The Literature Review Seminar

Tools

- Distinguish the major approaches of setting up tools for literature reviews
- Practice the use of an open-synthesis platform (CoLRev)
- Appreciate how AI and genAI/LLM may change the conduct of literature reviews

Start the demo

Start the demo (account and login on GitHub required)

Typical setups

Overall, there are many tools for literature reviews. The systematicreviewtoolbox.com alone lists over 340 tools.

There are two major approaches:

- **Self-managed approach**: Combine multiple tools, including a reference manager, and Excel
- **Platform**: Select a platform that handles the whole workflow and use functionality or extensions that are available

> Additional overviews were published by Cicero Jimenez et al., 2022.

Self-managed approach

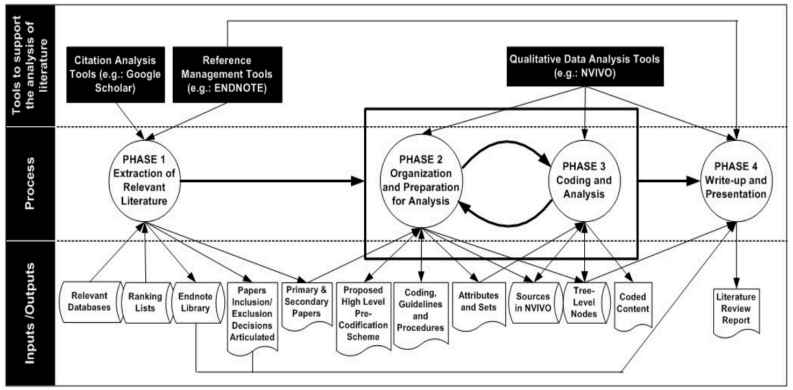
Common elements:

- **Reference manager** to import, deduplicate, screen, extract data, analyze, and cite search results (e.g., Zotero, Endnote, Citavi, Mendeley, Jabref)
- **Excel** can be used for the screen, data extraction, and analysis
- **Specialized tools** for individual steps (see next slide)
- **Word processor** for write-up

Self-managed approach: Reference manager

Key considerations:

- Keep a **separate copy of search results** (for reporting purposes)
- **Deduplication** is often inefficient → many **manual decisions** required
- Exporting to other tools or Excel requires care
  - Track **record IDs**
  - Use **tags/flags** for exported entries
- It is possible to complete most steps (including the pre/screening) **within** the reference manager
- Often, **one team member** takes ownership of data management tasks



See Bandara et al. 2015

Self-managed approach: Tools

Leading automation tools for literature reviews (Wagner et al. 2021):

Step	Research Tools
Search	*LitSonar*: Supports search query translation.
	litsearchr: Supports search strategy development.
	connectedpapers, inciteful: Support citation searches.
	TheoryOn: Supports construct searches.
Screen	*ASReview*: AI-supported screening (see intro).
Quality Assessment	*Robot Reviewer*: AI-supported quality appraisal (see intro).
Data Analysis	*Obsidian*: A tool for knowledge management and data extraction.
	RevMan: A tool to conduct meta-analyses.

> Note: We currently work on search-query for query validation, translation, and improvement.

Self-managed approach

Advantages:

- Low cost and quick setup
- Relatively high flexibility to use different tools and pursue different goals (review types)

Disadvantages:

- Data is handled manually: assigning IDs, sharing PDFs, keeping track of the status of records, data conversion, manual import and export
- Error-prone, especially when using Excel (see 1, 2)
- Individual tools may have limited compatibility
- Working in a team requires explicit and careful coordination
- Updating searches is challenging

Platforms

Aspect	CoLRev	LitStudy	BUHOS	Covidence
Review types	✓	✗	✗	✗
Supported steps	All steps	Partial	All steps	All steps
Automation/algorithms	✓	✓	●	●
Extensibility	High (102 extensions)	None	None	None
Search capabilities	APIs, updates	Limited	Limited	Manual only
Collaboration	Large teams supported	Limited	Limited	Limited
Transparency & validation	✓	✗	✗	✗
License	OSI	OSI	OSI	Proprietary
Technology	Python / CLI & API	Python / Jupyter	Ruby / Web UI	Web UI / SaaS
Development activity	commits 4.6k	commits 351	commits 845	NA

Platforms: CoLRev and open synthesis

We envision an open and extensible research platform supporting different types of literature reviews. Our focus is on the following aspects:

- Shared data structures and processes
- Open-Source license and extensibility through packages
- Transparent data management, enabling the collaboration of reviewers and algorithms, including Artificial Intelligence and Generative Artificial Intelligence
- Self-explanatory workflow

CoLRev
COLLABORATIVE LITERATURE REVIEWS

> Disclaimer: I am the lead developer of CoLRev.

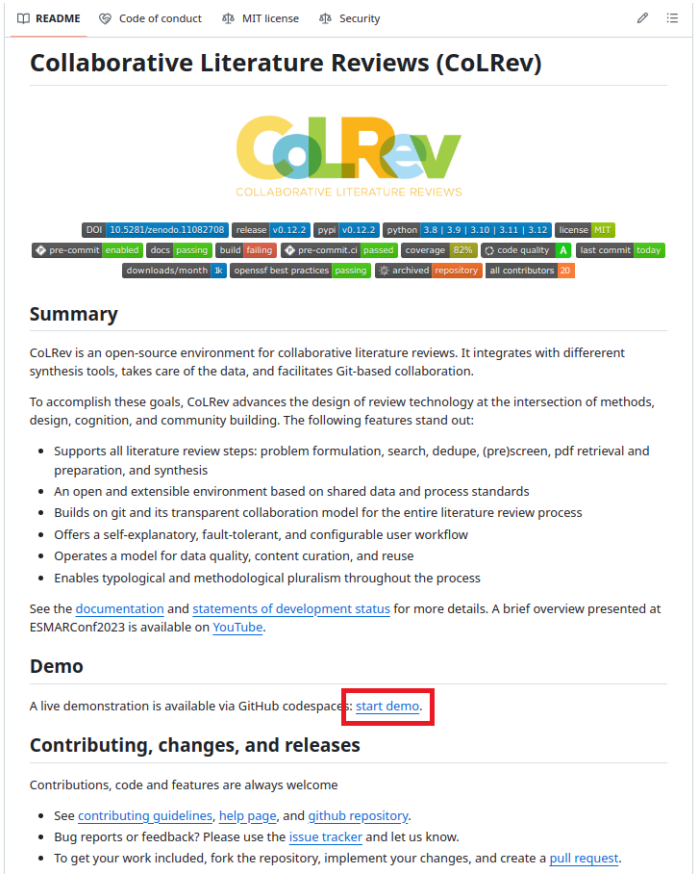
Platforms: CoLRev and open synthesis

- An open platform supporting all steps (see table below and demo in the documentation)
- Based on Git for data versioning and collaboration
- Extensible, offering different packages, e.g., packages for different types of reviews (not just "systematic reviews")

Step	Operations
Problem formulation	`colrev init`
Metadata retrieval	`colrev search`, `colrev load`, `colrev prep`, `colrev dedupe`
Metadata prescreen	`colrev prescreen`
PDF retrieval	`colrev pdfs`
PDF screen	`colrev screen`
Data extraction and synthesis	`colrev data`

Platforms: CoLRev and open synthesis

- Start the demo (account and login on GitHub required)
- Form small groups of 2-3 people
- Open the tutorial worksheet and complete it
- Consult the documentation whenever necessary



README Code of conduct MIT license Security

Collaborative Literature Reviews (CoLRev)

CoLRev
COLLABORATIVE LITERATURE REVIEWS

DOI 10.5281/zenodo.11082708 release v0.12.2 pypi v0.12.2 python 3.8 | 3.9 | 3.10 | 3.11 | 3.12 license MIT
pre-commit enabled docs passing build failing pre-commit.ci passed coverage 82% code quality A last commit today
downloads/month 1k opensssf best practices passing archived repository all contributors 20

Summary

CoLRev is an open-source environment for collaborative literature reviews. It integrates with differerent synthesis tools, takes care of the data, and facilitates Git-based collaboration.

To accomplish these goals, CoLRev advances the design of review technology at the intersection of methods, design, cognition, and community building. The following features stand out:

- Supports all literature review steps: problem formulation, search, dedupe, (pre)screen, pdf retrieval and preparation, and synthesis
- An open and extensible environment based on shared data and process standards
- Builds on git and its transparent collaboration model for the entire literature review process
- Offers a self-explanatory, fault-tolerant, and configurable user workflow
- Operates a model for data quality, content curation, and reuse
- Enables typological and methodological pluralism throughout the process

See the documentation and statements of development status for more details. A brief overview presented at ESMARConf2023 is available on YouTube.

Demo

A live demonstration is available via GitHub codespaces: start demo.

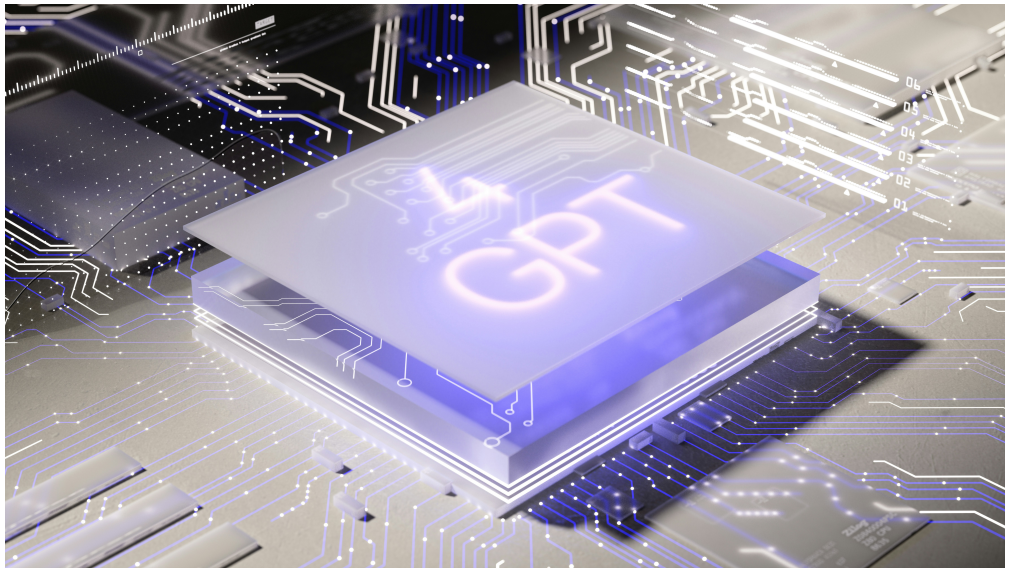
Contributing, changes, and releases

Contributions, code and features are always welcome

- See contributing guidelines, help page, and github repository.
- Bug reports or feedback? Please use the issue tracker and let us know.
- To get your work included, fork the repository, implement your changes, and create a pull request.

AI, genAI and the future(s) of literature reviews

Question: How would you use genAI-tools in a literature review?



Based on Wagner et al. 2021 and follow-up work (currently under review).

Abstract

Introduction

With the increasing accessibility of tools such as ChatGPT, Copilot, DeepSeek, Dall-E, and Gemini, generative artificial intelligence (GenAI) has been poised as a potential, research timesaving tool, especially for synthesising evidence. Our objective was to determine whether GenAI can assist with evidence synthesis by assessing its performance using its accuracy, error rates, and time savings compared to the traditional expert-driven approach.

Methods

To systematically review the evidence, we searched five databases on 17 January 2025, synthesised outcomes reporting on the accuracy, error rates, or time taken, and appraised the risk-of-bias using a modified version of QUADAS-2.

Results

We identified 3,071 unique records, 19 of which were included in our review. Most studies had a high or unclear risk-of-bias in Domain 1A: review selection, Domain 2A: GenAI conduct, and Domain 1B: applicability of results. When used for (1) searching GenAI missed 68% to 96% (median = 91%) of studies, (2) screening made incorrect inclusion decisions ranging from 0% to 29% (median = 10%); and incorrect exclusion decisions ranging from 1% to 83% (median = 28%), (3) incorrect data extractions ranging from 4% to 31% (median = 14%), (4) incorrect risk-of-bias assessments ranging from 10% to 56% (median = 27%).

ork/

LLMs, current challenges, and promises

Status quo: "Directly asking ChatGPT for research summaries does not produce compelling results"

- Language vs. knowledge and the problem of hallucination (fictitious references)
- Retrieval-augmented generation (APIs) as a potential remedy (e.g., Consensus)
- LLMs do not necessarily have access to paywalled research
- Need for human oversight, researchers need to understand nuances of review types, methods, and steps

Thought-provoking paper: Chen and Chan (2024) analyze to which degree **experts** and **novices** benefit from the use of LLMs in **ghostwriting** vs. **sounding board** modes.

- Using LLMs in ghostwriting mode was generally detrimental to the outcomes
- Using LLMs in sounding board mode was more effective (especially for non-experts)

💬 **Question**: How could LLMs be used in sounding board mode for a standalone literature review?

Which developments can be anticipated?

Review types

- Descriptive reviews may be the first to become obsolete given the summarizing capabilities of LLM
- For testing reviews, LLM can support different steps, including the generation of code for the analysis
- For reviews aimed at understanding or explaining, there may be different futures

Steps of the process

- LLM capabilities, or corresponding tools like litmaps, are particularly helpful for exploratory activities
- Language handling capabilities are useful for the design of queries in the systematic search phase (need to group synonyms)
- In the screen, restrictions of human cognitive capacities are one of the prime reasons to screen most of the papers based on the metadata (instead of the full-text). This could change with LLM, which can process full-text documents efficiently (possibly as part of the prescreen).
- Applications of LLM in the later steps have yet to be explored

Prompt example: Search query formulation

Best prompt identified by Wang et al. (2023):

You are an information specialist who develops Boolean queries for systematic reviews. You have extensive experience developing highly effective queries for searching the information systems literature. Your specialty is developing queries that retrieve as few irrelevant documents as possible and retrieve all relevant documents for your information needs. You are able to take an information need such as: “Review of IT Business Value” and generate valid Web of Science queries such as:
“TI=(IT OR IS OR …) AND TI=(value OR payoff OR …) AND TI=(firm OR business OR …)”.

Now you have your information needed to conduct research on “The effect of LLM on individual performance at work”, please generate a highly effective systematic review Boolean query for the information need.

⚠️ ChatGPT is useful for writing Boolean search queries in **high-precision reviews**, such as rapid reviews

Prompt example: Screen

Best prompt identified by Syriani et al. (2023):

Context: I am screening papers for a systematic literature review. The topic of the systematic review is the effect of generative AI on individual productivity for programmers. The study should focus exclusively on this topic.
Instruction: Decide if the article should be included or excluded from the systematic review. I give the title and abstract of the article as input. Only answer include or exclude. Be lenient. I prefer including papers by mistake rather than excluding them by mistake.

Task i:
• Title: “Twelve tips to leverage AI for efficient and effective medical question generation”
• Abstract: “Crafting quality assessment questions in medical education […]”

⚠️ Performance of LLM-based screening varies considerably across datasets, indicating **limited generalizability**
⚠️ The findings show that LLMs does not consistently perform better than random classification (in terms of recall)

Summary

- Carefully assemble your toolkit by considering the
  - Fit with the type of review
  - Need for collaboration in a team
  - Compatibility between tools (effort for data management and conversion)
- Consider open-synthesis platforms such as CoLRev
- Understand how AI and genAI/LLM may facilitate or change the process (especially in *sounding board* mode)

Thank you!

- Thank you for participating in the seminar
- If you would like to get feedback on a literature review, schedule a meeting
- Keep in mind: If you work on literature reviews, there are opportunities to reconnect!
- Help us spread the word to other students

References

Cierco Jimenez, R., Lee, T., Rosillo, N., Cordova, R., Cree, I. A., Gonzalez, A., & Indave Ruiz, B. I. (2022). Machine learning computational tools to assist the performance of systematic reviews: A mapping review. *BMC Medical Research Methodology*, 22(1), 322. doi:10.1186/s12874-022-01805-4

Clark, J., Barton, B., Albarqouni, L., Byambasuren, O., Jowsey, T., Keogh, J., ... & Jones, M. (2025). Generative artificial intelligence use in evidence synthesis: A systematic review. *Research Synthesis Methods*, 1-19. doi:10.1017/rsm.2025.16.

Bandara, W., Furtmueller, E., Gorbacheva, E., Miskon, S., & Beekhuyzen, J. (2015). Achieving rigor in literature reviews: Insights from qualitative data analysis and tool-support. *Communications of the Association for Information Systems*, 37(1), 8. doi:10.17705/1CAIS.03708

Syriani, E., David, I., and Kumar, G. 2023. “Assessing the Ability of ChatGPT to Screen Articles for Systematic Reviews,” arXiv. doi:10.48550/ARXIV.2307.06464.

Wagner, G., Lukyanenko, R., & Paré, G. (2022). Artificial intelligence and the conduct of literature reviews. *Journal of Information Technology*, 37(2), 209-226. doi:10.1177/02683962211048 20

Wang, S., Scells, H., Koopman, B., and Zuccon, G. 2023. “Can ChatGPT Write a Good Boolean Query for Systematic Review Literature Search?” in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1426–1436. doi:10.1145/3539618.3591703.