



HAL
open science

”#WorkInProgress : IA générative et outils de recherche de littérature académique”. URFISTinfo

Aline Bouchard

► To cite this version:

Aline Bouchard. ”#WorkInProgress : IA générative et outils de recherche de littérature académique”. URFISTinfo. 2025. hal-04960003

HAL Id: hal-04960003

<https://hal.science/hal-04960003v1>

Submitted on 21 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

#WorkInProgress : IA générative et outils de recherche de littérature académique

Aline Bouchard, URFIST de Paris, 20/02/2025

Article publié sur *URFISTinfo*, A. Bouchard. « #WorkInProgress : IA générative et outils de recherche de littérature académique ». *URFISTinfo*. 20 février 2025. <https://urfistinfo.hypotheses.org/4645>.

La recherche d'informations et la revue de littérature académique n'échappent pas à la récente vague d'intelligence artificielle générative, promesse d'assistance et de productivité. Durant les deux années écoulées, l'engouement autour de ChatGPT est venu mettre en lumière l'intégration de fonctionnalités d'intelligence artificielle (IA) générative dans les outils de recherche. Porté d'une part par les progrès sémantiques des grands modèles de langage (LLM), d'autre part par une diversification et de nouvelles capacités des outils, le paysage de la recherche d'information a déjà bien changé, entre tchatbots, moteurs de recherche et outils de recherche bibliographiques « augmentés ». Et les annonces de ces derniers mois, autour des modèles de raisonnement ou de recherche approfondie, augurent de nouveaux changements encore. Même si l'on manque encore de recul sur ces technologies récentes et évolutives, il est d'ores et déjà possible d'identifier un certain nombre de points-clés et de questions autour de l'intégration de ces nouveautés dans le cadre académique. État des lieux et perspectives deux ans après la sortie de ChatGPT...

TL;DR - résumé

- À l'instar de ChatGPT, le modèle de tchatbots basés sur des grands modèles de langage (LLM) représente une nouvelle famille d'outils pour la recherche documentaire.
- Les limites de ces outils observées en 2023-2024, dont les fameuses hallucinations et le manque de connaissances en temps réel, sont amenées à diminuer avec l'intégration de données plus larges et des capacités plus poussées (recherche web, capacités de raisonnement et IA agentique).
- Les fonctionnalités d'IA et autres assistants permettant d'interroger en langage naturel et d'obtenir des réponses synthétiques deviennent une fonctionnalité de plus en plus standard tant dans les moteurs de recherche que dans les bases de données bibliographiques, nous faisant passer de la recherche traditionnelle par mots-clés à une recherche générative, sémantique et élargissant la recherche d'information à l'extraction d'informations.
- Avec les fonctionnalités *deep research*, les outils de recherche, devenus ces dernières années de plus en plus des moteurs de réponse, deviennent désormais des agents de synthèse.
- Les fondamentaux de la recherche documentaire dans un cadre académique évoluent peu : méthodologie, esprit critique, usage responsable. L'intégration d'IA générative pose néanmoins de nouvelles questions sur la fiabilité et la transparence des résultats ainsi que la dépendance aux outils.
- Entre promesses et prudence, nous manquons encore de retours critiques pour prendre toute la mesure du changement dans un paysage en cours de structuration.

TABLE DES MATIERES

La recherche documentaire à l'heure sémantique et conversationnelle	2
L'IA pour dépasser les limites des outils traditionnels	2
IA, quelles intelligences artificielles ?	3
Limites des tchatbots de type ChatGPT pour la recherche d'information	4
Vers une recherche générative du quotidien ? Exemple des moteurs de recherche.....	10
L'IA générative : une fonctionnalité de plus en plus standard.....	13
LLM, RAG et recherche de littérature	13
Intégration d'IA générative dans les outils de recherche bibliographique.....	13
Une nécessaire réflexion sur les compétences de recherche	15
Changements dans la recherche documentaire.....	15
Développement de la recherche sémantique générative.....	17
« <i>Illusions of understanding</i> » et autres problèmes épistémiques	20
« <i>Not for the lazy</i> »	23
Un marché en cours de structurations	26
L'or noir des données académiques	26
<i>Business as usual</i> ?.....	27
Une offre en devenir.....	29
Conclusion provisoire pour les formateurs	31
Déclarations d'usages IA et remerciements	33
Pour aller plus loin	34

LA RECHERCHE DOCUMENTAIRE A L'HEURE SEMANTIQUE ET CONVERSATIONNELLE

L'IA pour dépasser les limites des outils traditionnels

L'un des objectifs de l'intégration de l'IA dans les outils traditionnels est d'abord de « dompter » la littérature académique ([Andy Extance](#)). On le sait : que ce soit pour faire une revue de littérature ou compléter une bibliographie, la recherche d'informations académiques est un processus long et potentiellement fastidieux. C'est *a fortiori* le cas dans le [contexte académique actuel](#), de plus en plus concurrentiel, pris entre l'augmentation constante du nombre de productions annuelles et l'injonction à la publication. Dans ce contexte, comment avoir une vue rapide sur un sujet qu'on ne connaît pas et en repérer les articles plus importants ? ; comment faire face au déluge informationnel et identifier non seulement les travaux, mais aussi des tendances ou des chercheurs, de manière rapide et efficace ? ; comment identifier des pistes moins couvertes ?

Mais l'autre objectif de l'intégration de l'IA dans les outils de recherche bibliographique est de **faire face aux limites des outils documentaires traditionnels**. La méthode de recherche documentaire, avec ses champs prédéfinis et ses mots-clés, ne correspond pas à la démarche scientifique, construite avant tout sur des hypothèses et des interrogations. Pour pouvoir chercher dans une base de données bibliographiques, un moteur de recherche scientifique ou un catalogue de bibliothèques, il convient de transformer son hypothèse en vocabulaire spécifique et en équation de recherche. Et si la recherche simple, plein texte ou fédérée, s'est imposée depuis une quinzaine d'années, elle n'en pose pas moins la question de l'adaptation du chercheur à l'outil : comment être sûr d'avoir employé les bons termes, contrôlés ou non, et de ne pas passer à côté des résultats les plus pertinents découverts ? ; comment chercher à partir d'hypothèses ? ; comment faire quand on ne connaît pas le vocabulaire spécifique ou quand la terminologie n'est pas encore totalement arrêtée pour des sujets nouveaux ?

Dans le domaine de la recherche documentaire, comme dans les autres domaines de la recherche scientifique (cf. parmi d'autres [OCDE, 2023](#) ; [European Research Council, 2023](#) ; [Commission européenne, 2023](#) ; [The Royal society, 2024](#)), la promesse de l'IA est d'« **augmenter** » les **capacités** de traitement de l'information, d'« assister » le chercheur dans son travail quotidien et donc de lui assurer une meilleure efficacité et une plus grande productivité. En facilitant et en accélérant la revue de littérature, l'intégration de l'IA dans les outils documentaires serait alors un moyen de découvrir de nouvelles idées et, finalement, de faire progresser la recherche [plus rapidement](#) : « Même si chaque outil dessert une niche spécifique, ils fournissent tous aux chercheurs un regard différent sur la littérature académique par rapport aux outils conventionnels comme PubMed et Google scholar » ([Andy Extance](#)). Par ailleurs, dans le cadre d'une revue de littérature large, ce développement d'[assistants de recherche avec de l'IA](#) (*AI research assistants*) serait un moyen de lever un certain nombre de biais humains [en termes de rigueur et de précision](#) dans la démarche.

IA, quelles intelligences artificielles ?

Le succès public et médiatique de ChatGPT depuis son ouverture fin 2022 ne doit pas faire oublier l'histoire déjà ancienne de l'intelligence artificielle. Si l'on regarde les seules bases de données bibliographiques numériques, les suggestions automatiques ou le classement par pertinence sont déjà des illustrations anciennes des capacités des systèmes informatiques à extraire des données, à les mettre en relation et à les utiliser – Google scholar par exemple a ouvert en 2004. Plus récemment, le milieu des outils bibliographiques a connu dans les années 2010 une vague d'**IA sémantique**, illustrée par l'ouverture du bien nommé [Semantic scholar](#) en 2015. Au moment où le moteur de recherche Google développait son *Knowledge graph*, Microsoft ouvre alors une nouvelle version de son moteur de recherche académique [Microsoft academic search](#), basé sur des technologies de recherche sémantique comme la compréhension du langage naturel (NLP – *natural language processing*) et la mise en relation de ces données dans des graphes de connaissances (*knowledge graphs*). Le vrai saut qualitatif pour les outils de recherche, ce n'est plus de fournir des références qui correspondent aux mots-clés saisis mais bien de comprendre le contexte et l'intention de la requête (que cherche l'internaute ?) et le sens des documents (ces documents répondent-ils à la question ?). En somme, les idées plutôt que les mots. On en trouve des prémices, par exemple avec les fonctionnalités de [Semantic scholar](#) ou de [Scite](#) permettant de situer une citation dans son contexte.

De tels progrès n'ont bien sûr été rendus possibles qu'avec le développement des contenus numériques et des technologies, notamment autour du traitement automatique des langues et de l'apprentissage automatique, des années 2010. D'une part, le développement d'internet fournit ainsi toujours plus de masses de corpus linguistiques (« [web as corpus](#) ») pour entraîner les machines. D'autre part, le développement des grands modèles de langage (**LLM** ou *large language models*) à partir des années 2017-2019 permet de dépasser les simples rapprochements au niveau des termes et de comprendre le fonctionnement des langues et toutes leurs subtilités. Alors que les outils précédents recherchaient les mots-clés de manière lexicale selon une logique de [query expansion](#) (chaîne de caractère identique, lemmatisation, racinisation / *stemming*), les [nouvelles technologies](#) permettent une approche plus sémantique et syntaxique, notamment grâce à des [plongements lexicaux](#) (*word embedding*), afin de comprendre le sens mais aussi le contexte des mots, notamment leur ordre et leurs relations. C'est ce qu'exprime bien l'un des premiers outils académiques gratuits de ce type, l'[initiative BioMed Explorer](#) de Google sortie pendant la pandémie de Covid-19. Basé sur le LLM BERT, « BioMed Explorer applique une compréhension sémantique du contenu des publications pour en tirer des réponses et sélectionner des extraits et des preuves pour l'utilisateur. En d'autres termes, l'outil s'efforcera de mettre en évidence les passages qui répondent réellement à votre

question, et non pas seulement de renvoyer des extraits correspondant à des mots-clés, qui peuvent être aléatoires » ([Google](#)).

La particularité des outils comme ChatGPT et de l'**IA générative**, c'est d'appliquer ces LLM [à deux niveaux complémentaires](#) : la « compréhension » du texte et la génération de contenu. A cette fin, ces IA établissent des modèles statistiques et probabilistes du langage humain et sont alors capables de produire de nouveaux contenus à la volée. Comme les tchatbots antérieurs de type Alexa ou Siri et contrairement aux bases de données bibliographiques de l'époque, elles sont capables d'interpréter des requêtes en langage naturel, par exemple une question. En revanche, elles vont plus loin que les tchatbots première génération dans la mesure où leurs bases de données d'entraînement sont beaucoup plus larges et où leur fonctionnement probabiliste leur permet de produire du contenu inédit au-delà de leurs données d'entraînement. Par ailleurs, leurs capacités itératives et de mémorisation en font de vrais outils conversationnels.

GPT (OpenAI), Gemini (Google), Claude (Anthropic), LLaMA (Meta), Mixtral (Mistral AI)... Les modèles de langage se multiplient et sont souvent accessibles de différentes manières. Si l'on prend le seul cas du modèle GPT, on peut y accéder, par exemple, par le tchatbot du site officiel, *via* les outils Microsoft (Copilot pour Bing ou dans la suite Office), par d'autres plateformes comme Poe ou encore *via* des API dans d'autres écosystèmes. On retiendra cependant trois choses : la sortie régulière de nouveaux modèles chez les producteurs, des modèles financiers généralement de type freemium et, dans le cas des tchatbots, la nécessité d'avoir un compte – permettant ainsi de récupérer les données personnelles de l'utilisateur et de limiter le nombre de requêtes moyenne.

Derrière le mot fourre-tout « IA », ce sont souvent **plusieurs modèles différents qui co-existent** dans un même service. Par exemple, ce ne sont [pas moins de 20 systèmes d'IA différents](#) qui tournent dans le moteur de recherche académique Consensus. Ces services en utilisent à [différents niveaux](#) : traitement de la requête en langage naturel, similarité sur les mots-clés ou sur l'*abstract*, compréhension du contexte d'une citation, synthèse, réponse en langage naturel... Les outils n'utilisent pas nécessairement l'ensemble des techniques possibles en fonction des besoins.

Limites des tchatbots de type ChatGPT pour la recherche d'information

Basées sur de grands modèles de langage, les IA génératives textuelles sont spécialisées dans le traitement et la « compréhension » du langage naturel. Elles sont donc capables de **mener à bien des tâches linguistiques** : rédiger du texte à partir d'une simple question, classifier des concepts, d'analyser, traduire et synthétiser des textes, de manière autonome. Preuve de cette diversité d'usage, une étude récente identifie trois catégories d'usages principaux chez les chercheurs : l'IA générative comme « cheval de trait », chargé de tâches répétitives ou fastidieuses, l'IA générative comme « assistant linguistique » pour l'aide rédactionnelle et éditoriale, l'IA générative comme « accélérateur de recherche » pour gagner en productivité ([Jens Peter Andersen et al., 2025](#)). Ce ne sont pas des outils de connaissances ; ce ne sont pas des outils de recherche d'information ; ce ne sont pas des outils de calcul ; ce ne sont pas des outils de logique. Aussi nombre d'erreurs attribuées à ChatGPT comme lui demander de faire des calculs ou de compter le nombre de lettres dans un mot sont en fait des erreurs d'usage. De même quand lors de la sortie de ChatGPT, on s'était aussi beaucoup amusé à lui demander le nom des [5 femmes Première ministre en France](#).

En aucun cas, ces techniques de LLM n'ont été pensées pour remplacer, seules, Google. Contrairement aux moteurs de recherche ou aux bases de données, elles ne s'appuient pas sur des index préconstruits pour produire des réponses individuelles sourcées, mais utilisent leurs données d'entraînement pour fournir des contenus linguistiques qui semblent les plus probables par rapport à ces données. Elles n'ont pas non plus vocation à contrôler les informations

fournies. Sans même évoquer les questions que ces outils posent en termes sociaux, écologiques ou stratégiques, les limites souvent citées concernant les LLM pour la recherche d'information sont les suivantes :

- d'un point de vue général :
 - des données d'entraînement **biaisées** : ces données étant massivement issues d'internet, elles reprennent donc les biais des contenus publiés en ligne, majoritairement anglophones et occidentaux ; les autres aires linguistiques et géographiques sont moins représentées. Par ailleurs, les contenus sont souvent bridés lors de la phase d'entraînement pour ne pas répondre sur des sujets sensibles. Enfin, ces outils pourraient favoriser une certaine homogénéisation du fait de leur rendu probabiliste des résultats, voire [amplifier les biais déjà existants dans la recherche](#) comme l'effet Matthieu,
 - des données d'entraînement **datées** : la version GPT-3.5, fermée en juillet, avait des données s'arrêtant en 2021 ; la version GPT-4 qui l'a remplacée a des données s'arrêtant [en octobre 2023](#),
 - des réponses potentiellement **superficielles** : selon les sujets, les modèles de langage peuvent disposer de plus ou moins de données d'entraînement, d'où des réponses manquant de profondeur, notamment sur des sujets de niche très pointus. Cependant, il est impossible pour l'internaute de le savoir *a priori*,
 - des réponses **non reproductibles** : les LLM ne fonctionnent pas sur des index préétablis mais produisent leurs réponses à la volée, de manière statistique probabiliste aléatoire (les fameux « perroquets stochastiques » d'[Emily Bender et al.](#)). Une même question peut donc amener des réponses différentes, tout comme deux questions similaires, mais rédigées sensiblement différemment : « Non seulement, [ChatGPT] est par construction indifférent à la vérité de ses assertions, mais de plus, il les adapte à son interlocuteur. Il n'y a donc, même pas, de cohérence interne. C'est le relativisme à l'état pur » ([Claire Mathieu et Jean-Gabriel Ganascia](#)),
 - des réponses potentiellement incorrectes voire **fausses** : si les LLM peuvent être puissants pour des éléments linguistiques de concept, de vocabulaire, d'organisation ou de traduction, ils pèchent souvent pour les informations factuelles, souvent fausses ou incomplètes. Les fameuses « hallucinations » de ChatGPT sont la simple conséquence de son [fonctionnement probabiliste](#), rassemblant parfois des éléments de réponses sans aucun rapport entre eux. Lors de la sortie de ChatGPT, on avait ainsi beaucoup parlé des [œufs de vache...](#) D'autant que les LLM ne sont souvent pas capables de reconnaître leur ignorance et répondent même en l'absence d'informations vraies, simplement parce que la probabilité existe sur l'échelle des possibles. Et si la situation s'améliore ce n'est [pas encore pleinement satisfaisant](#)... Si les LLM ont fait des progrès ces derniers mois, il est toujours bon de se rappeler ce que dit [Alexei Grinbaum](#) : « Toute information donnée par un tchatbot ne peut être vraie que par accident ». On le voit aussi avec le [phénomène de régénération successives](#) ; les LLM sont encore [incapables de s'auto-corriger](#). « Le taux d'erreur du meilleur logiciel actuel, GPT-4, se situe [entre 2,5% et 25%](#) » ([Thibault Prévost](#)). Par ailleurs, cette pratique probabiliste peut entraîner des biais en direction d'opinions populaires voire de théories scientifiquement fausses ou simplement humoristiques, fréquentes dans les données internet utilisées mais néanmoins reprises dans l'entraînement du modèle et [ses réponses](#), malgré des systèmes de modération humaine (RLHF ou *reinforcement learning from human feedback*). Enfin, la multiplication de contenus de piètre qualité voire faux créés

- par des IA sur internet risque de polluer les futures données d'entraînement et amener une [régression des LLM](#),
- un **biais de confiance ou d'ancrage pour l'internaute** : dans la mesure où les modèles de langage produisent du contenu bien écrit et bien présenté, on est enclin à faire confiance à l'outil, [indépendamment de la justesse et de la précision des résultats](#). Et plus la confiance dans l'outil augmente, [plus la pensée critique décroît](#). C'est particulièrement vrai avec des contenus éditoriaux. C'est ce que décrit parfaitement [Olivier Ertzscheid](#) : « Il serait trop long / complexe de vérifier les différents éléments présentés et qui peuvent poser isolément problème mais qui ainsi concaténés ensemble paraissent faire globalement sens ». Cette impression peut être renforcée si ledit outil s'est révélé pertinent sur des échanges précédents. Or, comme indiqué précédemment, les capacités des LLM varient selon les sujets ;
 - d'un point de vue plus académique :
 - des données d'entraînement avant tout **grand public** : les fournisseurs de LLM gardent généralement le secret sur le détail des sources utilisées, notamment parce qu'elles auraient souvent été récupérées [sans autorisation](#). Les données disponibles pour [GPT-3](#) donnent cependant une certaine idée des sources concernées, comme par exemple, le corpus Common Crawl, les forums Reddit et Wikipédia anglais. Néanmoins, si un corpus comme Common Crawl intègre bien des [données de sites académiques](#) (ex. : stanford.edu, mit.edu, harvard.edu, cnrs.fr, inria.fr, arxiv.org) et si le jeu de données C4 intègre des publications de [PLOS et Frontiers](#), il convient de savoir [ce qui est réellement utilisé dans ce corpus](#) car [intelligible](#), les sociétés d'IA ne fournissant généralement aucune information sur les corpus couverts (« *It's a product. Not science* », [Prithviraj. Ammanabrolu](#)). Par ailleurs, il s'agit normalement de données du web visible et crawlable. Comme pour les moteurs de recherche, toute une part de la recherche scientifique, accessible derrière des *paywalls*, échappe donc aux LLM. A côté des problèmes de dates déjà mentionnées, cela signifie que ces outils ne sont pas du tout représentatifs des actualités et des derniers développements dans un domaine de recherche ou un autre. On notera donc avec intérêt qu'à l'instar des partenariats signés au premier semestre entre OpenAI et des médias comme *Le Monde*, un certain nombre d'éditeurs ont [annoncé à l'été 2024](#) une ouverture de leurs catalogues à des entreprises IA afin d'en assurer une plus grande visibilité et contribuer à améliorer des modèles IA – nous y reviendrons ci-dessous. Comme [on le dit chez Wiley](#) : « Il est dans l'intérêt public que ces technologies émergentes soient entraînées sur de l'information de haute qualité et fiable »,
 - des **assertions probabilistes** : non contents de mettre en valeur les réponses les plus probables au détriment de réponses plus pointues indépendamment de leur qualité intrinsèque réciproques, ces [outils procèdent d'une approche inductive](#) (i.e. la généralisation de cas particuliers), en décalage avec l'approche logique et déductive (*evidence-based*) de la démarche scientifique,
 - des **assertions non sourcées** : les modèles de langage ne sont nullement des bases de connaissance. En interagissant avec eux, on n'interroge pas des savoirs structurés comme des index ou des listes de références dans un moteur de recherche ou une base de données bibliographiques. Leur fonction n'est pas de fournir des informations vérifiables, avec URL ou références bibliographiques à l'appui, mais simplement de produire du contenu textuel vraisemblable, présenté dans une langue correcte. Par ailleurs, les données d'entraînement sont décomposées en de telles petites unités (les *tokens*) qu'il n'est pas toujours

possible d'identifier les sources d'origine dans la reconstitution probabiliste finale,

- des **citations et des références scientifiques fausses** : comme pour le texte, les modèles de langage se révèlent incapables, utilisés seuls, de produire des références scientifiques réelles. « ChatGPT est un bon poète, mais un mauvais bibliographe » ([Véronique Mesguich](#)). La situation est d'autant plus préoccupante que l'on voit désormais se multiplier des listes de références inexistantes probablement tirées tout droit de ChatGPT, chez les étudiants comme [chez les chercheurs](#). A mesure que ces fausses références alimenteront les *preprints* voire les publications peu regardantes et qu'elles seront reprises par les moteurs de recherche académiques comme Google scholar, c'est une véritable [épidémie de fake references](#) qui est en germe, posant des questions pour l'intégrité de la recherche scientifique mais aussi l'[identité numérique des chercheurs](#),
- des **réponses interprétées** : jusqu'ici les moteurs de recherche et les bases de données bibliographiques, fournissaient en priorité des sources variées (liens URL, références bibliographiques) et c'était à l'internaute de faire le travail d'analyse et de synthèse, de construire son parcours et son interprétation. Avec le développement de l'IA générative, ce travail est délégué à l'outil, devenu médiateur de savoirs construits. Ce sont le tchatbot et ses modèles de langage qui rédigent la réponse, qui éditorialisent les contenus dans un monde sans source. Dès lors comment parvenir à sortir de ce cadre « monophonique », comment « produire un contre-discours ou [...] en discuter le fondement [quand] le cadre énonciatif initial postule que je ne dispose pas des connaissances me permettant de le faire (puisque ma requête est l'aveu même de cette – relative – ignorance) » ([Olivier Ertzscheid](#)) ? Dans un cadre scientifique, cela pose notamment la question de la représentativité des réponses par rapport à l'ensemble de la recherche existante et de la pluralité des points de vue. Cela peut également entraîner biais de confirmation et bulles de filtre : la requête initiale, le prompt, peut refléter ses propres croyances et l'IA a tendance à proposer du contenu n'allant que dans un sens particulier en s'adaptant à l'intention perçue, voire à conforter l'internaute dans ses positions,
- enfin, élément préoccupant : le **mésusage de ces outils**. Alors que ces outils ne sont pas des outils de recherche d'informations pleins et entiers, les IA génératives sont très utilisées pour la recherche d'informations. L'usage le plus fréquent dans la population française générale serait pour faire des recherches comme dans un moteur de recherche (31% des usages, [enquête IFOP-Talan, 2024](#)) et 42 % des étudiants au moins les utiliseraient comme une source de documentation pour rédiger leurs travaux ([sondage Compilatio, 2023](#)). Autre élément préoccupant, les jeunes générations sont plus nombreuses à penser que les IA génératives ne nécessitent pas de travail humain de relecture et de vérification (30 % des 18-24 ans contre 19 % de la population, [enquête Impact AI – Vivavoice, 2023, p. 82](#)) et reprennent les résultats sans les modifier (61 % des 25-34 ans contre 44 % des utilisateurs en général, [enquête IFOP-Talan, 2024](#)). Comme le dit ChatGPT lui-même : « ChatGPT peut faire des erreurs. Envisagez de vérifier les informations importantes ». La prudence est donc de mise en utilisant ce genre d'outils et il convient de savoir bien [distinguer les cas d'usages](#) adéquats ou totalement inadéquats.

En l'état actuel des choses, il convient donc de bien comprendre ce que sont et ce que peuvent les IA génératives de type ChatGPT pour la recherche d'information au moment où des IA de

type ChatGPT se développent à destination du monde universitaire, comme l'offre [ChatGPT Edu](#). Ce sont de nouveaux éléments dans la panoplie d'outils professionnels qui ne doivent pas faire perdre les bonnes pratiques recherche d'information, avec leurs atouts (ex. : [expression des besoins](#), [champs lexicaux](#), [définition d'une stratégie de recherche](#)). Mais aussi avec leurs précautions d'usages (ex. : connaissance des outils, méthodologie documentaire, évaluation de l'information), sous peine de favoriser une [recherche de mauvaise qualité](#).

Si ces éléments sont encore vrais début 2025, **quelles sont les perspectives pour les tchatbots** ? De fait, l'IA générative telle que l'on connaît depuis l'ouverture de ChatGPT fin 2022 est « la pire IA que nous n'ayons jamais utilisée » ([Katherine Aydelott, 2024](#)) et les récentes actualités du secteur pourraient porter de nombreuses promesses pour la recherche – nous indiquons ici en priorité des exemples issus d'OpenAI, mais tous les acteurs vont dans le même sens :

- **meilleure compréhension sémantique** : arrivés au bout du web *crawlable* facilement, les fournisseurs d'IA se tournent désormais vers les secteurs de niche à forte valeur ajoutée ([médias](#), [éditeurs scientifiques](#)) pour intégrer des sujets plus pointus, d'actualités et académiques, dans leurs modèles linguistiques. Des progrès techniques sont également nécessaires pour lutter contre les biais probabilistes et permettre une meilleure « compréhension » du monde ([world models](#)) ;
- meilleures capacités de raisonnement et de vérification : les LLM apparus dans le sillage de ChatGPT fonctionnent sur le même modèle : ils produisent du contenu probabiliste, sans compréhension intellectuelle. Afin de limiter les erreurs et approximations inhérents à ce fonctionnement, les chercheurs ont alors développé des techniques permettant d'améliorer leur raisonnement en une approche par étapes ([principe des chaînes de pensée, chains of thought – CoT](#)), par exemple en les entraînant [sur des tâches de raisonnement du monde réel](#). Comme l'illustrent les [séries de modèles o1 et o3](#) d'OpenAI, le modèle [Gemini 2.0 Flash Thinking](#) de Google ou [encore d'autres](#), il s'agit de « passer plus de temps à réfléchir avant de répondre » à des requêtes complexes, notamment dans les domaines scientifiques ou mathématiques. Plus performants que les LLM classiques (*raw LLM*), ces **modèles dits « de raisonnement »** (*reasoners / reasoning models*) semblent apporter des réponses plus longues et construites. Surtout, en termes d'évaluation de l'information, ils font [baisser le pourcentage d'hallucinations](#) et font [évoluer la manière actuelle de rédiger les prompts](#). Mais pur l'instant, une interrogation demeure : « Peut-on raisonner sans vraie compréhension ? » ([Kai Riemer](#)). De fait, « ce n'est pas exactement un nouveau modèle de langage, mais un modèle de "raisonnement" ajouté au modèle de langage actuel (GPT-4o). [...] Cela reste un "transformer", c'est à dire un modèle de langage autorégressif (qui prédit sa réponse). Et Il est toujours susceptible de se tromper » ([Benoît Raphaël](#)) ;
- meilleure intégration dans le quotidien des utilisateurs : le tchatbot tel qu'on l'a connu avec ChatGPT et ses données textuelles est de moins en moins courant. D'une part, les formats pris en charge tendent à se diversifier dans la même interface (texte, PDF, images, sons...) – le « o » de la série de ChatGPT (OpenAI o1, o3) vaut d'ailleurs pour « omnimodal ». D'autre part, les principaux acteurs (ChatGPT, Claude, Le Chat de Mistral) proposent des fonctionnalités d'éditorialisation de contenu directement dans l'interface (fonctionnalité « *canvas* »), voire la possibilité de personnaliser des assistants (fonctionnalité « *agent* »). Mais la principale modification est le progressif **déploiement de fonctions de moteurs de recherche** directement dans

l'interface du tchatbot (fonctionnalité « *search* »), alliant données des LLM et navigation internet. Une telle évolution vient désormais brouiller la distinction entre le tchatbot n'utilisant que les données fermées du LLM et le moteur de recherche « augmenté » et son index internet – nous y reviendrons plus bas ;

- meilleures interactions avec l'utilisateur : après le buzz autour de l'IA générative en 2023-2023, on ne compte plus cet hiver les articles autour de l'[IA « agentique » \(*agentic AI*\)](#), capables non plus seulement de générer des réponses (*outputs*), mais d'agir : dans le cas des tchatbots, « les agents d'IA sont des systèmes autonomes alimentés par de grands modèles de langage (LLM) qui, à partir d'instructions de haut niveau, peuvent planifier, utiliser des outils, effectuer des étapes de traitement et agir pour atteindre des objectifs spécifiques. Ces agents exploitent des capacités avancées de traitement du langage naturel pour comprendre et exécuter efficacement des tâches complexes » ([Mistral](#)). Si les cas d'usage des IA agentiques sont extrêmement nombreux, on assiste au développement d'[agents plus spécialisés pour la recherche d'information](#), mêlant l'IA générative pour la création de contenus et l'IA agentique pour les actions. Et si les exemples d'agent jusque-là présentaient souvent des usages grand public, les tout récents [Deep research dans ChatGPT](#), [Deep research dans Gemini](#) ou encore [STORM de Stanford](#) se situent bien au croisement de la recherche d'information sur internet et d'une recherche à étapes, d'un niveau plus avancé. Chez ChatGPT, on mêle ainsi notamment LLM, modèles de raisonnement, navigation web et outils Python : « Vous lui fournissez un prompt, et ChatGPT va trouver, analyser et synthétiser des centaines de sources en ligne pour créer un rapport complet d'un niveau d'un analyste de recherche », tout en étant capable de s'adapter en fonction des informations trouvées ([OpenAI](#)). En opposition avec un tchatbot « classique », il est alors capable d'indiquer ses sources ainsi qu'un résumé de la réflexion suivie (chaîne de pensée). Il est vrai cependant que ce travail prendra plus de temps qu'une interaction avec ChatGPT classique, de 5 à 30 minutes selon la complexité de la tâche, mais c'est une prérequis pour une meilleure qualité.

Néanmoins, si l'on voit plus clairement les pistes de développement pour l'utilisateur, la situation est un peu plus compliquée du côté des **acteurs** mêmes. Coup de tonnerre en effet fin janvier quand sort le tchatbot DeepSeek, et son modèle de raisonnement R1, dans un paysage jusque-là dominé par OpenAI, Meta ou Google. Performance de même niveau que ses concurrents, codes en *open source* pour un coût de production et un poids écologique moindres, [il n'en faut pas plus](#) pour enthousiasmer les communautés IA et faire trembler les marchés financiers. De fait, les [innovations techniques derrière DeepSeek](#) (RL – *reinforcement learning*, [MoE](#) – *mixture of experts* et MLA – *multi-headed latent attention*, prédiction multi-tokens) ne sont [pas nécessairement nouvelles](#) mais montrent que « de petits acteurs [comparés aux GAFAM] peuvent faire des choses plus économes et innovantes » ([Laurent Daudet](#)). Deux questions se posent avec l'arrivée de cet outsider chinois. D'une part, les modèles économiques de ces services : là où OpenAI avait [annoncé](#) n'être toujours pas rentable et prévoir une augmentation de ses tarifs d'abonnement, DeepSeek montre que l'on peut produire des LLM nouvelle génération à moindre coût. D'autre part, cela remet en lumière la question de l'éthique de l'IA générative, entre [censure](#), [biais culturels](#) des IA, [fiabilité](#), sort des données personnelles et récupération de données sensibles, à l'heure où l'expérience française [Lucie](#) a fait long feu et où [seul Mistral](#) représente encore l'Europe [dans la course, pour l'instant](#).

Vers une recherche générative du quotidien ? Exemple des moteurs de recherche

Signe révélateur : à côté du simple « IA », surfant sur la mode actuelle, nombre de ces outils prennent le nom d'« assistant » ou de « copilote ». De fait, l'IA générative est de plus en plus intégrée dans l'écosystème de l'internaute, d'autres environnements applicatifs, des bases de connaissance, des navigateurs internet, des réseaux sociaux, etc., marquant l'entrée dans une **nouvelle ère de l'assistance**. On le voit avec Copilot (basé sur des GPT) chez Microsoft auquel on accède aussi bien par un site particulier, que par le moteur Bing, le navigateur Edge, Office 365 et certains PC Microsoft.

Du côté des moteurs de recherche, cela fait [plus de 10 ans](#) que les acteurs ont l'ambition de fournir des réponses plutôt que de simples listes d'URL. C'est l'ambition à l'œuvre derrière les évolutions sémantiques, illustrées par les *knowledge graphs* et les *rich snippets*, du milieu des années 2010, et qui ont progressivement transformé les moteurs de recherche en moteurs de réponses factuelles. C'est cette ambition aussi qui a poussé Google à développer des technologies permettant une vraie recherche conversationnelle (comprendre le contenu des documents et l'intention de l'internaute), illustrée par la technologie des *transformers* – le « T » du sigle GPT – [introduite en 2017](#). Dans la foulée de l'ouverture de ChatGPT, on a alors vu les **moteurs de recherche** intégrer de l'IA générative, comme des résumés instantanés synthétisant plusieurs résultats de recherche et indiquant les sources utilisées. De plus en plus, les moteurs de réponses (factuelles) deviennent des « moteurs de synthèse ». A côté d'*outsiders* (Brave, Qwant, Perplexity), les grands acteurs de la recherche et de l'IA sautent tous progressivement le pas mi-2024. En mai, Google annonce le déploiement progressif au niveau mondial des [AI overviews](#) (AIO ou Aperçus IA, anciennement SGE – *Search generative experience*) dans son moteur de recherche ; en juillet, Bing annonce sa [recherche générative](#), nouvelle interface du moteur de recherche mettant en avant du contenu IA ; en juillet encore, OpenAI, la maison-mère de ChatGPT, lance son prototype de moteur de recherche [SearchGPT](#), basé partiellement sur l'index Bing et des [technologies propres](#), qui doit combiner les capacités actuelles de conversation à une recherche web mais dont le prototype [n'échappe pas davantage](#) aux erreurs – [ChatGPT search](#) a depuis été déployé pour tous, même sans compte ChatGPT. En s'appuyant sur des sources web, ChatGPT est de plus en plus capable de distinguer l'intention de l'internaute : s'agit-il d'une requête pour rechercher de l'information (sous forme de question) ou d'une requête pour mener à bien une tâche (sous forme de prompt) ?

Si cette fonctionnalité est faite pour rester, au vu des évolutions récentes des moteurs de recherche, il n'en reste pas moins **trois questions centrales pour la recherche d'information**, qui viennent renforcer les questions que l'on pouvait se poser sur les moteurs de recherche eux-mêmes :

- les **données disponibles** : les algorithmes utilisés pour les tâches spécifiques du *search* et la couche LLM ne peuvent à eux seuls produire des données justes : l'outil est forcément tributaire des données disponibles sur internet, de leurs biais, de leurs ambiguïtés, de leur représentativité ou de leur complétude – d'où là aussi le problème des contenus derrière des ère des *paywalls* ;
- la couche générative : la couche générative pose deux questions : celle des contenus sélectionnés pour la réponse d'une part, celle de la compréhension et de la génération du langage naturel par LLM d'autre part. Au niveau du choix des contenus, quels sont les critères de sélection ? Même si [on affirme chez Google](#) que la technologie mise en place « identifie des résultats web de haute qualité qui corroborent les informations présentées dans le résultat », comment être certain que les sources sélectionnées pour la synthèse sont les plus pertinentes, les plus neutres et informatives ou qu'elles représentent différents points de vue sur le sujet ? Au niveau du langage naturel, on sait déjà que les LLM ont tendance à « halluciner ». Mais les LLM sont aussi sujets à de

[mauvaises interprétations](#) de la requête ou des nuances linguistiques. **L’outil n’est pas fiable à 100 %**, notamment parce qu’il n’est pas capable de discriminer intellectuellement les contenus, par exemple en cas de simple rumeur, de satire ou d’humour ([la fameuse pizza à la colle](#)), ni de vérifier lui-même les informations données ([les fameux chats sur la Lune](#)). Et encore actuellement, [Perplexity](#) ou [ChatGPT search](#) fournissent des réponses erronées ou incomplètes, notamment pour des sources fiables comme des médias. C’est d’ailleurs pour cette raison que Google, par exemple, ne prévoit pas de répondre sur des sujets sensibles comme les conseils médicaux ou financiers (principe du YMYL — *Your Money or Your Life*). Or, en proposant des contenus éditorialisés permettant l’économie d’aller aux sources d’origine, ces outils ont une responsabilité accrue par rapport aux simples listes de résultats ;

- la position de l’usager : l’ambition de Google est de « faire le travail pour vous » à l’heure du « [zéro clic](#) » ; : “*Google will do the searching, the researching, the planning, the brainstorming and so much more. All you need to do is just ask*” ([L. Reid](#)). Même son de cloche chez Perplexity : « Notre produit consiste à prendre des informations et à vous les présenter sans que vous ayez à aller nulle part » ([Henry Modisett](#)). Mais comment et peut-on se contenter de ces réponses ? [Selon Google](#), les premiers utilisateurs auraient été plus satisfaits de l’*AI overviews* que de la recherche classique et les liens présentés y seraient plus cliqués que sur la page de résultats. De son côté, Perplexity ne présente qu’un nombre restreint de réponses, illustrant ce qu’[Olivier Ertzscheid](#) nommait « un enfermement toujours plus ancré dans des “**silos**” de réalités interprétées ». Or, dans le même temps, certaines études laisseraient entendre qu’un outil comme Google *AI overviews* pourrait mettre en avant des pages moins visibles dans les résultats de recherche ([étude Authoritas](#) sur des mots-clés commerciaux) et favoriserait des visites sur une plus grande diversité de sites et plus longues, tout particulièrement chez les 18-24 ans ([Google](#)).

Quelles sont les **perspectives actuelles du développement** de l’IA générative dans les moteurs de recherche ? Comme pour les tchatbots, plusieurs axes sont concernés :

- les données interrogées : les acteurs du secteur savent en effet qu’il convient d’avoir un contenu important et de qualité pour espérer concurrencer Google. Par conséquent, plusieurs acteurs de l’IA signent des partenariats avec des fournisseurs de contenus : après les partenariats d’OpenAI courant 2024, c’est Perplexity qui multiplie ainsi des partenariats, comme [Numerama en France](#) – il est vrai qu’il [était menacé de procès](#) par certains journaux ; de son côté, Mistral signe [avec l’AFP](#). Du côté des contenus, on note aussi certains projets, [comme chez Perplexity](#), de pouvoir intégrer à la recherche RAG ses propres sources et données (bases de données internes, *cloud*, référentiels documentaires) au-delà de l’index du moteur de recherche pour une recherche globale ;
- le modèle économique : tout aussi crucial, il convient de trouver un système viable alors que l’utilisation même de l’IA augmente les coûts de fonctionnement – une recherche avec IA générative coûterait à Google [10 fois plus](#) qu’une recherche classique. Mais quel positionnement tenir alors que Google a imposé depuis 25 ans le principe du moteur de recherche gratuit ? Si ces outils proposent quasi-systématiquement des abonnements payants, ceux-ci sont rarement rentables. La tentation est forte alors de développer aussi des **publicités ciblées**, [comme Perplexity](#). Il est vrai cependant que le développement de modèles *open source* et les chutes considérables de coût de fonctionnement (de l’ordre de [-97 % en 18 mois](#) pour les tokens, chez [différents acteurs](#)), devraient permettre une intégration des LLM moins onéreuses dans les outils tiers ;
- le développement des outils d’IA agentique, comme mentionné précédemment avec les outils *Deep research* dans ChatGPT et Gemini. Perplexity, à son tour, vient d’annoncer [son Deep research](#). Dans l’immédiat, [il semblerait](#) que ces outils se fassent encore

l'écho de leur domaine de spécialité initial. L'outil Google, grâce à son index, est meilleur pour la recherche internet et les actualités ; [moins agile que son concurrent](#), il ne propose ainsi pas d'étape préalable pour mieux définir le contexte et les objectifs et reste plutôt statique dans sa démarche. L'outil OpenAI, lui, grâce à son modèle o3, est meilleur pour l'analyse et le raisonnement ; par son analyse préalable et son dynamisme, il se rapproche davantage du travail d'un expert. Dans l'immédiat, comme pour toute intelligence artificielle, l'outil connaît encore de [nombreuses limitations](#) (présence d'erreurs, d'approximations et d'hallucinations, absence de contenus derrière des *paywalls*) et fonctionne mal pour les éléments factuels. De fait, [comme on le dit](#) chez Corpora, [auto-déclaré « research engine »](#) – et [tout aussi fautif](#) : “It is NOT A SEARCH ENGINE”. Dans l'immédiat, quelle plus-value attendre de *Deep Research* et d'une manière générale ces agents « raisonneurs » pour les chercheurs ? De fait, si l'humain est toujours présent dans le processus, en amont (définition des besoins et des objectifs) et en aval (évaluation du rapport), le reste du processus est géré de manière autonome par l'agent. Dans l'immédiat, les avis sont partagés : si certains espèrent voir entrer la recherche scientifique dans une nouvelle ère de l'IA générative, favorisant la réflexion (“*The End of Search, The Beginning of Research*”, [Ethan Mollick](#)), d'autres sont plus nuancés, rappelant les limites inhérentes aux IA, même basées sur des sources identifiées et le risque général pour la science (“*Deep Research, Deep Bullshit, and the potential (model) collapse of science*”, [Gary Marcus](#)). Par ailleurs, si les sources en accès libre peuvent être crawlées, restera à voir comment les sources académiques seront prises en compte pour un cadre scientifique et remonteront dans les synthèses. La question est d'autant plus stratégique pour les chercheurs et le grand public, au risque de développer une [fausse démocratisation de l'expertise](#).

Plus généralement, là comme ailleurs, l'introduction d'un DeepSeek non-censuré dans Perplexity montre que les [lignes ne sont pas encore arrêtées](#), entre innovation, *open source*, conformité aux cadres locaux et collaboration entre les acteurs. D'un point de vue européen, enfin, on attend le développement des *AI overviews* dans Google, jusque-là [bloqué en Europe en raison de du Digital Markets Act \(DMA\)](#), visant à limiter la concurrence et la transparence en ligne. La guerre est donc loin d'être terminée dans le domaine des moteurs de recherche. Pour la première fois depuis longtemps, Google est descendu [en dessous des 90 % des parts de marché](#) durant l'automne, et les nouveaux acteurs Perplexity et ChatGPT fournissent [de plus en plus de trafic](#) aux sites internet. Il conviendra de suivre avec attention le match entre les différents acteurs pour identifier [les domaines où ils sont les plus pertinents](#), Google [continuant à tenir sa place de leader dans la plupart des cas](#). Rappelons pour terminer que l'aspect conversationnel et synthétique de l'IA n'est pas pertinent pour tous les besoins d'information. Si, pour reprendre la [taxonomie de Broder](#), on peut identifier trois besoins principaux (informationnel, navigationnel, transactionnel), seul le besoin navigationnel semble [hors du champ](#) de la recherche générative. Restera à voir comment les outils [seront capables d'identifier](#) les besoins des internautes et si nous aurons toujours la possibilité de forcer un mode de recherche sans IA générative alors qu'un Google par exemple semble [de plus en plus confiant](#) dans ses réponses. Restera à voir comment Google résistera à l'[image de marque acquise par ChatGPT](#) face à des [générations de plus en plus familières](#) du réflexe ChatGPT – « chatGPTisant » comme on « googlisait » avant. D'autant qu'en regard du tchatbot ChatGPT qui s'élargit au moteur de recherche, Google [plancherait pour intégrer](#) du tchatbot à son moteur de recherche. Restera enfin à évaluer « ce que nous sacrifions quand nous formatons les systèmes de cette manière » ([Chirag Shah et Emily M. Bender](#))...

LLM, RAG et recherche de littérature

À l'instar d'autres écosystèmes, l'IA générative « à la ChatGPT » s'invite désormais dans les bases de données bibliographiques et outils de recherche documentaires. Passé l'ère des outils simplement « sémantiques » pour comprendre, bienvenue dans les outils « IA générative », « GPT » et autres « assistants » pour également générer du nouveau contenu ! L'enjeu d'utiliser des LLM et de la génération de contenus textuels dans ce contexte est multiple :

- améliorer la pertinence des réponses par la recherche sémantique et le traitement du langage naturel (compréhension de la requête et des contenus) ;
- générer des réponses synthétiques, avec les références voire des citations (documentation de la réponse) ;
- extraire des données des sources (tableaux) ;
- converser avec l'outil.

Néanmoins, comme pour les moteurs de recherche de type Google *AI overviews*, l'IA générative est alors limitée aux fonctions LLM sémantiques et génératives et n'est dans ce cas qu'une couche de plus ajoutée à la base de connaissance. On pourrait comparer cela à un Google Scholar + ChatGPT. [Comme le dit](#) l'un de ces services : « Consensus est d'abord un moteur de recherche, avec de l'IA « autour ». Cela signifie que chaque réponse est basée sur de la recherche scientifique ». Ils ne fournissent pas des réponses issues des données d'entraînement générales comme ChatGPT, mais s'appuient sur des bases de références et des documents spécifiques. C'est ce qu'on appelle le **RAG** (*retrieval augmented generation* ou génération augmentée par récupération), une architecture qui combine la fonction générative de LLM comme GPT, à une fonction de recherche dans un corpus ou une base de données bien défini. Élément clé du « futur de la recherche documentaire » ([Emmanuelle Bermès](#)), cela permet d'accroître la pertinence de la réponse et donc de [limiter les hallucinations](#) ([sans totalement les supprimer](#)), au prix il est vrai d'un certain nombre de limites ([Emmanuel Barthe](#), [Aaron Tay](#)) que nous évoquerons plus loin.

La précision des réponses et l'**adéquation à la recherche scientifique** est permise par trois éléments principaux par rapport aux LLM généralistes comme ChatGPT :

- des données issues de la littérature académique : ces outils ne cherchent pas sur le web en général, mais s'appuient sur les fournisseurs classiques de données bibliographiques académiques (bases comme Semantic scholar, partenariats avec des éditeurs scientifiques) ;
- des modèles d'IA entraînés (« *fine-tuned* », « *customized* ») sur des tâches spécifiques à l'outil afin de [limiter les hallucinations](#), avec éventuellement de la supervision humaine ;
- des données vérifiables : ces outils sont généralement de [type « extractif »](#), non pas seulement « génératif ». Ils permettent de remonter à la source de l'information (notice bibliographique, texte intégral, citation dans son contexte).

Intégration d'IA générative dans les outils de recherche bibliographique

Fin 2022, un *outsider*, OpenAI, prend tout le monde de court en ouvrant le premier au grand public un outil génératif, ChatGPT. Échaudés par l'échec retentissant du tchatbot Tay de Microsoft en 2016, les grands acteurs de la recherche étaient eux [plus frileux](#). Si la technique des LLM était développée aussi du côté de Google, la société de Mountain View restait alors prudente sur le déploiement d'un outil maison grand public, en raison des enjeux de fiabilité de la technique et de risque réputationnel trop grand. Ce n'est que suite au succès de ChatGPT que Microsoft et Google sont passés à la vitesse supérieure, le premier en passant un partenariat

avec OpenAI, le second en ouvrant dans une certaine urgence son propre outil Bard (désormais Gemini). De même, ce sont des *outsiders* ou autres *start-up* qui les premiers sortent publiquement des fonctionnalités d'IA générative dans les outils de recherche documentaires courant 2022-2023, avant que les acteurs historiques n'emboîtent le pas, suite à la déferlante ChatGPT à partir de 2023. Fin 2024, deux grandes tendances sont visibles dans l'intégration de l'IA actuelle dans les outils documentaires : d'un côté, des acteurs déjà existants dans l'écosystème académique (fournisseurs d'outils documentaires ou de services) qui développent de nouvelles fonctionnalités autour de la recherche documentaire. De l'autre, des start-up créant de zéro des outils documentaires boostés à l'IA. Parmi ces *outsiders* et start-up, trois acteurs sortent du lot et sont régulièrement cités par les chercheurs, les étudiants et les professionnels de l'information. Ils sont utilisés ici à titre d'illustration de la situation actuelle. D'autres services existent ou se développent régulièrement, voir par exemple [ici](#) et [là](#).

Site	Société	Corpus bibliographique	Méthode IA générative	Nombre d'utilisateurs
Consensus	Consensus NLP (US)	200 M. réf. (Semantic scholar)	custom fine-tuned LLM, GPT-4o	2 M.
Elicit	Elicit Research (US)	125 M. réf. (Semantic scholar)	divers LLM et GPT	2 M.
SciSpace	PubGenius (IN / US)	270 M. réf. (?)	?, GPT	1 M.

Onglet « *assistant* » ou « *copilot* », paragraphe « *key insights* » ou simplement mention « IA » ou d'une petite étoile... Durant l'année écoulée, l'IA a aussi été développée et mise en scène dans nombre d'**outils documentaires plus anciens**. Citons, par exemple, des bases de données bibliographiques ([Dimensions summarization](#), [Scopus AI](#), [Web of science Research assistant](#)), des fournisseurs de contenus ([Cairn](#), [les plateformes d'EBSCO](#), [JSTOR interactive research tool](#), [Lexis+ AI](#), [nouvelle expérience Factiva](#), [ProQuest](#)), des catalogues de bibliothèques ([Primo Research assistant](#)), des bases de données ([Statista Research AI](#)), etc. Seul Google scholar semble encore prudent sur le sujet, proposant pour l'instant un [PDF Reader](#) dans le navigateur de recherche Chrome, destiné à favoriser la lecture des PDF. A ce stade, il n'a pas encore implémenté de l'IA générative à la ChatGPT dans son interface de moteur de recherche, et chercherait encore, selon son co-fondateur, une solution pour synthétiser de multiples articles d'une manière qui « préserver tout le contexte important » ([Anurag Acharya](#)).

Derrière la communication commerciale, le terme d'IA est bien souvent à interroger, puisqu'il ne s'agit pas nécessairement d'IA générative, mais aussi d'[IA sémantique plus « traditionnelle »](#). On pourrait distinguer 3 niveaux d'intégration de l'IA dans les outils actuellement, permettant d'élargir ou d'accélérer la recherche :

- une simple amélioration de l'existant, par exemple en proposant des synonymes, des sujets connexes, des concepts, dans la continuité des suggestions automatiques ;
- de nouvelles fonctionnalités de recherche, en proposant la recherche en langage naturel et multilingue et générant des synthèses de réponses, tant au niveau des références bibliographiques simples que d'une liste de références ;
- une nouvelle expérience utilisateur avec le développement des assistants de recherche, chargés d'interagir avec l'utilisateur, voire de faire une partie du travail de recherche à sa place.

Les outils de recherche documentaire ont tendance à devenir plus complexes, faisant cohabiter les anciennes interfaces et méthodes de recherche avec les nouvelles fonctionnalités IA. Une très bonne illustration est donnée par Scopus. A côté de la base bibliographique

historique, Scopus propose un onglet dédié « [Scopus AI](#) » pour chercher en langage naturel, dont la fonctionnalité [Copilot](#) est capable de transformer la requête en langage documentaire avec mots-clés et booléens.

De la même façon que l'IA se déploie progressivement au plus près des usages bureautiques (lecteurs de PDF, navigateurs web, outils de visioconférence, etc.). Comme l'illustre aussi les multiples fonctionnalités autour des résumés de textes d'un [Scholarcy](#), il est logique de voir les outils bibliographiques boostées à l'IA sortir de leurs sites internet dédiés pour gagner d'**autres environnements de travail**. Si les extensions de navigateurs (comme celles de [SciSpace](#) ou [Scite](#)) sont un incontournable de la recherche d'information, on voit progressivement se développer des extensions permettant d'appréhender rapidement le contenu d'un document ([Google scholar PDF Reader](#) et son nouveau Plan IA) intégrables dans des traitements de textes (comme [Keenious](#) pour Microsoft Word et Google docs), voire un véritable écosystème de traitement de texte IA (comme l'[AI Writer de SciSpace](#) ou [Jenni AI writer](#)), fournissant des références bibliographiques et des synthèses en fonction du contenu déjà rédigé. Autre type d'intégration, celle des GPTs, ces plugins disponibles dans ChatGPT, accessibles désormais gratuitement avec le déploiement durant l'été de GPT-4 pour tous (ex. : [Consensus GPT](#), [Dimensions Research GPT](#), [ScholarAI](#)). Dans l'optique de développer la productivité des chercheurs, ces outils permettent de déléguer à l'outil IA le flux de travail de recherche de littérature, depuis l'identification jusqu'à l'analyse et la synthèse. Ce qu'on exprime clairement chez [Clarivate](#), maison du Web of science : "*Intelligent discovery to reclaim research time*".

UNE NECESSAIRE REFLEXION SUR LES COMPETENCES DE RECHERCHE

Changements dans la recherche documentaire

Les chiffres d'usage annoncés par les *outsiders* sont encourageants : [2 millions de chercheurs](#) pour Elicit, [1 million](#) pour SciSpace, [2 millions](#) pour Consensus... Derrière l'effet de mode autour de l'IA, ces outils semblent répondre à une attente forte de la part des chercheurs, celle de pouvoir faire sa revue de littérature rapidement (ex. : [sondage SciSpace](#) et [sondage Research Solutions](#)). Néanmoins derrière les chiffres, on constate que les [retours d'expérience](#) élogieux sont le plus souvent le fait de **jeunes chercheurs** : ce sont eux les publics cibles de ces services. Comme l'exprimait bien Elsevier en début d'année : Scopus AI a pour but de "*help early-career academics and researchers working across different disciplines to navigate and understand academic content. It assists users in comprehending unfamiliar academic fields*" ([Elsevier, 01/2024](#)) : « L'idée est de fournir un guide fiable » parmi les données de Scopus ([Elsevier, 02/2024](#)). L'utilisation de la recherche sémantique permet en effet de chercher des concepts similaires, quand on ne connaît pas les mots-clés précis, qu'on découvre un sujet ou que la terminologie d'un domaine n'est pas encore bien figée, tandis que l'utilisation du langage naturel permet de faire l'économie des équations booléennes ([Aaron Tay](#)). Si ces outils peuvent se révéler suffisants pour « aider les chercheurs à cheval sur plusieurs disciplines essayant de se familiariser avec un sujet donné rapidement » ([M. Khan, Elsevier](#)), les retours des chercheurs connaissant bien leurs sujets sont eux [plus mitigés](#). Comme on le précise [chez Elicit](#) : « Les résultats d'Elicit doivent être pris comme un point de départ pour vos revues [de littérature] et évaluation ultérieures ». Il est d'ailleurs intéressant de constater que suite aux retours de communautés académiques sur Scopus AI, les fonctionnalités de Scopus AI n'ont pas été fusionnées avec la recherche « Documents » mais font l'objet d'un onglet spécifique, permettant de distinguer des profils d'utilisateurs différents. Pour Elsevier, il s'agit d'offrir une [porte d'entrée](#) aux contenus Scopus, plus intuitive à des usagers n'ayant pas encore toutes les compétences nécessaires à la recherche documentaire traditionnelle. Du côté du Web of science Research assistant, lui aussi distingué de l'outil historique, on a le choix entre « Ask

a research question », « Understand a topic » et « Literature review ». En privilégiant la recherche en langage naturel et en fournissant des contenus synthétisés au lieu de simples références bibliographiques, l'objectif est de faciliter l'accès à la recherche scientifique pour tous ([Consensus](#)). On retrouve ici la division bien connue entre la recherche (*search*) et la découverte (*discovery*) – rappelons que la recherche active n'est qu'une manière de trouver des productions pertinentes ([28,6 % par searching](#) et [33,7 % par browsing](#)).

De fait, les possibilités permises par ces outils sont très larges. Par la recherche en langage naturel et conversationnelle, ils élargissent les moyens d'évaluer des hypothèses sur des sujets de recherche, faire une rapide revue de littérature, trouver des références soutenant ou invalidant une thèse particulière, identifier des lacunes de la littérature. Par leur utilisation de la sémantique, ils facilitent la découverte de sujets approchants ou permettent d'identifier le contexte de citations. Par leur extraction d'informations, ils permettent de rechercher plus en profondeur dans les *abstracts*, les PDF ou encore de fournir des tableaux de résultats (« *matrix* »). En plus des fonctionnalités standard des bases de données bibliographiques (comme les tris et filtres, l'accès au PDF disponible, l'export des références, etc.), les **fonctionnalités « boostées à l'IA »** les plus fréquemment présentes sur ces outils sont les suivantes :

- exemples de requêtes ;
- interrogation en langage naturel ;
- synthèse de réponse à partir des résultats les plus pertinents (*top*), avec leurs références ;
- liste de références bibliographiques classées en fonction de leur pertinence sémantique ;
- court résumé de la publication (TL;DR : *too long; didn't read* ; *insights, snapshots, takeaways* : aperçus) ;
- liste de concepts associés à la requête ;
- suggestions de questions complémentaires ;
- citations d'extraits des documents ;
- présentation des résultats sous forme de liste ou de tableau (*matrix*) personnalisable ;
- fonction « Assistant » ou « Copilot » pour interagir avec les résultats à partir d'instructions ;
- recherche multilingue.

Hormis les questions découlant spécifiquement de la recherche sémantique sur laquelle nous reviendrons juste après, trois questions sont centrales pour toute **analyse d'un outil de recherche** –les éléments soulevés ici pourraient être évoqués pour un certain nombre d'outils plus anciens, à commencer par Google scholar :

- la couverture : nombre des nouveaux entrants sont de véritables **boîtes noires**, n'indiquant pas les sources précises de leurs données (ex. : Scite, SciSpace, ScholarAI), empêchant de bien identifier leur représentativité. Quant aux autres comme Elicit ou Consensus, qui utilisent Semantic scholar, il convient de noter les [lacunes de Semantic scholar](#), base par exemple plus axée sur les sciences, techniques, médecine et les articles scientifiques anglophones plutôt que les sciences humaines et sociales ou d'autres types de publication. Il conviendra de voir comment (et si) peuvent être intégrées d'autres bases bibliographiques comme OpenAlex. Enfin, il est tout aussi difficile d'avoir des informations sur les conditions de mises à jour de ces bases, tant pour le nettoyage des données que pour la fraîcheur des contenus : privilégie-t-on la mise à jour automatique en quasi temps réel par rapport à la base de données bibliographiques en utilisant l'API (léger, peu cher mais potentiellement plus lent) ou la mise à jour en décalé, par téléchargement des données et nettoyage avant déploiement dans l'outil (plus rapide, plus flexible mais limitant la fréquence des mises à jour) ? C'est un truisme, mais les traitements algorithmiques, *a fortiori* avec de la recherche sémantique, ne sont bons que

si les données utilisées sont bonnes – c’est le fameux adage remis au goût du jour avec les IA actuelles « *Garbage in, garbage out* » ;

- l’interface de recherche : dans ce domaine, les choses sont en train de changer puisque courant 2024, on voit progressivement les outils proposer une vraie recherche multilingue, avec une interface dans d’autres langues que l’anglais. Dans le même temps, la recherche en langage naturel dans d’autres langues est également de mieux en mieux prise en compte ;
- les fonctionnalités de recherche avancée : dans une logique d’assistant de recherche, la priorité de l’outil est de fournir un résumé synthétique et des pistes bibliographiques. On constate alors l’éventuelle disparition des possibilités de filtres en amont et de tris en aval, rendant difficile toute exploration fine rapide, par exemple pour la date ou le nombre de citations : c’est lors de la requête qu’il convient d’être le plus précis possible.

Développement de la recherche sémantique générative

Bien sûr, les outils documentaires « historiques » n’étaient pas à l’abri de [certaines limites](#), sur leur couverture ou leur manque de transparence, ce qui a amené au développement récent de [nouveaux outils de recherche bibliographique](#). Mais l’intégration de fonctionnalités d’IA sémantique et générative dans les outils de recherche vient poser de **nouvelles questions** sur les données utilisées, la démarche de recherche et les résultats fournis.

La première interrogation porte sur les **données véritablement utilisées** par l’IA. Le fait même d’utiliser de la recherche sémantique contraint ces outils pour des raisons techniques et légales. Concernant les données interrogées, ces outils utilisent les métadonnées classiques comme les titres, auteurs, mots-clés, etc. Ces métadonnées sont par ailleurs complétées par les *abstracts*, nécessaires aux algorithmes IA pour établir au mieux le contexte des publications. Or, si ces *abstracts* sont rendus de plus en plus accessibles grâce au mouvement des [open abstracts](#) ces dernières années, nombre de publications anciennes ou des publications d’[éditeurs phares récentes](#) ne proposent pas de résumé librement accessible et sont donc potentiellement exclues de cette recherche sémantique, quand bien même ces publications seraient pertinentes. Ainsi, lors de tests, Semantic scholar, base de références de plusieurs de ces outils, ne propose pas d’[abstracts](#) pour 25 à 40 % des articles, ne laissant que le titre pour identifier le contenu. On notera par exemple que les données utilisées par Scopus AI [ne commencent qu’en 2003](#) ou qu’Elicit [exclut volontairement les livres](#) car il a été conçu pour les *papers*. D’une manière générale, ces outils ont tendance à [privilégier les articles](#). Par ailleurs, quelle confiance donner à une recherche sémantique basée principalement sur des résumés quand on sait que ces résumés [ne reflètent pas toujours](#) le contenu exact des textes intégraux et que chaque discipline a ses [propres pratiques de résumé](#) ? Cela fait d’ailleurs un certain temps que [des chercheurs appellent](#) à une indexation des textes tout autant que des résumés. Or, des questions à la fois légales, techniques et économiques limitent une indexation large des textes intégraux. D’une part, pour des questions de droit d’auteur et de licences, malgré les [progrès de l’open access du côté des éditeurs](#), nombre de textes intégraux sont encore accessibles sous conditions uniquement. D’autre part, l’indexation sémantique des textes intégraux coûte plus cher que les seuls résumés, posant des questions de rentabilité des services. Quant au résumé de textes longs, il est encore soumis aux capacités de mémorisation (*context window*) des IA de type ChatGPT, dont les [limites actuelles](#) favorisent les approximations et les hallucinations. Pour ces raisons de coût et d’efficacité, ces outils ont alors tendance à limiter le nombre de références réellement utilisées pour les résultats résumés (*summary, answer*). Ainsi, même si [davantage de résultats](#) sont souvent considérés (par exemple, 25 par défaut sur Scite Assistant, avec possibilité de monter à 100), la synthèse proprement dite ne se contente que de quelques *top papers* : 4 (8 en version payante) sur Elicit, 5 (10 en version payante) sur SciSpace. Comme [le résumé Consensus](#) : « Ce résumé n’est qu’un aperçu de certaines des recherches pertinentes auxquelles

nous avons accès, et non une analyse exhaustive de toutes les recherches concernant votre question ». Si bien que l'on interroge sur des contenus bien souvent hétérogènes : ainsi [Elicit indique](#) bien extraire les données « du plein texte s'il est disponible, sinon, de l'*abstract* ».

La seconde interrogation porte sur la **manière d'interagir avec l'outil**. Le développement de ChatGPT a fait émerger la notion de « prompt », d'invite conversationnelle. Pour favoriser le contenu le plus adapté, l'utilisateur du tchatbot doit fournir suffisamment d'informations à l'IA générative pour qu'elle « comprenne » le contexte et l'intention de la demande ([méthode ACTIF](#), [méthode ETAPE/APRES](#)), ce qui est [déjà assez proche](#) d'une logique documentaire. Est-ce que cela change nécessairement quelque chose aux outils de recherche bibliographique permettant l'interrogation en langage naturel ? A vrai dire, pour la revue de littérature, pas tant que ça. Comme [le dit Consensus](#) : « Rappelez-vous toujours que Consensus est un moteur de recherche – pas un tchatbot. [...] Consensus sera le plus efficace quand il y aura une question de recherche ou un sujet de recherche au « cœur » de votre demande ». L'utilisation de RAG permet de limiter l'indication de contexte en centrant cette recherche sur de la littérature académique. On peut donc poser sa requête directement sous forme de questions (exemples fournis par Consensus [ici](#) et [là](#)) : simples questions oui/non (ex. : *“Does the death penalty reduce crime?”*), questions ouvertes (ex. : *“benefits of mindfulness meditation”*, *“How do you increase local voter turnout?”*), relations entre des concepts (ex. : *“What is the impact of climate change on GDP?”* ; *“Review the literature about the relationship between socioeconomic status and healthcare outcomes in America”*)... Comme l'outil est capable de comprendre le langage naturel, « il n'y a pas une seule façon « correcte » de structurer une recherche Consensus » ([Consensus](#)). Il conviendra de [bien distinguer](#) alors le RAG proprement dit, pour la recherche bibliographique augmentée, et les fonctions d'assistance pour mener à bien des tâches. Il [convient d'être très prudent](#) sur la formulation de sa requête sous forme d'un prompt du même type que les tchatbots. Ce sont surtout les fonctionnalités de tchat, souvent signalées sur le nom Assistant ou Copilot, qui peuvent permettre une interrogation plus proche du prompt, plus conversationnelle, sous forme d'instructions (ex. : *“Discuss the various supplements that reduce nausea, do it in very simple language”* ; *Conduct an umbrella review around pandemic early warning systems*). Il convient donc aussi de bien regarder les exemples fournis sur chacun des services utilisés : une requête peut bien fonctionner sur un outil et pas sur un autre.

La troisième interrogation porte sur les **résultats fournis**. C'est peut-être dans ce domaine-là que l'on constate le plus de différences avec les outils historiques. Premier problème : la sélection et le mode de classement des résultats. De fait, le premier critère de classement des résultats est la similarité sémantique par rapport à la requête, [et non la date ou le nombre de citations](#) (voir [ici](#) pour l'exemple d'Elicit), donnant ainsi plus de poids à des contenus plus récents et moins connus. Et si le fonctionnement des bases de données bibliographiques les apparente souvent à des boîtes noires, prenant en compte ou pas métadonnées, *abstracts* ou citations, le recours à des fonctionnalités sémantiques rajoute une nouvelle couche de mystère au fonctionnement de l'outil et donc des problèmes pour l'évaluer correctement : « Le plus gros désavantage du point de vue du bibliothécaire de la recherche de type sémantique [...] est que vous abandonnez l'interprétabilité et la prédictibilité » ([Aaron Tay](#)). Du temps de la recherche par mots-clés, quand on n'obtenait pas de réponse ou peu de résultats pertinents, le réflexe était alors, en l'absence de résultats pertinents, de changer les mots-clés, voire l'outil. A l'heure sémantique, sauf contre-indication explicite donnée à l'outil [comme dans Consensus](#), des réponses peuvent être fournies même en cas de similarité sémantique très faible. Or, peu d'outils proposent des [scores de pertinence](#) qui permettraient de situer la proximité des résultats obtenus par rapport à la question – actuellement, par exemple Scopus AI, suite aux retours utilisateurs. Le risque est alors de ne se contenter que de ce premier résultat, tout imparfait qu'il soit, sans faire de nouvelle recherche. Par ailleurs, dans le cas des

réponses résumées, le nombre de références est limité et seules quelques réponses sont présentées. Il est souvent impossible d'obtenir les résultats moins pertinents sauf à refaire une autre recherche. Et dans le cas où beaucoup de références seraient sémantiquement similaires, sur quels critères s'appuie l'outil ? Ne risque-t-on pas un effet bien connu dans les bases de données comme Google, l'effet Mathieu ou le principe du « *rich get richer* », selon lequel on trouve avant tout le plus évident, déjà bien cité, au détriment des publications plus rares ? Enfin, les outils sémantiques ne favorisent pas l'analyse de leur démarche. Du temps de la recherche par mots-clés, les données traitées pour la recherche étaient évidentes : c'étaient les mots-clés saisis. Avec la recherche sémantique, [on ignore bien souvent](#) comment la question a été traitée (sémantique ?, mots-clés et recherche booléenne ? [combinaison des deux](#) ?). Rares sont encore les outils qui, comme Scopus AI, affichent une partie au moins de leur méthodologie. Son outil [Copilot](#) en effet indique ainsi les différentes étapes de la recherche, la recherche en langue naturelle, mais aussi la recherche par mots-clés effectuées. En revanche, l'outil le fait de manière autonome : c'est lui qui « regarde le contenu de votre requête et décide de lancer une recherche vectorielle [sémantique] et/ou une recherche par mots-clés ». De son côté Scite Assistant, indique aussi la *search strategy* suivie (requêtes et sources utilisées). Il est néanmoins difficile d'identifier et expliquer comment a fonctionné la recherche sémantique : qu'est-ce qui a été pris en compte, qu'est-ce qui a été laissé de côté, les termes cherchés sont-ils exhaustifs pour le sujet ? On sait aussi que de plus en plus d'outils utilisent même les LLM pour générer la requête booléenne ([CoreGPT, Scite assistant](#)), alors même que les LLM ne sont [pas encore totalement satisfaisants](#) en ce domaine. Beaucoup plus que dans le cas de la recherche par mots-clés, il est somme toute difficile d'évaluer et de contrôler correctement les résultats obtenus. Sans compter que ces outils, dans la droite ligne de ChatGPT et des LLM, fournissent des [réponses aléatoires](#), [interrogeant la reproductibilité d'une requête](#) au sein d'un même outil. Elicit reconnaît ainsi une « [imprévisibilité inhérente](#) » dans leurs réponses et invite à considérer l'outil comme un assistant humain, donnant des réponses différentes selon le moment... Au temps cependant pour la promesse de gain de temps, si l'on doit passer davantage de temps pour analyser les résultats, par exemple dans le cadre de projets d'équipe.

Au final, ces outils apportent-ils une réelle **plus-value dans la démarche de recherche de littérature** ? Tout dépend sans doute du contexte même de cette recherche de littérature et du profil de l'utilisateur. Si ces outils peuvent apporter des pistes pour explorer un nouveau sujet, ils ne sont pas suffisants pour une recherche rigoureuse. Certes, ils permettent de gagner du temps mais ils ne permettent pas de faire le tour d'un sujet d'une manière exhaustive, par exemple pour une revue de littérature complète et systématique, depuis la recherche de littérature, jusqu'à l'extraction et la synthèse des données (voir par exemple [ici](#)). Un certain nombre de chercheurs, comme [ici](#) et [là](#), estiment à l'heure actuelle que ces outils ne peuvent se suffire à eux-mêmes : ce sont avant tout des outils de recommandation ou de suggestions complémentaires, dont l'utilisation doit être [bien située dans l'offre d'outils existants](#). La plupart de ces chercheurs notamment signalent le caractère encore bien souvent incontournable [de Google scholar](#), ainsi que [de Scopus ou du Web of science](#). Même avis du côté des spécialistes de la recherche d'information, comme [ici](#), [là](#) ou [là](#), pour qui la recherche sémantique (*semantic search, vector search, embedding-based search*) vient plutôt compléter les méthodes de recherche conventionnelles, dominées jusque-là par deux axes : d'une part, la recherche lexicale par mots-clés (*keywords search*) et booléens (*boolean search*) et d'autre part la recherche de co-citations (*citations chaining, snowball search*). La recherche traditionnelle par mots-clés et interfaces classiques serait encore [plus fiable, a fortiori](#) pour les sujets rares ou émergents, ou les questions complexes ne pouvant se résumer en une simple hypothèse de recherche. Néanmoins, si la recherche sémantique ne peut se substituer à une recherche rigoureuse, méthodique et reproductible, elle peut permettre de [trouver des références complémentaires](#) pertinentes pour un sujet, [non présentes dans les outils classiques](#),

probablement aussi en raison de couvertures différentes des outils. « L'habilité de l'IA à analyser de grands jeux de données et à discerner des schémas ou des relations qui ne seraient pas facilement visible par les chercheurs pourrait révolutionner la méthodologie des revues de littérature, en particulier dans les disciplines marquées par une croissance rapide du savoir et une recherche interdisciplinaire » ([Przemysław Tomczyk et al.](#)). On pourrait alors identifier deux profils de recherche : un profil « débutant » qui commencerait sa recherche par des outils IA, plus intuitifs, avant d'aller vers des méthodes plus poussées et un profil « avancé » qui sait déjà utiliser les outils classiques et qui utiliserait les outils IA en complément. C'est en tout cas ce qu'illustre la coexistence de plusieurs interfaces parfois au sein d'un même outil de recherche (onglets Scopus AI dans Scopus, Research assistant dans Web of science, Assistant dans Scite). Du côté des producteurs de ces bases historiques, il n'est pas question de supprimer les interfaces de recherche structurée classique, pour permettre aux experts de rechercher au mieux. Et du côté des nouveaux entrants, on peut aussi proposer des éléments de requête avancée traditionnels – Elicit les appelle même des « [fonctionnalités cachées](#) », identifiables uniquement en passant par l'aide. La difficulté pour l'utilisateur sera de savoir quelles sont les fonctionnalités et les différences de ces deux modes de recherche (recherche par mots-clés, recherche sémantique). Sans même parler de la possibilité d'[identifier clairement](#) si l'outil a fait une recherche de mots-clés, sémantique ou les deux. Les lignes bougent et, on l'a vu, un outil comme le Copilot de Scopus AI peut faire une recherche hybride mélangeant les deux, en fonction du contexte. De même, du côté du Web of science, l'outil combine à la fois [combine](#) similarité sémantique sémantique et recherche par mots-clés, avant de passer le tout à un score de similarité. L'une des difficultés à venir dans l'utilisation des outils de recherche bibliographique sera de pouvoir associer [précision, complétude](#) et explicabilité et d'avoir toujours la possibilité de rechercher uniquement sur les mots-clés, sans sémantique.

« *Illusions of understanding* » et autres problèmes épistémiques

La communication des outils de recherche boostés à l'IA est généralement bien rôdée et promet gain de temps et efficacité. De même, [d'après un sondage auprès de ses utilisateurs](#), Elicit permettrait d'économiser en moyenne 1,4 h par semaine et jusqu'à 5 h par semaine pour 10 % d'entre eux. Dans le cadre de revues systématiques, l'outil serait plus efficace que des assistants humains à la fois pour identifier les références et pour extraire les données. Par ailleurs, [dans des projets pilotes](#), il a permis des gains de coûts mais aussi de temps de l'ordre de 50 % en automatisant l'extraction de données. Si bien qu'on [affirme](#) : « Elicit coûte 50-80 % moins cher que des assistants de recherche humains ». Dans le même temps, [une étude portant sur l'indexation MeSH automatique](#) de contenus dans PubMed et Medline révèle un taux de 47 % de problèmes par rapport à un indexation humaine, soit par la présentation de mauvais résultats, soit par l'absence de résultats pertinents ; et [du côté de Clarivate](#), on ne manque pas de signaler la supériorité du travail d'experts face aux LLM pour résumer des brevets, à la fois pour leur exactitude et leur cohérence. Dès lors, on comprend mieux que des voix invitent à utiliser avec prudence ces outils au risque de nuire à la science elle-même, à la fois pour des questions de méthodes et de construction des savoirs scientifiques. A l'instar d'autres outils d'information, ces bases de données avec IA peuvent favoriser une plus grande personnalisation de l'accès à l'information mais aussi une homogénéisation du contenu ([Noora Hirvonen et al.](#)). Si une partie d'entre eux s'appliquent au développement de l'intelligence artificielle dans la recherche en général, certains sujets d'inquiétude peuvent concerner directement les outils documentaires :

- l'inéquation **des moteurs de réponse dans le cadre académique** : même si ces outils [soulignent](#) eux-mêmes l'intérêt de développer des moteurs de synthèse, le principe même de « *search engine answers* » [ne peut être pleinement appliqué](#) à la science qui est cumulative et qui ne peut arriver à une conclusion définitive, *a fortiori* appuyée sur

un très petit nombre de publications. Par ailleurs, les synthèses proposées se contentent souvent d'aligner les points essentiels des articles les uns après les autres, [sans vraiment faire de synthèse](#) entre les résultats des publications. L'approche statistique probabiliste et automatique n'est pas nécessairement pertinente pour du contenu académique qui doit prendre en compte aussi des contenus rares ou nouveaux. Néanmoins, [certains enseignants](#) en soulignent la vertu pédagogique, en première approche, pour interroger la validité d'hypothèses et « pointer l'état mouvant de la science » même sur les questions bien établies et Elicit [invite](#) aussi à chercher les différentes perspectives d'une question. Cela permettrait également de [détecter des tendances](#) de fond ou au contraire des apports plus originaux – néanmoins, en l'état actuel des choses, ces outils LLM ne semblent capables d'identifier que les pistes, les hypothèses ou encore les lacunes exprimées dans les productions scientifiques et non proposer de nouvelles pistes originales ;

- la **sélection** (« *cherry picking* ») de citations : en présentant directement dans les interfaces des citations sémantiquement similaires à la question, de tels outils favoriseraient des pratiques discutables comme des « citations de pure forme et des argumentations bâclées » et accroîtraient les biais de confirmation ([S. P. J. M. Horbach et al.](#)). [Cela pourrait détourner](#) d'une vraie lecture critique des publications d'origine, en cherchant uniquement des citations qui conviennent au discours. Si [certaines de ces remarques se retrouvent dans d'autres outils de recherche](#), cela limiterait les points de vue contradictoires et l'ouverture à d'autres communautés disciplinaires – c'est d'ailleurs l'objectif du [Consensus meter](#) que d'identifier le consensus scientifique (positif ou négatif) sur une hypothèse ou celui des [citation statements de Scite](#) que d'indiquer des citations en appui ou en opposition à une publication ;
- le **manque de connaissances scientifiques des IA génératives** : l'une des promesses de ces outils est de [pouvoir chercher en langage naturel](#), sans même connaître le jargon académique. Or, on le sait, des IA génératives comme ChatGPT, qui sont souvent utilisées dans les outils documentaires pour la couche LLM, sont avant tout entraînées sur des données grand public et non académiques. Si les LLM sont de plus en plus utilisées pour la revue de littérature, l'extraction de données et des échanges conversationnels, leurs limites sont [bien documentées](#). Difficultés à comprendre le contexte, à gérer des informations implicites ou conceptuelles, à distinguer des idées établies ou seulement émergentes ou encore à mener à bien des tâches nuancées complexes comme les hypothèses ou les inférences, les « [bords irréguliers](#) » (*jagged edges*) [des LLM dans les RAG](#) sont nombreux. Un problème récurrent est par exemple celui du vocabulaire : les travaux de [Guillaume Cabanac et al.](#) sur les « expressions torturées » montrent bien, par exemple, comment l'utilisation d'IA généralistes peut produire des expressions inattendues au lieu des expressions consacrées dans la communauté académique (ex. : *counterfeit consciousness* ou *man-made power* pour *artificial intelligence*). Malheureusement, l'on retrouve ces expressions « *nonsensical* » [aussi dans les contenus générés par IA dans les outils documentaires](#). Le risque est particulièrement important pour les auteurs qui utiliseraient ces outils documentaires dans une langue qu'ils ne maîtrisent pas totalement ou pour un domaine dont ils ne sont pas très familiers. Cela explique pourquoi ces outils s'appuient toujours sur des [modérateurs humains](#) pour vérifier les résultats, notamment sur les domaines sujets à débat, comme pour le [Web of science Research assistant](#). Comme [on le reconnaît chez Consensus](#) : « A chaque fois qu'un modèle génératif est utilisé, il est possible qu'il crée une réponse qui n'est pas basé sur la réalité », ce qu'illustre parfaitement [l'échec de l'outil Galactica de Meta débranché seulement 3 jours après son ouverture](#). C'est aussi le cas pour les [résumés automatiques](#), [des citations](#) ou [des extractions](#) parfois incomplets

ou erronés par rapport aux publications d'origine, en l'absence de système à même de contrôler la qualité des réponses – [risque sans doute accru](#) dans le cas de nouveaux sujets ou de sujets de niche comme en médecine ou en droit. On peut constater aussi des résultats, vrais factuellement mais absents des sources, le LLM venant apporter des informations implicites voire absentes dans les sources d'origine (« [hallucinations extrinsèques](#) »). Enfin, ces outils sont incapables de prendre en compte la qualité d'une recherche. Consensus [indique ainsi](#) que « la qualité de la recherche ne fait pas partie de l'analyse », tandis qu'Elicit [rappelle](#) que son apport se limite à l'aspect technique et non qualitatif : « Elicite résume les résultats d'une mauvaise étude de la même façon qu'il résume les résultats d'une bonne étude » ([source](#)). Ce n'est pas propre aux seuls outils contenant de l'IA générative, mais [ces outils](#) donnent parfois le même poids à des niveaux de preuve différents (une méta-analyse et une étude unique) ou à des méthodologies variées (des publications théoriques, des études de cas ou des résultats d'expérience), contiennent des éditeurs prédateurs ou sélectionnent des contenus réfutés scientifiquement. Lors de son lancement, Consensus avait notamment mis en valeur une publication, pourtant rétractée, affirmant un [lien entre vaccins et autisme](#). Outre les biais linguistiques possibles, [il est à craindre](#) des réponses trompeuses, présentant des éléments ne correspondant pas aux données d'origine ou passant à côté d'éléments importants ;

- les **risques d'« intrusion épistémique »** : la notion d'« *epistemic trespassing* » ([Nathan Ballantyne](#)) désigne le fait de porter des jugements dans des domaines où l'on manque d'expertise ; si le décloisonnement est important pour l'avancée de la recherche, ce risque est néanmoins renforcé avec le développement des études interdisciplinaires. Appliquée à ces outils documentaires, le [problème est double](#). D'une part, ces outils établissent leurs synthèses à partir de documents de [différentes disciplines](#) au risque de gommer les particularités et les nuances de chacune, ou d'être moins adaptés pour certains domaines. Ainsi, pour Elicit, l'outil [s'est centré](#) sur la recherche empirique, par exemple en biomédecine, « de façon à appliquer les leçons de la revue systématique » et [fonctionne donc moins bien](#) pour les éléments factuels ou les domaines théoriques. Quelle est la valeur d'une synthèse couvrant le même type de contenus d'un point de vue strictement sémantique, mais différent d'un point de vue épistémique ? De même, comment être sûr que le vocabulaire dans un domaine ait les mêmes implications dans un autre ou même dans différentes écoles de pensée ? D'autre part, au niveau du chercheur, l'expertise et le jugement dans une discipline ne peuvent s'appliquer à l'identique dans une autre : dans ses domaines de spécialités, on connaît les codes et les points d'évaluation (réputation des revues, chercheurs proéminents, compréhension du jargon, etc.). Comment faire dans les domaines dont on est moins familier ? ;
- l'**illusion de la compréhension et la perte d'expertise** : avec leurs citations et leurs réponses immédiates et prêt-à-penser, ces outils sont un exemple de « découverte » (*discovery*) et non de « compréhension » (*understanding*) scientifique ([Mario Krenn et al.](#)). On peut ainsi les comparer à un « oracle » capable de résumer et de fournir du savoir scientifique au détriment d'une réflexion construite et poussée ([Lisa Messeri et M. J. Crockett](#)). De plus en plus d'études soulignent les risques des assistants IA de « délégation cognitive » (« [cognitive offloading](#) »), pour les apprenants mais aussi pour les experts (voir par ex., [ici](#), [ici](#) et [là](#)). Or, peut-on comprendre un concept, une nouvelle question de recherche par un simple résumé et quelques références choisies par similarité ? Résumer et synthétiser, [ce n'est pas](#) analyser et réfléchir. La crainte est alors de voir se développer « un nouveau type d'intermédiaire entre les lecteurs et la littérature académique » ([Tracy Bergstrom et Dylan Ruediger](#)), ainsi que des « *illusions of understanding* » qui nous poussent à surestimer nos connaissances d'un sujet et au

final d'« introduire une phase de la recherche scientifique où l'on produit plus mais l'on comprend moins » ([L. Messeri et M. J. Crockett](#)) : « Ces outils encourage une culture du « pratique » et de l'« efficacité » plutôt que du « bien pensé » et de la *slow science* » ([S. P. J. M. Horbach et al.](#)). L'ambition de la recherche, toujours imprévisible et itérative, serait alors amoindrie et réduit à la simple équation : « Je cherche, je trouve » ([Richard Jefferson](#)). Néanmoins, si elles ne remettent pas en cause les limites des IA génératives, certaines voix veulent y voir des vertus pédagogiques : « Ce sont plus comme des pointeurs vers des idées / de la littérature que vous devez investiguer vous-mêmes » ([John Horton](#)), ce que signalait dès le début [Mark Carrigan](#), appelant à utiliser les IA génératives « d'une manière réflexive et dialogique » ;

- une **orientation biaisée des recherches futures** : utiliser ces outils sans connaître leurs biais risque, à terme, de produire l'« illusion de l'étendue de l'exploration » (« *illusion of exploratory breadth* », [L. Messeri et M. J. Crockett](#)). En outre, cela risque de produire une recherche moins diverse. En effet, alors que des outils bibliographiques récents comme OpenAlex souhaitent favoriser une meilleure inclusivité de toutes les recherches, tant linguistiques, géographiques que disciplinaires, les outils intégrant de l'IA générative ne peuvent bien fonctionner qu'avec des contenus interrogeables par les LLM, soit les textes intégraux et les *abstracts*, laissant ainsi de côté tout un pan de la littérature sur un sujet. Autre motif de crainte, le risque de voir une homogénéisation et un appauvrissement des sujets de recherche, en privilégiant les sujets couverts par ces outils ou les [sujets susceptibles d'être les plus cités](#). Sous réserve cependant que les sujets émergents soient bien couverts par les outils ! En effet, Consensus [nous invite](#) à privilégier des sujets vraisemblablement couverts par des articles de recherche académiques, tandis que le [Consensus meter](#) « a besoin d'au moins 5 oui/non prédictions à afficher », faute de quoi, il invite simplement à « essayer de poser une question de recherche populaire ». Ces outils « peuvent être un outil utile s'il[s] aide[nt] les gens à plonger plus profond dans la littérature scientifique. Le problème est que cela pourrait juste les faire s'arrêter » ([Seth Baum](#)). « Qu'est-ce qui arrive à l'innovation quand tout le monde utilise le même outil avec les mêmes biais, qui est essentiellement construit pour nous offrir plus de choses que nous aimons / connaissons déjà ? Est-ce la recette pour une homogénéisation et une stagnation similaire de la science et de la construction du savoir en général ? » ([David Crotty, 2024](#)). L'accélération de la recherche de littérature ne remet-elle pas en cause notamment [la sérendipité](#) ? ;
- la **pression sur les chercheurs** : évoqué comme l'un des risques majeurs dans un [rapport de la Royal Society](#), l'un des risques de la course à l'IA serait de pousser les chercheurs à incorporer de plus en plus de technologies IA pour des motifs de coûts et de productivité au détriment des méthodes plus traditionnelles, développant des chercheurs « bons en IA » plutôt que « bons en science ». Il ne faudrait pas que le développement de ces outils contribue au développement de nouvelles inégalités numériques (techniques, informationnelles et financières) au sein même des communautés académiques.

« *Not for the lazy* »

Là encore, la communication des outils est bien huilée : « Nous nous efforçons de permettre aux scientifiques d'arrêter de chercher [de la littérature] et de retourner faire de la vraie science » ([Consensus](#)). Or, les outils bibliographiques avec IA générative illustrent bien à leur manière le « paradoxe de la numérisation » du travail scientifique ([Barbara Ribeiro et al.](#)) car ils nécessitent toujours un important « travail du savoir banal » (« *mundane knowledge work* ») d'analyse et de vérification. [« Not for the lazy »](#) (« Pas pour les paresseux ») : c'est d'ailleurs bien l'un des principaux messages passés par Elsevier lui-même lors de la sortie de

Scopus AI. Comme tout outil de recherche bibliographique, ils doivent s'utiliser avec un **vrai regard critique**, tant sur l'outil lui-même que sur les résultats fournis et ses propres connaissances disciplinaires, afin de développer de vraies idées originales. De fait, [chacun de ces outils a ses particularités](#) (diversité de couverture, longueur des résumés, facilité de vérification, reproductibilité, biais). Par ailleurs, les erreurs existent toujours. Si certaines sont liées aux [métadonnées utilisées](#), nombre d'entre elles sont issues des [LLM eux-mêmes](#). Tant [Elicit](#) que [Consensus](#) estiment que 90 % de l'information fournie est correcte, ce qui laisse quand même 10 % de marge d'erreurs, et ils invitent les utilisateurs à toujours [faire preuve d'esprit critique](#), en travaillant bien les questions, en vérifiant les résultats fournis et en diversifiant les sources. Cependant la situation est plus délicate qu'avec les hallucinations d'un ChatGPT : « La « menace » IA pour la revue de littérature n'est pas qu'il peut citer des papiers fictifs mais parce qu'il (les systèmes de type RAG) citera de vrais papiers de manière inappropriée (mais pas des papiers inventés). Le premier est facile à détecter. Le second est presque impossible, surtout quand l'erreur est subtile » ([Aaron Tay, 09/2024](#)).

Par ailleurs, la **recherche d'information en soi est un processus** participant [de la réflexion scientifique](#), de la construction de compétences informationnelles et plus généralement situant les chercheurs comme des producteurs et non pas seulement des consommateurs de savoir. Que perd-on, d'un point de vue épistémique et informationnel, en s'appuyant majoritairement sur l'IA générative dans les outils bibliographiques ? La [recherche extensive](#) n'est pas seulement instrumentale – identifier des sources –, mais a une valeur intrinsèque – construire un cadre conceptuel et théorique. La manière de poser une question n'est pas un problème seulement sémantique ; pour poser les bonnes questions, les chercheurs ne doivent-ils pas avoir un minimum de bases sur le sujet et les méthodes de la discipline ? Or, nombre de limites des LLM [découlent directement de leur manque de connaissances](#) disciplinaires et méthodologiques. D'une manière générale, en l'état actuel des choses, ces outils [passent à côté](#) des éléments clés d'une revue de littérature : « La connaissance du domaine, l'esprit critique et la capacité à faire des liens entre des études ». L'automatisation de ce travail [pourrait diminuer la valeur](#) du travail scientifique, accélérant le processus, au détriment de la profondeur de la réflexion.

Par ailleurs, comment utiliser ces outils de manière responsable [à l'aune des valeurs académiques actuelles](#) ? De nombreux acteurs scientifiques se sont en effet prononcés ces deux dernières années sur l'**utilisation responsable** des IA génératives dans le processus d'écriture – le Code de conduite européen pour l'intégrité dans la recherche a d'ailleurs intégré ce volet dans sa [version revue en 2023](#). Comme le formule fort bien la [Charte sur l'utilisation de l'intelligence artificielle au sein de l'université d'Orléans](#), « il ne s'agit pas là d'une prise de position morale mais bien d'un positionnement pratique et pragmatique dicté par les principes de l'intégrité scientifique ». On pensera aussi aux recommandations au niveau européen, [Living guidelines on the responsible use of generative AI in research](#) de la Commission européenne ou au niveau d'établissements, du [CIRAD](#) et de l'[INRAE](#). Concernant l'utilisation de ces outils pour trouver des références bibliographiques, cela ne semble pas poser de problème à l'instar des bases de données bibliographiques antérieures ; en revanche, une utilisation plus extensive dans le processus serait à déclarer. Les [Guidelines de la Commission européenne](#) indiquent ainsi que « les chercheurs, pour être transparents, détaillent les outils d'IA générative ont été utilisés substantiellement dans leur processus de recherche » et « Par exemple, l'utilisation d'IA générative comme outil de base d'aide à l'auteur n'est pas un usage substantiel. Cependant [...], mener à bien une revue de littérature [...] peut avoir un impact substantiel ». La question se pose aussi pour la production de synthèses avec ces outils. [Certaines sociétés de droits d'auteur et éditeurs](#) demandent ainsi que les IA ne soient pas utilisées par le monde académique pour résumer les productions de tiers, de crainte notamment de récupération des données pour les fournisseurs de service. La question se pose aussi de la réutilisation des parties de synthèse

fournies par l'outil, au niveau de l'article ou de plusieurs articles, par exemple dans un état de l'art. Trois éléments sont à prendre à compte : les conditions générales d'utilisation des services ; la propriété intellectuelle ; l'utilisation responsable de ces outils. Du point de vue de l'outil, la situation est plus nuancée que les tchatbots comme ChatGPT qui autorisent souvent la réutilisation des contenus produits (*terms* de [ChatGPT](#)). Pour les outils bibliographiques, leurs conditions générales d'utilisation peuvent attribuer une licence sur le contenu, mais généralement limitée : le contenu de [Consensus](#) est ainsi réservé à un usage personnel non-commercial et ne doit être publié ; la position est assez similaire chez [SciSpace](#). Du point de vue de la propriété intellectuelle, il ne s'agit pas de plagiat puisque ce sont des contenus produits par une machine et non des œuvres de l'esprit. En revanche, du point de vue de l'éthique, s'attribuer des contenus qui ne sont pas les siens relève d'une méconduite scientifique. D'ailleurs, pour [Elicit](#), les résumés fournis « ne sont pas prévus pour être utilisés comme votre propre texte ».

En l'état actuel, même si les positions sont variables [chez les éditeurs notamment](#), les [consignes données vis-à-vis des IA génératives de type ChatGPT](#) devraient [s'appliquer](#) à ces contenus produits par les IA génératives des RAG. On trouve au cœur deux notions fondamentales : **la transparence et la responsabilité des usages** :

- les auteurs sont responsables de ce qu'ils signent (*accountability*). Ils doivent vérifier les contenus fournis ; les erreurs laissées dans un texte après utilisation de l'IA sont de leur responsabilité (*human oversight* et *verification*) : « Rien dans les services IA ou l'*output* ne remplace votre expertise indépendante et votre jugement professionnel » ([SciSpace](#)) ;
- les LLM ne peuvent être considérés comme des auteurs ou co-auteurs (*authorship*). Les utilisateurs ne doivent pas présenter comme leurs les contenus générés par l'outil ([SciSpace](#)). En ce sens, [il n'est pas correct](#) de citer ou de créditer des IA : les LLM n'ont pas d'autorité épistémique et ne sont pas propriétaires de leurs productions. De la même façon, sauf dans certains cas, les réponses des LLM ne sauraient être considérées comme des sources d'un travail ;
- les usages amenant à la production de contenus inédits substantiels doivent être déclarés (*transparency* et *disclosure*). Outre cette référence (*attribution*), on demande de plus en plus que ces usages soient documentés (*documentation* : nom, version, date) avec les requêtes utilisées (*input*) et les réponses obtenues (*output*) et discutés à des fins de preuves et de traçabilité (*reproductibility* et *robustness*), [notamment dans le contexte de la science ouverte](#) ;
- les chercheurs [doivent faire attention](#) à l'intégration de données confidentielles, sensibles, sous droit ou brevetables dans leurs interactions avec ces outils (*privacy*, *confidentiality* et *intellectual property rights*). On sait que les *inputs* des IA génératives généralistes peuvent être conservés et utilisés pour l'entraînement des modèles. Dans le cas de ces outils bibliographiques, on manque globalement d'informations sur le sort donné aux informations soumises par l'utilisateur mais l'on constate qu'à l'instar des réseaux sociaux, nombre de ces services s'attribuent en effet une licence de réexploitation sur les contenus du possesseur du compte (*terms* de [Consensus](#), d'[Elicit](#), de [SciSpace](#), de [Scite](#)) En revanche, du côté de Scopus, [on assure](#) que, même si Scopus AI utilise ChatGPT, les requêtes des utilisateurs ne sont pas conservées ou utilisées pour entraîner ChatGPT, et [dans le Web of science](#), les LLM sont déployés sur des serveurs privés. Se pose aussi la question des PDF sous droits, y compris sous licence ouverte, déposés par le chercheur lui-même pour se constituer des collections ou tchater avec le PDF puisque cela peut rentrer dans le cadre du DMCA (*Digital Millenium Copyright Act*) – il est vrai qu'[Elicit précise](#) cependant que les documents téléchargés resteront privés.

L'or noir des données académiques

On connaît la valeur des *data* pour l'IA. A l'heure de l'IA, plus que jamais, « *content is king* ». Dans le cas des RAG, l'IA générative n'est qu'une couche associée à des index bibliographiques. Il est donc crucial pour les fournisseurs de ces services de disposer au préalable de suffisamment de **contenus pertinents** pour la partie « recherche » (le R de RAG). Les [grands index bibliographiques ouverts](#), en développement depuis 10 ans, viennent apporter une solution. C'est notamment le cas de [Semantic scholar](#), dont les données soutiennent de [nombreuses applications](#) et se retrouvent dans Elicit et Consensus. Néanmoins, toutes les références ne sont pas exploitables par les LLM (métadonnées insuffisantes, absence d'*abstracts* ou de texte intégral) : cela explique pourquoi Elicit annonce seulement une base de données de 125 millions de références sur les 220 millions de Semantic scholar. Par ailleurs, les limites de Semantic scholar [sont connues](#) : couverture privilégiant les sciences dures, anglophones, sous forme d'articles... Il est possible cependant qu'à l'avenir d'autres bases bibliographiques ouvertes soient également prises en charge. Keenious utilise ainsi OpenAlex, d'une couverture plus large et plus diversifiée que Semantic scholar, même si, là encore, la [base de données est moindre](#) que l'index d'origine (111 millions contre 250 millions) : articles et *preprints*, avec suffisamment de métadonnées (DOI). Afin d'accroître les données disponibles et améliorer la couverture de l'outil, certains acteurs complètent ces index ouverts par des partenariats avec les éditeurs. Scite, par exemple, a des [agrément d'indexation](#) avec des éditeurs – Google Scholar ne fait pas autrement pour son index. Or, bien souvent, ces [tiers](#) restent dépositaires du *copyright* sur les *abstracts* et les textes intégraux, empêchant toute utilisation par les LLM sans accord préalable. Ainsi, [Scopus AI – et sans doute d'autres](#) – ne dispose pas de licence pour accéder au texte intégral et doit se contenter des *abstracts* obtenus dans le cadre du service Scopus.

Pour favoriser le développement des LLM sur les contenus scientifiques, il existe deux voies : la première concerne l'**ouverture des informations de recherche scientifiques** ; la seconde concerne le développement de partenariats avec les éditeurs. Concernant l'ouverture des données, on connaît l'essor évident de l'*open access* et de la science ouverte ces quinze dernières années. Dans la lignée de l'appel aux éditeurs à rendre accessibles les citations ([Initiative for Open Citations IAOC](#)), de plus en plus d'acteurs appellent désormais les éditeurs à rendre accessibles également les résumés ([Initiative for Open Abstracts – IAO4](#) et [Déclaration de Barcelone sur l'information de recherche ouverte](#)). Après les initiatives individuelles et les négociations avec les éditeurs, le développement de l'IA générative serait le [troisième âge de l'accès ouvert](#). Mais si le mouvement est bien implanté désormais pour les citations, [nombre d'éditeurs](#), à commencer par Elsevier, Taylor et Francis ou encore IEEE, sont encore réticents à ouvrir les résumés. Cette timidité à l'égard de l'ouverture a des [fondements stratégiques et économiques](#) : l'IA n'est qu'un moyen supplémentaire de monétiser les contenus éditoriaux et leur expertise disciplinaire à une époque où les fournisseurs de ces contenus deviennent de plus en plus des fournisseurs de services complets et où l'intermédiation de l'IA entre les productions scientifiques et les éditeurs risque de [faire baisser le lectorat monétisable direct](#). Les fournisseurs de données de qualité comme les grandes maisons d'édition seraient parmi les « gagnants inattendus » ([Alex Zhavoronkov](#)) de la révolution générative. De la même façon que les IA génératives généralistes tendent à renforcer le poids des GAFAM, l'implémentation généralisée de fonctionnalités IA dans les outils de recherche académique pourrait finalement renforcer ou maintenir le poids des principaux acteurs du domaine, possesseurs de contenus éditoriaux de niche essentiels aux futurs développements qualitatifs des LLM.

Les données scientifiques de qualité sont importantes non seulement pour les contenus interrogés mais aussi pour l'**entraînement des LLM**, *a fortiori* pour des LLM ayant à interagir

avec des contenus scientifiques (le G de RAG). Problème : non seulement les données d'entraînement de modèles généralistes comme GPT couvrent mal les contenus scientifiques, mais on semble [arriver au bout](#) du web interrogeable, faisant craindre un « effondrement du modèle » (« *model collapse* », [Illia Shumailov et al.](#)). Il est devenu crucial pour les IA génératives de couvrir désormais les contenus de niche, souvent de haute qualité, à commencer par les médias et les publications scientifiques. C'est ce qui explique les partenariats noués entre OpenAI et des journaux comme *Le Monde* ou le *Financial Times*. Après une [période de refus](#), nombre de fournisseurs de contenus font le pari de partenariats avec les fournisseurs d'IA, selon une logique gagnant-gagnant : fiabilisation des réponses de l'IA d'un côté et ressources financières complémentaires de l'autre. Si le mouvement était déjà bien engagé du côté des médias, l'annonce durant l'été 2014 de [plusieurs éditeurs scientifiques](#) d'ouvrir leurs données à l'IA a fait l'effet d'une bombe pour la communauté académique. Taylor & Francis, Wiley et Oxford University press ont ainsi annoncé des rapprochements avec des sociétés d'IA, tandis que Cambridge University press propose désormais une option *opt-in* à ses auteurs. Comme dans le [cas du Monde](#), il [s'agit pour ces éditeurs](#) de « protéger l'intégrité » des chercheurs, de « limiter la reproduction *in extenso* des textes » ou encore de « défendre le rôle vital des chercheurs » dans un monde d'IA. Mais il s'agit tout autant de diversifier le modèle économique : du côté de [Taylor & Francis](#), on escompte ainsi 58 millions de livres sterling ; 44 millions de dollars chez [Wiley](#). Dans l'immédiat, tous les détails ne sont pas encore connus, mais [le mouvement semble lancé](#) et l'on pourrait envisager [deux modes d'utilisation](#) de ces contenus : pour [l'entraînement des LLM et leur fine-tuning](#) (accès aux contenus derrière *paywall*) et pour le [RAG à la volée](#) (résumé des contenus permettant ainsi leur diffusion et leur réutilisation indépendamment du texte intégral). En outre, le RAG permet de [mieux conserver](#) les sources d'origine des données.

« A cause des intelligences artificielles, le Web se ferme de plus en plus » ([Le Monde, 13/09/2024](#)). Ce qui est vrai pour le web en général est également vrai pour le web scientifique et c'est tout le paradoxe de la période. A l'heure du développement d'index bibliographiques ouverts mondiaux, le développement parallèle des IA génératives dans les outils de recherche risque à terme de renforcer un **web scientifique à deux vitesses**, nouvelle illustration des jardins fermés (« *walled gardens* ») de Tim Berners-Lee et du souhait des éditeurs de se placer de plus en plus comme des [fournisseurs de données](#) : d'un côté un web ouvert et gratuit, interrogeable par les IA mais risquant d'être de plus en plus pollué par les dérives actuelles ([paper mills et éditeurs prédateurs en tête](#)), de l'autre, un web payant pour certains contenus et fermé à l'interrogation des IA – entraînant un [raidissement des chercheurs](#) et [remettant en question](#) des années d'ouverture de la science. Preuves en sont le [retrait de certains abstracts d'Elsevier et Springer Nature](#) dans OpenAlex, l'[opt-out de certains fournisseurs de Primo research assistant](#) ou encore les [positions des éditeurs](#) sur la réutilisation de leurs contenus dans le cadre d'IA institutionnelles. On peut s'interroger aussi [sur le futur positionnement des éditeurs](#) face aux fonctionnalités *Deep research* en cours de déploiement : accepteront-ils les *crawlers* de ces services ? Le développement des LLM et des IA génératives devient une [nouvelle source de revenus](#) pour les éditeurs, au prix d'[inégalités entre les acteurs eux-mêmes](#).

Business as usual ?

Le modèle « *freemium* » gagne de plus en plus de plateformes web et de [moteurs de recherche](#) ; les moteurs académiques avec de l'IA générative ne font pas exception. Non seulement faut-il généralement disposer d'un compte, mais aussi d'un abonnement pour profiter de l'ensemble des fonctionnalités. C'est le cas pour Elicit, SciSpace ou encore Consensus, tandis que Scopus AI et le Web of science Research assistant nécessitent un abonnement supplémentaire par rapport aux bases historiques. Et même les initiatives ouvertes développent progressivement un **business model** : ainsi d'OpenAI passé d'une organisation à but non

lucrative à une société valorisée à + 150 MM. \$; ainsi d'Elicit, issue d'une [organisation à but non lucratif](#) à une société [levant 9 M. \\$](#) ; ainsi de Consensus d'abord [financé](#) par l'Allen Institute for Artificial Intelligence, levant [+ 11 M. \\$ de capitaux risques](#) ; ainsi enfin de SciSpace issu d'un [hackathon](#). Le fait que le service donne son nom à l'entreprise est d'ailleurs révélateur, tout comme les rachats par des acteurs plus importants ; ainsi de Scite, récipiendaire de *grants* du National Science Foundation et du National Institute on Drug Abuse of the National Institutes of Health, acquis par Research Solutions [fin 2023](#).

Outre les infrastructures nécessaires à tout service numérique, l'utilisation de l'IA générative accroît de fait les **coûts de fonctionnement**, [balançant entre qualité et coût](#) : chaque requête pourrait coûter ainsi [plusieurs centimes](#), soit un coût marginal très élevé pour chaque utilisateur. Par ailleurs, le coût ([et les émissions en CO₂](#)) diffère selon l'action demandée : les fonctions de recherche de littérature et d'extraction de donnée consomment ainsi moins que les recherches plus poussées ([listes des concepts](#), [fonction Copilot](#)). Dans ces conditions, le principe même d'un abonnement mensuel fixe peut vite toucher ces limites, et l'on voit certains acteurs privilégier un système de crédits de consultation, qui n'est pas sans rappeler le paiement à l'acte des bases de données des années 1980 ([exemple pour Elicit](#) jusqu'en 07/2024 et pour [ScholarAI](#)). Il peut alors revenir à l'outil ou à l'utilisateur de définir dans quels cas utiliser des fonctionnalités plus poussées, mais plus coûteuses (ex. : [high-accuracy model chez Elicit](#)) ; l'outil peut aussi mettre en place un système qui limite la « lecture » des documents, en ne proposant les documents moins pertinents qu'à la demande (ex. : [Undermind](#)). Se pose aussi la question des modalités d'accès aux modèles LLM, souvent fournis par d'autres acteurs. Sur le modèle de Bing, dont le [prix de l'API avait bondi](#) en 2023, faut-il envisager un bond conséquent du prix de l'API d'OpenAI, sous-couche LLM traditionnelle de nombre d'outils, alors que l'abonnement à ChatGPT [est amené à augmenter rapidement](#) et que les derniers modèles sont [souvent plus coûteux](#) que leurs prédécesseurs ? On comprend alors mieux le souhait de gros acteurs, de développer leurs propres modèles LLM pour être autonomes : [Scopus](#) a ainsi en instance de brevet sa propre technologie RAG. Enfin, quid de coûts pour d'éventuels partenariats avec des éditeurs ? Dans ces conditions, malgré les progrès techniques, les coûts d'accès à ces outils risquent d'augmenter et l'on risque d'assister à la même situation que les moteurs de recherche au début des années 2000 ou des réseaux sociaux académiques à la fin des années 2000 : une fois la mode passée, quand les financeurs se lasseront, le paysage va se simplifier, en lien aussi avec [la confiance](#) que l'on peut accorder à ces acteurs. D'autant que les utilisateurs ne sont peut-être pas prêts à payer pour de tels outils. Consensus annonçait ainsi seulement [12 000 utilisateurs payants sur 400 000 utilisateurs mensuels](#). Dans ces conditions, on peut aussi envisager le développement de publicités sponsorisées : c'était déjà l'un des choix de ResearchGate ; on l'a vu pour [Perplexity](#) et la [réflexion est en cours chez OpenAI](#).

Dernier point récurrent pour ce genre d'initiative : l'ouverture de services encore souvent mal testés. Rappelons, pour l'anecdote, que ChatGPT est sorti fin 2022 malgré tous les [appels à la prudence émis précédemment](#) sur les hallucinations des tchatbots. C'est aussi le cas avec ces outils bibliographiques, sortis « inconfortablement en bêta pour itérer rapidement avec les retours des utilisateurs » ([Elicit](#)), « avec beaucoup de points à améliorer » ([Consensus](#)). Mais c'est aussi le cas chez des acteurs historiques : JSTOR présente ainsi son assistant IA en [version bêta](#). Du côté de Scopus et du Web of science, en revanche, une attention particulière a été donnée à des tests usagers avant l'ouverture générale. Dans la mesure où la technologie des LLM est très jeune, ces sociétés [manquent encore d'informations pour évaluer](#) leurs propres plateformes. Face aux questions voire aux risques que ces IA peuvent représenter pour la recherche et l'expertise des chercheurs, les **appels à des audits** indépendants se multiplient « pour assurer le futur de l'accès au savoir » ([Michael Gusenbauer](#)).

Une offre en devenir

La multiplication et la démocratisation des outils bibliographiques boostés à l'IA générative s'accompagnent d'une **progressive homogénéisation** des principales fonctionnalités proposées. Il est donc important de **noter les différences et les pistes de développement** entre les outils :

- du point de vue des utilisateurs :
 - maîtrise de l'utilisateur sur le processus : l'interface de recherche se présente souvent comme une simple fenêtre de requête simple. Rares sont les outils comme [Scite assistant](#), qui permettent de définir des paramètres de recherche *a priori*. En outre, l'utilisateur devrait pouvoir intervenir à toutes les étapes du processus,
 - présentation des résultats sous forme de liste, mais aussi sous forme de tableaux, notamment dans le contexte de [développement des revues de littérature systématique](#),
 - offre « *Teams* » pour la collaboration,
 - guides d'utilisation : [encore trop souvent](#), ces outils proposent peu d'aide aux utilisateurs sur leur fonctionnement (données et technologies utilisées, mises à jour, construction de la requête, pièges et limites à éviter...),
 - prise en compte de différents profils d'utilisateurs, [distinguant](#) par exemple entre une simple découverte et des questions plus précises sur un corpus ;
- du point de vue des données :
 - critères d'évaluation : il peut être intéressant d'avoir des scores de confiance des réponses. Il serait tout aussi intéressant que le système soit capable d'indiquer l'absence de réponses pertinentes en-deçà d'un certain seuil, ou la partie de la réponse entièrement produite par l'IA,
 - prise en compte des publications problématiques : parallèlement aux préoccupations actuelles sur les contenus douteux, il serait bon que ces outils intègrent des filtres spécifiques, notamment sur les publications ayant fait l'objet de signalement (errata, rétractations...) à l'instar des bases bibliographiques plus classiques. Combien, [à l'instar de Keenious à partir d'OpenAlex](#) se préoccupe de ne pas couvrir des contenus non-rétractés, voire des [prédateurs](#) ?,
 - [déconstruction de l'article](#) : Semantic scholar avait déjà illustré comment les documents peuvent être découpés en unités de sens (extractions des figures et tableaux, citations dans leurs contextes). Certains outils [comme ScholarAI](#) sont désormais capables d'extraire certains contenus de PDF, comme des figures et de les analyser. De même, on devrait pouvoir chercher sur certaines parties seulement des publications (méthodes, discussion...). Afin de favoriser l'accessibilité et le travail des IA sur les publications, [certains](#) appellent à [changer la forme](#) même des publications,
 - prise en compte de différents types de contenus : les LLM sont de plus en plus des outils multimodaux, capables de prendre en compte des contenus divers (textes, images, etc.). Pour l'instant, les outils bibliographiques se concentrent sur des documents textuels, mais il est possible d'envisager, à l'instar de feu Galactica, des outils prenant mieux en compte les formules chimiques ou les séquences de protéines, ou des outils capables de résolutions mathématiques, dans un [modèle « text-to-science »](#). Il est aussi possible d'envisager à terme une [diversification du type de données](#) disponibles (présentations, podcasts, vidéos, etc.), voire – dans la lignée du succès de Google NotebookLM –, une diversification des formats de sortie (podcasts, etc.) ;

- du point de vue des résultats :
 - transparence des résultats : alors que les [LLM gagnent même les techniques plus anciennes de query expansion](#), il faut pouvoir garder le contrôle sur la manière de chercher, par exemple en forçant la recherche sur les mots-clés saisis et non une recherche augmentée,
 - multilinguisme : bien qu'en progrès ces derniers mois dans les outils bibliographiques, la [recherche en différentes langues](#) (*cross-language information retrieval* – CLIR) présente souvent des résultats différents selon les langues. En outre, la prééminence de l'anglais notamment dans les recommandations de requêtes pourrait [générer des biais](#) culturels,
 - fonctionnalités d'analyse et de visualisation de données : ces outils sont d'abord pensés comme des outils de recherches bibliographiques. Les représentations graphiques, plus souvent associées à des recherches bibliométriques, comme pour OpenAlex ou le récent [HAL monitor](#), sont absentes de ces outils encore très linéaires et textuels.. Ces outils ne sont pas davantage des [outils de cartographies de citations](#), autre famille d'outils qui se développe actuellement (ResearchRabbit, Connected Papers...) et qui fonctionne sur d'autres principes (cartographie de réseaux). Dans l'immédiat, ce sont surtout Scopus AI et le Web of science Research assistant qui en proposent, dans la lignée des tableaux de bord déjà disponibles dans leurs interfaces (ex. : [cartes des concepts](#) pour le premier ; [visualisations contextuelles](#) dans le second avec graphique chronologique, carte de concept et réseaux de citations),
 - technologies de mémoire : l'outil doit pouvoir se référer à des échanges précédents, par exemple pour compléter ou mettre à jour des contenus déjà présentés. De là à envisager des outils personnalisés suggérant des pistes de recherche ou proposant des résumés en fonction du profil du chercheur ([Elicit](#), p. 220), il n'y a qu'un pas – au risque d'accroître toujours plus la variabilité des réponses d'un chercheur à l'autre et donc le manque de reproductibilité de la démarche,
 - technologies pour fiabiliser les résultats : si les LLM entraînent nécessairement des erreurs et des hallucinations, de nombreux travaux sont en cours pour permettre une meilleure compréhension des contenus scientifiques, notamment dans le cadre des revues de littérature systématiques (voir par ex., [ici](#), [ici](#), [ici](#) et [là](#)), dans la lignée des évolutions mentionnées plus haut : [progrès sémantiques](#) pour mieux prendre en compte les contenus académiques (modèles de langage plus petits mais spécialisés ([SLM – small language models](#)), [taxonomies, ontologies et knowledge graphs](#)), progrès sur les données pour s'appuyer sur des contenus de qualité (RAG), progrès logiques pour dépasser la réponse probabiliste (chaînes de vérification et de raisonnement, arbres de pensée - cf. [ici](#) pour une présentation simple). On peut ainsi penser à [comparer les scores de plusieurs modèles LLM](#) ou [mettre en place un système de vérification automatisé](#). Il faudrait aussi donner plus de poids au niveau de preuve, par exemple en donnant plus de poids à des méta-analyses qu'à des études uniques ou en dépassant les métriques classiques (métriques associées à la revue, citations reçues) pour se concentrer sur la publication elle-même.

Les principales pistes d'évolution semblent cependant se tourner vers une intégration de ces outils au plus près du *workflow* du chercheur. Les fonctions « copilot » ouvrent la voie vers une véritable « **ère de l'assistance** ». Ce concept démocratisé par [Google dès 2017](#) doit faire sortir des outils de recherche pour une vraie assistance personnalisée et individuelle. Deux modèles sont d'ores et déjà visibles :

- le « copilote », dans une logique d'[intégration verticale](#) : dans ce modèle, la recherche d'informations n'est qu'une étape d'un processus, et se situe au plus près d'autres outils ou modalités (logiciel de traitement de texte, analyse de données...). C'est le parti-pris d'un outil comme [ScholarAI](#), d'abord un simple GPT, complété ensuite d'une interface web, association de fonctionnalités de recherche et de découverte (outil bibliographique, résumés et *insights*, recommandations), d'édition (rédaction, gestion de références bibliographiques, visualisation de données, multilinguisme) et d'apprentissage (production de contenus personnalisés, aide à la prise de notes). On [pourrait alors envisager](#) des outils capables d'analyser des *drafts* et de suggérer de nouvelles pistes bibliographiques ;
- l'« agent », spécialisé sur la recherche bibliographique : en associant le potentiel de recherche contextuel du RAG et d'exécution autonome de l'« [agentic AI](#) », l'outil s'adapte au fur et à mesure de la recherche comme le ferait un assistant humain personnel. Sur le modèle du *Deep research* de ChatGPT, l'utilisateur indique ses objectifs et l'agent se charge de mener à bien la recherche, en décidant par exemple quoi et comment chercher. Ce faisant on se rapproche de vraies [recommandations personnalisées](#), permettant au chercheur de se centrer davantage [sur la conception et la réflexion critique](#). C'est le parti pris d'[Undermind](#), qui ne se contente pas de donner des réponses, mais accompagne la démarche de recherche dans une [démarche itérative par étapes](#) (aide à l'explicitation de la question, prédiction du nombre de papiers existants et représentativité de la réponse, score de pertinence, extension possible de la recherche aux papiers moins pertinents, rapport de recherche). Ce faisant, un [premier benchmark](#) montrerait une nette amélioration par rapport à une simple requête Google scholar, au prix, il est vrai, d'une durée de traitement un peu plus longue. Bien évidemment, on peut s'attendre à ce que de plus en plus d'acteurs bibliographiques proposent ce genre de service. [SciSpace](#) vient tout juste d'annoncer son *AI-powered research agent*.

CONCLUSION PROVISOIRE POUR LES FORMATEURS

Outils de découverte permettant de tester rapidement des hypothèses, de prioriser quoi lire et de faire un pas de côté transdisciplinaire ? Outils de lecture rapide permettant de gagner du temps pour favoriser le reste du processus académique ? Ou au contraire, illustration d'une recherche scientifique privilégiant la vitesse et la productivité au détriment de la qualité ? Sans doute, tout cela à la fois. Bien sûr, nombre de limites signalées ici préexistent au développement des LLM dans les outils bibliographiques : manque d'exhaustivité des outils, sélection de citations appropriées, déformation de citations, outils propriétaires dans un environnement académique... Ces interrogations sont au cœur de la recherche documentaire et de la méthode scientifique. Comme bien souvent, l'IA vient surtout mettre en valeur les interrogations sur la manière de faire de la recherche.

Au croisement des questions disciplinaires, informationnelles et épistémiques, nous ne sommes encore qu'au début de l'intégration de ces IA génératives dans les outils documentaires et nous manquons de recul sur les usages et les enjeux éthiques. Néanmoins, « les fondements de ces développements sont déjà en place, faisant des avancées significatives une question de « quand » et non de « si » » ([Mark Glickman et Yi Zhang](#)). Il convient dès lors de prendre en compte dès maintenant la mesure des changements en cours et de savoir accompagner étudiants, [chercheurs](#), personnels d'appui à la recherche et [décideurs](#). Et ce, alors même que les étudiants [sont en attente](#) d'accompagnement et de cadre institutionnel clair pour utiliser ces outils. Car derrière la facilité d'utilisation de ces outils, la vigilance reste de mise pour vraiment les comprendre et les utiliser. Bien sûr, le développement de ChatGPT ne semble souvent que réactiver des réflexions et des craintes antérieures. Le développement de Google au tournant

des années 2000 et l'apparition de Wikipédia quelques années plus tard nous ont habitués à s'adapter dans un paysage documentaire en constante évolution et [nous fournissent des éléments de réflexion](#) pour la situation actuelle. Les [pistes d'action perçues actuellement face à l'IA générative](#) sont **dans la continuité de 25 ans de réflexion autour des compétences informationnelles**. Il n'en reste pas moins que les professionnels de l'information sont pris entre [plusieurs mouvements et injonctions contradictoires](#), reflet aussi des préoccupations sociales et académiques actuelles – on notera notamment que la récente mise à jour du [Référentiel des compétences informationnelles de l'ADBU](#) mentionne spécifiquement de savoir « utiliser l'IA de façon éthique et responsable ».

En complément du [décalogue du formateur en usages LLM](#) de Christophe Deschamps, complété des remarques d'Emmanuel Barthe, on notera l'**importance des points** suivants pour la recherche documentaire académique :

- le développement d'une nouvelle famille d'outils et de nouvelles techniques de recherche pour la recherche documentaire (tchatbots et prompt) : il est tout aussi important de [savoir quand l'utiliser que comment](#) ;
- l'intégration des LLM dans de nombreux écosystèmes numériques, de manière de plus en plus invisibilisée. C'est notamment le cas [dans les outils pour la revue de littérature](#), de la question de recherche à la recherche bibliographique et aux outils d'assistance ;
- le rôle central des RAG dans ce contexte, pour limiter les hallucinations en s'appuyant sur des sources identifiables, malgré [les questions que cela pose](#) à tous les niveaux. Dans ce cadre, on se reportera aux conseils donnés par les services pour construire sa requête : contrairement aux tchatbots, ces services s'interrogent normalement en langage naturel, et non avec des prompts ;
- la [différence](#) entre une opération de recherche d'information et une opération de synthèse générative, qui est souvent plus sélective. En ce sens, il convient de distinguer les outils de recherche classique (Google Scholar, Scopus, Web of science, OpenAlex...), des outils avec de l'IA (Elicit, Consensus...) : ce ne sont ni les mêmes données ni les mêmes méthodes de recherche. De même, Scopus AI et le Web of science Research assistant ont des particularités différentes par rapport aux bases d'origine ;
- le développement d'une recherche générative (*generative search*) ou sémantique (*semantic search*) dépassant la traditionnelle recherche par mots-clés (*keyword search*), et [les compléments](#) que peuvent apporter la recherche avec IA aux méthodes plus classiques (mots-clés + booléens, co-citations), alors même qu'aucun outil ou méthode [n'est suffisant en lui-même](#) ;
- de [nouvelles possibilités de recherche](#) d'information, d'exploration et de réflexion, grâce à l'extraction de données (tableaux), l'aspect conversationnel et les fonctionnalités *deep research* ;
- [le rôle central des sources](#), tant les sources derrière les IA que la mention de l'usage d'IA ;
- la nécessité de distinguer les messages et les outils selon les niveaux (licence, master, doctorat) : [selon Aaron Tay](#), par exemple, il ne faudrait pas présenter les outils bibliographiques augmentés à l'IA avant le master, une fois acquis les bases disciplinaires et méthodologiques de la recherche ;
- le rôle toujours central de la [méthodologie documentaire et de la démarche d'investigation](#), avec son aspect itératif et critique. Le recours aux tchatbots et outils automatisant la revue de littérature entraîne en effet une « moins bonne compréhension de ce qu'est un état de l'art » et une distanciation avec le sujet ([Fabienne Rosier et Christelle Grima](#)). [A l'image du pilote automatique](#), il faut

savoir d'abord piloter sans l'outil pour l'utiliser au mieux. Dans la mesure où ces outils dépassent les mots-clés au profit de questions en langage naturel, cela signifie insister notamment sur [le contexte et les besoins préalables d'une recherche](#), tout autant que l'utilisation des outils eux-mêmes (analyse et problématisation du sujet, stratégie et planification de la recherche). Ce sera encore plus vrai avec les fonctions *deep research* ;

- le responsabilité humaine : si elle est fréquente dès que l'on parle d'IA, la notion d'« *human-in-the-loop* » est toute aussi pertinente pour la recherche documentaire et rappelle la nécessité d'intégrer la réflexion humaine à l'ensemble du processus. Là encore, la vigilance sera encore plus vraie avec les fonctions *deep research* ;
- l'intégration de ces outils dans une réflexion plus large autour d'une [generative AI literacy](#), couvrant aussi leur aspect éthique, leur coût écologique, leur sécurité et leurs liens à l'intégrité scientifique ;
- enfin, les progrès en cours en direction d'outils plus puissants et fiables, dans la lignée de l'IA agentique, permettant de réaliser de [vraies revues de littérature](#).

Dès lors, [comment intégrer l'IA aux formations](#) ? Que faut-il adapter ? De fait, les [formateurs ont parfois du mal à se positionner](#) : ils ne sont pas toujours eux-mêmes utilisateurs de ces outils au quotidien, et il est difficile de faire face aux constantes évolutions et multiplication d'outils IA. En l'état actuel des choses, nous ne sommes qu'au début de l'analyse de ces outils ; il [convient désormais d'approfondir](#) notre connaissance aux questions disciplinaires, d'une part, et aux besoins et aux compétences des usagers, d'autre part, tout [en respectant quelques principes méthodologiques clés](#) afin de comparer ce qui est comparable. S'il revient aux professionnels de l'information de connaître et de **savoir former aux grands principes et aux limites de ces outils**, la meilleure adaptation serait « l'exploration collaborative (avec les étudiants) » ([Sheila Webber](#)), notamment par des travaux de groupe, chacun apportant ses connaissances, ses interrogations et ses pratiques. Cela permet non seulement d'expérimenter, mais aussi de se situer au plus près des usages, de lutter contre les inégalités entre participants, et de [mieux évaluer les offres commerciales](#) des outils, voire [les intégrer](#) dans les outils mis à disposition. Il sera tout aussi intéressant de se rapprocher des experts disciplinaires, enseignants et chercheurs, pour évaluer au mieux les outils. Ce sera aussi l'occasion de suivre les développements, annoncés et prévisibles, de l'IA agentique. Mais ceci est une autre histoire...

DECLARATIONS D'USAGES IA ET REMERCIEMENTS

Aucun chat(bot) n'a été martyrisé pendant la rédaction de ce billet. D'autres technologies IA ont en revanche été utilisées ou testées (moteurs de recherche [Google, Brave search, Perplexity], bases de données bibliographiques [Scite, SciSpace, Consensus, Elicit pour quelques références et bien sûr, toujours, Google scholar] pour la recherche documentaire ; DeepL Translator pour une aide ponctuelle pour les traductions). Ainsi que beaucoup d'intelligence (humaine) collective 😊.

Aussi merci à tous les collègues et participants à nos actions de formation et sensibilisation pour les échanges, leur veille et leurs questions. Et s'il ne fallait n'en citer qu'un : le travail de [l'incontournable Aaron Tay](#).

POUR ALLER PLUS LOIN

Top 5 pour une sélection subjective :

- sur le contexte général de l'IA pour la science : [The Royal Society. *Science in the age of AI. How artificial intelligence is changing the nature and method of scientific research.* 2024.](#)
- sur le paysage actuel de la recherche d'information et les compétences associées pour les professionnels de l'information : [Aaron Tay. « Learning about information retrieval and "AI search" in 2025 ». *Musings about librarianship.* 12/01/2025.](#)
- sur les limites des LLM dans les RAG : [Stuart Leitch. « Guest Post: Jagged Edges of Conversational Interfaces Over Scholarly and Professional Content ». *The scholarly kitchen.* 30/05/2025.](#)
- sur l'usage responsable des IA : [Commission européenne. *Living Guidelines on the responsible use of generative AI in research.* 2024.](#)
- pour un support de formation : [Aster Zhao. *Emerging AI Tools for Literature Review.* 04/2024-](#), dont [support de présentation, 11/2024.](#)

L'URFIST de Paris met à disposition sur son site des ressources sur le sujet (supports commentés, bibliographies) :

- sur le contexte général et les enjeux des IA génératives et des moteurs de recherche [ChatGPT et les autres : recherche d'information et intelligence artificielle](#) (2025) ;
- sur les outils académiques : [Au-delà de ChatGPT : recherche d'informations académiques et intelligence artificielle](#) (2025) et la présentation [Assistance ou risque pour la science : l'IA générative au cœur des outils de recherche de littérature académique](#) ;
- pour les formateurs : [Former les usagers à l'heure de ChatGPT : IA et compétences informationnelles \(formation de formateurs\)](#) (2025) et la présentation [Assistant ou alternative aux compétences informationnelles ? IA générative et pratiques documentaires des usagers.](#)