

# Recherche d'informations sur Internet (perfectionnement) méthodologie et outils disponibles

2024



## Recherche d'informations sur internet (perfectionnement)

A. Bouchard (URFIST de Paris)  
MAJ 06/2024

 Rappels méthodologiques

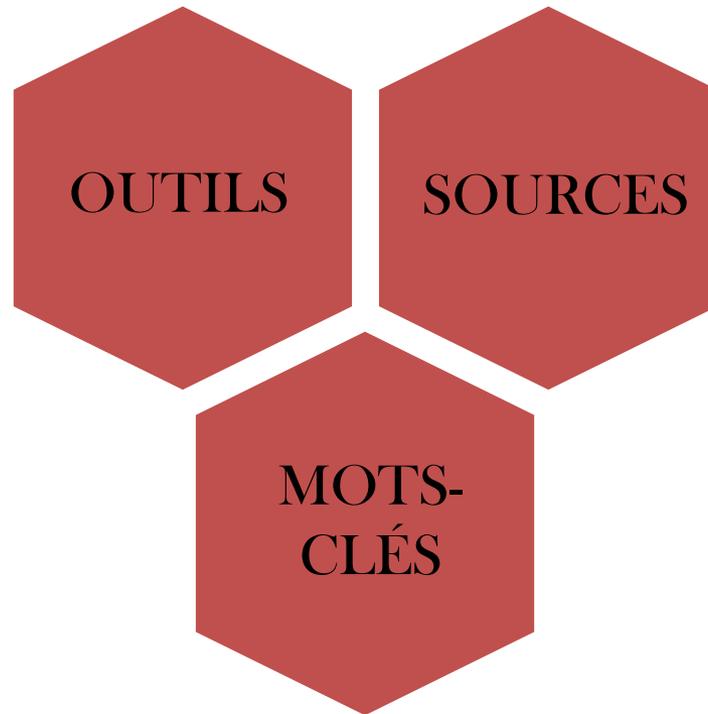
Je cherche des pages web, des informations ponctuelles,  
des personnes, des types de ressources particuliers...

Je cherche des publications scientifiques

<https://framindmap.org/c/maps/446941/public>

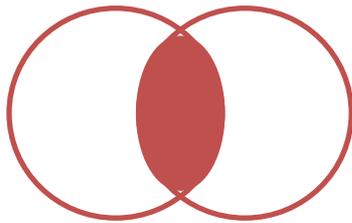
# Méthodologie de recherche

- contexte



# Méthodologie de recherche

- opérateurs booléens

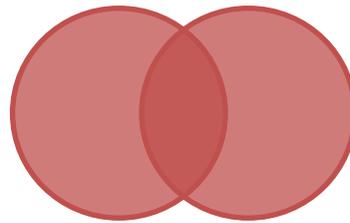


AND  
ET

= réunion

le document doit comprendre  
tous les termes

[parc ET jardin]

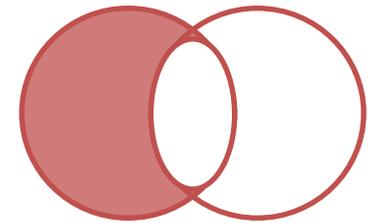


OR  
OU

= union

le document peut comprendre  
tous les termes  
ou seulement certains d'entre eux

[parc OR jardin]



NOT / -  
SAUF

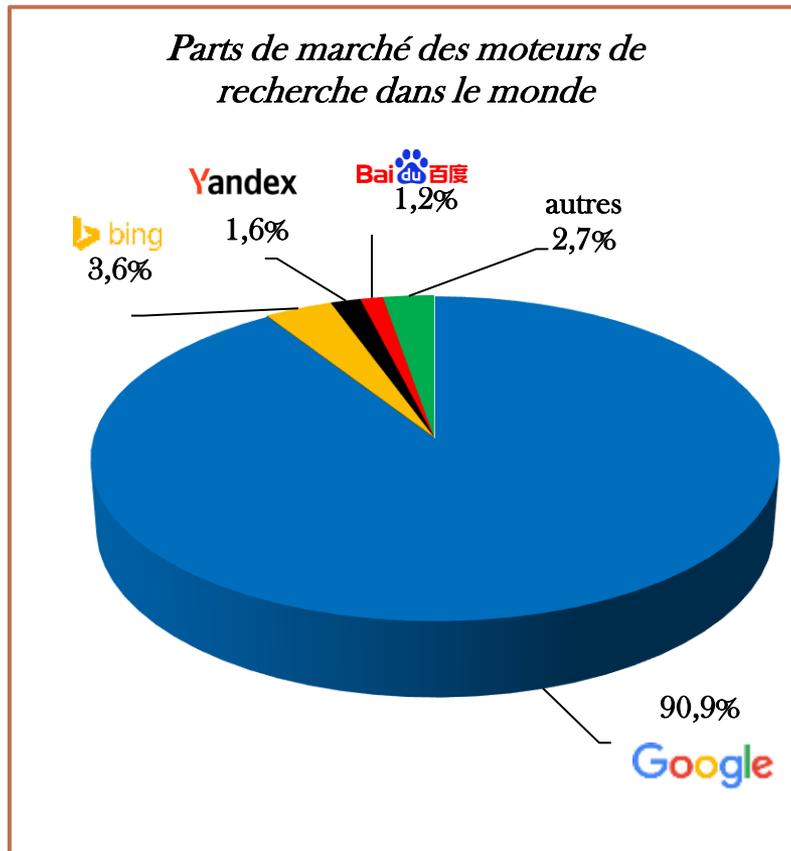
= exclusion

le document ne doit pas  
comprendre certains termes

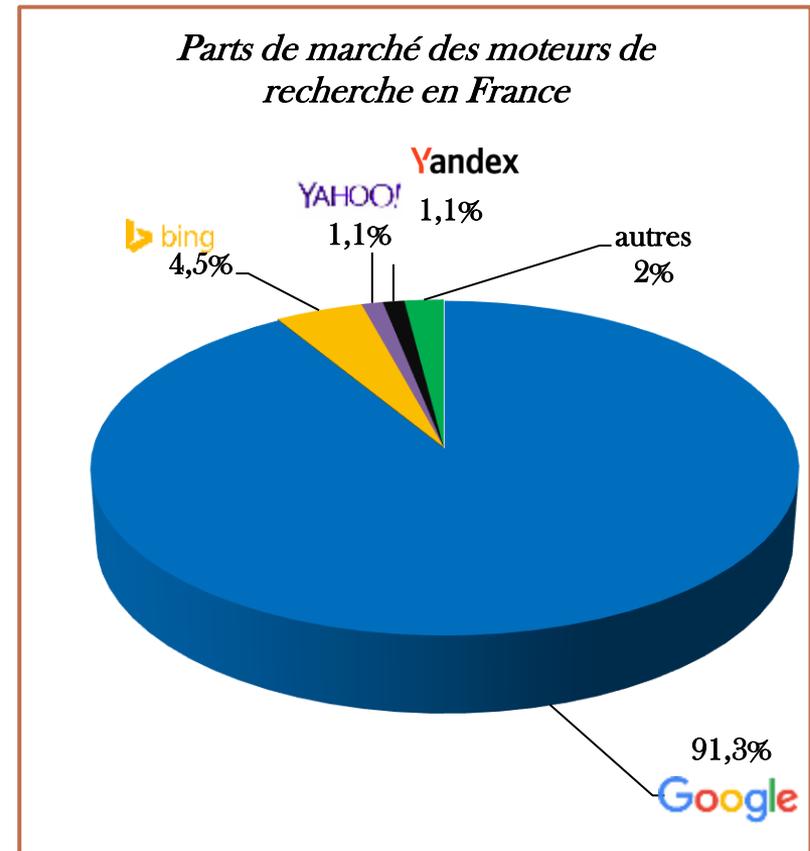
[parc NOT jardin]  
[parc -jardin]

# Moteurs de recherche

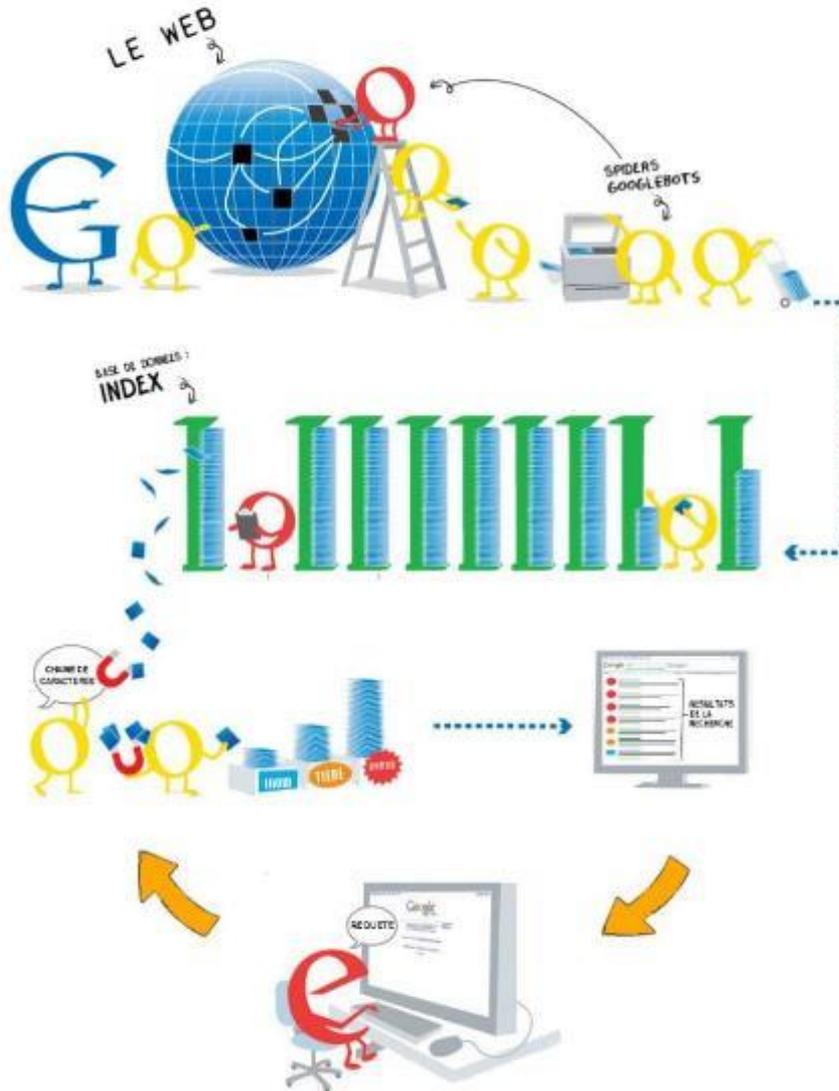
## Monde



## France



# COMMENT FONCTIONNE LA RECHERCHE GOOGLE



## ROBOT D'EXPLORATION

Des logiciels d'exploration, appelés "crawlers" ou "spiders", explorent régulièrement le web en parcourant les réseaux de liens. Ils scannent les pages qu'ils trouvent.

## ROBOT D'INDEXATION

Le robot d'indexation extrait les données des pages scannées (URL, titre, mots-clés sous forme de chaînes de caractères), puis les classe pour former l'index de la base de données de Google

## MOTEUR D'INTERROGATION

Interface entre l'internaute et l'index, le moteur d'interrogation récupère la requête tapée dans le formulaire, la compare aux données de l'index, puis recueille les résultats qu'il classe avant de les afficher. (voir l'algorithme de recherche Google)

- Google en 2024
  - une baisse de qualité ? (oui, mais pas le seul)
  - la disparition de certaines fonctionnalités (cache)
  - le développement de l'intelligence artificielle générative dans le *search* (cf. Google I/O 2024)

## Quizz

Comment limiter la personnalisation sur Google ?

ne pas être connecté

gérer son navigateur

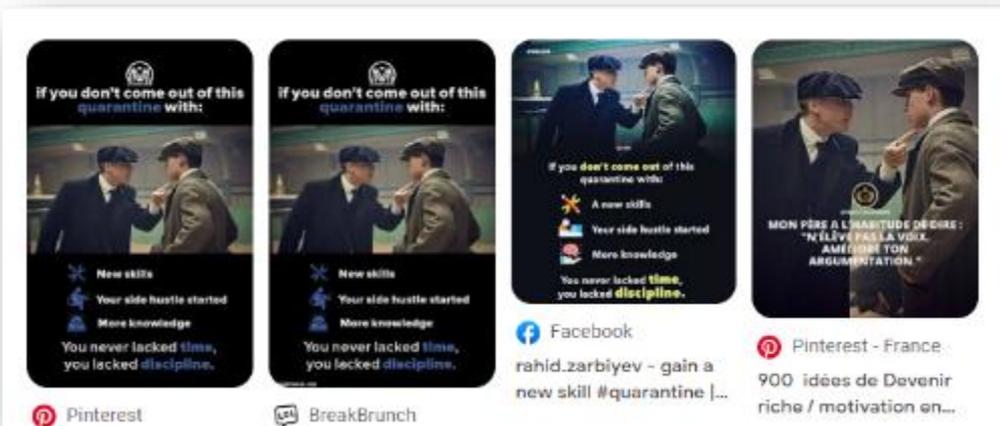
changer de moteur

- les moteurs de recherche en 2024
  - des alternatives aux *Big tech* ?
  - modèle économique et données personnelles ?
  - des outils favorisant les biais à la désinformation (étude)
  - le développement des réseaux sociaux (étude) et de l'IA
  - des moteurs de recherche aux assistants personnels
  - la nécessaire diversification des outils

# Focus : les images



*recherche inversée*



# Focus : les images

CREATIVE COMMONS LICENSES		 COPY & PUBLISH	 ATTRIBUTION REQUIRED	 COMMERCIAL USE	 MODIFY & ADAPT	 CHANGE LICENSE
	PUBLIC DOMAIN	✓	✗	✓	✓	✓
	CC BY	✓	✓	✓	✓	✓
	CC BY-SA	✓	✓	✓	✓	✗
	CC BY-ND	✓	✓	✓	✗	✗
	CC BY-NC	✓	✓	✗	✓	✓
	CC BY-NC-SA	✓	✓	✗	✓	✗
	CC BY-NC-ND	✓	✓	✗	✗	✗

				
You can redistribute (copy, publish, display, communicate, etc.)	You have to attribute the original work	You can use the work commercially	You can modify and adapt the original work	You can choose license type for your adaptations of the work.

Licences  
Creative Commons

BY : Attribution ;  
NC : Noncommercial ;  
ND : NoDerivs ;  
SA : Share Alike

An iceberg floating in the ocean. The tip of the iceberg is above the water surface, and the much larger, jagged base is submerged underwater. The sky is blue with some clouds, and the water is a deep blue.

**Web  
visible**

**Web  
invisible  
(ou profond)**

**“Dark web”**

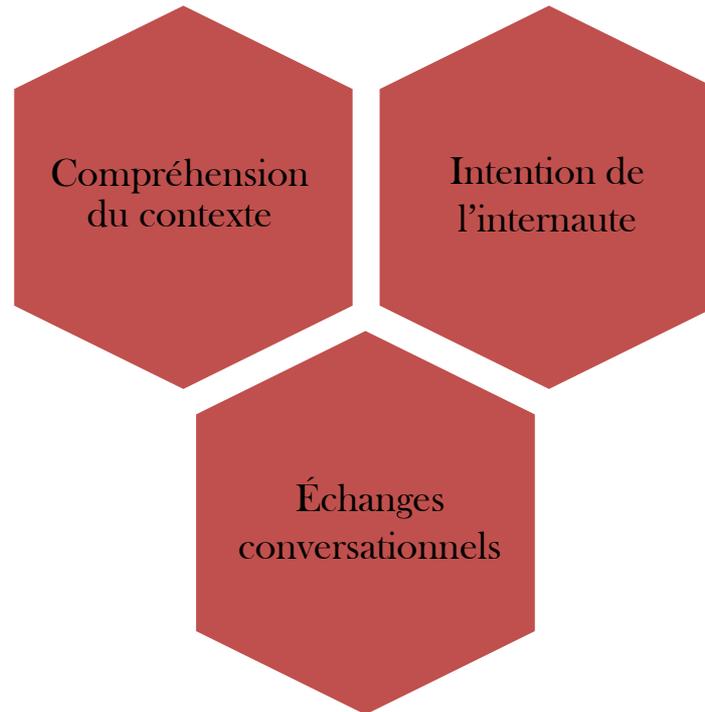
# Les réseaux sociaux

moteurs de recherche « classiques »	réseaux sociaux
<p data-bbox="426 615 745 651">indexation du web</p> <p data-bbox="423 793 749 829">filtre algorithmique</p> <p data-bbox="394 948 774 983">popularité (<i>PageRank</i>)</p> <p data-bbox="365 1100 807 1136">confiance dans les sources</p>	<p data-bbox="1016 572 1667 708">interrogation des contenus spécifiques d'une site en temps réel (fraîcheur et tendances)</p> <p data-bbox="1230 793 1450 829">filtre humain</p> <p data-bbox="1107 948 1574 983">sérendipité (signaux faibles)</p> <p data-bbox="1095 1100 1586 1136">confiance dans les personnes</p>

- recherche d'information en 2024
  - des informations complémentaires des moteurs de recherche
  - le maintien de X-Twitter et le développement de LinkedIn ?
  - des services de plus en plus fermés (comptes)
  - le développement d'un usage des réseaux sociaux comme moteurs de recherche (ex. : TikTok et Snapchat)
  - des outils de moins en moins adaptés à la veille ?

<b>moteurs de recherche « classiques »</b>	<b>assistants conversationnels de type ChatGPT</b>
<p>indexation du web régulière</p> <p>approche par mots-clés (et de plus en plus langage naturel)</p> <p>requête (avec booléens, filtres et tris)</p> <p>liste de résultats (et de plus en plus réponses), par filtres algorithmiques</p> <p>pas de conservation de l'historique d'une requête à l'autre (itérations)</p> <p>→ moteurs de recherche</p>	<p>base de données d'entraînement, sans toujours de connexion à internet</p> <p>approche par énoncé (langage naturel)</p> <p>« prompt » plus ou moins précis</p> <p>réponse (parfois sans source), par traitement sémantique statistique</p> <p>recherche conversationnelle (échanges)</p> <p>→ moteurs de réponses</p>

- contexte



→ pertinence des résultats

- des moteurs de réponse

Google hauteur de la tour de pise

Images Maps Calcul Vidéos Actualités Livres Vols Finance

Environ 1000 résultats (0,51 secondes)

**réponse directe**

**57 m**

La **Tour de Pise** s'est affaissée dès le début de sa construction : le sol sur lequel elle fut bâtie était trop meuble. Symbole de la puissance de la république maritime de Pise au Moyen Âge, le campanile médiéval haut de 55,86 m côté sud et 56,71 m côté nord s'est mis à pencher dès le début de sa construction en 1173.

National Geographic  
<https://www.nationalgeographic.fr/tourisme/la-tour-d...>

La Tour de Pise est en train de se redresser

Recherches associées

- Cathédrale Santa Maria del Fiore 114 m
- Cattedrale di Pisa 48 m
- Campanile di Giotto 85 m

**suggestions**

Autres questions

- Qu'est-ce qu'il y a à l'intérieur de la tour de Pise ?
- Pourquoi la tour de Pise s'appelle la tour de Pise ?
- Est-ce que la tour de Pise a toujours été penché ?
- Qui a eu l'idée de construire la tour de Pise ?

**résultats naturels**

Wikipédia  
[https://fr.wikipedia.org/wiki/Tour\\_de\\_Pise](https://fr.wikipedia.org/wiki/Tour_de_Pise)

**Tour de Pise**

↑ Sa situation géographique est de 43° 43' 23" N en latitude, 10° 23' 47" E en longitude et à environ 2 mètres au-dessus du niveau de la mer. - ↑ Site de ...

**Knowledge graph**

Tour de Pise

Site Web Itinéraire Enregistrer

4,7 ★★★★★ 141 408 avis Google

Tour à Pise, Italie

VOIR LES BILLETS

La tour de Pise est le campanile de la cathédrale Notre-Dame de l'Assomption de Pise, en Toscane. Elle est située à proximité du chevet de la cathédrale et fait partie des monuments de la piazza dei Miracoli, inscrit au patrimoine mondial de l'UNESCO. Sa construction débute en 1173. Wikipédia

Emplacement : Piazza dei Miracoli

Adresse : Piazza del Duomo, 56126 Pisa PI, Italie

Horaires : Ouvert - Ferme à 20.00 -

Téléphone : +39 050 635011

Hauteur : 57 m

Début de la construction : 9 août 1173

Architecte : Bonanno Pisano

Province : Pise

Étages : 8

Date d'ouverture : 1372

Suggérez une modification

Entrée

À propos de ces résultats

Tour de Pise Site d'entrée 20,00 €

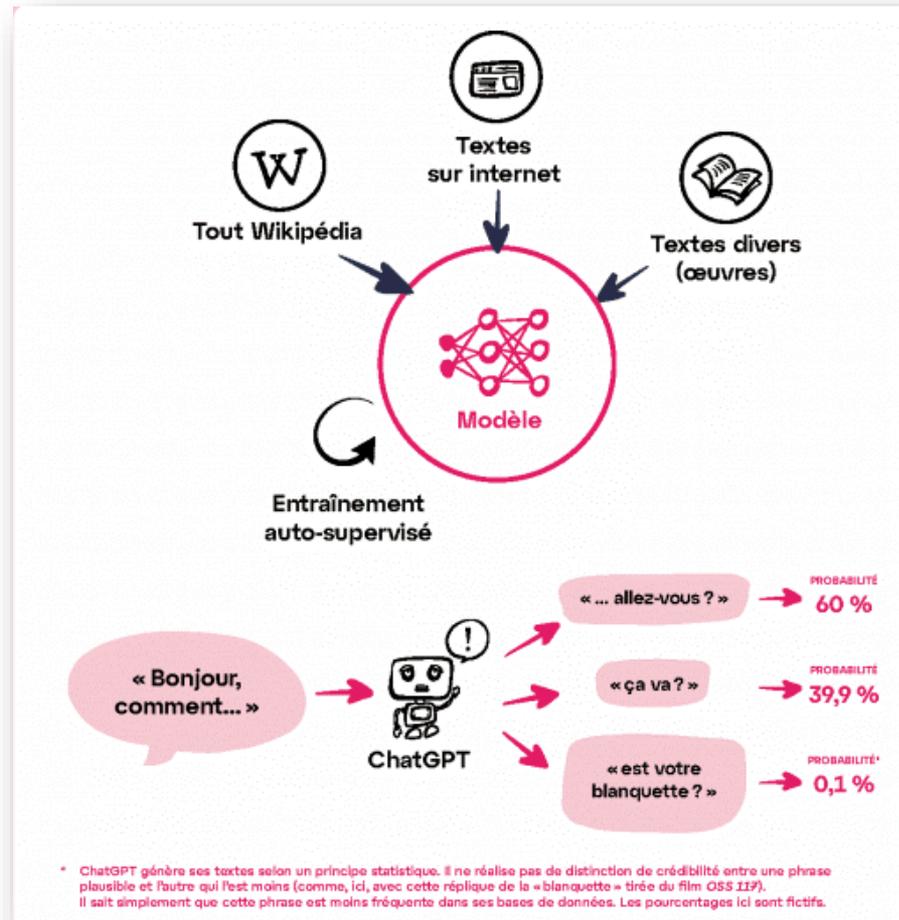
- les IA génératives

données

algorithmes

entraînement

prompt

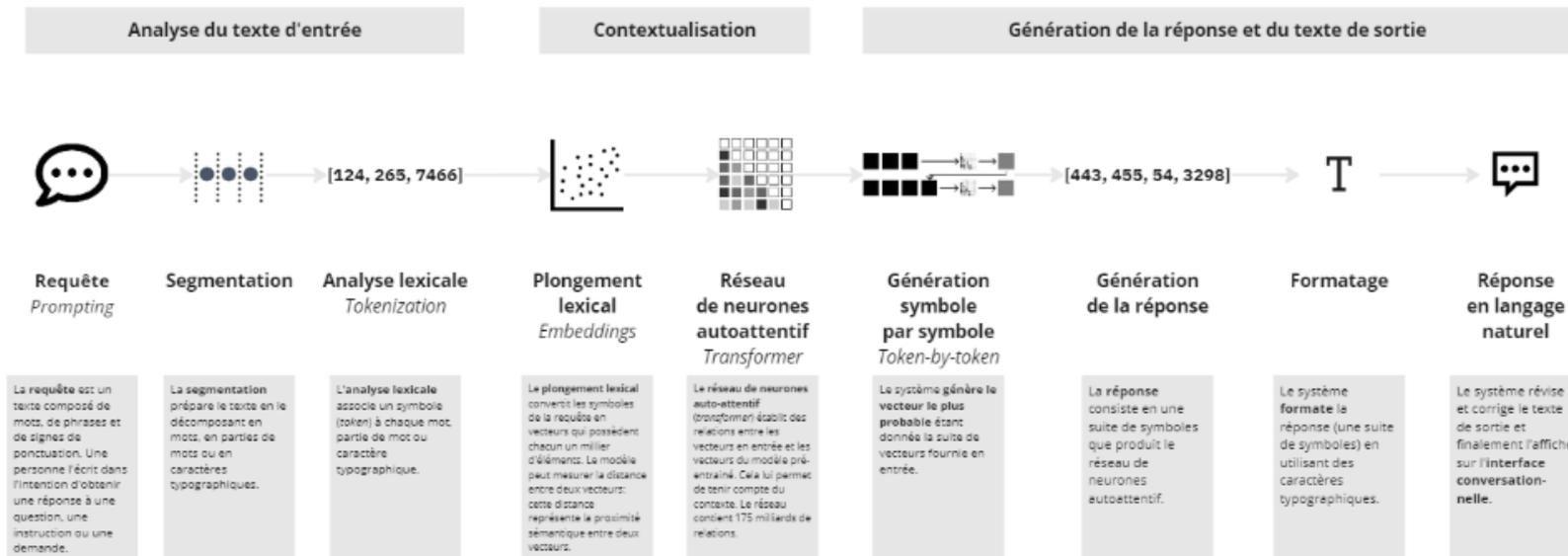


contenu inédit

- IA génératives et LLM (*large languages models*)



## Comment fonctionne ChatGPT de la requête à la réponse



**Requête Prompting**  
La requête est un texte composé de mots, de phrases et de signes de ponctuation. Une personne l'écrit dans l'intention d'obtenir une réponse à une question, une instruction ou une demande.

**Segmentation**  
La segmentation prépare le texte en le décomposant en mots, en parties de mots ou en caractères typographiques.

**Analyse lexicale Tokenization**  
L'analyse lexicale associe un symbole (token) à chaque mot, partie de mot ou caractère typographique.

**Plongement lexical Embeddings**  
Le plongement lexical convertit les symboles de la requête en vecteurs qui possèdent chacun un millier d'éléments. Le modèle peut mesurer la distance entre deux vecteurs: cette distance représente la proximité sémantique entre deux vecteurs.

**Réseau de neurones auto-attentionnel Transformer**  
Le réseau de neurones auto-attentionnel (transformer) établit des relations entre les vecteurs en entrée et les vecteurs du modèle pré-entraîné. Cela lui permet de tenir compte du contexte. Le réseau contient 175 milliards de relations.

**Génération symbole par symbole Token-by-token**  
Le système génère le vecteur le plus probable étant donné la suite de vecteurs fournie en entrée.

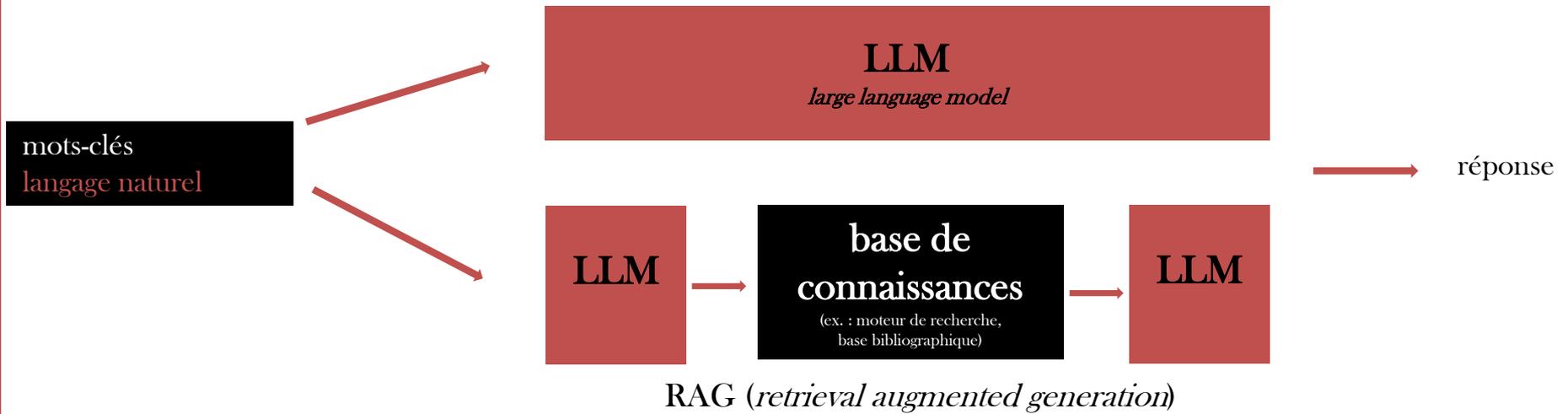
**Génération de la réponse**  
La réponse consiste en une suite de symboles que produit le réseau de neurones auto-attentionnel.

**Formatage**  
Le système formate la réponse (une suite de symboles) en utilisant des caractères typographiques.

**Réponse en langage naturel**  
Le système révisé et corrige le texte de sortie et finalement l'affiche sur l'interface conversationnelle.

- le prompt
  - rôle
  - contexte et objectifs
  - requête et instructions
  - format de sortie
  - [exemples]

- limites des IA génératives généralistes
  - des données non scientifiques
  - un manque de profondeur
  - une absence de sources
  - des hallucinations
  - des biais cognitifs possibles



what is open access?

du Google SGE (Search Generative Experience) au AI overviews

Generative AI is experimental. Info quality may vary. Learn more

Open access (OA) is a set of principles and practices that make research outputs available online for free, without access charges or other barriers. OA content can include any kind of digital content, from texts and data to software, audio, video, and multi-media.

OA scholarly literature is free of charge and often has less restrictive copyright and licensing barriers than traditionally published works. OA content is open to all, with no access fees.

There are several types of open access:

- Diamond: Open access journals that do not require a fee to publish
- Bronze: Journals where the license or copyright is not clear

You can check if an article is open access by looking for an OA logo or OA statement on the publisher website. You can also add Unpaywall, Open Access Button, or Kopernio plugins to your Chrome or FireFox browsers.

How to determine Open Acces...  
24 May 2019

The fundamentals of open...  
30 Oct 2023

Types of Open Access - Levy Library Guides

Ask a follow up

What is the difference between free access and open access?

What is the benefit of open access?

- évolution de la requête ?
  - et de plus en plus la recherche interrogative (que..., qui..., quand... ?)

hauteur de la tour de pise

Tout Images Actualité Vidéos Goggles

## Hauteur de la tour de pise

La hauteur de la tour de Pise est de **58,36 mètres**, selon les informations fournies par la page Wikipédia. Cette hauteur est mentionnée dans le tableau de la page Wikipédia consacrée à la Tour de Pise.

Il est également mentionné que la tour a une inclinaison de 5,1 degrés et un diamètre extérieur de 14,95 mètres et intérieur de 9,52 mètres.

Il est important de noter que la hauteur de la tour de Pise est également mentionnée dans d'autres sources, notamment dans l'article "Combien mesure la tour de Pise ?" qui indique que la tour s'élève à une hauteur de 57 mètres. Cependant, la hauteur de 58,36 mètres mentionnée sur la page Wikipédia est considérée comme la hauteur officielle de la tour.

ⓘ Réponse générée par IA. Pensez à vérifier les informations importantes. En savoir plus

### Contexte

- combien.io Combien mesure la tour d...
- www.futura-sciences... Tour de Pise : quand a-t-él...
- www.leaningtowe... Tour de Pise : Quelle est la...
- + 4 more

Qwant+ hauteur de la tour de pise

All News Images Videos Shopping Filters

## Hauteur de la tour de Pise

- Tour de Pise — Wikipédia wikipedia
- Tour de Pise : Quelle est la hauteur de la... leaningtowerofp...
- La Tour de Pise pisa-tour

- La hauteur moyenne de la tour de Pise est de 57 mètres.
- Actuellement, la tour mesure 56,67 mètres de haut à son point le plus haut et 55,86 mètres de haut à son point le plus bas.
- L'inclinaison actuelle de la tour de Pise est de 3,99 degrés.



Summary by Qwant's artificial intelligence

fr.wikipedia.org  
fr.wikipedia.org › wiki › Tour\_de\_Pise

### Tour de Pise — Wikipédia

# IA et search

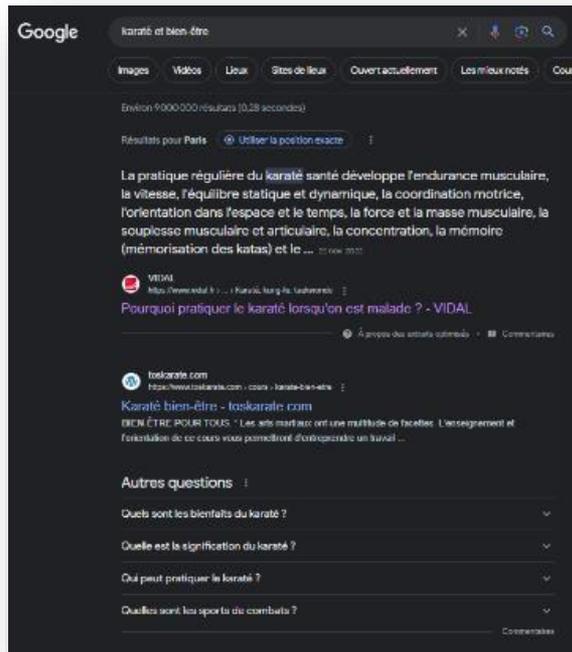
Google search for "arthur rimbaud paris photo". The search results include a snippet from "20 Minutes" dated Jun 16, 2023, with a red circle around the "..." icon. The snippet reads: "Intelligence artificielle : On vous raconte l'histoire de cette ... Jun 16, 2023 — Il ne s'agit pas d'une image inédite ou rare d'Arthur Rimbaud prise par « Ernest Ballhazar, un photographe de rue, à Paris le 1er novembre 1873". Below it is a Wikipedia result for "Arthur Rimbaud par Carjat". An "About this result" panel is open on the right, showing the URL "https://www.20minutes.fr/culture/4041433-20230615-intelligence-artificielle-raconte-histoire-lausse-photo-arthur-..." and source information: "20 minutes Newspaper". A description of 20 minutes as a free daily newspaper is provided.

A screenshot of a 20 Minutes article. The main image shows a young man in a dark coat standing on a cobblestone street. A context menu is open over the image, with a red circle around the three-dot menu icon. The menu options are: "Share", "Send feedback", and "About this image".

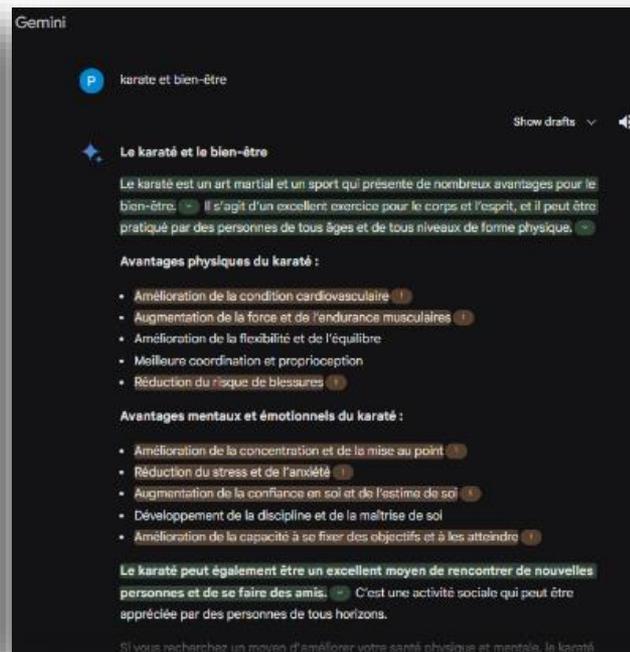
An "About this image" panel from Google. It states: "A version of this image is at least 5 months old". Below this is a small thumbnail of the image. Under "Web results", there is a result from "France's VRAI DU FAUX" dated Sep 5, 2023, with the headline "Une photo 'transmissio' d'Arthur Rimbaud trop belle...". The snippet reads: "Sep 5, 2023 — Une lausse photo de poète Arthur Rimbaud circule sur les réseaux sociaux depuis quelques mois, mais elle n'est vraie que pour la date, et pas pour le poète."

- limites de l'IA dans les moteurs de de recherche
  - des questions sur les sources utilisées et leur sélection
  - des questions sur la fiabilité et l'exhaustivité
  - des erreurs possibles dans la synthèse
  - une nécessaire transparence sur l'utilisation de l'IA et les sources

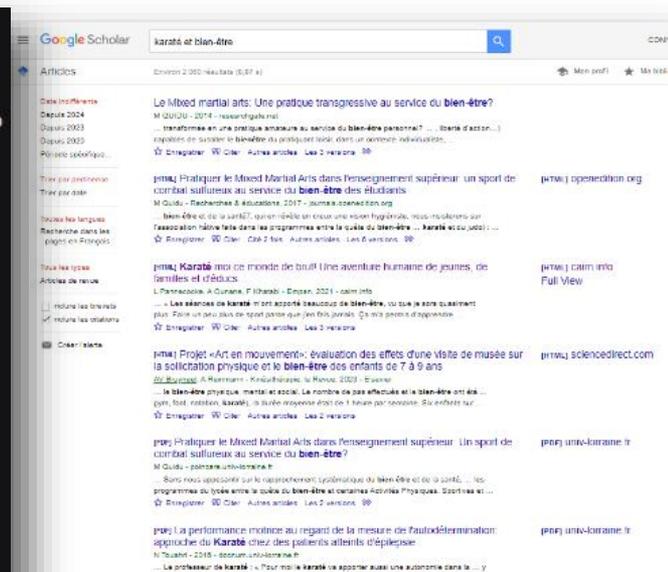
# moteur de recherche



# IA générative



# outil bibliographique



# Moteurs de recherche scientifiques

moteurs de recherche « classiques »	moteurs de recherche scientifiques
<p data-bbox="349 494 823 529">indexation du web régulière</p> <p data-bbox="272 646 900 736">fonctionnalités de recherche et filtres génériques</p> <p data-bbox="253 1053 919 1143">des acteurs généralement commerciaux (Google, Microsoft...)</p>	<p data-bbox="1025 468 1653 608">indexation de corpus spécifiquement académiques (références bibliographiques, texte intégral, etc.)</p> <p data-bbox="987 672 1692 965">fonctionnalités et filtres spécifiquement liés aux publications (auteur, titre...) ou aux types de documents scientifiques (<i>preprints</i>, brevets...), avec des nombreuses fonctionnalités académiques (mise en forme des citations, métriques...)</p> <p data-bbox="1045 1029 1634 1165">des acteurs plus divers (Google, fournisseurs de service, services de documentation)</p>

# Moteurs de recherche scientifiques

Quels points d'entrée privilégiez-vous pour accéder au texte intégral d'un article ?

<b>Google Scholar</b>		
Google Scholar	3184	5598
<b>Plateforme des éditeurs de revues</b>		
Springer Link ScienceDirect	2608	5598
<b>Archives ouvertes</b>		
HAL arXiv	2400	5598
<b>Réseaux sociaux</b>		
ACADEMIA ResearchGate	2204	5598
<b>Sites pirates</b>		
SCI-HUB Library Genesis™	1989	5598
<b>Moteurs de recherche généralistes</b>		
Google Microsoft Bing	1584	5598
<b>Web of Science</b>		
Clarivate Web of Science™	1482	5598
<b>Plateformes / agrégateurs de revues</b>		
CAIRN.INFO MATIÈRES À RÉVÉLER OpenEdition	1387	5598
<b>Autres bases bibliographiques</b>		
Dimensions NIH	1182	5598
<b>Scopus</b>		
Scopus	870	5598

## The Uptake of Open Science: Mapping the Results of a Systematic Literature Review

Hanna Shmagun<sup>1</sup>, Charles Oppenheim<sup>2</sup>, Jangsup Shim<sup>3</sup>, and Jaesoo Kim<sup>4\*</sup>

<sup>1</sup>the Korea University of Science and Technology (UST), the Korea Institute of Science and Technology Information (KISTI), Daejeon, South Korea

<sup>2</sup>the Robert Gordon University, Aberdeen, the United Kingdom

<sup>3</sup>the Woosong University, Daejeon, South Korea

<sup>4</sup>the Korea Institute of Science and Technology Information (KISTI), Daejeon, South Korea

**Abstract.** This paper contributes to “Open Science” theory, with a specific focus on Open Science data generated by scholars. To this end, a mixed-method systematic literature review, including science mapping techniques, was conducted. Our preliminary results reveal the potential of Open Science as a domain for interdisciplinary research. A keyword co-occurrence network analysis using the VOSviewer visualisation tool identified five clusters of interrelated sub-concepts within Open Science research. The key distinctive characteristics and the various categories of Open Science data have been identified. The relevant data platforms have been provided to exemplify each category of Open Science data. Finally, a distinction between Open Science data and Open Government data was explored and the convergence point between them was presented.

### 1 Introduction

The aim of this study is to propose some initial conceptual foundations of “Open Science” (OS) field of scholarship, with a particular focus on OS data. The relatively new phenomenon called “Open Science” has emerged due to the fundamental, revolutionary change government and science policies triggered by the development of Information Communication Technologies (ICT). ICT advancements along with authorities’ commitment allow both government and science to become transparent, participatory, and collaborative.

There are three ICT factors, which make up the phenomenon of OS. First, data sharing technologies are expanding the boundary of scientific knowledge dissemination to the public far beyond academic communities. Second, online collaboration tools, including those designed to engage with general public, contribute to the networked science, which turn speeds up scientific discoveries and knowledge creation. Finally, in the era of Big and Linked Open Data, data-driven intelligence enhances our ability to extract knowledge and create value under condition that the access to the world’s scientific sources is widely available and the data are linked.

\* Corresponding author: [jaesoo@kisti.re.kr](mailto:jaesoo@kisti.re.kr)

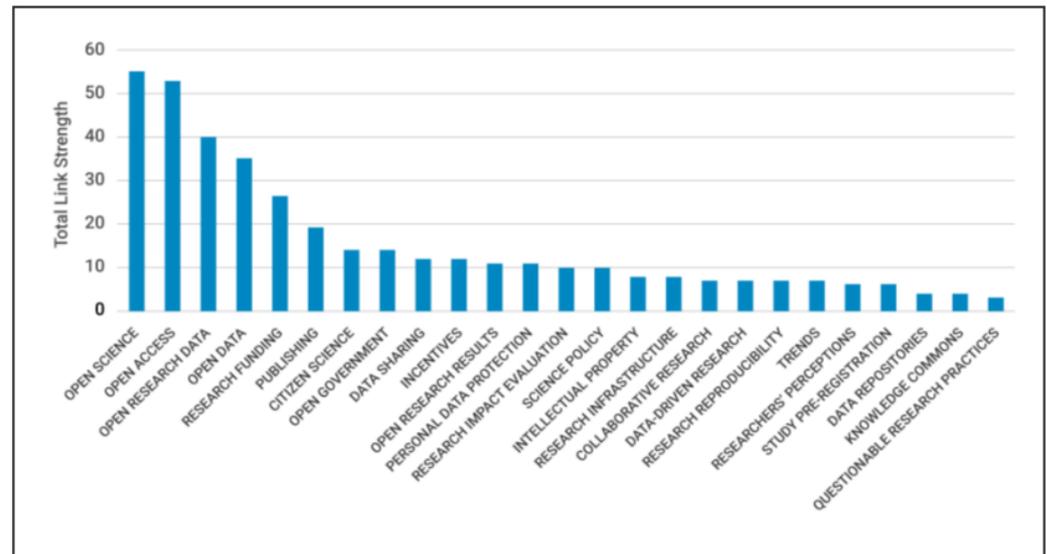


Fig. 3. The most important keywords (concepts) related to OS

## Ouverture des métadonnées académiques

- *open access*
- *open citations*
- *open abstracts*



# Moteurs de recherche scientifiques

recherche par mots-clés

mots-clés, booléens

recherche par co-citations

références et  
citations d'une  
publication

recherche sémantique

langage naturel

# Moteurs de recherche scientifiques

Text	Number of works	Open Access works	Citations	Price	Data Openness	Org structure
<a href="#">OpenAlex</a>	248M	52M	1.9B	Freemium	Fully open, CC0 license	Non-profit
<a href="#">Scopus</a>	90M	20.5M (ref)	1.8B	Subscription	Closed	For Profit
<a href="#">Web of Science (core)</a>	89M (ref)	24M (ref)	1.8B	Subscription	Closed	For Profit
<a href="#">Dimensions</a>	140M+	29M (ref)	1.7B	Freemium	Partly open, personal use	For Profit
<a href="#">Google Scholar</a>	389M (estimated)	?	?	Free	Closed	For Profit
<a href="#">Crossref</a>	145M	20M	1.45B	Free	Fully open, CC0 license	Non-profit

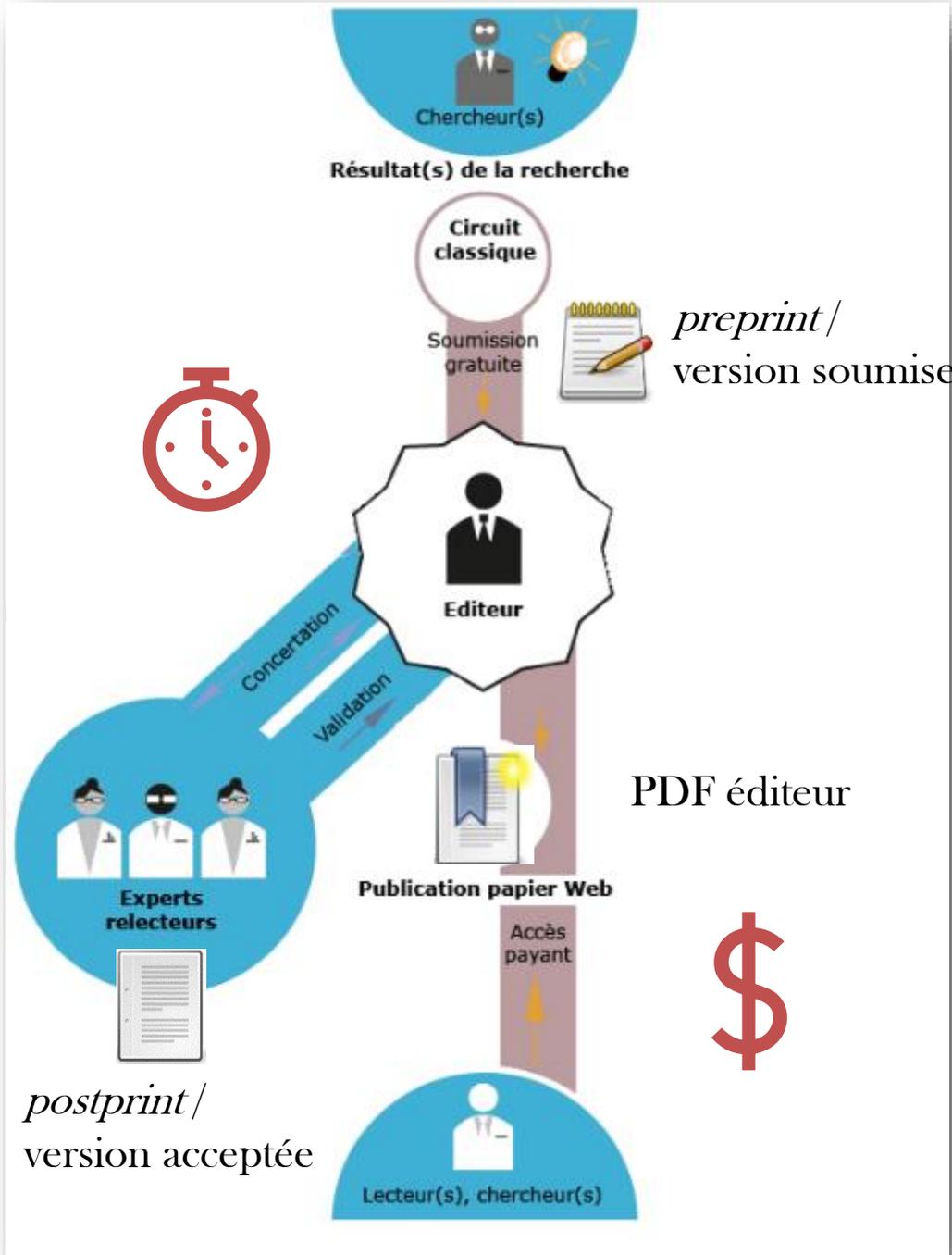
<https://help.openalex.org/coverage>

# Moteurs de recherche scientifiques

- les moteurs de recherche scientifiques en 2024
  - des *outsiders* qui cherchent à se démarquer
  - l'importance de diversifier les outils gratuits et payants (institution)
  - l'intérêt des extensions de navigateurs
  - l'importance de rester vigilant face à la notion d'*open access*
  - le développement d'outils utilisant de l'intelligence artificielle

# Focus : *Open access*







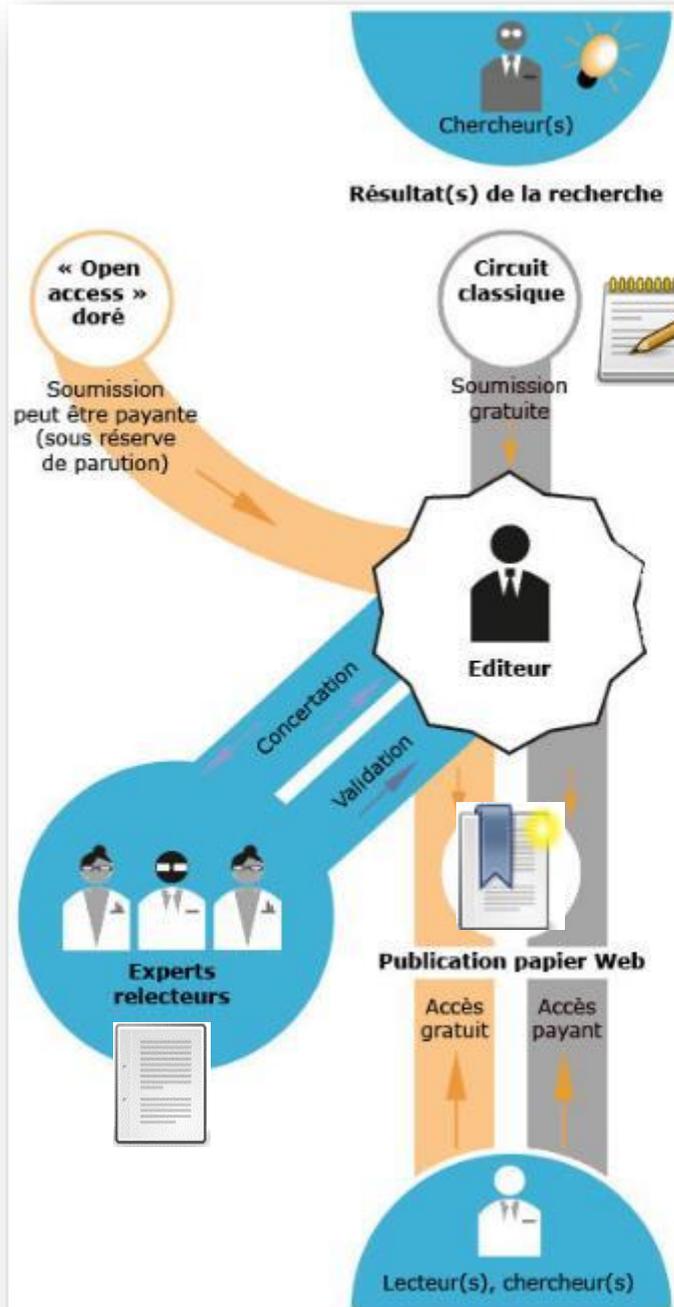
*gold open access*  
voie dorée  
→ éditeurs



*green open access*  
voie verte  
→ chercheurs (autoarchivage)



*gold open access*  
voie dorée  
→ éditeurs

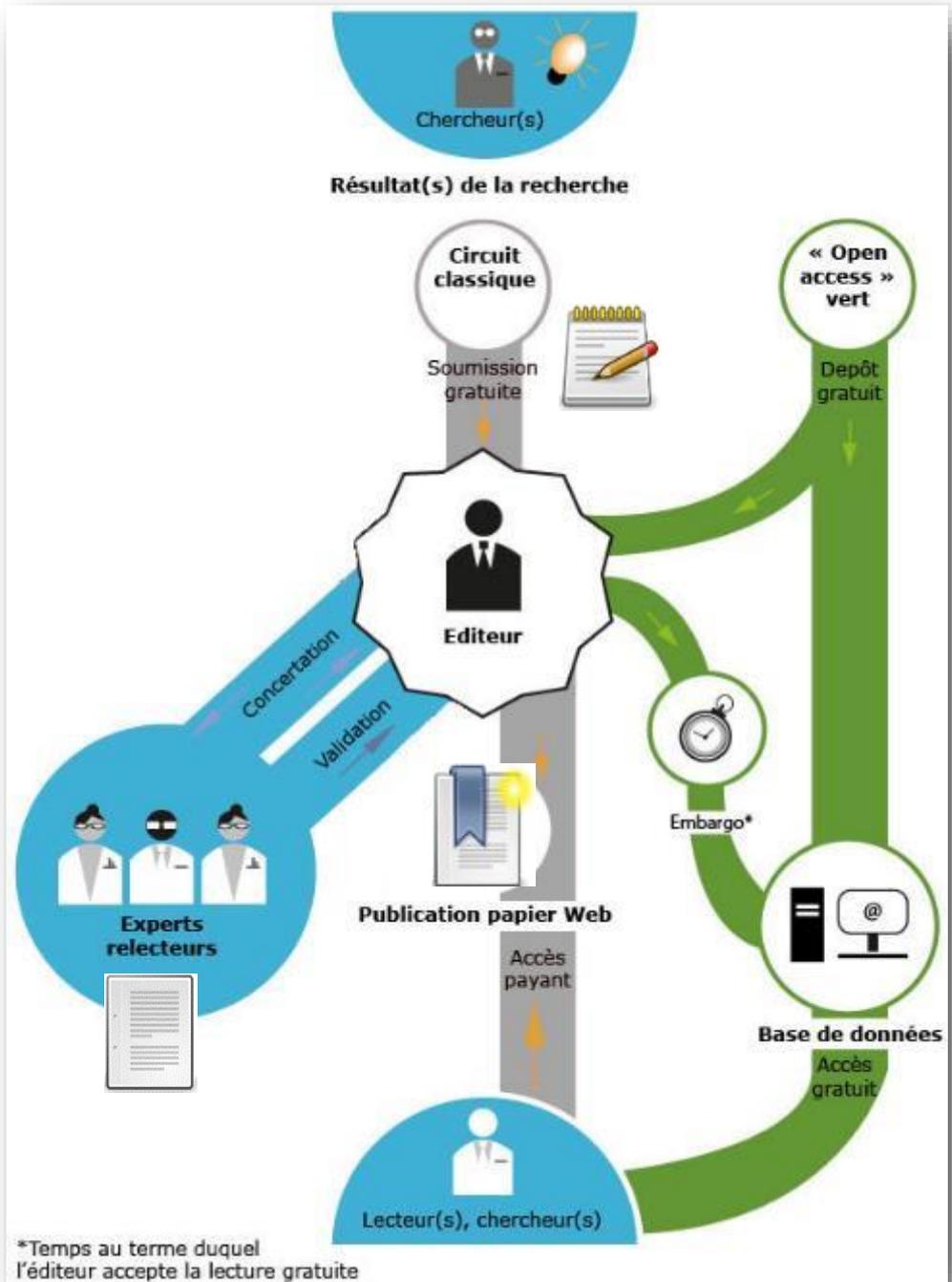




*green open access*

voie verte

→ chercheurs (autoarchivage)



# Focus : *Open access*

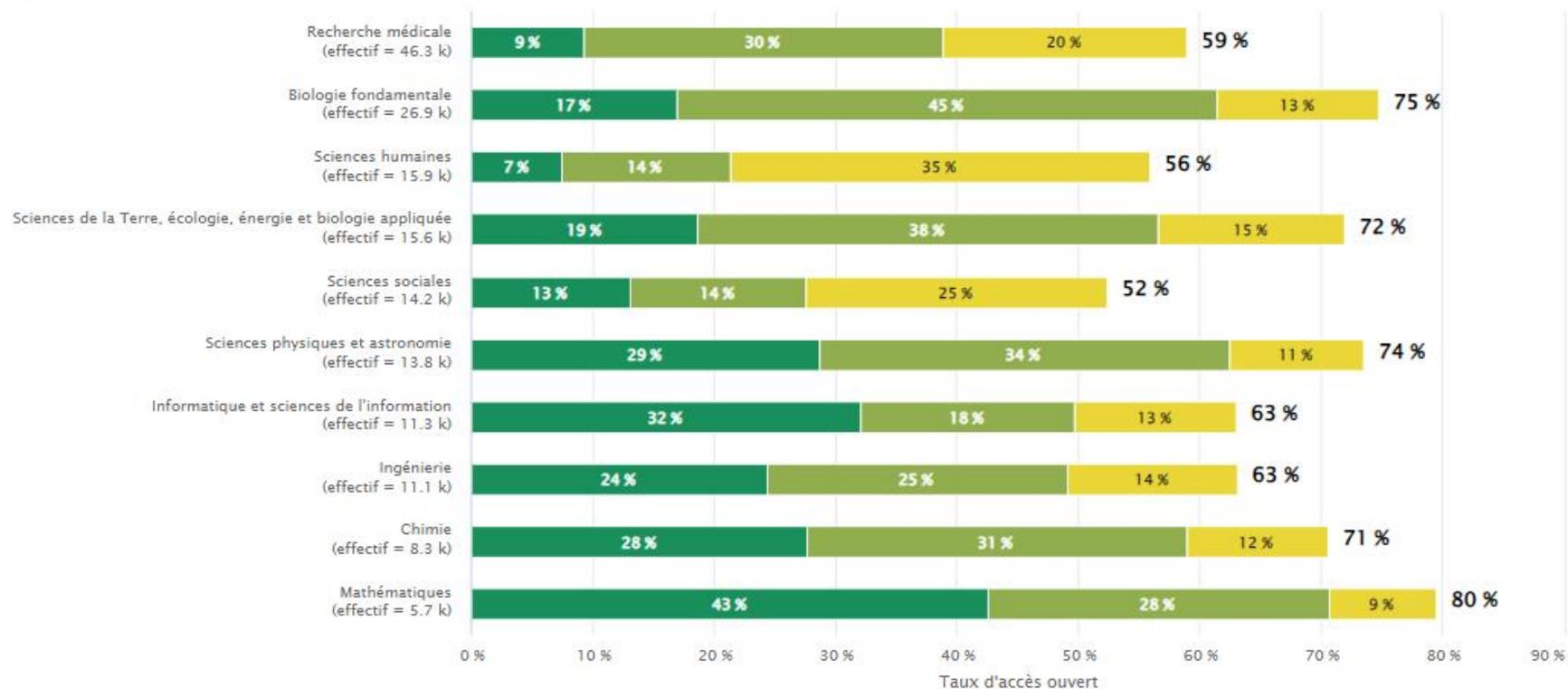
Répartition des publications françaises, avec un DOI Crossref, par voie d'ouverture pour chaque discipline (publications de 2022)

Trier par :

Plus grand effectif  Plus fort taux d'accès ouvert

Voie d'ouverture

Archive ouverte  Éditeur & archive ouverte  Éditeur



Baromètre français de la Science Ouverte - CC-BY MESR

## Web of Science

### Open Access

- All Open Access
- Gold
- Gold-Hybrid
- Free to Read
- Green Published
- Green Accepted
- Green Submitted

[en savoir plus](#)



Scopus

### Open access

- All open access
- Gold
- Green
- Bronze
- Hybrid gold

[en savoir plus](#)

arXiv is a free distribution service and an open-access archive for 1,842,477 scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics. Materials on this site are not peer-reviewed by arXiv.

### Subject search and browse:

Physics



Search

Form Interface

Catchup

### News

Read about recent news and updates on [arXiv's blog](#). (View the former "what's new" pages here). Read [robots beware](#) before attempting any automated download.

## Physics

- [Astrophysics](#) ([astro-ph new](#), [recent](#), [search](#))  
includes: [Astrophysics of Galaxies](#); [Cosmology and Nongalactic Astrophysics](#); [Earth and Planetary Astrophysics](#); [High Energy Astrophysical Phenomena](#); [Instrumentation and Methods for Astrophysics](#); [Solar and Stellar Astrophysics](#)
- [Condensed Matter](#) ([cond-mat new](#), [recent](#), [search](#))  
includes: [Disordered Systems and Neural Networks](#); [Materials Science](#); [Mesoscale and Nanoscale Physics](#); [Other Condensed Matter](#); [Quantum Gases](#); [Soft Condensed Matter](#); [Statistical Mechanics](#); [Strongly Correlated Electrons](#); [Superconductivity](#)
- [General Relativity and Quantum Cosmology](#) ([gr-qc new](#), [recent](#), [search](#))
- [High Energy Physics - Experiment](#) ([hep-ex new](#), [recent](#), [search](#))
- [High Energy Physics - Lattice](#) ([hep-lat new](#), [recent](#), [search](#))
- [High Energy Physics - Phenomenology](#) ([hep-ph new](#), [recent](#), [search](#))
- [High Energy Physics - Theory](#) ([hep-th new](#), [recent](#), [search](#))
- [Mathematical Physics](#) ([math-ph new](#), [recent](#), [search](#))
- [Nonlinear Sciences](#) ([nlin new](#), [recent](#), [search](#))

### COVID-19 Quick Links

See COVID-19 SARS-CoV-2 preprints from

- [arXiv](#)
- [medRxiv and bioRxiv](#)

**Important:** e-prints posted on arXiv are not peer-reviewed by arXiv; they should not be relied upon without context to guide clinical practice or health-related behavior and should not be reported in news media as established information without consulting multiple experts in the field.

## la nécessaire évaluation

- versions ?
- *peer-review* ?



### parts non couvertes

- disciplines
- langues
- types de publication
- DOI

### *free vs open access*

- accès
- réutilisation

## Focus : *Open access*

- limites de l'*open access* en recherche d'informations
  - versions accessibles (*preprints, postprints, PDF* éditeurs)
  - indexation et accès au texte intégral ou seulement aux métadonnées
  - accès à l'ensemble du texte intégral
  - droits de réutilisation
  - accroissement des productions problématiques (éditeurs prédateurs, *paper mills*)

# Focus : Co-citations

- intérêt des outils

- technique de la « boule de neige » (*snowball*) : au-delà des seuls mots-clés et sémantique
- des résultats faciles à naviguer (visualisations)

- limite des outils

- données utilisées ?
- nécessité de bien choisir les « *seeds* » / « *seminal papers* »
- possibilité de croiser les co-citations et d'autres *inputs* ?
- difficulté de toujours comprendre les critères d'affichage

# IA et bases de données bibliographiques

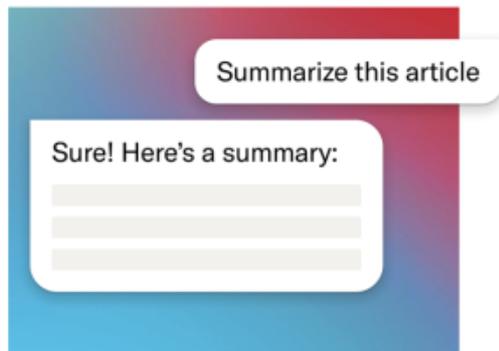
moteurs de recherche scientifiques « classiques »	moteurs de recherche scientifiques « boostés à l'IA »
<p>recherche classique (mots-clés, booléens)</p> <p>fonctionnalités de recherche avancée (tris, filtres)</p> <p>classements les plus fréquents : alphabétique, popularité</p>	<p>recherche sémantique (langage naturel)</p> <p>parfois peu de systèmes de recherche avancée, mais plutôt assistants de recherche et filtres <i>a posteriori</i></p> <p>classement selon la pertinence sémantique</p>

# IA et bases de données bibliographiques

## About JSTOR's AI research tool (beta)

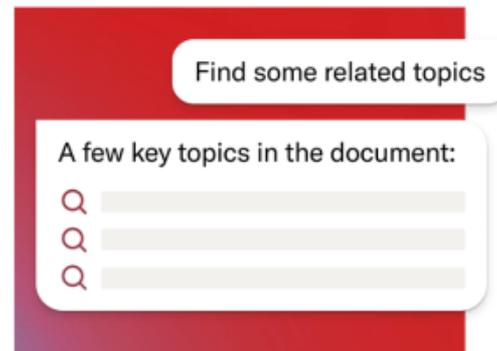
Our generative AI-powered research tool is designed to help people work more efficiently and effectively. This beta feature will appear on the content page for journal articles, book chapters, and research reports, and as an alternative to JSTOR's standard keyword search. The tool helps you do the following:

### Assess content relevance



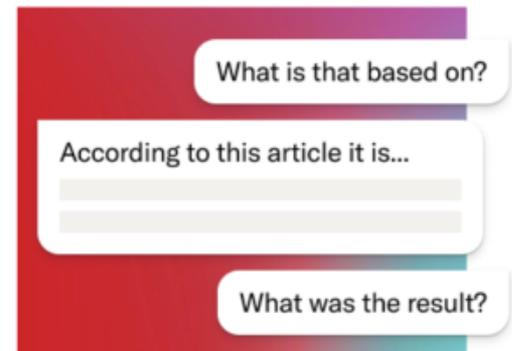
The tool generates a summary of what you're reading to help you quickly assess its relevance, and lets you know how it relates to your search terms

### Deepen your research



Discover related topics, enrich your reading with similar content from the JSTOR corpus, and try new ways of searching

### Be conversational



Use natural language to ask questions and get quick answers about what you're reading or researching

# IA et bases de données bibliographiques

REQUÊTE

BASE DE DONNÉES  
BIBLIOGRAPHIQUES

RÉSULTATS

Référence A

---

---

---

Référence B

---

---

---

*abstract*

---

---

---

Référence C

---

---

---

*abstract*

---

---

---

---

texte intégral

---

---

---

---

---

---

---

---

classement des résultats

*clustering* mots-clés

suggestions automatiques

visualisations

synthèses et résumés  
en langage naturel

prise en charge de  
différentes langues

extraction  
d'informations

conversation  
(Q & A)

mots-clés, booléens  
langage naturel



+ LLM  
*large language model*



# IA et bases de données bibliographiques

Find papers

**karate and well-being**

Searching more than 125 million academic papers  
 Summarizing the first 4 abstracts  
 Results

Summary of top 4 papers Copy

Research consistently shows that karate training can have a positive impact on well-being. [Fuller \(2019\)](#) and [Croon \(2014\)](#) both highlight the role of karate in promoting self-awareness, self-control, and psychological well-being. This is further supported by [Potoczny \(2022\)](#), who found that karate training can indirectly enhance satisfaction with life through improved self-control and emotion regulation. [Moore \(2020\)](#) also underscores the positive effect of martial arts training, including karate, on mental health outcomes, particularly in improving well-being and reducing symptoms associated with internalizing mental health.

+ Add columns Sort Filters

Paper	Abstract summary
<input type="checkbox"/> <b>Martial Arts and Well-Being</b> C. Fuller +1 2019 8 citations <a href="#">DOI</a>	Participating in martial arts can have a powerful influence on the construction of behavior patterns.
<input type="checkbox"/> <b>Self-Control and Emotion Regulation Mediate the Impact of Karate Training on Satisfaction With Life</b> Wlitor Potoczny +2 Frontiers in Behavioral Neuroscience 2022 3 citations <a href="#">PDF</a> <a href="#">DOI</a>	Karate training can play a powerful influence on personality characteristics.
<input type="checkbox"/> <b>Embodying martial arts for mental health: Cultivating psychological well-being with martial arts practice</b> Adam M. Croon 2014 27 citations	Martial arts practice may positively promote psychological well-being.
<input type="checkbox"/> <b>The effect of martial arts training on mental health outcomes: A systematic review and meta-analysis.</b> Brian Moore +2 Journal of bodywork and movement therapies 2020 10 citations <a href="#">DOI</a>	Martial arts training is an efficacious sports-based mental health intervention for improving well-being and reducing symptoms associated with internalizing mental health.

Find papers

**a paper that shows the benefits of karate for well-being**

Searching more than 125 million academic papers  
 Summarizing the first 4 abstracts  
 Results

Summary of top 4 papers Copy

A range of studies have highlighted the potential benefits of karate for well-being. [Mastnak \(2016\)](#) and [Bigs \(2018\)](#) both suggest that karate can improve musculoskeletal health and balance, which are important components of overall well-being. [Fuller \(2019\)](#) and [Croon \(2014\)](#) further explore the psychological and social benefits of karate, including increased self-awareness, positive emotion, and a sense of community. These findings collectively suggest that karate can have a positive impact on both physical and mental well-being.

+ Add columns Sort Filters

Paper	Abstract summary
<input type="checkbox"/> <b>Karate-based prevention of work-related musculoskeletal syndromes: a study on the possible benefits of martial arts in sports medicine and for occupational health</b> Wolfgang Mastnak Current Contents for Medicine	A modified and individually tailored form of karate are viable key factors for a sports-based decrease of work-related overuse risks.
<input type="checkbox"/> <b>Participating in martial arts can have a powerful influence on the construction of behavior patterns.</b>	Participating in martial arts can have a powerful influence on the construction of behavior patterns.
<input type="checkbox"/> <b>Hard martial arts seem to have potential to improve balance and cognitive functions that decline with age.</b>	Hard martial arts seem to have potential to improve balance and cognitive functions that decline with age.
<input type="checkbox"/> <b>Martial arts practice may positively promote psychological well-being in adults: a systematic review</b>	Martial arts practice may positively promote psychological well-being in adults: a systematic review
<input type="checkbox"/> <b>Embodying martial arts for mental health: Cultivating psychological well-being with martial arts practice</b>	Martial arts practice may positively promote psychological well-being

Find papers

**Can Karate improve well-being?**

Searching more than 125 million academic papers  
 Summarizing the first 4 abstracts  
 Results

Summary of top 4 papers Copy

Research consistently suggests that karate can have a positive impact on well-being. [Moore \(2018\)](#) found that a martial arts-based intervention improved resilience and self-efficacy in secondary school students, while [Moore \(2020\)](#) reported a small but significant positive effect on well-being and a medium effect on internalizing mental health. [Jansen \(2012\)](#) and [Jansen \(2017\)](#) both found that karate training led to improvements in emotional well-being and cognitive functioning in older adults, with the latter study also noting a decrease in anxiety. These findings collectively support the idea that karate can enhance well-being across different age groups.

+ Add columns Sort Filters

Paper	Abstract summary
<input type="checkbox"/> <b>Developing Wellbeing Through a Randomised Controlled Trial of a Martial Arts Based Intervention: An Alternative to the Anti-Bullying Approach</b> B. Moore +2 International Journal of Environmental Research and Public Health 2018 62 citations <a href="#">DOI</a>	A martial arts-based psycho-social intervention may be an efficacious method of improving wellbeing outcomes including resilience and self-efficacy.
<input type="checkbox"/> <b>The effect of martial arts training on mental health outcomes: A systematic review and meta-analysis.</b> Brian Moore +2 Journal of bodywork and movement therapies 2020 10 citations <a href="#">DOI</a>	Martial arts training is an efficacious sports-based mental health intervention for improving wellbeing and reducing symptoms associated with internalizing mental health.
<input type="checkbox"/> <b>Effects of Cognitive, Motor, and Karate Training on Cognitive Functioning and Emotional Well-Being of Elderly People</b> P. Jansen +1 Front. Psychology 2012 17 citations <a href="#">DOI</a>	The integrated involvement in Karate leads to a feeling of self-worth.
<input type="checkbox"/> <b>Effects of Karate Training Versus Mindfulness Training on Emotional Well-Being and Cognitive Performance in Later Life</b> P. Jansen +3 Research on Aging 2017 36 citations <a href="#">DOI</a>	Karate and mindfulness-based stress reduction showed only small training effects concerning the assessed emotional and cognitive parameters.

# IA et bases de données bibliographiques

- limites de l'IA dans les bases bibliographiques
  - sources utilisées et sélection
  - données (métadonnées, *abstracts*, texte intégral)
  - classement (pertinence sémantique)
  - fiabilité des réponses, notamment leur caractère scientifique
  - risque d'une « intrusion épistémique »

- un paysage de la recherche de plus en plus fragmenté
  - un outil toujours central : Google
  - une multiplication d'autres acteurs et des fonctionnalités qui bougent
  - des services de plus en plus fermés voire payants (fin du « tout gratuit »)
  - des questions sur les modèles économiques et leur viabilité

→ nécessité de diversifier les outils

- le développement de l'IA comme une fonctionnalité « standard »
    - comme outils spécifiques
    - dans les moteurs de recherche
    - dans les bases de données bibliographiques
- une extension du domaine de la « recherche »
- de nouveaux points de vigilance

- une recherche académique de plus en plus « hybride »
  - par mots-clés
  - par co-citations
  - par sémantique

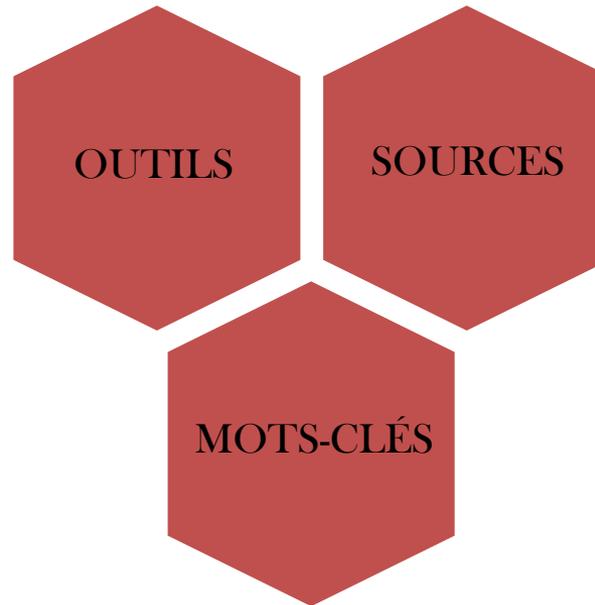
→ de nouveaux points de vigilance

- une évaluation de plus en plus centrale dans la démarche de recherche d'informations
  - résultats obtenus
  - sources utilisées
  - classement des résultats (crédibilité, classement, fraîcheur, pertinence, exhaustivité)
  - possibilités de recherche avancée (filtres, tris...)
  - fiabilité des résultats par rapport aux sources

→ un renforcement nécessaire de l'expertise ?

- le rôle toujours plus central de la méthodologie documentaire  
comment être à la fois pertinent et efficace ?

- analyse critique des outils  
et de leur fonctionnement  
- diversification des outils  
et des méthodes de  
recherche



- évaluation des résultats  
- vérification des sources

- identification des mots-clés  
et concepts

+ documentation de sa recherche

# Pour aller plus loin

- méthodologie et outils

- *Recherche d'informations sur internet (perfectionnement)*. MAJ 2021, à compléter par la carte : <https://framindmap.org/c/maps/446941/public>. <https://urfist.chartes.psl.eu/ressources/recherche-d-informations-sur-internet-perfectionnement>.
- Véronique Mesguich. *Rechercher l'information stratégique sur le web. Sourcing, veille et analyse à l'heure de la révolution numérique*. 2<sup>e</sup> éd. ADBS-éd. De Boeck, 2021. 236 p.
- id. *Rechercher l'information stratégique sur le web. Sourcing, veille et analyse à l'heure de l'IA*. 3<sup>e</sup> éd. ADBS-éd. De Boeck, 2024. 240 p.

## focus particuliers

- *Rechercher des publications scientifiques en accès libre* (2024): <https://urfist.chartes.psl.eu/ressources/exploiter-l-open-access-en-recherche-d-informations>
- *ChatGPT et les autres : recherche d'information et intelligence artificielle* (2024): <https://urfist.chartes.psl.eu/ressources/chatgpt-et-les-autres-recherche-d-information-et-intelligence-artificielle>
- *Au-delà de ChatGPT : recherche d'informations académiques et intelligence artificielle* (2024): <https://urfist.chartes.psl.eu/ressources/au-dela-de-chatgpt-recherche-d-informations-academiques-et-intelligence-artificielle>

- pour suivre l'actualité du domaine, les incontournables :

- ressources de FLA Consultants, plus ou moins gratuites : revues *Bases* et *Netsources*, <https://www.bases-netsources.com/> ; brèves : <https://www.bases-netsources.com/breves-de-veille> et compte X-Twitter : <https://x.com/BasesNetsources>
- blog d'Aaron Tay, <https://musingsaboutlibrarianship.blogspot.com>