées réellement dans le prolongement de l'article uploadé pour permettre un approfondissement de la réflexion.

		Cette expérience décevante a pu être reproduite (avec d'autres fichiers chargés à la source). Par contre, Scholar GPT s'est avéré très utile pour réaliser une analyse comparative, qui a comparé l'étude de cas fournie (par upload) à d'autres études ou approches, en soulignant les similitudes, différences et les enseignements à en tirer. Les résultats étaient là à la hauteur des attentes. Autre performance de Scholar GPT, la réalisation de cartes mentales des articles uploadés. Ces résultats positifs n'étant toutefois pas jugés pertinents dans le cadre d'une analyse en lien avec la qualité et la plus-value en matière de recherche documentaire à proprement parler, la qualité des sources est jugée faible.
Transparence des sources	☐ 100 % des références sont complètes et traçables ☒ Au moins 80 % des références sont complètes et traçables ☐ Moins 80 % des références sont complètes et traçables	Le « Au moins 80% » est une estimation maximaliste : l'un des membres du groupe est même arrivé à moins de 80% lors d'un de ses tests.
Présence de sources hallucinées	☒ Oui ☐ Non	Exemples constatés. Difficile d'en estimer la fréquence.
Transparence du raisonnement	☒ Communiquée ☐ Non communiquée	La réponse relative à la transparence et à la reproductibilité mérite d'être nuancée. Scholar GPT peut fournir des réponses structurées et argumentées lorsqu'un prompt demande explicitement une justification ou une explication méthodologique. Toutefois, le raisonnement affiché correspond à une reconstruction explicative destinée à l'utilisateur, et non à l'intégralité du processus interne du
Reproductibilité (si transparence du raisonnement accessible)	☐ Recherche reproductible ☐ Non reproductible	

		modèle. L'accès au raisonnement détaillé dépend en outre de la configuration de l'outil et de l'interface, certaines pouvant volontairement limiter ou synthétiser ces explications. La reproductibilité conceptuelle (contenu, structure, conclusions générales) est généralement élevée, mais la reproductibilité textuelle exacte ne peut être garantie, en particulier à plusieurs semaines d'intervalle.
Capacité auto-critique	☐ Détection de biais ☐ Suggère contre-arguments ☐ Propose des questions de clarification ☒ Non pertinent	
Ergonomie et usage		
Accessibilité linguistique	☒ Français ☒ Anglais	Si Scholar GTP gère les questions et réponses en français et anglais, il semble toutefois que la qualité des réponses est supérieure si cela se passe en anglais.
Interface utilisateur	☒ Intuitive ☐ Assez complexe ☐ Complexe	
Option de personnalisation	☐ Oui ☐ Limitée ☒ Non	
Aide et support en ligne	☒ Disponible ☐ Non disponible	
Interopérabilité & compatibilité		
Formats acceptés (input)	☒ PDF ☒ DOCX ☒ BibTeX ☒ RIS ☒ DOI ☒ URL	

Formats acceptés (export)	☒ PDF ☒ DOCX ☒ EXCEL/CSV ☒ BibTeX ☒ RIS	Autres formats d'export: .txt, .json, .py (Python) .ipynb (Jupyter Notebook)
Connecteurs disponibles	☐ API ☐ Extensions ☐ Export vers bibliographie ☐ LibKey ☐ Autre :	Aucun connecteur. Le passage à des fichiers .bib ou .ris est par exemple nécessaire.
Compatibilité écosystème académique	☒ PubMed ☒ ArXiv ☐ Scopus ☒ Autre : bioRxiv, Europe PMC	Il existe aussi des compatibilités partielles, comme avec CrossRef, CORE et Semantic Scholar, ou moyennant des opérations manuelles de recherche, comme HAL et les archives institutionnelles des universités de la FWB (pas d'accès à l'API pour exploitation automatique des contenus).
Aspects éthiques & juridiques		
Respect des données personnelles	☒ RGPD respecté ☐ Anonymisation ☐ Non respecté ☐ Aucune information trouvée	Selon ScholarGPT, ils respectent le RGPD. Politique de confidentialité https://openai.com/fr-CA/policies/privacy-policy/
Biais algorithmiques connus	☐ Explicités ☐ Non explicités	

CRITÈRES	VALEURS	Commentaires
Informations générales		
Editeur/fournisseur	Typeset – Saikiran Chandha - 2016	Plateforme développée par une startup axée sur l'IA et l'Open Science
Descriptif / Fonctionnalités	Plateforme dite "tout-en-un" conçue pour aider à naviguer dans la littérature scientifique. Fournit des outils pour découvrir des articles, comprendre des recherches complexes et améliorer le processus d'écriture de recherche. Affirme rendre la science plus ouverte, collaborative et accessible. 4 domaines globaux pointés : recherche académique, rédaction scientifique, revues de la littérature, études interdisciplinaires	Literature review, Chat with PDF, copilot, extract data, citation generator, paraphraser, AI Writer L'outil se positionne comme un assistant complet, mais certaines fonctionnalités (comme l'AI Co-editing) sont réservées aux versions payantes. L'efficacité dépend fortement de la qualité des sources indexées.
Public ciblé	☒ Étudiants (BA) ☒ Ma-Doc ☒ Chercheurs-Enseignants	Adapté à tous les niveaux académiques, mais les fonctionnalités avancées (comme la "deep review") sont limitées aux utilisateurs premium.
Domaines couverts	☒ Multidisciplinaire ☐ Sciences & Techniques ☐ Sciences humaines et sociales ☐ Sciences de la santé	Pas de spécialisation disciplinaire claire. La couverture dépend des sources indexées (Semantic Scholar, OpenAlex, etc.), qui peuvent avoir des biais disciplinaires.
Couverture (sources)	Sources indexées : Semantic Scholar, Open Alex, crawlers web de dépôts institutionnels et prépublications	Pas de transparence complète sur l'exhaustivité de la couverture disciplinaire ni les partenariats éditoriaux
Actualisation des données	☒ Mise à jour ☐ Corpus figé	A priori en continu avec des ajouts de fonctionnalités et des optimisations

		Mises à jour fréquentes, mais pas de détails sur la fréquence ou la méthodologie de sélection des nouvelles sources.
Tarification & modèle économique		
Modèle économique	☒ Gratuit ☐ Freemium ☐ Payant personnel ☐ Payant institutionnel ☒ Premium	Modèle freemium agressif : les fonctionnalités clés (comme l'analyse approfondie) sont verrouillées. La version gratuite est très limitée (ex. : 5 articles max par requête).
Tarif indicatif	Premium : 20\$/mois Advanced : 90\$/mois	Prix compétitif comparé à d'autres outils (ex. : Elicit, ResearchRabbit), mais justifié par l'accès à des fonctionnalités avancées et plus diversifiées que pour d'autres outils. SciSpace entreprise access : devis demandé en août (pour le périmètre d'une des 3 grandes universités), mais aucun suivi n'a été accordé à notre demande. Depuis cette démarche, SciSpace a mis en place un formulaire de contact dédié aux demandes d'accès institutionnel. Ce nouveau formulaire n'a toutefois pas été utilisé.
Limites d'usage (du modèle testé)	☒ Nombre de requêtes limité : illimité Nombre restreint de recherches de haute qualité par période ☐ Volume limité : maximum 5 articles analysés par requête dans la version gratuite ☐ Durée limitée : ☐ Fonctionnalités verrouillées dans la version gratuite : AI Co-editing, summarization avancée, deep review, PDF to video ☐ Autres limites : ☐ Aucune limite	SciSpace a été testé avec un compte gratuit et Premium. Nombre de requêtes notées comme illimitées mais basées sur des "crédits" qui eux ne sont pas illimités. 1200 crédits par mois pour l'utilisation de l'agent Scispace. Revue de la littérature : illimité. Le système de crédits est opaque. Les utilisateurs peuvent être surpris par l'épuisement rapide des crédits pour des tâches complexes (ex. : revue de littérature).

Performance		
Qualité des sources (pertinence de la source)	☐ Très bonne ☒ Bonne ☐ Moyenne ☐ Faible	Recherche sémantique performante mais pertinence variable selon la complexité de la requête. Résulte de la limitation à 5 articles pour l'analyse.
Transparence des sources	☐ 100 % des références sont complètes et traçables ☒ Au moins 80 % des références sont complètes et traçables ☐ Moins 80 % des références sont complètes et traçables	Métadonnées parfois incomplètes ou erronées (ex. : auteurs manquants, années incorrectes). Pas de mécanisme de signalement des erreurs.
Présence de sources hallucinées	☐ Oui ☒ Non	
Transparence du raisonnement	☐ Communiquée ☒ Non communiquée	Aucune explication sur les algorithmes de recommandation ou de synthèse. Impossible de vérifier la logique derrière les résultats.
Reproductibilité (si transparence du raisonnement accessible)	☐ Recherche reproductible ☒ Non reproductible	Sans transparence algorithmique, les résultats ne sont pas reproductibles. Problématique pour un usage scientifique rigoureux.
Capacité auto-critique	☐ Détecte biais ☐ Suggère contre-arguments ☐ Propose des questions de clarification ☒ Non pertinent	L'outil ne présente pas de fonctionnalités d'analyse critique intégrée
Ergonomie et usage		
Accessibilité linguistique	☒ Français ☒ Anglais	Interface : en anglais Requêtes : accepte les requêtes en français Réponses : génération en anglais par défaut (paramétrage possible) ==> Cette option est disponible uniquement

		pour les fonctionnalités suivantes: Literature Reviews, Chat with PDF, Paraphraser, AI Detector, and AI Writer. Support multilingue
Interface utilisateur	☒ Intuitive ☐ Assez complexe ☐ Complexe	Personnalisations disponibles : filtres de recherche, colonnes personnalisables dans les tableaux de résultats, organisation en collections thématiques Limitations : pas de paramétrage avancé de l'algorithme de recherche. Options de personnalisation étendues réservées aux versions payantes.
Option de personnalisation	☐ Oui ☒ Limitée ☐ Non	Personnalisation basique (filtres, organisation). Pas de paramétrage avancé (ex. : ajustement des poids sémantiques dans la recherche).
Aide et support en ligne	☒ Disponible ☐ Non disponible	FAQ, tutoriels vidéo, support par mail. Support réactif mais pas de communauté d'utilisateurs ou de forum
Interopérabilité & compatibilité		
Formats acceptés (input)	☒ PDF ☐ DOCX ☐ BibTeX ☐ RIS ☐ DOI ☐ URL	Limite : pas de support pour DOCX, BibTeX ou RIS, ce qui restreint l'intégration avec d'autres outils (ex. : Zotero, EndNote).
Formats acceptés (export)	☒ PDF ☐ DOCX ☐ EXCEL/CSV ☐ BibTeX ☐ RIS	Markdomwn. Pas d'export en BibTeX ou RIS, ce qui complique l'intégration dans les workflows bibliographiques standards.
Connecteurs disponibles	☐ API ☐ Extensions ☐ Export vers bibliographie	Pas d'API accessible en version gratuite

	☒ LibKey ☐ Autre :	
Compatibilité écosystème académique	☐ PubMed ☐ ArXiv ☐ Scopus ☐ Autre :	Sources d'indexation : Semantic Scholar, Apen Alex et crawlers web. Pas d'intégration directe avec PubMed ou Scopus.
Aspects éthiques & juridiques		
Respect des données personnelles	☐ RGPD respecté ☐ Anonymisation ☐ Non respecté ☒ Aucune information trouvée	Pas d'info trouvée sur la politique de confidentialité et la conformité RGPD (ou partielle). Prudence recommandée pour l'usage avec des données de recherche sensible ou personnelles.
Biais algorithmiques connus	☐ Explicités ☒ Non explicités	Aucune documentation sur les biais potentiels (ex. : sur-représentation de certaines disciplines ou langues).

Scite

(sur base des études menées par UCLouvain et ULB)

CRITÈRES	VALEURS	Commentaires
Informations générales	Scite est un outil de recherche documentaire basé sur l'IA (LLMs). Il repose sur une base de données de littérature scientifique grâce à des accords conclus avec des éditeurs. En particulier, Scite a recours à des citations intelligentes (« smart citations »), lesquelles affichent le contexte de la citation et décrivent si l'article fournit des preuves à l'appui ou à l'opposé de ce qui est soutenu. Scite offre aussi un outil « Assistant » qui répond à des questions en se basant sur les informations disponibles dans sa base de données.	
Editeur/fournisseur	Research Solutions	
Descriptif / Fonctionnalités	Fonctionnalités testées : • Search (par titre, auteur, mots-clés ou DOI) • Visualizations (visualiser les articles liés) • Scite assistant (répondre à des questions) • Reference check (à partir d'un document téléversé) • Custom dashboard (regroupe les informations de Scite sur un ensemble d'articles) • Smart citations (regroupe les informations de Scite sur un ensemble d'articles)	
Public ciblé	☐ Étudiants (BA) ☐ Ma-Doc ☒ Chercheurs-Enseignants	
Domaines couverts	☒ Multidisciplinaire	Prédominances de STM

	☐ Sciences & Techniques ☐ Sciences humaines et sociales ☐ Sciences de la santé	
Couverture (sources)	https://scite.ai/journals	
Actualisation des données	☒ Mise à jour ☐ Corpus figé	
Tarification & modèle économique		
Modèle économique	☒ Gratuit ☐ Freemium ☒ Payant personnel ☒ Payant institutionnel	Gratuit limité
Tarif indicatif	Personnel : 16\$/mois // institutionnel : cher	+20.000 FTE : +/- 50.000\$
Limites d'usage (du modèle testé)	☒ Nombre de requêtes limité : ☐ Volume limité : ☐ Durée limitée : ☒ Fonctionnalités verrouillées : ☐ Autres limites : ☐ Aucune limite	
Performance		
Qualité des sources (pertinence de la source)	☐ Très bonne ☒ Bonne ☐ Moyenne ☐ Faible	Plus faible en SHS
Transparence des sources	☐ 100 % des références sont complètes et traçables ☒ Au moins 80 % des références sont complètes et traçables ☐ Moins 80 % des références sont complètes et traçables	
Présence de sources hallucinées	☒ Oui ☐ Non	Pas vraiment des hallucinations mais des biais de données ou de couverture d'étude

Transparence du raisonnement	☒ Communiquée ☐ Non communiquée	
Reproductibilité (si transparence du raisonnement accessible)	☒ Recherche reproductible ☐ Non reproductible	
Capacité auto-critique	☐ Détecte biais ☐ Suggère contre-arguments ☒ Propose des questions de clarification ☐ Non pertinent	
Ergonomie et usage		
Accessibilité linguistique	☐ Français ☒ Anglais	
Interface utilisateur	☐ Intuitive ☒ Assez complexe ☐ Complexe	
Option de personnalisation	☐ Oui ☒ Limitée ☐ Non	
Aide et support en ligne	☒ Disponible ☐ Non disponible	
Interopérabilité & compatibilité		
Formats acceptés (input)	☒ PDF ☒ DOCX ☐ BibTeX ☒ RIS ☒ DOI ☐ URL	
Formats acceptés (export)	☐ PDF ☐ DOCX ☒ EXCEL/CSV ☐ BibTeX ☒ RIS	
Connecteurs disponibles	☐ API ☐ Extensions ☒ Export vers bibliographie ☐ LibKey ☐ Autre :	Zotero

Compatibilité écosystème académique	☒ PubMed ☒ ArXiv ☒ Scopus ☐ Autre :	
Aspects éthiques & juridiques		
Respect des données personnelles	☐ RGPD respecté ☐ Anonymisation ☒ Non respecté ☐ Aucune information trouvée	Utilisation des données personnelles à des « fins commerciales légitimes »
Biais algorithmiques connus	☐ Explicités ☒ Non explicités	

Undermind

CRITÈRES	VALEURS	Commentaires
Informations générales	Undermind.ai est un système qui effectue des recherches (et une synthèse) d'articles scientifiques trouvés dans Semantic Scholar, et fournit un rapport organisé, présentant les articles clés classés par pertinence et des tableaux récapitulatifs.	
Editeur/fournisseur	https://www.undermind.ai/	fondé en 2023 par Harthe & Ramette.
Descriptif / Fonctionnalités	▪ Recherche sémantique avancée : exploration autonome, recherches successives, requêtes complexes ▪ Découverte d'articles ▪ Analyse de pertinence ▪ Assistance à la recherche personnalisée ▪ Récupération d'informations ▪ Prise en charge de requêtes complexes ▪ Synthèse documentaire automatisée ▪ Vérification des sources ▪ Estimation de complétude ▪ Interface de données relationnelles : chronologie et réseau de citations ▪ Évolution dans le traitement du sujet ("Trends, Patterns, Milestones") et pistes de recherche futures ("Implications for the State of the Field and Future Research"; "Summary Table: Primary Milestones and Shifts") ; sujets en débat ▪ Possibilité d'alertes sur les requêtes ; sauvegarde des requêtes ▪ Envoi de rapport par courriel : nombre d'articles pertinents trouvés, score de correspondance sur 100	

	pour chaque article, réseau de citations, chronologie ; articles classés en fonction de leur «influence relative » (fréquence de citation au sein du réseau); fonctionnalité » Rédiger un court article de synthèse »: récit factuel des points clés des articles + citation ; possibilité de lancer des nouvelles requêtes à partir des articles déjà analysés dans le rapport. On peut aussi télécharger ce rapport en PDF.	
Public ciblé	☐ Étudiants (BA) ☒ Ma-Doc ☒ Chercheurs-Enseignants	
Domaines couverts	☒ Multidisciplinaire ☐ Sciences & Techniques ☐ Sciences humaines et sociales ☐ Sciences de la santé	SHS présentes dans une moindre mesure mais mieux représentées que dans d'autres outils.
Couverture (sources)	Semantic scholar, Pubmed, ArXiv, IEEE, ACM	
Actualisation des données	☒ Mise à jour ☐ Corpus figé	
Tarification & modèle économique		
Modèle économique	☒ Gratuit ☐ Freemium ☒ Payant personnel ☒ Payant institutionnel	
Tarif indicatif	Academic : • Free • Pro : 16\$/mois si facturation annuelle ; 20\$/mois si facturation mensuelle • Team : 15\$ par personne par mois si facturation annuelle ; 18\$ par mois si facturation mensuelle (avec calcul jusqu'à 200 membres d'équipe)	Les modèles testés étaient le gratuit l'Academic Pro.

	Industry • Free • Pro : 60\$/mois si facturation annuelle ; 75\$/mois si facturation mensuelle • Team : 60\$ par personne par mois si facturation annuelle ; 75\$/mois si facturation mensuelle (avec calcul jusqu'à 200 membres d'équipe) • Entreprise : pas de données tarifaires, mais une option de contact	
Limites d'usage (du modèle testé)	☒ Nombre de requêtes limité : ☒ Volume limité : ☐ Durée limitée : ☒ Fonctionnalités verrouillées : ☐ Autres limites : ☐ Aucune limite	Accès à certains articles coupé (couvert) par de la publicité en version gratuite.
Performance		
Qualité des sources (pertinence de la source)	☐ Très bonne ☒ Bonne ☐ Moyenne ☐ Faible	plutôt pertinentes ; le degré de pertinence est annoncé dans le résultat (et semble adéquat), ainsi que le nombre de citations ; trouve aussi des sources anciennes ; mais est loin d'être complet ; certaines références sont masquées par une publicité pour la version "Pro" ; beaucoup d'articles chinois pour des thématiques sans précision géographique.
Transparence des sources	☐ 100 % des références sont complètes et traçables ☒ Au moins 80 % des références sont complètes et traçables ☐ Moins 80 % des références sont complètes et traçables	presque toutes les sources citées sont bien référencées et retrouvables ; le rapport signale en quoi certaines

		sources sont moins proches du sujet (“Ce qui manque / Mises en garde”).
Présence de sources hallucinées	☒ Oui ☐ Non	Mais nous avons pu trouver des références inventées : bien rédigées mais inexistantes.
Transparence du raisonnement	☐ Communiquée ☒ Non communiquée	Il est sommairement affiché lors de la préparation du rapport ; des points d’attention sur la pertinence des résultats sont signalés. Le processus lui-même reste non-transparent.
Reproductibilité (si transparence du raisonnement accessible)	☒ Recherche reproductible ☐ Non reproductible	Variable en fonction des interactions avec l’IA pour préciser la requête ; dès lors différences dans les résultats. Algorithmes adaptatifs – modifie les résultats selon l’apprentissage progressif.
Capacité auto-critique	☒ Détecte biais ☒ Suggère contre-arguments ☒ Propose des questions de clarification ☐ Non pertinent	
Ergonomie et usage		
Accessibilité linguistique	☒ Français ☒ Anglais	Question : FR/EN ; réponse : EN
Interface utilisateur	☒ Intuitive ☐ Assez complexe ☐ Complexe	
Option de personnalisation	☒ Oui ☐ Limitée ☐ Non	alertes, sauvegarde, requêtes
Aide et support en ligne	☒ Disponible ☐ Non disponible	vidéo + FAQ

Interopérabilité & compatibilité		
Formats acceptés (input)	☐ PDF ☐ DOCX ☐ BibTeX ☐ RIS ☐ DOI ☐ URL	Limité
Formats acceptés (export)	☒ PDF ☐ DOCX ☐ EXCEL/CSV ☐ BibTeX ☒ RIS	Formats standards (CSV, BibTeX and RIS).
Connecteurs disponibles	☐ API ☐ Extensions ☒ Export vers bibliographie ☐ LibKey ☐ Autre :	Citation manager : Zotero, EndNote; screening platform (e.g., Covidence).
Compatibilité écosystème académique	☒ PubMed ☒ ArXiv ☒ Scopus ☐ Autre :	
Aspects éthiques & juridiques		
Respect des données personnelles	☒ RGPD respecté ☐ Anonymisation ☐ Non respecté ☐ Aucune information trouvée	Collecte les infos personnelles pour communiquer avec les utilisateurs (envoi du rapport) et répondre aux questions (effectif au lancement du produit ; pas de réponse à des mails envoyés durant le test) ; possible partage des données (e-mail, cookies, journaux) avec des tiers “si nécessaire “ (confirmé par la <i>Privacy Policy</i>).
Biais algorithmiques connus	☐ Explicités ☒ Non explicités	Articles non pertinents ; références fausses.