

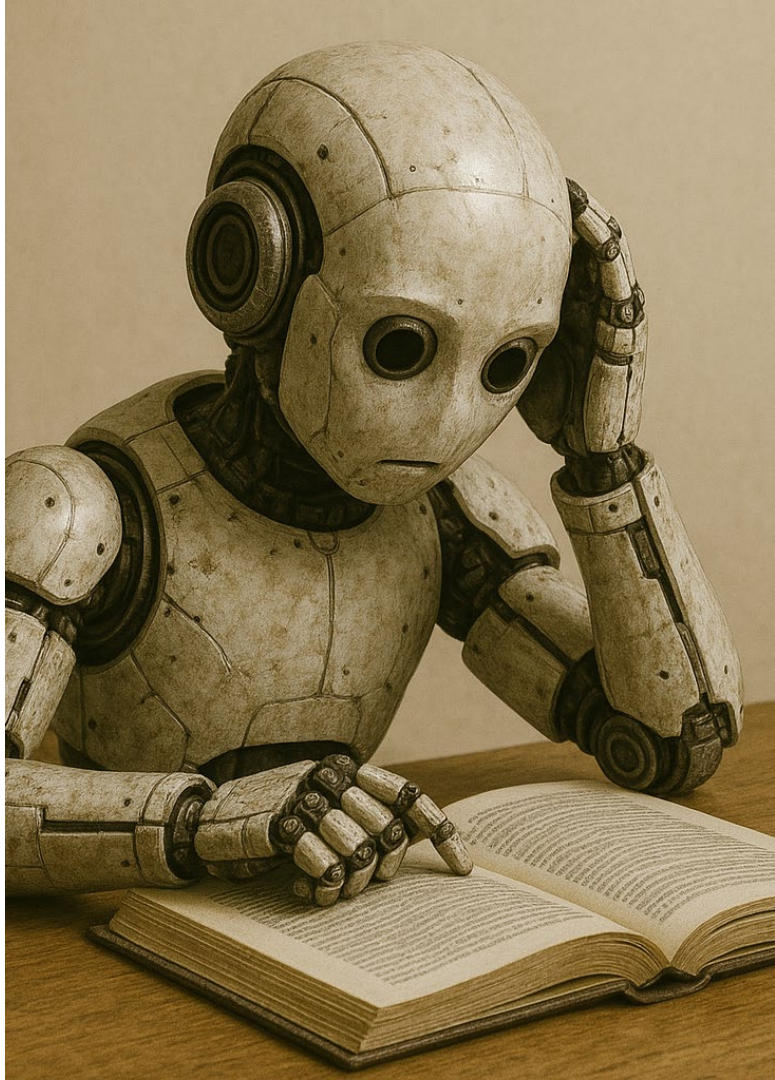


Reshaping Information Research with AI

ENG-619 Day1: Searching

Vincenzo Palatella

EPFL
31.10.2025



Agenda

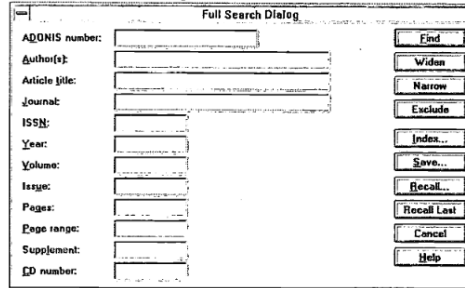
- **A Short History of Information Retrieval**
- **GenAI and Information Retrieval**
- **Research Assistants : Test and Share**
- **Conclusions + Q&A**



A Short History of Information Retrieval

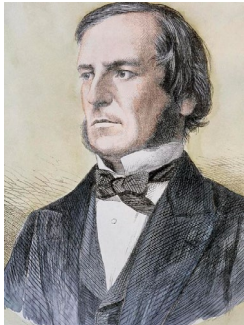
A short history of Information Retrieval

1847
1854



1997
1998

(Pereira 1996, Orchard 1988)



(Wikipedia 2024)

1987



But...

- We've been using the search equation as is for almost **30 years**
- Bibliographic databases (e.g. Scopus and WoS) work well with **simple keywords**
- Every tool has its own **peculiarities** (filters, proximity operators, etc.)

Moreover...

- Traditional tools are «**going AI**» ([Scopus AI](#), [WoS Research Assistant](#))...
- ...and AI tools are «**going academic**» ([Perplexity and Wiley](#))
- **Search engines are dead (?)**
(Wolff 2023)





GenAI and Information Retrieval

Message ChatGPT



ChatGPT can make mistakes. Check important info.



Claude can make mistakes. Please double-check responses.

Reply to Claude...



Claude 3.5 Sonnet ▾  Choose style ▾

Keep in mind:

- They're tools, not reliable sources per se
- They use patterns, no full understanding of the subject

Ask yourself:

- Where the information comes from?
- Can we retrieve the sources?
- Are the sources correct?



- Fine-tune language/style
- Get new ideas (brainstorm) if you get stuck
- Create pictures
- Always inform the final reader about the tool you used

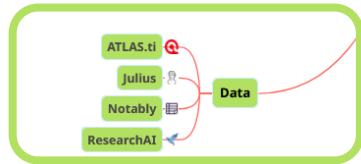
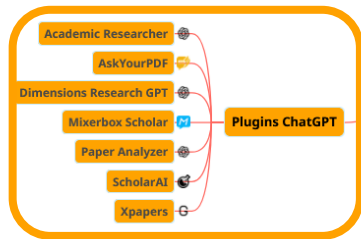
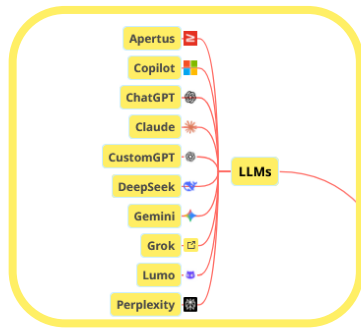
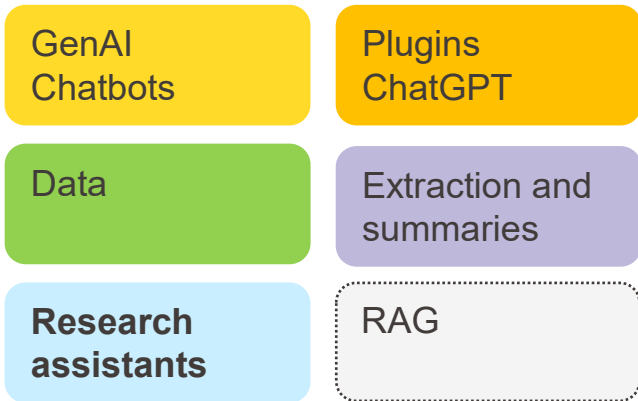


- Get a basic idea about a new subject (but always double check the information)
- It can generate “hallucinations”, false information presented as facts (including bibliographic references)

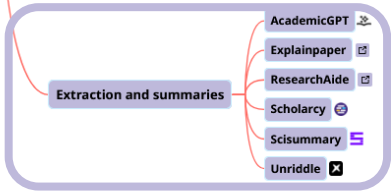


- Write a paper from scratch
- Use it as the only source of information
- Cite a generated text in a paper
- Beware of made-up references

Beyond GenAI: Research Assistants



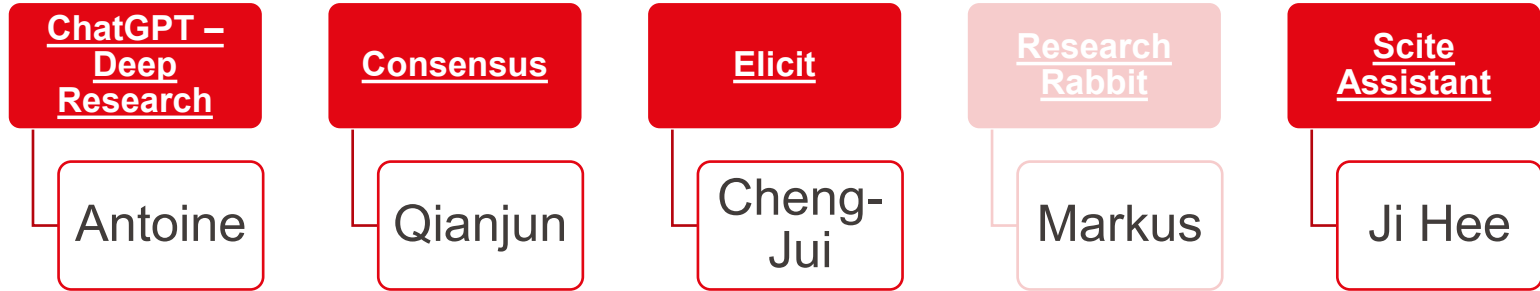
AI in information research



Palatella, Vincenzo. 'AI in information research', 2025.
<https://www.mindomo.com/fr/mindmap/ai-in-information-research-1a5de88798ec453e9be2fd3abf9205d8>.



Test and Share!



Question

Is there an alternative to the use of lead in Perovskite solar cells?

Test the assigned tool, trying to answer these questions:

- How does it work? Which are the main features?
- Where does the information (corpus) come from?
- What's better than searching with traditional tools? What's worse? Can we compare them?
- What's the best possible use in academic research?

Time

- 20' to test / 5' (each) to present

Tool	Main features	Corpus	Best use case	Comparison with «classic» tools
ChatGPT (Deep research)				
Consensus				
Elicit				
Scite Assistant				

Main Features

- Citation context analysis (supporting/contrasting)
- AI Assistant for research questions
- Citation alerts and tracking

Corpus Used

- 1.2B+ citations across 200M+ sources

Best Use Case in Academic Research

- Evaluating research credibility and impact
- Understanding how papers are being cited in context
- Tracking research influence over time

Comparison with "Classic" Research Tools

- ➡ Provides unprecedented citation context analysis
- ➡ Shows how research is actually being used and interpreted
- ➡ Helps assess research credibility beyond citation counts
- ⚠ Database coverage may not include every scientific article
- ⚠ Some fields or sub-disciplines may be better represented than others
- ⚠ May exhibit biases in citation categorization

Main Features

- Multi-step autonomous web research
- Comprehensive report generation with citations
- Real-time web browsing and PDF analysis

Corpus Used

- Real-time web search across public internet

Best Use Case in Academic Research

- Complex, multi-faceted research questions requiring diverse sources

- Time-sensitive research requiring current information

Comparison with "Classic" Research Tools

- 📌 Generates comprehensive, cited reports automatically
- 📌 Accesses real-time information beyond academic databases
- 🚫 Can sometimes hallucinate facts or make incorrect inferences
- 🚫 Limited to publicly available web content (but it can be extended via plugins)

Main Features

- Evidence-based synthesis from scientific research
- "Consensus Meter" showing research agreement

Corpus Used

- Over 200 million academic papers from Semantic Scholar corpus (updated monthly)

Best Use Case in Academic Research

- Quick consensus assessment on research topics
- Identifying areas of scientific agreement/disagreement

- Rapid literature synthesis for grant proposals

Comparison with "Classic" Research Tools

- 🟢 Provides quick scientific consensus assessment
- 🟢 Generates evidence-based summaries with citations
- 🔴 May not include the most recent academic publications
- 🔴 Limited to peer-reviewed literature only
- 🔴 Synthesis may oversimplify complex research areas

Main Features

- Automated literature reviews
- Data extraction from research papers
- Semantic search across abstracts and full-text

Corpus Used

- 125 million academic papers from Semantic Scholar corpus
- Covers all academic disciplines

Best Use Case in Academic Research

- Systematic reviews
- Data extraction from large volumes of literature
- Literature discovery without perfect

keyword matching

Comparison with "Classic" Research Tools

- 👉 Reduces systematic review time
- 👉 Automates tedious literature screening processes
- 👉 Finds relevant papers without perfect keyword matching
- 👎 Limited to academic databases
- 👎 Requires manual verification
- 👎 May miss very recent publications due to indexing delays

- **One size doesn't fit all:** each tool has its own features, the key is understanding how it works and how it can support our research. Some tools work better with some disciplines, some are focused on specific aspects of research (i.e. [Elicit](#) for literature reviews).
- The **corpus** is key! No tool covers everything and quantity doesn't always rhyme with quality.
- **Play** with them to understand them 😊

- 'George Boole'. In *Wikipédia*, 20 February 2024. [https://fr.wikipedia.org/w/index.php?title=George Boole&oldid=212637018](https://fr.wikipedia.org/w/index.php?title=George_Boole&oldid=212637018).
- Orchard, Constance. 'ADONIS and Electronically Stored Information: An Information Broker's Experience'. *The Serials Librarian* 15, no. 3–4 (1 December 1988): 85–91. https://doi.org/10.1300/J123v15n03_09.
- Pereira, Monica. 'ADONIS: One Library's Experience with a CD-ROM Document Delivery System', 17 May 1996. <https://eric.ed.gov/?id=ED399965>.
- Wolff, Tristan. 'Search Engines Are Dead. Right?' *Predict* (blog), 24 March 2023. <https://medium.com/predict/search-engines-are-dead-right-6cb08cc784a7>.

All images (slide 5 and 21 aside) have been generated with ChatGPT 4o.

Prompts used:

- Slide 1: *Create a realistic image where a robot teacher is teaching the subject "human beings" to a class of young robots*
- Slide 2: *In the same style, depict one of the young robots studying on a table (whatever studying for a robot means)*
- Slide 4: *Create an image about information retrieval in 2200*
- Slide 6: *Depict the same character in the same style but way older and reading a book in a library full of spiderwebs*
- Slide 7: *Create an image of young robots watching an instructional video about human beings*
- Slide 11: *Create a realistic images of 4 people discussing in front of a flat screen; Change the caption of the graph into "use of research assistants over time"*



Guidelines for an Ethical Use of AI in Research

The guidelines

- 1. Environmental Impact:** Select AI tools with transparent energy consumption reporting and consider computational efficiency in research design to minimize carbon footprint.
- 2. Algorithmic Bias & Fairness:** Document and critically evaluate potential biases in AI systems used in research, implementing verification procedures to identify and mitigate their impact.
- 3. Researcher Agency:** Maintain final intellectual responsibility for all research outputs while establishing clear institutional policies that recognize appropriate AI use and contribution.
- 4. Source Quality:** Independently verify all AI-generated citations and implement rigorous validation protocols for factual claims from AI tools.
- 5. Intellectual Property:** Select AI tools with transparent training data provenance and ensure proper attribution for all AI-assisted components of research.
- 6. Equity & Accessibility:** Prioritize open-access and open-source AI tools to promote equitable research opportunities across institutions and geographical regions.
- 7. Reproducibility & Transparency:** Document all AI tool parameters, prompts, and methodologies used in research to ensure reproducibility and accountability.



formations.bib@epfl.ch



library.epfl.ch
[Our trainings](#)



[Book a librarian](#)