Waipapa Taumata Rau
University of Auckland

AI tools for literature reviews

We will get started soon!

Meanwhile, introduce yourself, faculty and your research area in the chat

Sept 2025

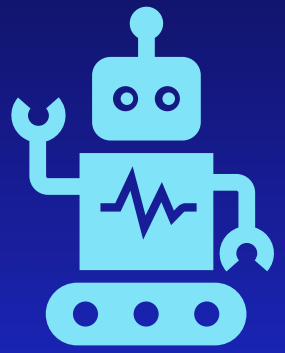
Ngā Ratonga Manaaki Rangahau | Research Services
Student and Scholarly Services

Session roadmap

1. Using AI for research – navigating policy
2. Literature review overview
3. AI for literature discovery
4. AI for screening and data extraction
5. Evaluating - use case, tool, output

Pātai

How have you been using AI for study, research or personal applications?



AI for
Your research

What guides acceptable use?

National

Publishers

Discipline practices

Institution

What guides acceptable use?

National

Publishers

Discipline practices

Institution

Governmental policy

AI for the Public Service

Royal Society Te Apārangi

Guidelines for the best-practice use of generative artificial intelligence in research in Aotearoa New Zealand (Online)

- Te Tiriti o Waitangi
- Māori, Pacific and Indigenous data sovereignty
- Research integrity

What guides acceptable use?

National

Publishers

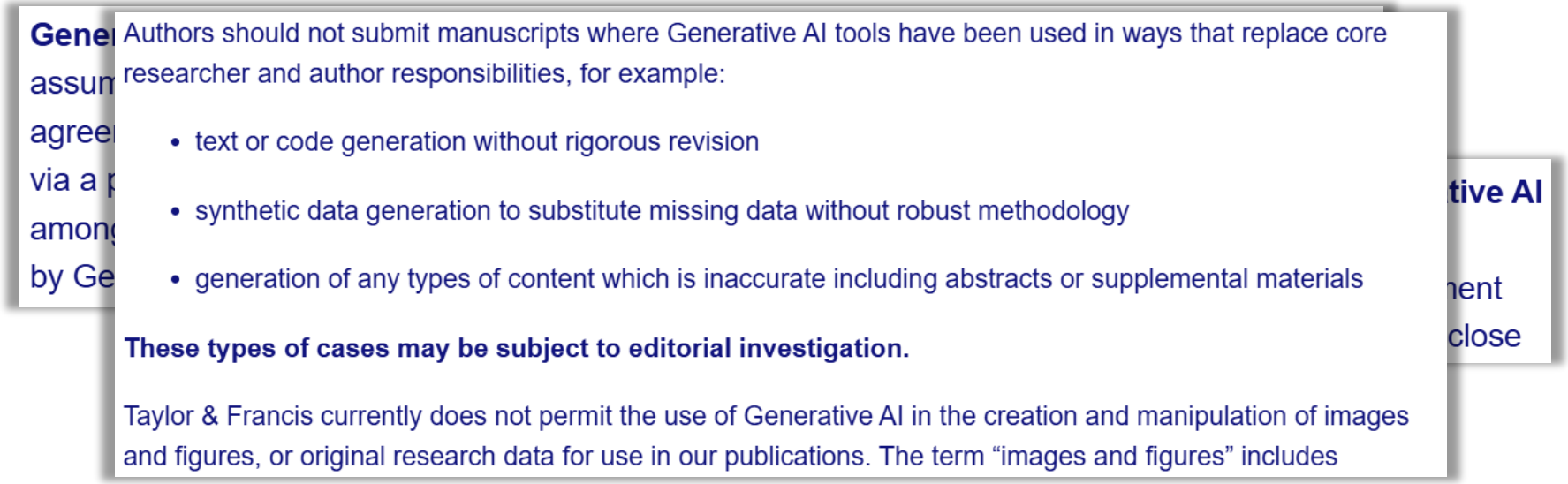
Discipline practices

Institution

Publishers

Not a co-author, disclosure and specific stipulations

Taylor & Francis AI policy
as of 01/08/2025



Authors should not submit manuscripts where Generative AI tools have been used in ways that replace core researcher and author responsibilities, for example:

- text or code generation without rigorous revision
- synthetic data generation to substitute missing data without robust methodology
- generation of any types of content which is inaccurate including abstracts or supplemental materials

These types of cases may be subject to editorial investigation.

Taylor & Francis currently does not permit the use of Generative AI in the creation and manipulation of images and figures, or original research data for use in our publications. The term “images and figures” includes

What guides acceptable use?

National

Publishers

Discipline practices

Institution

NZ Law Society

- Generative AI guidance for lawyers
- Generative AI guidance for lawyers: balancing the opportunities and risks

Medicine

- AI in Health Research Network
- National Artificial Intelligence and Algorithm Expert Advisory Group

Ministry of Education

- Guidance and resources for education professionals on the use of artificial intelligence in schools.

Department of Conservation

- Long-Term Insights Briefing

What guides acceptable use?

National

Publishers

Discipline practices

Institution

What reports and research are coming out in your discipline?

What do governing bodies or relevant research societies have on AI?

Discuss current applications for research with your peers and supervisors.

What guides acceptable use?

National

Publishers

Discipline practices

Institution

Gen AI

AI at the University
(guidance & resources)

Generative AI Usage Standard
(no SSO-required via TeachWell)

Research data

Data Classification Standard

Research integrity & good research practice

Research Integrity Policy

Good Research Practice Guidance

Authorship and Publication Guidelines

Copyright

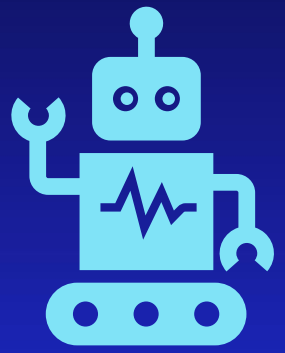
Academic integrity

Student Academic Conduct Statute

Doctoral policies and guidelines

Steps to take for acceptable use - theses

1. Discuss with your supervisor if AI use is appropriate and what are the benefits, risks and limitations (how these will be acknowledged/addressed).
2. Follow the Generative AI usage standard & prioritise using secure UOA-approved tools.
3. Log your interactions with AI (prompts, inputs, outputs) to support research integrity.
4. A thesis is a “substantial piece of original work” – the final content should be your work, not AI’s.
5. Cite your use



AI for
Your literature review

Pātai

What is the purpose and outcomes of a literature review?

- Get an overview of the existing literature
- Form evidence-based conclusions
- Identify research gaps & build research topic
- Critically evaluate research
- Understand methodologies in your discipline
- Build critical thinking, reading and writing skills

Pātai

What could you gain from using AI to do all or some of this work?

What could you lose from using AI to do all or some of this work?

Review process

- Different reviews, different methodologies, different steps
- Unbiased, reliable, reproducible, transparent

Should you use AI for your review?

Purpose

Methodology

Tools

Should you use AI for your review?

Purpose

Methodology

Tools

Different uses – different regulations

What is acceptable use and what is not?

Theses

- ✓ Personal learning about a topic
- ✓ Brainstorming
- ✓ Finding seed articles
- X Substantial pieces of work
- ? What does your supervisor advise

Publications

As per publisher guidelines, discipline standards, University policy

Should you use AI for your review?

Purpose

Methodology

Tools

What is the level of rigour needed?

Meticulousness, systematic and transparent approach to the review process, ensuring the reliability, trustworthiness, and credibility of the findings.

Narrative reviews — Systematic reviews

It might only be appropriate for certain steps

- Research topic development
- Initial scoping
- Finding key articles
- Hand searching
- Screening assistants
- Supervised data extraction

Should you use AI for your review?

Purpose

Methodology

Tools

Are there appropriate tools?

General tools like large language models vs task-specific tools tailored for reviews

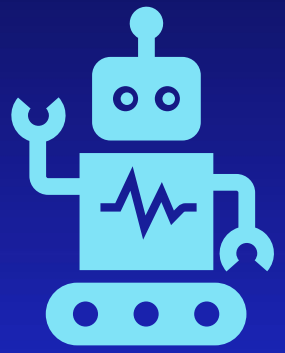
Assistant vs automation

No tools specific to reviews provided via UoA

Tools available to staff and students include:

- Copilot (use for public, internal & sensitive data)
- Gemini (use for public & internal data)
- NotebookLM (use for public & internal data)

See intranet pages Getting started with AI for updates and Data Classification Standard for more information on data types.



AI for
Literature discovery

Traditional database searching

How is artificial intelligence being implemented for medical literature reviews?

Concepts	Artificial intelligence	Literature reviews	Medical
Synonyms	AI	Natural language processing	Systematic Reviews	Narrative Reviews	Health
	Machine learning	Neural networks	Scoping reviews	Cochrane	JBI
Subject terms	Artificial Intelligence		Review Literature as Topic		

Traditional database searching

How is artificial intelligence being implemented for medical literature reviews?

Concepts	“Artificial intelligence”	“Literature reviews”	Medical
Synonyms	AI	“Natural language processing”	“Systematic Reviews”	“Narrative Reviews”	Health
	“Machine learning”	“Neural networks”	“Scoping reviews”	Cochrane	JBI
Subject terms	Artificial Intelligence/	exp Review Literature as Topic/	

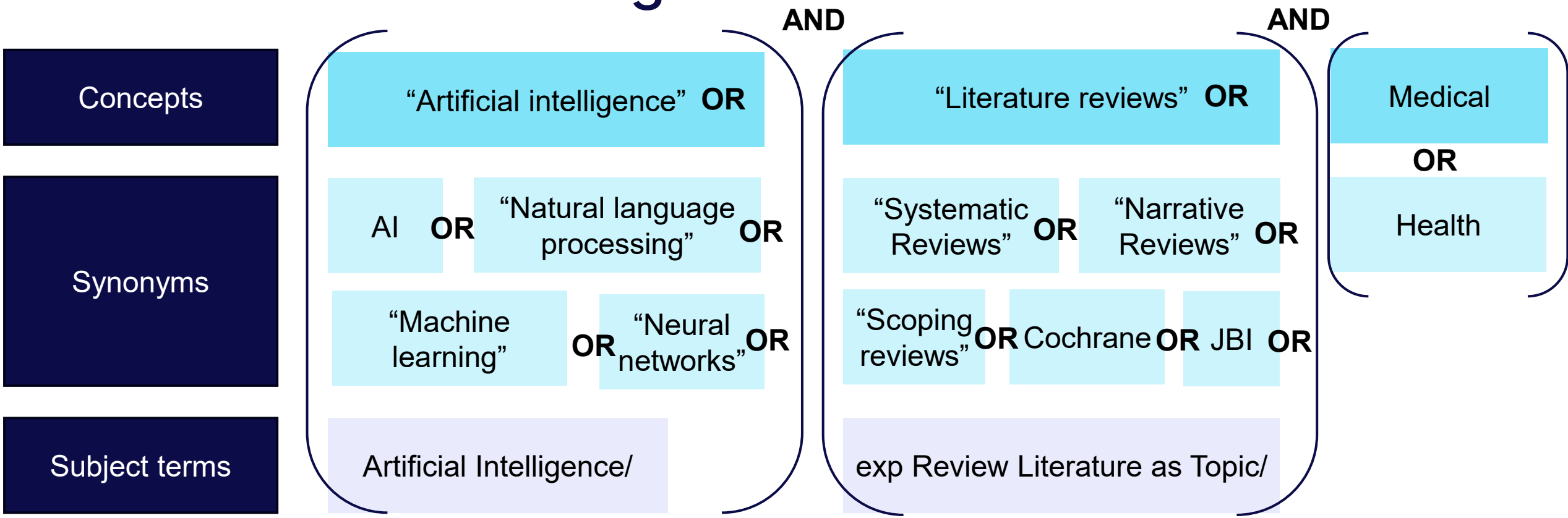
Traditional database searching

How is artificial intelligence being implemented for medical literature reviews?

Concepts	“Artificial intelligence” **OR**	“Literature reviews” **OR**	Medical **OR**
Synonyms	AI **OR** “Natural language processing” **OR**	“Systematic Reviews” **OR** “Narrative Reviews” **OR**	Health
	“Machine learning” **OR** “Neural networks” **OR**	“Scoping reviews” **OR** Cochrane **OR** JBI **OR**	
Subject terms	Artificial Intelligence/	exp Review Literature as Topic/	

Traditional database searching

How is artificial intelligence being implemented for medical literature reviews?



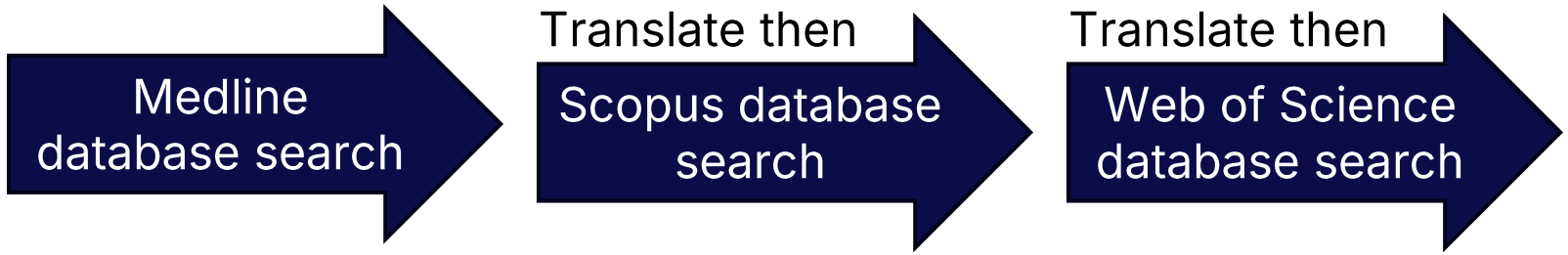
Traditional database searching

How is artificial intelligence being implemented for medical literature reviews?

("Artificial intelligence" OR AI OR "Natural language processing" OR "Machine learning" OR "Neural networks" OR Artificial Intelligence/)

AND ("Literature reviews" OR "Systematic Reviews" OR "Narrative Reviews" OR "Scoping Reviews" OR Cochrane OR JBI OR exp Review Literature as Topic/)

AND (Medical OR Health)



Pātai

Have you heard of or tried any AI tools for finding literature?

Demo #1: Creating search strategies with Copilot

Platform:

Copilot with University login

Prompt:

Create a Boolean search strategy for Medline via Ovid for my research question, "How is artificial intelligence being implemented for medical literature reviews?". Include MeSH terms where relevant.

Discussion:

- LLMs generate related terms, synonyms, and subject terms, such as "AI," "Machine learning," and "Natural language processing.“
- Creates a basic search string of variable quality
- Transfer to database, results examined
- Initial AI-generated searches may require further refinement and integration of subject matter expertise (knowledge of database search tips, of relevant AI tools, etc) to improve results.
- Research shows (Clark et al., 2025) search quality is not good even with next-level prompting (GPT4 studies)
  - Return rate of 13% of relevant results compared to a human-made search by researchers well-practised in review methods
  - Lower precision, meaning more results to sift through before getting relevant results.

Demo #2 - Chatbots with deep research

Platform:

Gemini with University login

Prompts:

1. How is artificial intelligence being implemented for medical literature reviews?
2. How is artificial intelligence being implemented for medical literature reviews? Use only published research or preprints for sources.
3. Create a Boolean search strategy for Medline via Ovid for my research question, "How is artificial intelligence being implemented for medical literature reviews?". Include MeSH terms where relevant.
4. Create a literature review introduction section for my upcoming research article on "How is artificial intelligence being implemented for medical literature reviews?". My target journal is Journal of Librarianship and Information Science so make sure it complies with the journal guidelines.

Discussion:

- Method of deep research tools
  - Some provide initial research plans before starting to sense check what it is going to do
  - Multi-step autonomous working
  - Uses more comprehensive web searching
  - Reasoning its next move based on what it finds
  - The output is based on the results it finds – this is called *retrieval augmented generation* or RAG, where the results are the sources pulled from exclusively
- Cautions
  - Still need to check the sources
  - Use still needs to abide by policy
  - Outputs will often need thorough reworking
  - Can you use the outputs for your use case? i.e. thesis as original work, generative AI outputs as is are not your original work.

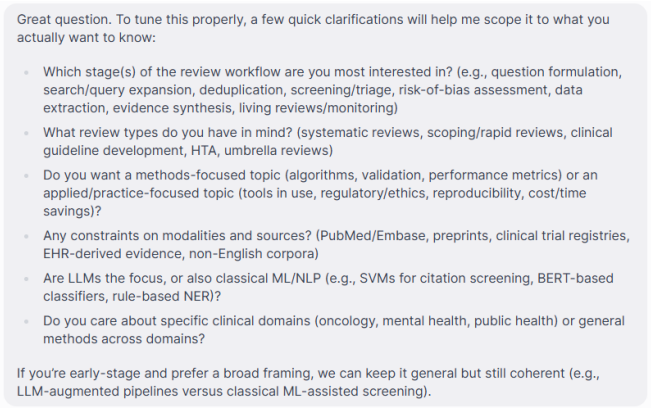
Demo #3 - Academic search engines

Platform:

Elicit / Undermind free accounts

Prompts:

1. Elicit “Find papers” search - How is artificial intelligence being implemented for medical literature reviews?
2. Undermind - How is artificial intelligence being implemented for medical literature reviews? Subsequent iteration with Undermind’s prompt:



Great question. To tune this properly, a few quick clarifications will help me scope it to what you actually want to know:

- Which stage(s) of the review workflow are you most interested in? (e.g., question formulation, search/query expansion, deduplication, screening/triage, risk-of-bias assessment, data extraction, evidence synthesis, living reviews/monitoring)
- What review types do you have in mind? (systematic reviews, scoping/rapid reviews, clinical guideline development, HTA, umbrella reviews)
- Do you want a methods-focused topic (algorithms, validation, performance metrics) or an applied/practice-focused topic (tools in use, regulatory/ethics, reproducibility, cost/time savings)?
- Any constraints on modalities and sources? (PubMed/Embase, preprints, clinical trial registries, EHR-derived evidence, non-English corpora)
- Are LLMs the focus, or also classical ML/NLP (e.g., SVMs for citation screening, BERT-based classifiers, rule-based NER)?
- Do you care about specific clinical domains (oncology, mental health, public health) or general methods across domains?

If you’re early-stage and prefer a broad framing, we can keep it general but still coherent (e.g., LLM-augmented pipelines versus classical ML-assisted screening).

Discussion:

- Scope of tools varies
  - Elicit & Undermind data is from Semantic Scholar, Elicit will pull from full texts if they are available whereas Undermind is abstracts only unless you pay for pro account
- Paid functionality improves performance but is quite expensive
  - Elicit Pro = more columns for table data extraction, a more comprehensive review module, more comprehensive RAG summaries, export results, etc
    Undermind Pro = full texts, exporting references, more chats and better models, etc
  - Raises an equity issue
  - Free use is usually paid for in other ways, i.e., trade off with data privacy

Demo #4: Citation mapping tools

Platform:

Research Rabbit

(others include Litmaps, Connected Papers, Citation Gecko)

Prompts:

AI for literature reviews paper list from

- Uploaded reference library
- DOI search
- Papers found by ResearchRabbit added back into library

Discussion:

- Network from the provided seed articles
- Provides citation tracking (cited and citing papers)
- Some AI-based semantic (contextual/meaning) matching for related papers
- Further publications from the authors

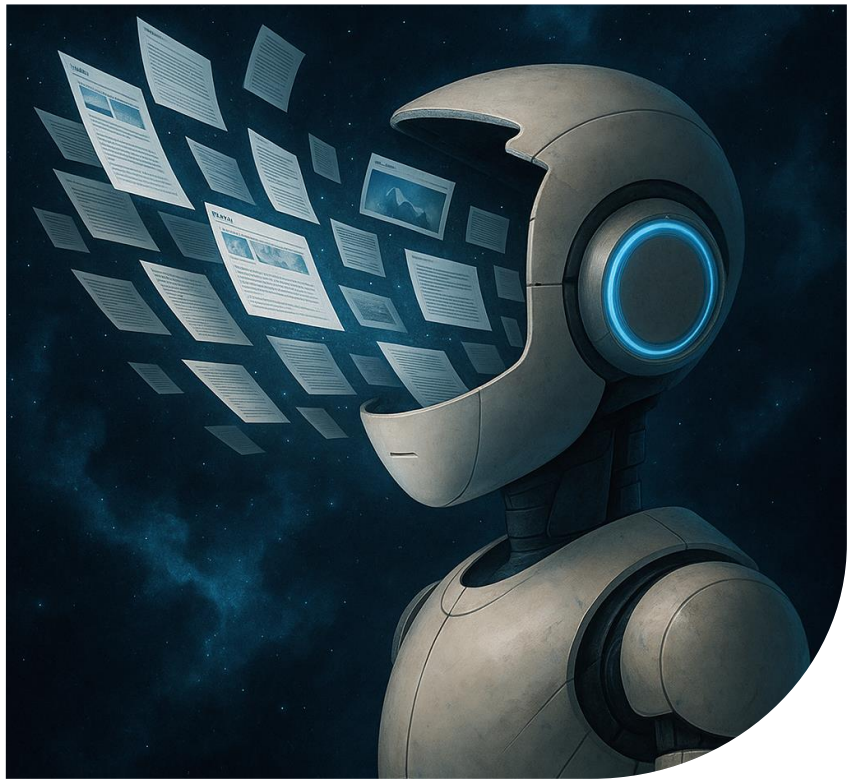
What information are these tools working with?

Sources

- Public metadata (titles, abstracts)
- Open full-texts
- Preprints
- General web content
- Your own papers

Is the data used:

- Quality
- Comprehensive
- Unbiased

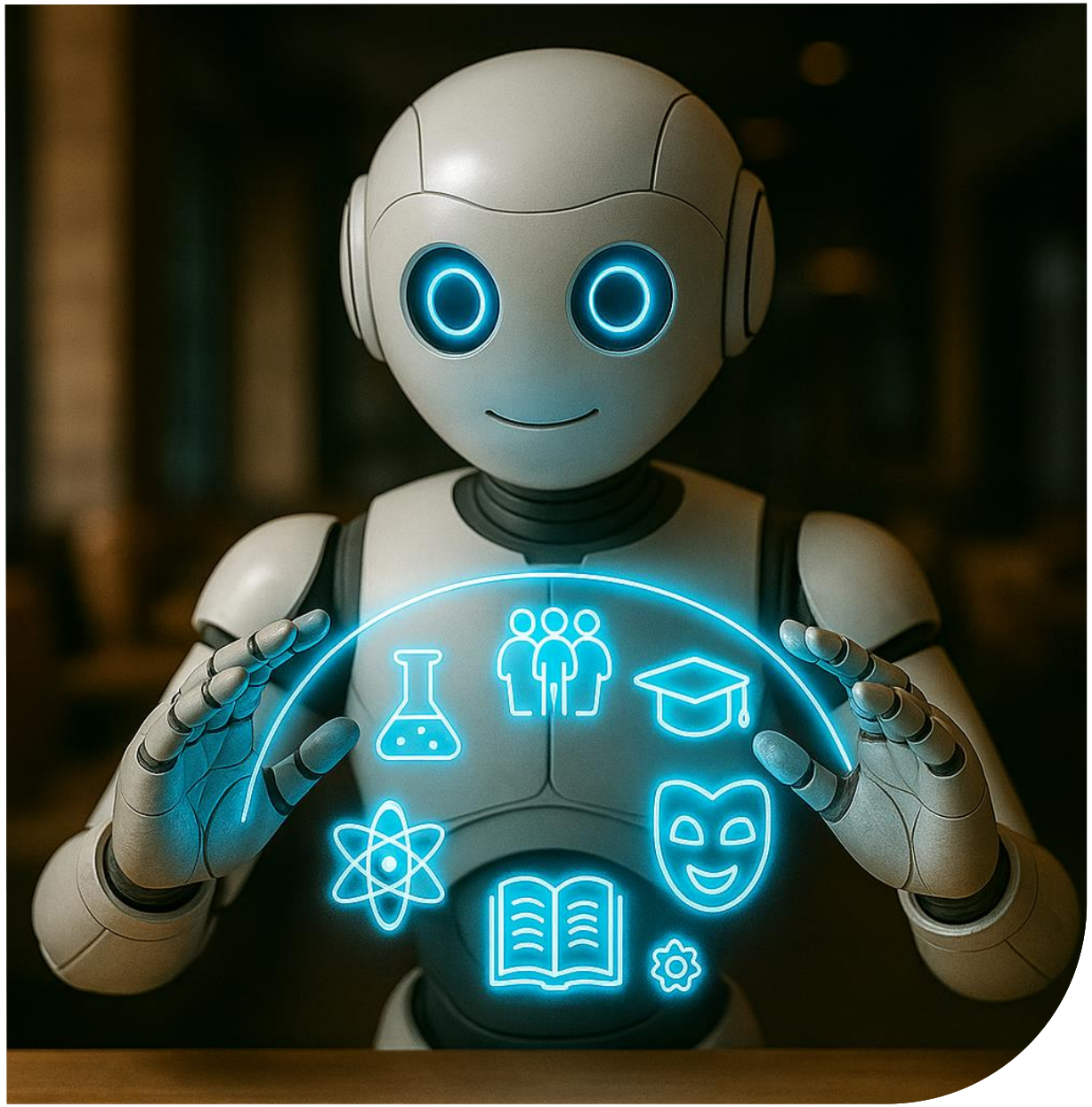


Limitations for literature discovery

Quality

Comprehensive

Biases



Suggestions on when/how to use AI

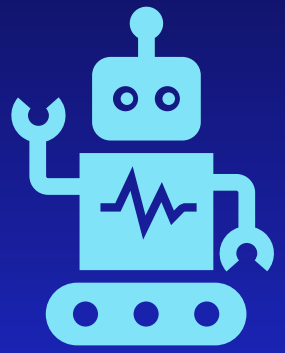
When the required rigour is lower

Where outputs have expert evaluation or validation

Personal knowledge building

Keeping up to date

Record your use



AI for
Screening & data extraction

Pātai

Have you used AI to pull out information from a research article before?

Demo #5

Screening

- Discussion: There is a lot of potential in screening to increase consistency and reduce subjectivity
  - Cao et al., 2025 study of *otto*-SR tool had AI assistance at 96.7% sensitivity (correctly includes relevant; human performance 81.7%) and 97.9% specificity (correctly excludes irrelevant, human performance 98.1%)
  - Many established assistive tools like Rayyan, Covidence
- Automation vs Assistance
- ASReview lab
  - Upload reference library
  - Titles/abstracts
  - Assistant rather than automation
- TERA
  - Another one of the functionalities provided
- Elicit as an example of automation
  - Systematic Review functionality is automated (but subscription based) and is beta, admitting to 70-80% accuracy

Demo #6

Data extraction

- Discussion:
  - Efficacy of large language models (common model studied GPT4) at 72-100% of that of humans performing data extraction.
  - Again, *otto*-SR got to 93.1% accuracy compare to human (79.7%) and data extraction tool Elicit (74.8%) Cao et al., 2025.
- Notebook LM (?)
  - Doesn't always exclude/acknowledge retracted works if you provide the
  - No validation studies on the efficacy for data extraction
  - 50 paper limit, anecdotal reports of loss of efficacy around 20+ papers.
  - Potential use for
    - one-off rapid scans, drop a paper in and ask basic questions
    - "first-pass" triage for narrative/low rigour reviews, add ~20 papers, ask which include relevant concepts/inclusion criteria and follow up with reading full text
  - Insufficient for other review types, with a cap, insufficient reporting, it's closed (i.e. without API functionality), making integration into other software of a workflow harder.
  - Example of information discovery for NotebookLM validation studies or grey literature via Undermind
- Elicit
  - Upload your reference library.
  - However, quality extraction relies full text information, and uploading full texts is a bit rife in terms of copyright.

Navigating copyright, licensing & full text sharing

What articles are covered by copyright?

All articles except public domain information

Includes paywalled subscription content and open access, free-to-read content

Who owns the copyright?

Publishers

Authors

What are library licenses?

Paid access to subscription database content that is negotiated between the publisher & the library.

Determines what University members can do with the publisher’s copyrighted materials

Database usage is monitored by the publishers

Copyright laws for AI & full texts

Exceptions in the NZ Copyright Act allow copies made for private research or study as fair dealings.

Sharing articles with a closed AI is likely covered under this exception.

Closed = tool does not re-use submitted data for training or on-share to third parties. Check the tool's terms of use & privacy policy.

See Copyright at Auckland or discuss your use case with our Copyright Officer.

Library licenses & AI full texts

Whilst the University could potentially not be liable due to copyright exceptions, publishers can cancel University access to databases if you go beyond what is allowed under our licenses.

Text mining projects will need to be negotiated individually between the research group and the publisher.
e.g. IEEE *Xplore* Text and Data Mining (TDM) Options

Example of library license

IEEE license, section 3c

Restrictions. Except as expressly permitted in this Agreement, Licensee and its Authorized Users may not: **(1) download, reproduce, retain or redistribute the Licensed Products, in its entirety, or any journal or issue of a journal in the Licensed Products in any substantial or systematic manner**, including, but not limited to, accessing the Licensed Products using a robot, spider, crawler, screen scraping or similar technological device; (2) **electronically distribute, via e-mail or otherwise, any Article or eBook**; (3) abridge, modify, translate or create any derivative work based upon the Licensed Products without the prior written consent of IEEE; (4) **display or otherwise make available any part of the Licensed Products to anyone other than Authorized Users**; (5) sell, resell, rent, lease, **license, sublicense, assign or otherwise transfer any rights granted under this Agreement**, including, but not limited to, use of the Licensed Products for document delivery, fee-for-service or any other substantially similar commercial purpose; ..... (continues)

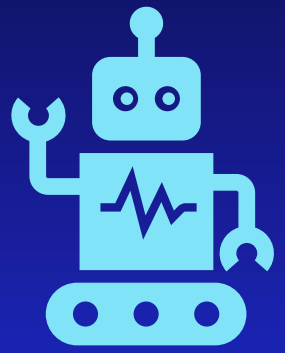
Suggestions for screening/data extraction

Limit your uploading and stay in closed systems without user data reuse.

Or, stick to public data for now with tools that support title/abstract screening & extract data manually.

Verify extracted information

Document use



Considerations around implementing AI

Evaluating your use case

Consider these questions for your use case to judge whether using AI is appropriate for your application.

- **What is the purpose of your review and is it appropriate to use AI?**
  i.e. following regulations, discipline-specific standards
- **Which specific tasks within your literature review are you considering using AI for?**
- **Is there a suitable tool for this task?**
  (Next slide discusses tool evaluation)
- **How might using AI for these tasks impact your learning and skill development as a researcher?**
  e.g., discipline knowledge, critical thinking and communication skills
- What are the potential **benefits, risks** and **costs** for using AI for this task?

Evaluating a tool

Consider these questions for a tool you might want to use to judge whether it is suitable

- Does the tool suit your **review method and level of rigour**?
- If you are doing a rigorous review, has it been **validated** for your use case?
- Is the **data source** the tool uses comprehensive, high-quality, and unbiased?
- **What kind of data are you inputting into the tool? Do you have to use university-approved tools?** Check out the Data Classification Standard and the Getting started with AI (staff intranet) to understand when and where to share data.
- **Check the terms of service - how are your inputs used by the tool and do you have the authority to give those permissions?** e.g. subscription article uploads aren't allowed under library license terms
- Does this tool work in a way that aligns with **institutional, publisher, and national policies**?

Evaluating an output

Consider these questions to judge whether the an AI's output is of high quality

- Is the information accurate and factually correct based on verified sources?
- Are all claims supported by credible evidence and properly cited?
- Does the output contain "hallucinations" or fabricated information? How can you verify the accuracy of outputs?
- Does it genuinely reflect your ideas and research question, or does it seem generic/biased/influenced by AI?
- How relevant are the results to your research question and inclusion criteria?
- Is the list comprehensive? Are key papers or concepts missing?
- If so, could this or other biases (e.g., in selection, ranking, or training data) make results unrepresentative, and how will you mitigate/minimise bias?
- Do summaries or data extractions accurately reflect original papers?

- **What are your next steps to build upon and verify this output?**

Summary

1) AI for literature reviews is a developing space
2) Augment rather than replace established practices
3) Be careful with sharing content; you are responsible for ethical conduct and safe data sharing practices
4) Understand what guidelines, policies and practices you need to adhere to
5) Critical evaluation of your use, tools and outputs is needed
6) Maintain transparency with documentation and acknowledgement

Questions?

More resources:

- Research AI on ResearchHub
- Literature review learning resources
  - Finding information
  - Research writing
- Contact the library via AskUs

Feedback:

https://tinyurl.com/mvaecx87