

L'IA ET LA RECHERCHE DOCUMENTAIRE

En sciences de la santé

Marianne Ruel, MSI.

Mars 2025

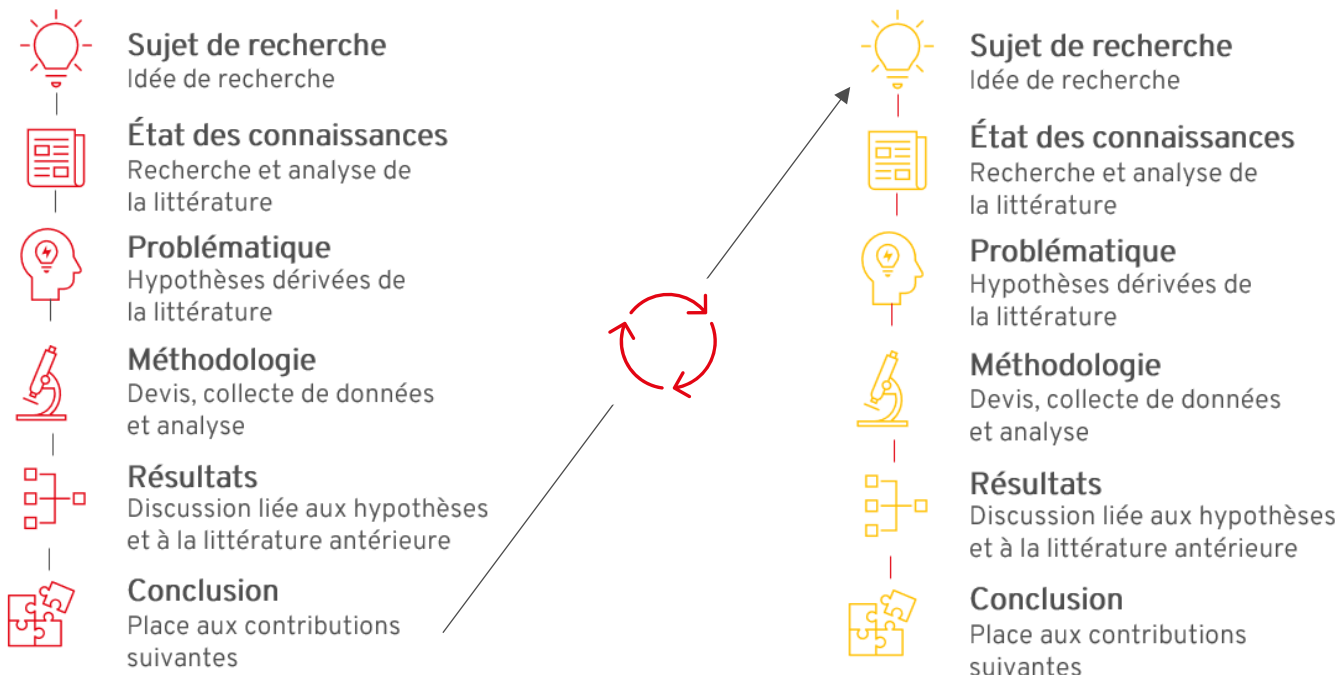
OpenAI. (2025). Image générée par DALL-E

Bibliothèque

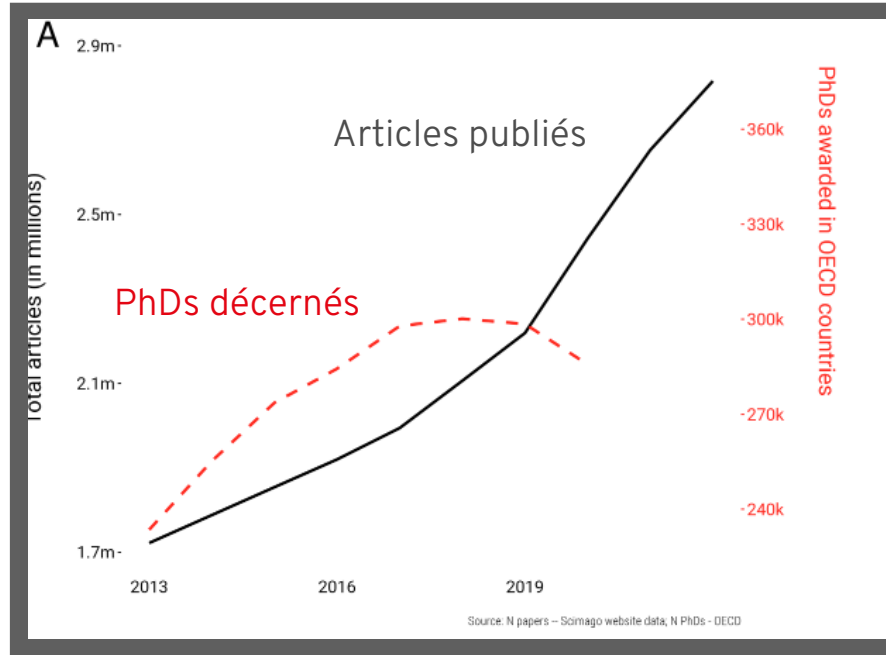


UNIVERSITÉ
LAVAL

La recherche documentaire au cœur de la Recherche

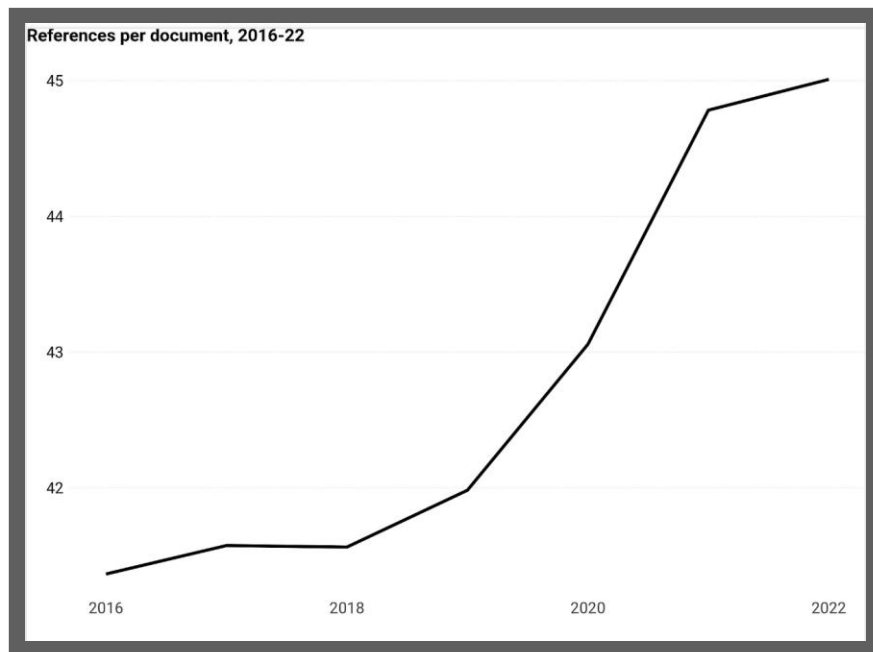


Croissance exponentielle des produits de la recherche



Évolution du nombre d'articles publiés et de doctorats décernés dans les pays de l'OCDE (2013-2019)

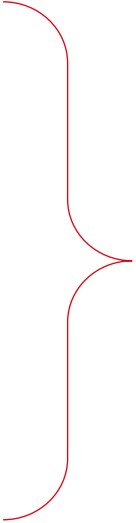
Croissance exponentielle des produits de la recherche



Références par documents publiés
entre 2016-2022

Croissance exponentielle des produits de la recherche

- › Plus de 3 millions de documents sont publiés en anglais annuellement.
- › En 6 ans, le nombre de publications indexées dans Scopus et Web of Science a augmenté de près de 50% (1,92 millions en 2016 et 2,82 millions en 2022).



!

Incidence directe sur la
recherche documentaire
et le **traitement de**
l'information

La recherche documentaire

Aujourd'hui



- › Les résultats issus de la recherche documentaire sont de plus en plus nombreux.
- › La recherche documentaire devient plus complexe.
- › Le traitement de l'information prend un temps considérable.

L'IA à la
rescousse ?



Objectifs de la présentation

- › Examiner les opportunités offertes par l'intelligence artificielle (IA) pour optimiser la recherche documentaire et le traitement de l'information.
- › Comprendre les limites de l'IA en recherche documentaire et les précautions à prendre pour une utilisation adéquate dans le domaine des sciences de la santé.

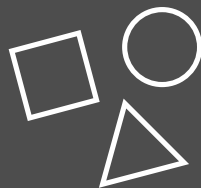


Plan de la présentation

- › L'ABC de l'IA
- › L'IA en recherche documentaire
- › Enjeux et limites
- › Discussions

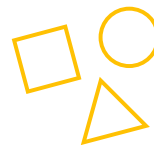
Point de vue d'une spécialiste de
l'information et non de l'IA !

L'ABC DE L'IA



Intelligence artificielle

Définition



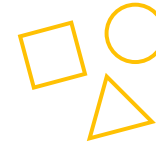
« Capacité de la machine à reproduire l'intelligence humaine » (John McCarthy, 1956)

- > Capacité de la machine à réaliser une activité habituellement associée à un effort cognitif (ex. interprété des instructions en langage naturel).
- > Définition subjective liée aux avancées technologiques :
 - L'IA du siècle dernier n'est pas la même que celle d'aujourd'hui.
 - Lorsqu'elle n'impressionne plus, elle n'est plus considérée comme étant de l'IA (ex. les programmes de jeu d'échecs).
 - L'IA d'aujourd'hui est souvent associée à **l'apprentissage automatique.**

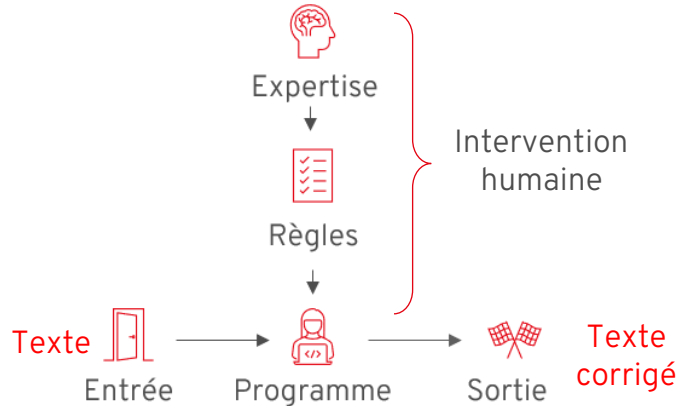
Plusieurs définitions selon les approches ou les disciplines

Apprentissage automatique

Définition

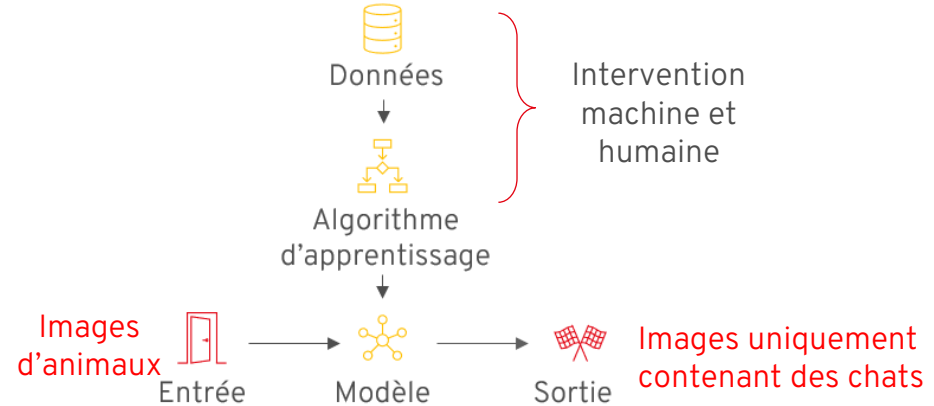


Programmation traditionnelle



Ex. Correction automatique d'un terme (règle = pas écrit de la même manière que dans le dictionnaire). Programmation manuelle des règles possible.

Apprentissage automatique

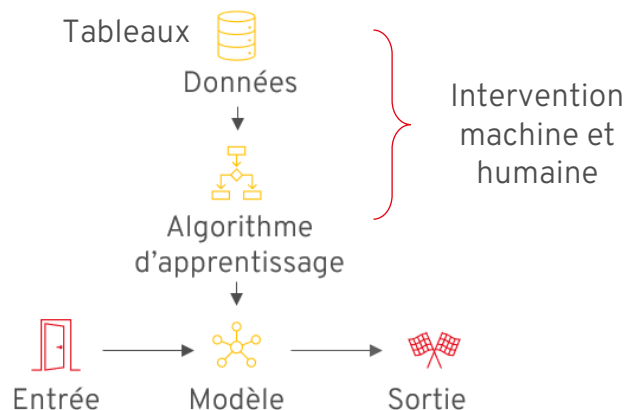


Ex. Identifier l'animal sur une image (règles = difficile à définir). La machine infère les règles à partir de données caractérisées fournies en exemple et crée elle-même les règles (modèle).

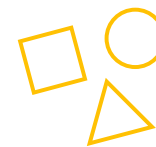
Apprentissage profond

Définition

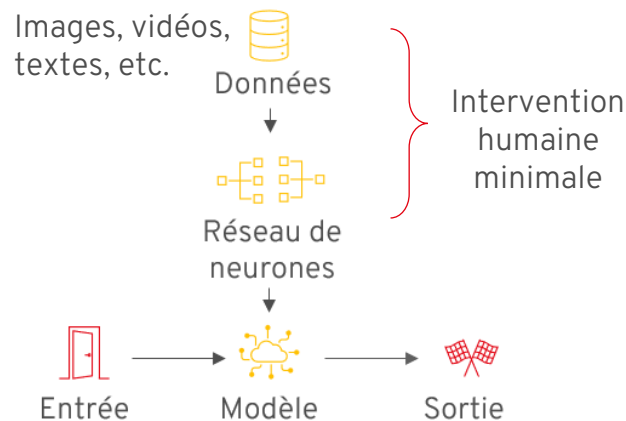
Apprentissage automatique



Ex. Identifier l'animal sur une image (règles = difficile à définir). La machine infère les règles à partir **de données structurées et portant des attributs** (création manuelle d'attributs à partir des données brutes).



Apprentissage profond

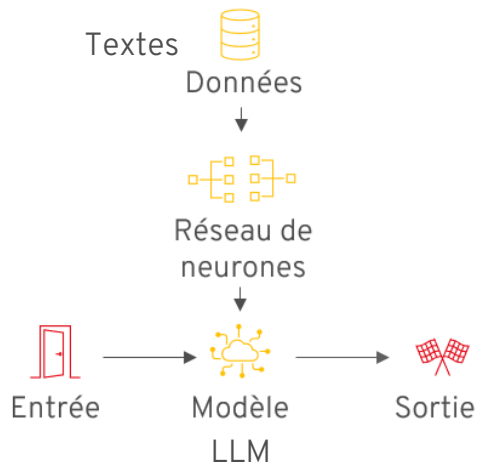


Ex. Identifier l'animal sur une image (règles = difficile à définir). La machine infère les règles à partir **de données brutes et non structurées** (le réseau de neurones apprend directement des données brutes).

Grands modèles de langage

Définition

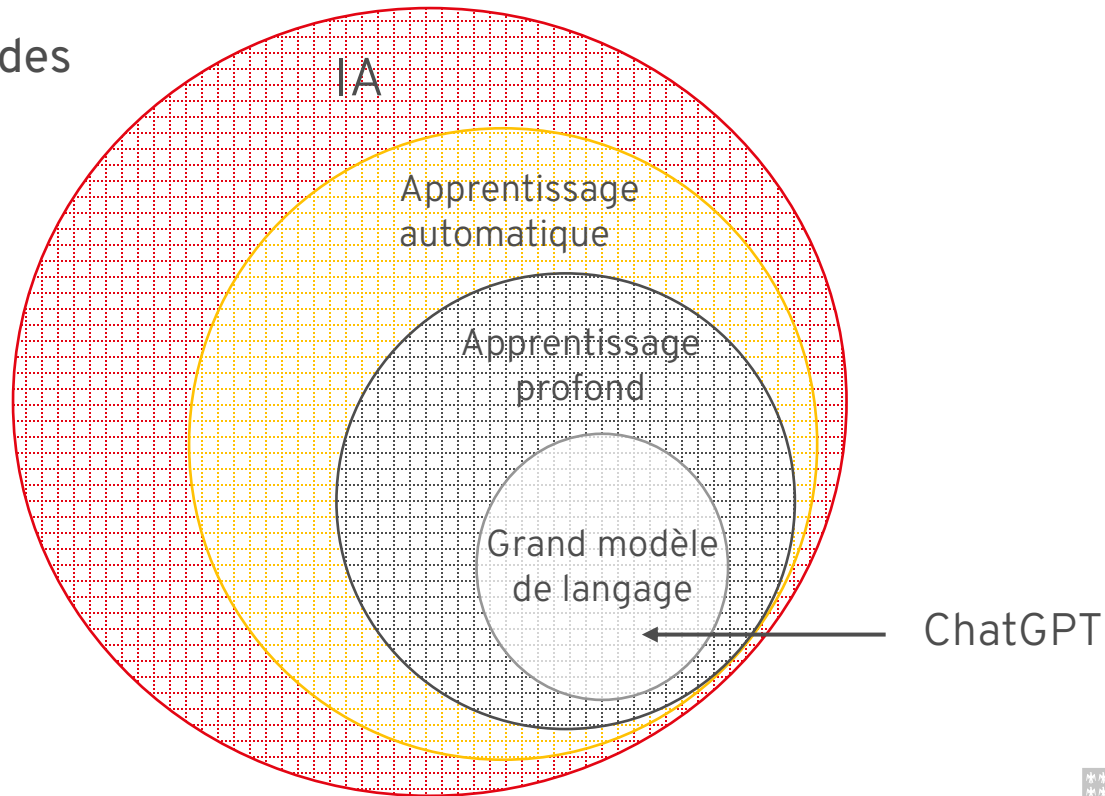
Les grands modèles de langage sont des systèmes d'intelligence artificielle **conçus pour comprendre et générer du texte en langage naturel**. Ils sont entraînés sur de vastes quantités de textes provenant de livres, d'articles, de sites web, et d'autres sources afin d'apprendre les structures, le vocabulaire et le sens des phrases.



Aussi nommé *Large Language Models (LLMs)*

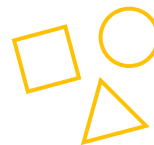
L'ABC de l'IA

Schématisation des
concepts



ChatGPT

Caractéristiques



- > Grand modèle de langage qui intègre des techniques d'apprentissage profond et de traitement du langage.
- > Utilise un grand réseau neuronal pour **prédire le mot** suivant dans une phrase en tenant compte du contexte fourni par les mots précédents.
- > Formule des réponses :
 - Qui **semblent** naturelles et pertinentes;
 - Selon des **probabilités**, donc sans en saisir la signification;
 - Avec **confiance**, même lorsque les **informations sont inexactes**;
 - Marquées par les **biais** présents dans les données d'entraînement.



Utile lors de recherches factuelles ?

L'IA EN RECHERCHE DOCUMENTAIRE



ChatGPT pour la recherche documentaire ?

"Plusieurs études, même conclusion"

> J Pers Med. 2023 Sep 30;13(10):1457. doi: 10.3390/jpm13101457.

Navigating the Landscape of Personalized Medicine: The Relevance of ChatGPT, BingChat, and Bard AI in Nephrology Literature Searches

 Noppawit Aiumtrakul¹, Charat Thongprayoon², Supawadee Suppadungsuk², Pajaree Krisanapan^{2, 4}, Jing Miao², Fawad Qureshi², Wisit Cheungpasitporn²

Affiliations + expand

PMID: 37888068 PMCID: PMC10608326 DOI: 10.3390/jpm13101457

 Download PDF 

> JMIR Med Inform. 2024 May 14;12:e51187. doi: 10.2196/51187.

The Use of Generative AI for Scientific Literature Searches for Systematic Reviews: ChatGPT and Microsoft Bing AI Performance Evaluation



Yong Nam Gwon¹, Jae Heon Kim¹, Hyun Soo Chung², Eun Jee Jung², Joey Chun^{1, 3}, Serin Lee^{1, 4}, Sung Ryul Shim^{5, 6}

Affiliations + expand

PMID: 38771247 PMCID: PMC1107769 DOI: 10.2196/51187

 Download PDF  View Complete Issue

Review > JB JS Open Access. 2024 Sep 5;9(3):e24.00099. doi: 10.2106/JBJS.OA.24.00099.

eCollection 2024 Jul-Sep.

ChatGPT-4 Knows Its A B C D E but Cannot Cite Its Source



Diane Ghanem¹, Alexander R Zhu², Whitney Kagabo¹, Greg Osgood¹, Babar Shafiq¹

Affiliations + expand

PMID: 39238880 PMCID: PMC11368215 DOI: 10.2106/JBJS.OA.24.00099

 Download PDF  View Complete Issue

(1) <https://pubmed.ncbi.nlm.nih.gov/37888068/>

(2) <https://pubmed.ncbi.nlm.nih.gov/39238880/>

(3) <https://pubmed.ncbi.nlm.nih.gov/38771247/>

ChatGPT pour la recherche documentaire ?

"Plusieurs études, même conclusion"

L'utilisation des agents conversationnels comme ChatGPT n'est pas recommandée pour la recherche documentaire.

- Aiumtrakul, N. (2023) : 38% des références générées par ChatGPT sont **valides**.
- Ghanem, D. (2024) : 57% des références générées par ChatGPT ne sont **pas valides**.
- Gwon, Y.N. (2024) : 0,05% des références identifiées à partir de ChatGPT sont **pertinentes**.

Pas très utile pour repérer des références de qualité en sciences de la santé



ChatGPT pour la recherche documentaire ?

Non, mais ...

Les agents conversationnels peuvent être utiles en complément d'une recherche documentaire, notamment pour :

- Générer des idées préliminaires avant de lancer une recherche dans une base de données spécialisée;
- Clarifier ou vulgariser un sujet de recherche;
- Explorer des concepts ou définir du vocabulaire.

Utile à la planification de la recherche



Qu'en est-il des outils d'IA
conçus spécifiquement pour
la recherche ?



Outils de recherche avec IA

De la recherche booléenne à la recherche sémantique

Transforme le processus de recherche documentaire en :

- › Intégrant les grands modèles de langage;
- › Simplifiant considérablement la création de requêtes;
- › **Outrepassant la recherche booléenne traditionnelle pour laisser la place à la recherche sémantique;**
- › Permettant l'analyse des résultats en langage naturel directement dans le processus de recherche;
- › Plus encore ...



On interroge ces outils à partir d'une **question de recherche textuelle** simple ou complexe

Outils de recherche avec IA

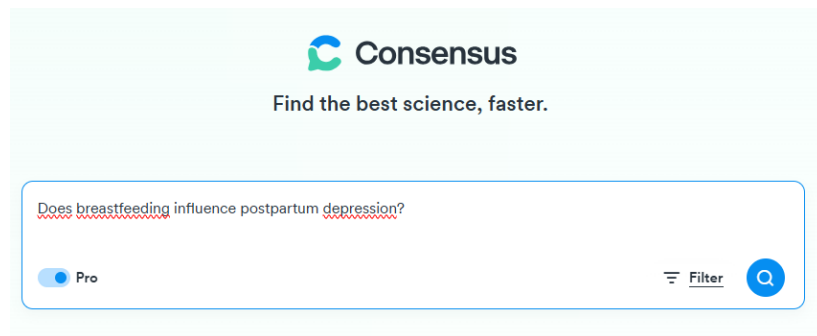
De la recherche booléenne à la recherche sémantique

Quel est l'impact de la **dépression post-partum** sur l'**allaitement** maternel ?

Recherche PubMed

Search	Query	Results
#7	Search: #3 AND #6	638
#6	Search: #4 OR #5	48,474
#5	Search: "Breast Feeding"[Mesh] Sort by: Most Recent	46,447
#4	Search: ("breastfeed*" [Title/Abstract] OR "breast feed*" [Title/Abstract] OR "breastfed*" [Title/Abstract] AND "breast fed*" [Title/Abstract])	5,431
#3	Search: #1 OR #2	11,244
#2	Search: "Depression, Postpartum"[Mesh] Sort by: Most Recent	8,362
#1	Search: ("postpartum depress*" [Title/Abstract] OR "post-partum depress*" [Title/Abstract] OR "post-natal depress*" [Title/Abstract])	7,466

Recherche dans Consensus



Outils de recherche avec IA

Exemples



Find the best science, faster.



✦. Research Assistant

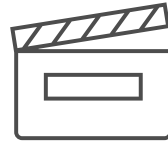
Web of Science™



Spécialisés dans la recherche
de littérature scientifique

Non spécialisé pour la littérature
scientifique, mais utile pour les
recherches sur le Web (littérature
grise)

Elicit Démonstration



Démonstration Elicit



Recent

Library

Help

M Marianne

Get a research report

Start a systematic review **PRO**

Find papers

Ask a research question



Quick tools

Extract data from PDFs

List of concepts

Recent

Elicit

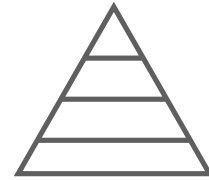
Retour sur les fonctionnalités

- › Recherche sémantique
- › Personnalisation des résultats (filtres et Citations trails)
- › Extraction automatique (à partir des résumés, mais aussi des PDFs)
- › Analyse d'un corpus en mode conversationnel



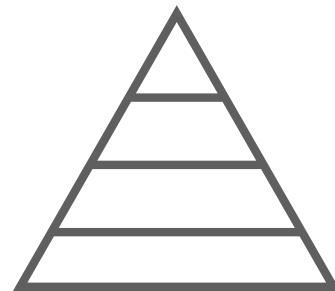
Recherche documentaire,
tri des résultats, **extraction**
des données, analyse et
synthèse des résultats ...
Revue de littérature ?

L'IA et les synthèses de la connaissance



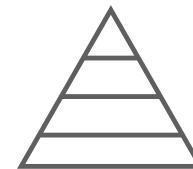
Synthèses de la connaissance

- › Devis d'étude qui **collecte, évalue et intègre** la littérature de manière systématique pour créer une compréhension globale et objective du sujet.
- › Demande une recherche documentaire **exhaustive et systématique**.
- › Influencent la recherche, la santé, les politiques, etc.



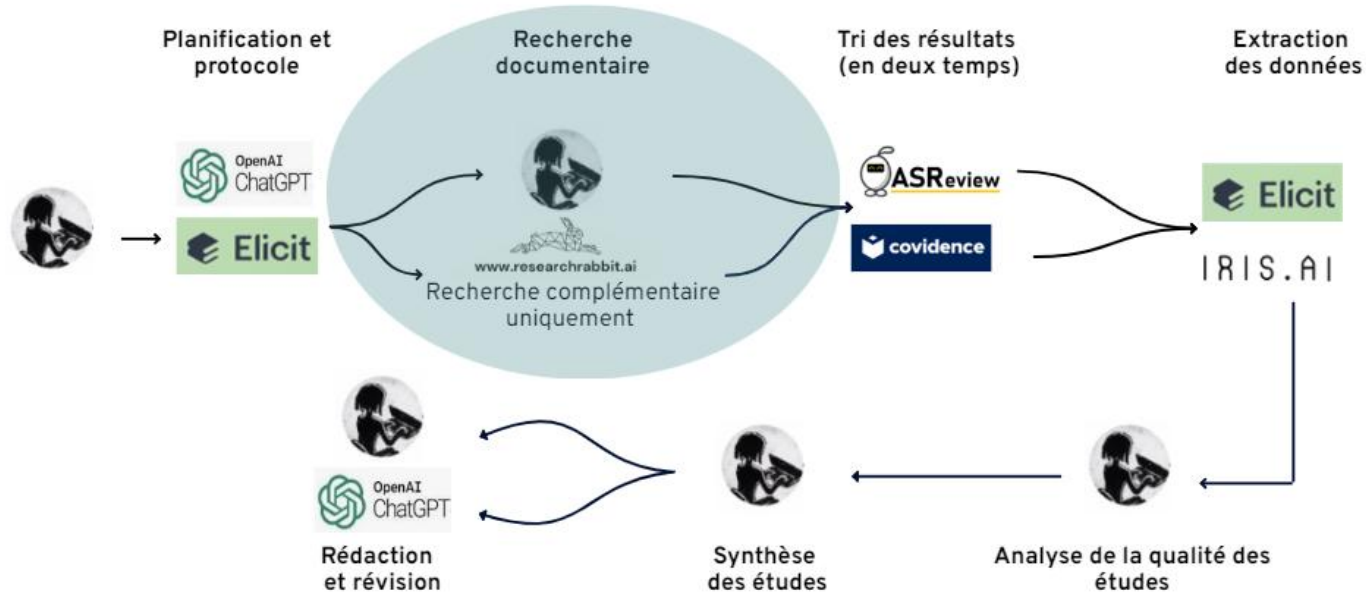
Synthèses de la connaissance

- › L'analyse complète et objective de la littérature garantit que les conclusions tirées sont **fiables, non biaisées** et crédibles.
- › Comme pour tout devis, il faut des données de qualité pour arriver à des résultats de qualité.
- › Les données proviennent de la **recherche documentaire**.



Données de qualité :
une recherche
documentaire **exhaustive,**
transparente et
reproductible.

IA et synthèses de la connaissance



ENJEUX ET LIMITES DE L'IA EN RECHERCHE DOCUMENTAIRE



Limites de l'IA en recherche documentaire

Couverture des outils

La couverture complète de ces outils est souvent inconnue, mais certaines tendances peuvent être observées :

- › Biais en faveur des contenus disponibles en libre accès
- › Références indexées de qualité variable (pas toujours des articles de revues scientifiques)
- › Résultats de qualité et fiabilité variables
- › Meilleure représentation des disciplines empiriques



Aucune certitude que les sources les plus importantes dans la discipline sont couvertes par un outil

Limites de l'IA en recherche documentaire

Résultats inexplicables

La recherche dans ces outils n'est jamais :

› Exhaustive

- Ces outils ne sont pas conçus pour fournir une liste exhaustive de résultats sur un sujet.
- Ces outils ne pourront jamais offrir des résultats exhaustifs en raison de leur couverture limitée.

› Reproducible

- Une même requête retourne des résultats différents d'une fois à l'autre.

› Transparente

- L'utilisation de l'IA rend la recherche opaque.
- Il est impossible d'expliquer pourquoi certains résultats ont été retrouvés et d'autre non.



Pour ces raisons, il n'est
**pas possible d'effectuer
une revue de littérature**
seulement avec ces outils

Enjeux de l'IA en recherche documentaire



Biais - exemples

Les IA sont susceptibles de présenter des biais et peuvent altérer les résultats de la recherche :

› Biais de confirmation

- Favoriser des résultats qui correspondent à des vues ou des hypothèses populaires ou prédominantes : limite la diversité des résultats.

› Biais algorithmique

- Préférences implicites de l'outil en fonction dont le modèle a été entraîné : influence la sélection des résultats.

› Biais de couverture

- Inclusion limitée de sources : limite la portée de la recherche et la pertinence des résultats.

› Biais d'omission

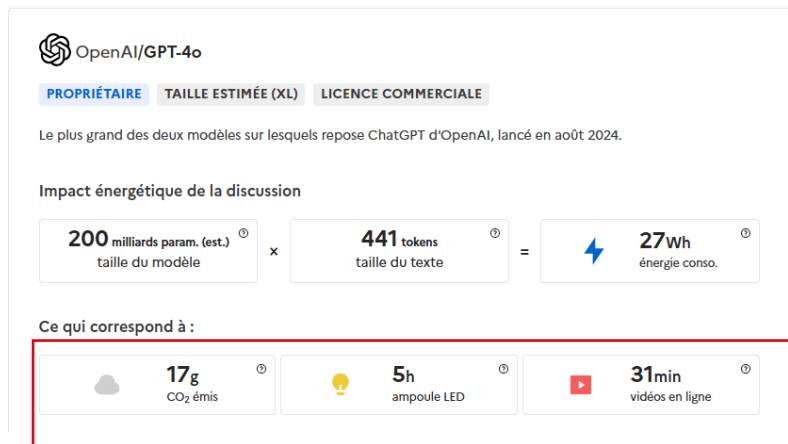
- Limitation d'accès à certains contenus en raison de restrictions techniques, de droit d'auteur : exclusion de perspective importante au besoin.

Enjeux de l'IA en recherche documentaire



Impacts environnementaux

- › Ces outils ont une **consommation énergétique** très importantes.
- › L'utilisation des modèles d'IA, comme ChatGPT, nécessite des centres de données gourmands en énergie.
- › Le traitement d'une seule requête peut émettre plusieurs grammes de CO₂ en raison des besoins techniques de ces installations.



<https://www.comparia.beta.gouv.fr/>

Bonnes pratiques



Bonnes pratiques

Esprit critique

Utiliser ces outils avec **nuance** et **esprit critique** :

- › Bien comprendre leurs **points forts**
 - Recherche rapide et efficace retournant des résultats ciblés.
- › Bien comprendre leurs **points faibles**
 - Recherche non exhaustive, transparente, reproductible.
- › Bien évaluer la **qualité et la fiabilité des sources** repérées
 - Évaluer la crédibilité des résultats avant de les utiliser.



Bien comprendre
son besoin de
recherche pour
utiliser ces outils de
la bonne manière

Bonnes pratiques

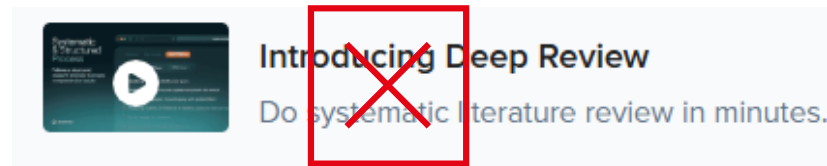
Utilisation optimale

Ces outils sont particulièrement utiles lorsque le besoin d'information est sommaire :

- › Recherche **exploratoire**
- › Recherche **succincte**
- › Recherche **complémentaire**



Attention aux fonctionnalités trompeuses !



Bonnes pratiques

Sobriété numérique

La sobriété numérique est une démarche qui vise à **réduire l'impact environnemental** du numérique par une **approche responsable de l'utilisation des outils numériques** pour contribuer à la préservation de l'environnement et à une utilisation durable des ressources.

- › **Prendre conscience de l'impact environnemental**
 - L'IA et le numérique consomment énormément d'énergie.
- › **Privilégier des solutions alternatives lorsque possible**
 - Un outil plus simple peut-il répondre au besoin ?
- › **Remettre en question la nécessité d'utiliser l'AI**
 - L'outil est-il indispensable pour la réalisation de cette tâche ?

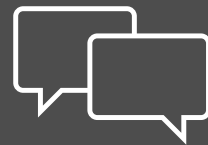




Perspectives pour l'avenir

- › Intégration de l'IA dans les bases de données de qualité en sciences de la santé assurant des résultats de meilleure qualité.
- › Développement de solutions adéquates à l'automatisation des revues de littérature.
- › Intégration de l'IA dans les outils cliniques d'aide à la prise de décision.
- › Plus encore ...

DISCUSSIONS



Outils de recherche avec IA mentionnés en exemples

- Consensus : <https://consensus.app/search/>
- Elicit : <https://elicit.com/>
- Perplexity : <https://www.perplexity.ai/>
- Scispace : <https://scispace.com/>
- Scite AI : <https://scite.ai/>
- Web of Science Research Assistant : <https://www-webofscience-com.acces.bibl.ulaval.ca/wos/research-assistant>



MERCI !

Marianne.ruel@bibl.ulaval.ca

Mars 2025

OpenAI. (2025). Image générée par DALL-E

Bibliothèque



UNIVERSITÉ
LAVAL