What Are We Thinking?

AI in Research and Supervision

SPEAKER

Steven Denney

Leiden University

MODERATOR

Aron van de Pol

Leiden University

21 October 2026 | 15:00–17:00

DH Lab, Huizinga 0.09

ROADMAP

This afternoon

1 What the surveys say

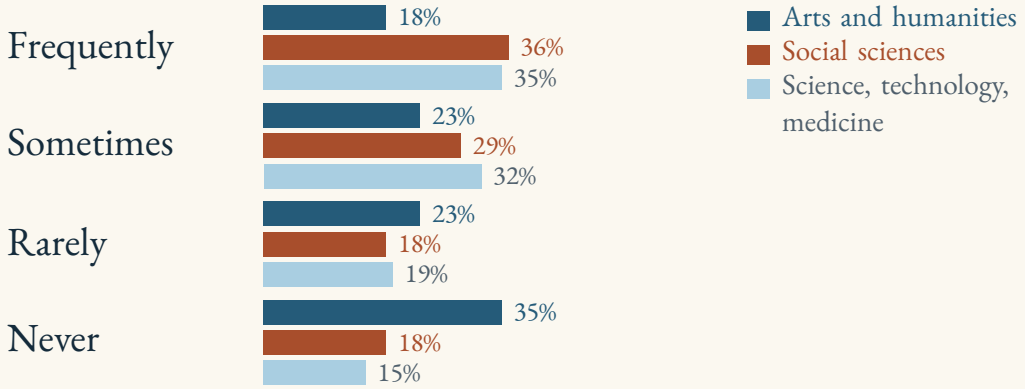
2 Automate, guard, collaborate

3 A research knowledge base

4 Discussion

How often researchers use AI

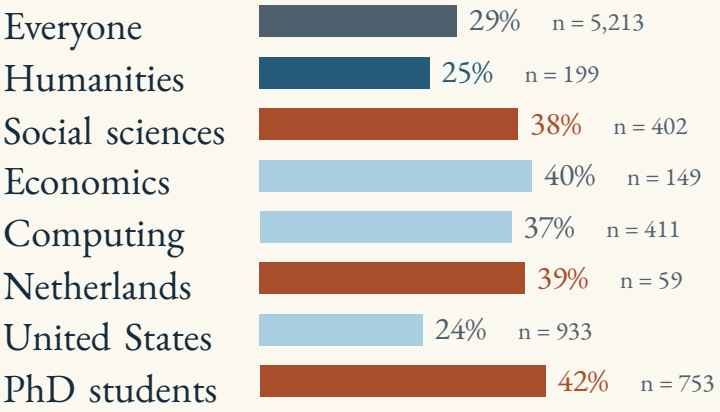
"How often do you use AI tools in your academic research?"



Online survey of 2,606 academic researchers worldwide, April 2026, weighted by region. Oxford University Press (2026), *Disclosing AI Use in Academic Research*.

Who has used AI to edit a draft

Wrote a draft article, then used generative AI to help edit it



Nature survey of 5,229 researchers worldwide, Feb–Mar 2025: authors, a publisher panel, newsletter readers. Fields could overlap. Kwon, *Nature*, 14 May 2025.

WHAT THE SURVEYS SAY

We are behind

20%

of PhD students have talked about AI with their supervisors

23%

of universities do not know how their PhD candidates use it

74%

of researchers find the rules on disclosure unclear

SGSSS (2026), 308 UK social-science PhD students. EUA-CDE (2026), 217 European universities. OUP (2026).

Five terms

LLM	A large language model. Trained on large amounts of text, it produces text by predicting what comes next.
Chatbot	An app for talking with an LLM, such as ChatGPT or Claude.
Agent	An LLM given tools and a goal. It opens files, runs code, and works in steps until the task is done.
CLI	The “terminal”: a text window where you type commands. Claude Code runs there.
Skill	Written instructions an agent follows for a task that comes up again.

THE FRAMEWORK

Automate, guard, collaborate

Three ways to work with AI

Three modes

Automate	Work we want to offload to AI.
Guard	Work we keep away from AI, to protect durable learning and core skills.
Collaborate	Work we do with AI to increase productivity, do better work, and do things we otherwise could not.

AUTOMATE

Willingly offload

Examples

Email	Read, triage, flag what is due, and summarize the inbox
Calendar	Review the schedule. “When am I free? Am I missing anything?”
Track	Tasks, milestones, and deadlines
Files	Name and sort documents, format references

My morning brief

Morning Brief · Wednesday 21 October

generated 06:30

Schedule

10:00 PhD supervision
15:00 Seminar, DH Lab

Deadlines

Thu Chapter feedback
Fri Referee report

Inbox

41 emails sorted overnight. 2 need a reply today.

ILLUSTRATION

Built each morning from my email, calendar, and task lists. The entries are invented.

GUARD

Protect core skills

What we're protecting

For ourselves

Writing, reading, and judging evidence. We keep these skills only by using them.

For our students

The development of core skills. They are learned by doing the work, and they will be scarce in an age of AI.

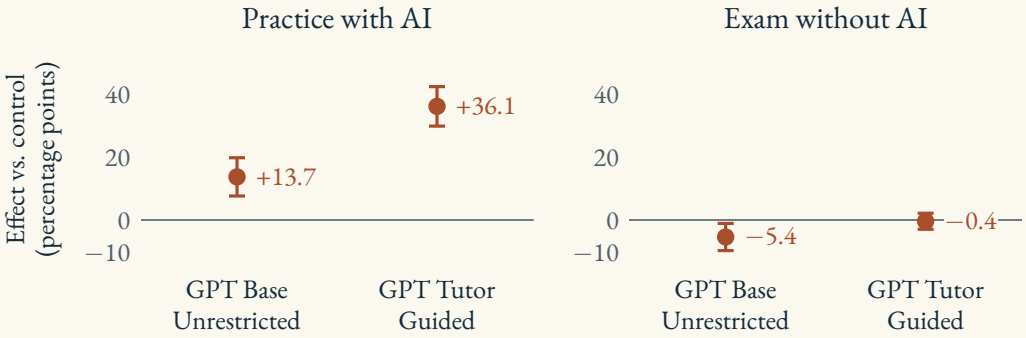
The goal is durable learning.

Offloading hurts durable learning

Study	Who and what	What they found
Melumad & Yun (2025)	10,462 adults. ChatGPT (GPT-3.5) or Google search.	Chatbot users reported learning less deeply than web searchers
Fernandes et al. (2026)	246 adults. Logic problems with ChatGPT-4o.	They thought they had done better than they had
Bassner et al. (2026)	275 students. Coding with or without GPT-4.	Higher exercise scores with AI, but no greater knowledge gain
Shen & Tamkin (2026)	52 programmers. A new code library, with GPT-4o.	Scored 17% lower on a later quiz taken without AI

Melumad & Yun: Experiment 1, self-reported depth, 3.43 vs 3.86 on a 1–5 scale. Bassner: a hint-first tutor and unrestricted ChatGPT, both on GPT-4. Shen & Tamkin: preprint.

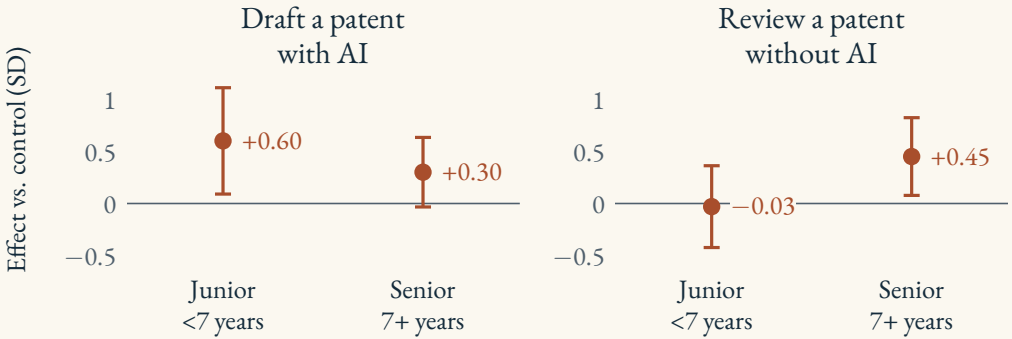
Practicing with AI and taking an exam without it



Bars show approximate 95% intervals. Bastani et al. (2025), PNAS 122(26), Table 1. Both tutors were built on OpenAI GPT-4.

Experts are in a position to gain

Patent lawyers used AI for three months. Later, they reviewed a patent without it.



Bars show approximate 95% intervals. Autor et al. (2026), NBER Working Paper 35720, Tables 4 and 6. The tool was an unreleased Google Labs writing assistant; the paper does not name its model.

COLLABORATE

Working with the machine

Why collaborate

Productivity

More done in the same time.

Quality

The same work done, but better, and with a supercomputer as fact checker.

Reach

Work we could otherwise not do: more languages, more sources, greater scale.

Where the gains are

Sources	Find, gather, and organize new material, in any language
Digitization	OCR turns scans and archives into searchable text
Scale	Text and image analysis across whole collections
New data	Collecting new types of data, with staff and students
Verification	An expert fact checker and source manager (the demo)

COLLABORATE

Working together to verify

	What the agents do	What we do
Sources	Files and converts them	Choose what belongs
Citations	Checks each reference	Confirm every flag
Claims	Finds the passages	Judge whether they support the claim

A sentence from my draft

> "Studies of unilateral policymaking show that citizens penalize executive action relative to legislative action"
>
> (Reeves & Rogowski 2016; Christenson & Kriner 2017)

In plain words: people support a policy less when a president makes it alone than when Congress passes it.

Do these two studies support the claim? Do they even exist?

Using ChatGPT

Chat

ILLUSTRATION

ME

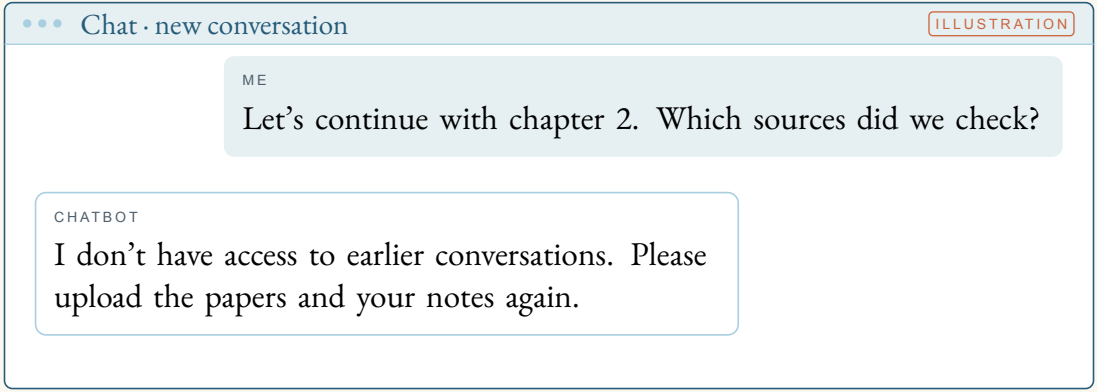
Do these two studies support this sentence?

CHATBOT

Yes. Both studies find that the public is more skeptical of executive action than of legislation, so your sentence is well supported.

No quotation. No page. Nothing to check.

The next week, starting again



Every session starts from zero.

COLLABORATION IN PRACTICE

A research knowledge base

How agents work with my research projects

WHERE THIS COMES FROM

Karpathy's LLM wiki

Sources	Put all your raw sources in one folder, and never change them.
Wiki	The model writes a summary of each source, and pages for recurring people, places, and ideas, with an index and a log.
Answers	Ask a question. The answer is written back into the wiki.
Accumulation	Repeat for the next question. Knowledge accumulates.

"The LLM is rediscovering knowledge from scratch on every question. There's no accumulation." — Karpathy (2026)

Karpathy, A. (2026), *LLM Wiki*, GitHub Gist, 4 April. Kenny (2026) builds a version for political scientists.

MY VERSION

The knowledge base idea

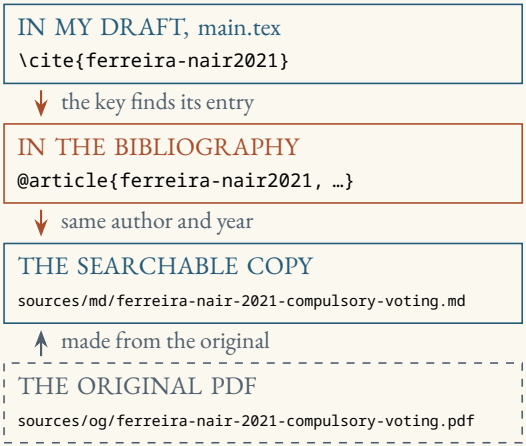
How I assemble it:

1. I keep the originals.
2. I create searchable text copies.
3. I generate bibliographic entries.
4. I bolt on additional skills.

This serves as the basis of my agent-based work.

Knowledge base scaffolding

```
research-project/
├── sources/
│   ├── og/                  untouched originals
│   ├── md/                  searchable copies
│   ├── unprocessed/         new files go here
│   ├── inbox/               notes for the wiki
│   └── references.bib       an entry per source
├── manuscript/
│   └── main.tex             my draft, in LaTeX
└── CLAUDE.md                agent instructions
```



If a link breaks, the citation cannot be checked.

How to incorporate a source in five steps




```
Download (3).pdf  →  ferreira-nair-2021-compulsory-voting.pdf

@article{ferreira-nair2021,
author = {Ferreira, Tomás and Nair, Priya},
title  = {Does Compulsory Voting Make Better Citizens?}, …}
```

SCANS NEED OCR FIRST

The original stays untouched. The agent works from the copy.

One example. Every source goes through the same steps. Sources and authors here are invented; the conversion problems are real.

Evaluating the question from before

The claim: *"Studies of unilateral policymaking show that citizens penalize executive action relative to legislative action."*

Reeves & Rogowski (2016)

"Only about a quarter of respondents in any of the surveys supported unilateral policy making"

Does not compare the two routes.

Christenson & Kriner (2017)

The route "has no significant influence on the probability of a subject backing the president"

Does not show a penalty.

Neither study shows a penalty. Unsupported as written.

I decide what to write

BEFORE

Studies of unilateral policymaking show that citizens penalize executive action relative to legislative action (Reeves & Rogowski 2016; Christenson & Kriner 2017).

AFTER

Studies of unilateral policymaking find that Americans express low generalized support for unilateral powers (Reeves & Rogowski 2016) while judging particular unilateral acts largely by whether they agree with them (Christenson & Kriner 2017).

The agent shows me the evidence. I write the sentence.

How to manage your knowledge base



Handoff note

Where I stopped, and what comes next.



Session log

What was done, in order. Added to, never rewritten.

Version history

Git records each change I commit. GitHub keeps a copy online.

When a session ends, `/oss:finished` writes the note and the log. When a new one starts, `/oss:sitrep` reads them.

I recommend Git for every project, with a private copy on GitHub.

Skills: procedures the agent follows

Four examples. There are others.

`research-repo`	Sets up the library and its rules
`process-source`	Files away new sources
`citation-check`	Does each reference exist, and match its record?
`fact-check`	Does the source say what I claim?

A skill is a written procedure the agent follows the same way every time. Open Science Skills, github.com/scdenney/open-science-skills.

Build your own

1. Install Claude Code and the Open Science Skills.
2. In an empty folder, run `/oss:research-repo` (library and Git).
3. Add a paper and run `/process-source`. Check the copy.
4. Run `/oss:citation-check` and `/oss:fact-check` on a draft.
5. End with `/oss:finished`. Start the next session with `/oss:sitrep`.

DEMO

What you are about to see

I prepared a draft about Koreans in Japan in 1946.

Sources	Kim 2021 (in Korean), Morris-Suzuki 2011 (in English), and a scanned 1946 order from SCAP, the Supreme Commander for the Allied Powers.
My draft	Cites all three. I planted four errors: a reversed claim, a mistranslation, an unsupported claim, a fake reference.
The agent	Files the sources into the knowledge base, then checks every reference and every cited claim against them.
Watch for	The scanned page with no text layer, the Korean text, and whether it catches all four.

The same setup serves any research that rests on sources you hold copies of: archives, scans, several languages.

DEMO

Live demonstration

scdenney.github.io/ai-for-research/archive-check

DISCUSSION

Your answers

What do you think?

What did you answer?

TO TRY

Resources

Demo	scdenney.github.io/ai-for-research/archive-check
Walk-through	scdenney.github.io/ai-for-research/knowledge-base
Skills	github.com/scdenney/open-science-skills
Writing	pixelsandpatterns.substack.com

APPENDIX

References

Surveys

AAC&U and Elon University (2026). *The AI Challenge*. 1,057 US faculty, October–November 2025.

Elsevier (2025). *Researcher of the Future*. 3,234 researchers in 113 countries, August–September 2025.

EUA Council for Doctoral Education (2026). *Policies in Doctoral Education*, survey report II. 217 universities in 37 countries.

Kwon, D. (2025). Is it OK for AI to write science papers? Nature survey shows researchers are split. *Nature*, 14 May. doi:10.1038/d41586-025-01463-8. 5,229 researchers. Groups computed from the published data.

Nature (2026). AI and the PhD student: friend or foe? *Nature* 651, 19 March. Nature Graduate Survey 2025, about 3,800 PhD students.

Oxford University Press (2026). *Disclosing AI Use in Academic Research*. 2,606 researchers, April 2026.

SGSSS (2026). *Generative AI and the Social Science PhD*. 308 UK PhD students, May–November 2025.

Wiley (2025). *ExplanAItions 2025*, press release. 2,430 researchers, August 2025.

Learning and judgment

Autor, D. H., Rodchenko, T., Martin, J., Iscenko, Z., Strand, S., Pearl, D., & Ferere, M. (2026). *Does AI Assistance Enhance or Erode Expertise? Evidence from a Three-Month Field Experiment in Patent Drafting*. NBER Working Paper 35720. doi:10.3386/w35720.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *PNAS*, 122(26), e2422633122. doi:10.1073/pnas.2422633122.

Bassner, P., Lenk-Ostendorf, B., Beinstingel, R., Wasner, T., & Krusche, S. (2026). Less stress, better scores, same learning: The dissociation of performance and learning in AI-supported programming education. *Computers and Education: Artificial Intelligence*, 10, 100537. doi:10.1016/j.caeai.2025.100537.

Fernandes, D., Villa, S., Nicholls, S., Haavisto, O., Buschek, D., Schmidt, A., Kosch, T., Shen, C., & Welsch, R. (2026). AI makes you smarter but none the wiser: The disconnect between performance and metacognition. *Computers in Human Behavior*, 175, 108779. doi:10.1016/j.chb.2025.108779.

Melumad, S., & Yun, J. H. (2025). Experimental evidence of the effects of large language models versus web search on depth of learning. *PNAS Nexus*, 4(10), pgaf316. doi:10.1093/pnasnexus/pgaf316.

Shen, J. H., & Tamkin, A. (2026). How AI impacts skill formation. Preprint, arXiv:2601.20245.

The example and the knowledge base

Reeves, A., & Rogowski, J. C. (2016). Unilateral powers, public opinion, and the presidency. *The Journal of Politics*, 78(1), 137–151. doi:10.1086/683433.

Christenson, D. P., & Kriner, D. L. (2017). Constitutional qualms or politics as usual? The factors shaping public support for unilateral action. *American Journal of Political Science*, 61(2), 335–349. doi:10.1111/ajps.12262.

Denney, S. (2026). *Governing Immigration by the Rules*. Archived manuscript draft.

Karpathy, A. (2026). *LLM Wiki*. GitHub Gist, 4 April.

Kenny, C. T. (2026). *Agentic AI for Political Science Research*. Center for the Study of Democratic Politics, Princeton University, 10 September. Code: csdp-llm-wiki.

Denney, S. (2026). *Build a Knowledge Base*. AI for Research. A walk-through with practice files.

Denney, S. (2026). *Open Science Skills*. Research procedures for Claude Code and Codex.

OpenDataLoader PDF. github.com/opendataloader-project/opendataloader-pdf. The PDF converter used at intake.