

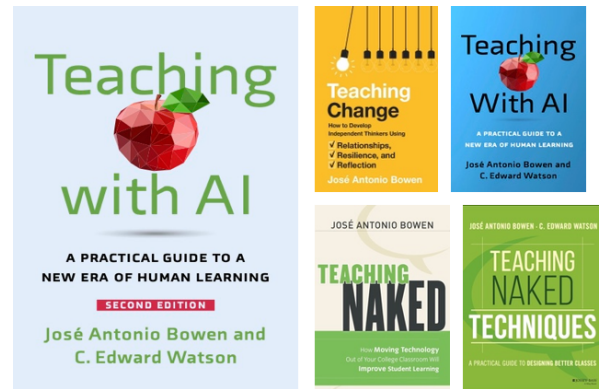
# RESEARCH with AI

## SLIDES & CITATIONS

José Antonio Bowen

Get the [2nd Edition of Teaching with AI](#).  
Discount code HTAI25 = 30% off at JHUP.  
WebSite: <https://WeTeachWithAI.com/>

► [Workshop Prompts & Materials](#)



### DOING & THINKING

AI as the new AVERAGE

COMMUNICATION & RELATIONSHIPS

CREATIVITY

AI as LABOR

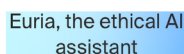
### NEW CAPABILITIES

Searching for Ideas

Expanding Perspective

Persuasion & Emotion

Customization



SEA-LION.AI  
Southeast Asian Languages in One Network



VULAVULA powered by LELAPA AI

Talkie  
(an LM from 1930)

Different AI models excel at different things. Open a fresh browser window and compare them; each tool opens in a new tab.

► [A range of AI models — WeTeachWithAI.com/models](#)

Claude (and probably all of the major models) respond very differently in different languages.

► <https://www.anthropic.com/research/claude-values-models-languages>

## WeTeachWithAI.com/models



### Searching for IDEAS

- Find anything like this..
- Has anything similar ever happened to a previous patient?
- Search for patents related to Y?
- What might be missing from this?
- Read our book and identify all of the places where we repeated an idea.

## Reasoning + Search (was "Deep Research")

- YOU are a new customer/student
- GO to our web pages
- TEST them like a naive user hoping to...
- REPEAT for our competitors' sites
- ANALYZE your findings in a brief report
- HIGHLIGHT the difficulties
- HOW might we make our website better?



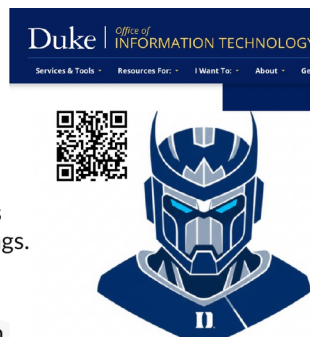
## Complete Prompt:

You are a high school senior hoping to apply to college.

Go to the University of X web page and test it like a naive user hoping to find out about majors and how to apply. Then go to ten other competitor universities and do the same thing. Collect your findings in a brief report that highlights the difficulties and how we might make using the University of X website better for new students.

## Devil's Advocate

- Reads citations and searches for newer or better papers to cite.
- Scans for 150 forbidden DEI words (like "orientation") and other red flags.
- Improves alignment with grant.
- Searches for little things so you can focus on the big things.

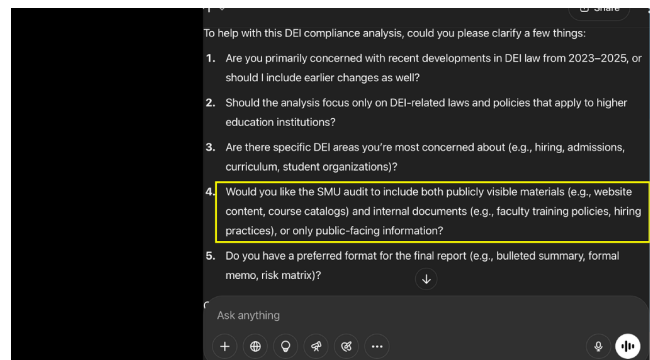


Devil's Advocate reads citations and searches for newer or better papers, scans for red flags, and improves alignment with the grant.

► [Devil's Advocate \(Duke OIT\)](#)

## Reasoning + Search (was "Deep Research")

- ANALYZE the new regulations around...
- SUMMARIZE the most important changes...
- SEARCH our web pages, courses and policies
- WRITE a report listing potential problems
- SUGGEST how we might remedy them
- CREATE an infographic based on this report
- USE our style guide and colors.



Example: a reasoning + search run on institutional DEI compliance. It begins by asking clarifying questions.

## II. Analysis of SMU's DEI-Related Policies, Courses, and Practices and Recommendations

### 1. Current SMU Initiatives

Based on an analysis of SMU's public-facing websites and related materials:

- **Office of Diversity and Inclusion / Office of Cultural Intelligence:**  
SMU has proactively rebranded its "Office of Diversity and Inclusion" to the "Office of Cultural Intelligence." This renaming effort appears designed to both reflect a broader educational mission and mitigate potential political backlash. However, the underlying functions—such as providing training on cultural competency and supporting inclusive practices—remain aligned with what many federal and state actors now label as DEI.

(See [smu.edu](#) for the Office of Cultural Intelligence.)

### • Curricular Requirements – Human Diversity:

SMU's common curriculum includes a "Human Diversity" requirement intended to help students understand cultural differences and structural inequalities. Recently, SMU paused this requirement for certain graduating cohorts following a U.S. Department of Education letter warning against "illegal" DEI programs that might jeopardize federal funding.





Shopping is the 27th most popular use of AI in 2026.

Zao-Sanders, M. (2026, June 1). How people are really using AI in 2026. Harvard Business Review.

<https://hbr.org/2026/06/how-people-are-really-using-ai-in-2026>

## AI Shopping is Here



- ✓ **46%** already made a major purchase or important decision based primarily on AI-generated information.
- ✓ **12%** trust the AI answer outright
- ✓ **55%** say top concern about AI-provided brand information is that it may be incorrect
- ✓ **17%** have switched brands because of AI-generated info

Skyword. (2026, June). AI-driven buyer insights. [vendor survey; n=1,000 U.S. adults, fielded by Dynata, April 2026]

<https://info.skyword.com/ai-buyer-research>

## Complete Prompt

A system (or general) prompt is a set of instructions for how you ALWAYS want your bot to respond. More templates on the website.



## SYSTEM PROMPTS

- Be aggressive in surfacing my assumptions
- Think globally: be vigilant about Western, gender, racial and other bias.
- Focus on accuracy. Check all of your sources and show citations, links and references. Always search in a variety of languages.
- Evaluate all information. Provide a probability score for facts and data that let me know your confidence level in the accuracy of the information. Doublecheck everything.
- Help me find unique insights and create new ideas. Provide alternative explanations and push me to think differently.

Explain concepts to me at the level of a college professor and always answer as an expert in the subject. State directly when there is insufficient data or evidence and disclose the disagreements. Highlight uncertainty and distinguish verifiable evidence from opinion.

Be direct, concise and honest and tell me when you think I am wrong. Be aggressive in surfacing my assumptions and be especially vigilant about Western, gender, racial and other bias.

Always think of our work in a global context.

I am very concerned with accuracy. Check all of your sources and show scholarly citations, links and references. Always search for information in a variety of languages. Doublecheck everything. Provide a reliability or probability score so I know your confidence level. If you are guessing, say you are guessing.

Help me find unique insights and create new ideas. Provide alternative explanations and push me to think differently. Say things like “the risk in your approach is...” and then provide specific problems. If I push back, do not fold if there is good evidence.

Start with the most useful thing you can say. No warm-ups. Delete empty sycophantic phrases. Do not flatter me.

If useful, begin by asking me three clarifying questions, each with numbered options so I can reply with numbers only.

► [More system prompt templates](#)



AI research tools

► [All tools — WeTeachWithAI.com/tools](#)

► [Consensus](#)

► [Elicit](#)

► [Research Rabbit](#)

► [Undermind](#)

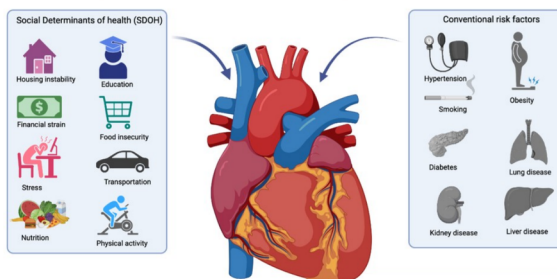
► [STORM \(Stanford\)](#)

► [Edison Scientific — an autonomous “AI scientist” for long, multi-step literature analyses](#)

## AI CAN DO (SOME) THINGS HUMANS CAN'T

### ECGs and Surveys from 280,323 patients

Mayo Clinic  
Dec 2025



ECGs and surveys from 280,323 patients, Mayo Clinic, December 2025.

Rajai, N., Medina-Inojosa, B. J., et al. (2025, December). Interplay of social determinants of health and traditional risk factors in predicting cardiac aging. Mayo Clinic Proceedings, 100(12), 2128–2139.  
[https://www.mayoclinicproceedings.org/article/S0025-6196\(25\)00106-5/fulltext](https://www.mayoclinicproceedings.org/article/S0025-6196(25)00106-5/fulltext)

### Worrying About Money and Food Ages the Heart Faster Than Traditional Risk Factors

### REDMOD detects 'invisible' pancreatic cancer 475 days before clinical diagnosis with standard CT scan



“AI can now identify the signature of cancer from a normal-appearing pancreas.”  
Dr. Ajit Goenka

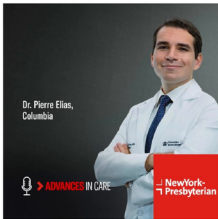
3x as accurate as human radiologists (68% vs 23%)  
2 years before clinical diagnosis.



Mukherjee, S., Antony, A., Patnam, N. G., et al. (2026, April 28). Next-generation AI for visually occult pancreatic cancer detection in a low-prevalence setting with longitudinal stability and multi-institutional generalisability. Gut. [published online first] <https://doi.org/10.1136/gutjnl-2025-337266>

## AI can do (some) things humans can't

### EchoNext



- "It detects diseases cardiologists can't."
- Trained on 1.2M Electrocardiogram (ECG/EKG) - Echocardiogram (Ultrasound) pairs.
- Consistent performance across different racial and ethnic groups.
- A (cheap) mammogram for your heart.

Poterucha, T.J., Jing, L., Ricart, R.P. et al. Detecting structural heart disease from electrocardiograms using AI. *Nature* 644, 221–230 (2025). <https://doi.org/10.1038/s41586-025-09227-0>

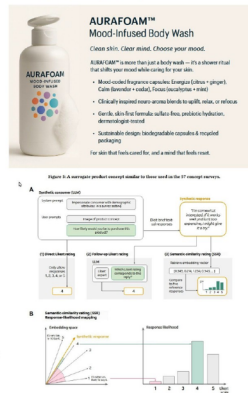
Poterucha, T. J., Jing, L., Ricart, R. P., et al. (2025). Detecting structural heart disease from electrocardiograms using AI. *Nature*, 644, 221–230. <https://doi.org/10.1038/s41586-025-09227-0>

Elias, P., & Finer, J. (2025). EchoNext: A dataset for detecting echocardiogram-confirmed structural heart disease from ECGs (v1.1.0). *PhysioNet*. <https://doi.org/10.13026/3ykd-bf14>

## 90% accuracy in predicting purchase interest

- Off-the-shelf LLM
- No fine-tuning
- Respond as a customer type X
- Only product image
- **Potential Students?!?**

Maier, B. F., Aslak, U., Fiaschi, L., et al. (2025, Oct 9) LLMs Reproduce Human Purchase Intent via Semantic Similarity Elicitation of Likert Ratings. *arXiv:2510.08338*



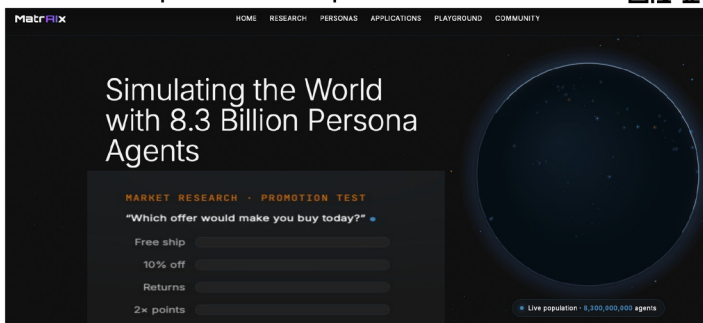
Maier, B. F., Aslak, U., Fiaschi, L., et al. (2025, Oct 9). LLMs reproduce human purchase intent via semantic similarity elicitation of Likert ratings. *arXiv [preprint]*. <https://arxiv.org/abs/2510.08338>

**MatrAlx:** Harvard and MIT researchers used real human profiles from public records.



► [MatrAlx](#)

Li, X., Hao, Y., Hou, J., et al. (2026, August 4). MatrAlx: Simulating the world with 8.3 billion persona agents. *arXiv [preprint]*. <https://arxiv.org/abs/2608.04205>



## Expand Perspectives

## Focus Groups & Empathy Interviews

- I am trying to gain a richer understanding.
- Respond as a trusting and honest Y.
- Deepen my knowledge.
- Question my assumptions.
- Build my empathy.

## Focus Groups & Empathy Interviews (slide 16)

### Complete Prompt:

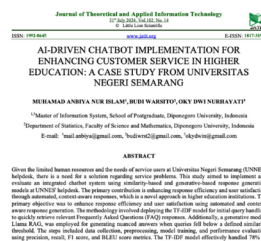
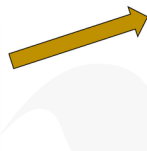
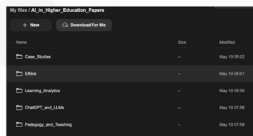
I am trying to gain a richer understanding.  
Respond as a trusting and honest Y. Deepen my knowledge. Question my assumptions.  
Build my empathy.

# AGENTS & AGENTIC BROWSERS

## AGENTS & BROWSERS

Research and find all of the academic papers on AI in Education in the last year and then find pdfs of every paper you can and download them into organized files.

- Massive Multi-Contextual Search
- Find pdf
- Download pdfs
- Organize
- Create a link to the files



A chatbot can only tell you things. An agent can do things: read your folder, run code on your data, open a database you are logged into, write a file, and come back with what it could not do.

## ChatBot



## Desktop Agent

Add AI Research Workflow to the prompting menu on my website after Vibe Coding.



Confirmed live. AI Research Workflow is now in Menu 1, inside the Prompting submenu, positioned directly after Vibe Coding. Linking to <https://weteachwithai.com/ai-research-workflow/>

## "AI has eliminated the doing and left the thinking"

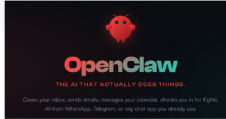
Dustin Bruzenak, CEO, Modern Logic



- Update all links on my website (Mondays)
- Check for newer or contradictory research
- Turn ppt into pdf and add citations
- Email students who have unfinished work
- Collect all records into a database

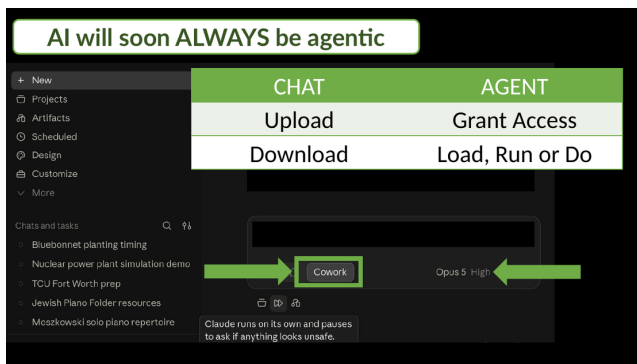
# AGENTS & BROWSERS

- Microsoft Edge Copilot Mode – Actions
- Comet (Perplexity)
- Genspark.ai, Manus.im
- Gemini Canvas, Claude Cowork, OpenAI Codex
- Atlas (req ChatGPT+) Dia, Opera, Gemini in Chrome
- Lindy, Zapier



Agentic browsers and automation tools shown on this slide:

- ▶ [Comet \(Perplexity\)](#)
- ▶ [Genspark](#)
- ▶ [Manus](#)
- ▶ [Claude Cowork](#)
- ▶ [OpenAI Codex](#)
- ▶ [Microsoft Edge Copilot Mode](#)
- ▶ [Zapier](#)
- ▶ [Lindy](#)



## AI RESEARCH WORKFLOW

### AI RESEARCH WORKFLOW

- 1 MOVE TO DESKTOP
- 2 LITERATURE REVIEW
- 3 HARVEST DATA
- 4 BS TESTING
- 5 ORGANIZE & UPDATE
- 6 FEEDBACK

### STAGE 1 · MOVE TO THE DESKTOP (slide 23)

Ten minutes. Once.

- ▶ [Claude Cowork — get started](#)
- ▶ [ChatGPT Codex](#)

Six stages. Stage 1 is a one-time, ten-minute set-up; the rest are prompts you reuse.

- ▶ [Copy, paste and modify all of these prompts](#)

### STAGE 1 MOVE TO THE DESKTOP

Ten minutes. Once.

Install a desktop agent

Create a Project

Connect a folder

Connect a browser  
Login to Databases

Claude Cowork



ChatGPT Codex

365 Copilot App  
Toggle to Cowork

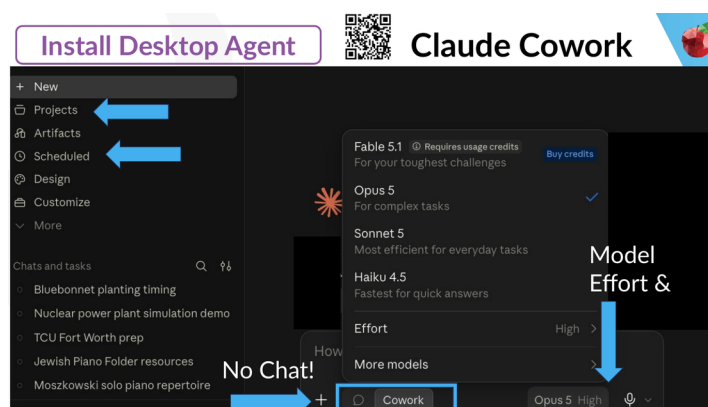


Gemini Spark

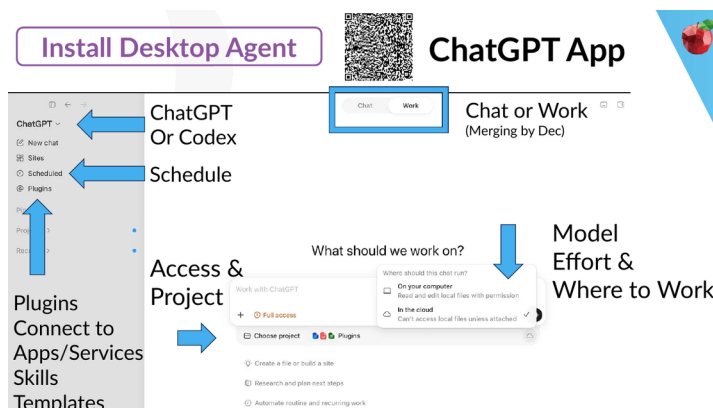
► [Microsoft 365 Copilot app \(then toggle to Cowork\)](#)

► [Gemini Spark](#)

1. Install a desktop agent — Claude Cowork, ChatGPT Codex, the Microsoft 365 Copilot app (toggle to Cowork), or Gemini Spark.
2. Make a Project per research project. A Project is a persistent container: instructions, files and memory that carry across sessions. Put your standing rules in it — evidence standards, citation style, what counts as a source in your field. Write these once and things start to happen automatically.
3. Connect a folder. Give the agent a real folder on your computer. Unless you are building a database with the AI, keep raw data or original files in one subfolder and let the agent write only to another. I have never had a problem, but you can add a rule like “always create a new file” to your project instructions.
4. Connect a browser. A browser extension (Claude in Chrome) or an agentic browser (Comet, Edge Copilot Mode) lets the agent work inside databases you are already logged into — library proxies, digitized archives, catalogues that block ordinary scrapers.



► [Claude Cowork — get started](#)



► [ChatGPT desktop app](#)

► [ChatGPT Codex](#)

## STAGE 2

## LITERATURE REVIEW

### Template 1

Create a research report that will... Make sure to examine... and include an analysis of... Begin with a critical review of ... and then...

Sources & Scope

Purpose & Framework

Audience



Elicit



Undermind

Edison Scientific

### Template 2

Literature Search

Map the Landscape

Group clusters of shared assumptions

List papers contradicted by others

Big Idea Lineage

List central claims and challengers

Mine the Gaps

Identify unanswered research questions

Summarize

What is known, believed and unknown?

Getting Started

List papers I should read first



## STAGE 2 · LITERATURE REVIEW

You now have as many research assistants as you want, but just like humans, you will need to guide them. You can do a lot with a single detailed prompt.

► [Copy, paste and modify a Deep Research template](#)

### Research Review Template 1

Create a research report that will illuminate/examine/explore X. Make sure to examine the questions A, B, and C and include an analysis of D & E. You should begin with a critical review of literature/practice/web and then provide a synthesis of the key ideas/controversies/concepts/case studies and a recommendation.

Sources & Scope: The research should draw from fields F & G; methodology H; focus on peer-reviewed journal articles/best practices/reputable studies/institutional sources; look for sector/Western/political/educational/gender bias in sources; seek global sources in language/culture I.

Purpose & Framework: Use K as a framework for understanding these issues. Focus on real-world applications and capabilities. Pay special attention to policy implications and government uses. Note any potential for L.

Audience: Write for an audience of M / for journal N or submission to conference O. Describe your findings with relevance to P.

### Research Review Template 2 — an iterative sequence

An iterative sequence can sometimes be better than one long prompt. Pasting each prompt in twice (before you hit return) also seems to improve results.

#### Literature Search

You can often use the prompt above or one of the API tools to do this within the Semantic Scholar database of published academic papers. Providing the actual papers (or links) improves the quality of what comes next.

#### Map the Landscape

Organize this list of papers. Group them into clusters of shared assumptions, claims, methodologies and/or data sets. Create a table. List the papers in column 1 and then list the core claim (in 50 words or less) in column 2. In column 3 list the key methodology or assumption that guides this paper. In column 4 list all of the ideas in that paper that are contradicted by other papers and cite those papers.

#### Big Idea Lineage

List the central claims and/or the most contentious issues or methods in this literature. First create a table that includes: (a) by whom and in what paper the idea was introduced, (b) who are/were the primary challengers, (c) summarize the positions on either side, (d) explain why they disagree and (e) tell me if there is now any consensus. Then also create a structured knowledge map or a family tree of this literature that shows how these ideas have interacted.

## Mine the Gaps

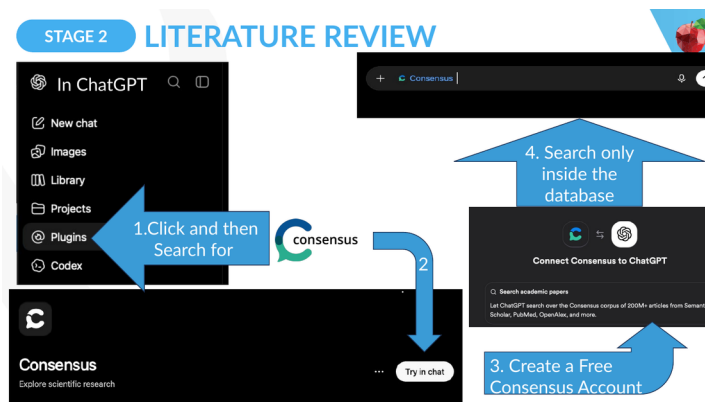
Based on all of these papers and this analysis identify 10 big research questions that are still unanswered. Describe the gaps and why they exist. Cite the papers which have come closest. What assumptions do most of these papers share, but do not explicitly justify or test? State these assumptions and cite a few of the important papers that rely on them most. What useful data or method is most underused?

## Summarize

Briefly summarize in less than 500 words what the field believes collectively, what is proven beyond a reasonable doubt, what remains contested and what is the single most important unanswered question. What would happen to the field if its most important assumption turned out to be wrong?

## Getting Started

Explain all of this in 300 words to a non-expert without jargon. Summarize what is known, what is unknown and where this matters in the real world. List the 3 most important papers I should read first to get a grip on this field.



Or search only inside a curated database: add the Consensus plugin to ChatGPT and create a free Consensus account.

► [Consensus](#)

## VERIFY

Fabricated citations in published papers

| 2023       | 2025     | 2026     |
|------------|----------|----------|
| 1 in 2,828 | 1 in 458 | 1 in 277 |

The audit was AI-assisted.

Assume the humanities rate is higher and harder to detect.

Topaz et al. (2026). The Lancet. - Orrall (2026). Retraction Watch.

Fabricated citations are measurably increasing in the published record.

Topaz, M., et al. (2026, May 7). Fabricated citations: an audit across 2.5 million biomedical papers. The Lancet.

[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(26\)00603-3/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(26)00603-3/fulltext)

Orrall, A. (2026, May 7). One in 277 PubMed-indexed papers in 2026 shows fabricated references, says analysis. Retraction Watch. <https://retractionwatch.com/2026/05/07/one-in-277-pubmed-indexed-papers-in-2026-shows-fabricated-references-says-analysis/>

**LIMITS.** The audit was itself AI-assisted. Science citations are easier to verify, so assume the humanities rate is higher and harder to detect. A third-party tool or a good search prompt can dramatically reduce fabricated citations, but a human still has to verify. The Stage 4 checks below are not optional.

## STAGE 3

## HARVEST DATA

## Leave it EMPTY.

Never infer. Never fill from general knowledge.

Name the  
archives  
yourself

URL + date for  
every field

Flag duplicates  
& never  
merge

Notes of what  
it could not  
verify

Then run the skew check before you use any of it.

## STAGE 3 · HARVEST DATA

AI agents are good at repetitive tasks but they are literal, and you need to give them clear rules. Start a new chat and first have a conversation about what is available: “My goal is X. I would like to build a database where I can do Y. Make a complete list of available sources and anticipate problems and where you might face difficulty gaining access to the best available data.” Push back and solve problems before you harvest anything.

## Harvest Prompt Template

HARVEST [items] from [named archives, catalogues, databases]. For each item record [fields A–D].

RULES: one row per item; record the source URL and the date you retrieved it for every field; if a field is unknown, leave it empty — never infer, never fill from general knowledge; flag duplicates and identity collisions (e.g. same name, different person) rather than merging them; list what you found and excluded, and why. OUTPUT a CSV plus a short memo of everything you could not verify.

## Duplicate and Check

FABLE writes the harvest handoff



SONNET harvests one small batch



FABLE checks it, writes the next handoff

Small batches are more accurate.

Small batches are more accurate. I asked Fable to write harvest handoffs for Sonnet in batches small enough to be accurate. Fable then checked every batch and wrote the next handoff — 47 in total.

## RUN A SKEW CHECK

Report the composition of what you have collected [by language, country, decade, source etc.].

Which of these is over-represented?

Is that a finding or an artifact of what happens to be digitized, easy to access or not behind a paywall?

What should the next pass target to correct it?

## RUN A SKEW CHECK

REPORT the composition of what you have collected [by language, country, decade, source etc.]. Which of these is over-represented? Is that a finding or an artifact of what happens to be digitized, easy to access, or not behind a paywall? What should the next pass target to correct it?

LIMITS. Anticipate corpus bias. Training data drawn from Common Crawl and similar sources is heavily skewed toward

English — even when you instruct it to search globally — and toward recently digitized web text. You, the human, need to be ever vigilant.

Vargas-Parada, L. (2025, November 27). Large language models are biased — local initiatives are fighting for change. Nature. <https://www.nature.com/articles/d41586-025-03891-y>

## Example: AI Literacy Programs

- SEARCH for all AI Literacy/Fluency programs, certificates, micro-credentials or badges at higher ed institutions
- ANALYZE their features
- CATEGORIZE them into types
- CREATE a Google Sheet I can share
- DOUBLE CHECK your work for accuracy



## Example: AI Literacy Programs

Complete prompt:

I want to make a comprehensive list of AI literacy programs, microcredentials, certificates and badges offered at universities anywhere in the world. Here are some examples... Think about the categories — name of the program, name of institution, format (online, self-paced, credit), length, required or optional (and if so, required of

what group and when), does it come with a digital badge or certificate? First do the comprehensive survey and collect the information and then create a spreadsheet.

► [The result — shared Google Sheet](#)

STAGE 4

## BS TESTING

Each catches a different lie. Run them one at a time.

- |    |                         |                                                                              |
|----|-------------------------|------------------------------------------------------------------------------|
| 4a | Analysis & Skew         | Don't summarize: Describe the shape<br>3 patterns you could defend           |
| 4b | Verify Citations        | Resolve every DOI, author, title<br>Note all retractions and revisions       |
| 4c | Check Claims            | What do the sources actually say?<br>Rate every source on an evidence ladder |
| 4d | Newer, Better, Contrary | Find failed replications, publication-bias<br>Work in other languages        |
| 4e | How wrong am I?         |                                                                              |

Each check catches a different lie. Run them one at a time — they can be in the same chat, but they should be separate prompts.

### 4a • Data Analysis & Skew Check

Ask for something a summary cannot fake. Insist on an audit trail and on testing the tool against itself. ADJUST BY DISCIPLINE: which repositories are digitized and which are not? Is provenance data important? What do you do with empty fields? Are there identity collisions (same name, different person)? Are quotations being checked against the right edition? Should you start in another language? What is the sampling frame?

Review the [corpus/dataset/set of texts]. Do not summarize it.

(1) Describe the shape of this data set: what is actually being counted, what is missing, what the sampling frame really is.

(2) Give me the three patterns you would defend to a hostile referee — each with the specific rows or passages that support it, and the strongest reason each might be an artifact of the data rather than a fact about the world.

(3) Give me one pattern you noticed that you cannot explain. Rate your confidence in each and tell me which parts I must check by hand.

(4) Apply my codebook/rules/guidelines [paste it] to these [interviews / field notes / documents]. Record the exact passage for every code applied. Do not invent codes — list passages that fit no

existing code separately as candidates. Then run the whole codebook a second time and show me where the two runs disagree.

**LIMITS.** The same prompt on the same text can produce different results, so run important things twice — disagreement between runs is a reliability estimate. Log and keep your prompts, model, version and date. Set temperature to 0 where the tool allows it. Disclosing AI use in your methods section is an obligation, at least in Europe.

European Commission. (2026, May 8). Living guidelines on the responsible use of generative AI in research (revised). [https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc\\_en](https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en)

Main, J. B. (2026). Thematic analysis with open-source generative AI and machine learning: a new method for inductive qualitative codebook development. *Humanities and Social Sciences Communications*, 13, 209. <https://doi.org/10.1057/s41599-026-06508-5>

#### 4b · Verify Citations

Check every reference: resolve the DOI or permalink and confirm authors, title, journal, year and pages against the publisher record. Mark every reference (or create a new document) VERIFIED / WRONG DETAIL / NOT FOUND. Then check each verified source for retractions, expressions of concern, published critiques, or failed replications. Do not correct anything silently — show me each discrepancy and let me decide. For any source you cannot verify through a publisher of record, tell me where it was actually published and what review it passed. If a paper's own provenance is doubtful, say so plainly and tell me not to cite its numbers — I may still want it as an example of the problem.

#### 4c · Check Claims Support Slide 34

Your goal is to find where I am wrong: counter-evidence to my existing claims is the most valuable thing you can produce. Examine each and every claim in my draft and the source attached to it. Does the cited source/tradition/theory actually support this specific claim, at this strength? Comment (using the comment function) on any claim that is overstated, unsupported or uncited. If there is a citation, verify it and mark any faulty citations. If there is a better or newer citation, put it in the comments. Always trace citations to the original study, never to the press release, the news write-up, or a secondhand citation. Where I have overstated, quote the sentence from the source that shows the real strength of the finding. If an assertion does not have a citation but should, find the best citation and list it in the comments.

FOLLOW-UP: I have now edited and revised this document. Add all of the new and revised citations to the bibliography.

#### 4d · Newer, Better, or Contradictory Citations

For each of my central claims and citations, search for: (i) work published since my citation that supersedes or qualifies it; (ii) the strongest published evidence against the claim; (iii) failed replications, null results, and publication-bias corrections in this literature; (iv) relevant work in languages other than English and from outside [region]. Report the contradictions first, before anything that agrees with me.

Rate every source using this rubric or evidence ladder [depending on your field: secondary source, randomized trial, systematic review or meta-analysis, quasi-experiment, correlational survey, preprint, vendor telemetry, practitioner report etc.]. Investigate sample size, setting, author reputation and so on. Flag any famous-but-fragile source whose citation count exceeds its evidentiary weight. Be vigilant about flagging low-quality evidence.

4b

## FIND WHERE I'M WRONG

Counter-evidence to my own claims  
is the most valuable thing you can produce.

SUPPORTED

OVERSTATED

UNSUPPORTED

UNCITED

Trace to the original — never the press release.



## 4e · How wrong could I be?

Here is my bibliography. Report what this evidence base is made of — by method, discipline, education level, country, language of publication, and year. What is over-represented? Which of my claims rest on a part of the literature that is thinner than the claim implies?

Draw a random sample of X [quotes, data points etc.] and hand-check them against the primary source, showing your work for each.

Report the error rate in your own verification, the kinds of errors you made, and which categories of item you are least reliable on.

Never bulk-apply a match on an ambiguous identifier. Shared surnames, variant spellings and namesakes get hand-checked individually, never flipped in a batch. One confident bulk edit can corrupt a dataset in ways that are very hard to find later.

When something cannot be verified, the correct output is a blank and a note — not a plausible guess. A guessed value is indistinguishable from a checked one once it is in the file.

### Example: Evidence for AI in Teaching



Famous-but-fragile

An EEG study & the field's most-cited paper carry weight beyond their design

Contrary evidence

A bias-adjusted reanalysis finds no effect

Skewed evidence base

Broad geography, but almost all quantitative STEM; little humanities



Example: running 4a–4e on the evidence base for AI in teaching. 4b flagged a widely circulated EEG study and the field's most-cited paper as carrying public weight beyond their designs; 4d found a bias-adjusted reanalysis showing no effect; 4e showed the literature is geographically broad but overwhelmingly quantitative and concentrated in STEM and language learning, with almost nothing from arts and humanities teaching.

► [Design principles for teaching in the age of AI](#)

## YOU TRY: Run a check on your own draft (slide 37)

Give it a draft you already trust.

Run 4c Check Claims or 4e How wrong could I be?

Ask what has been published since.

## ORGANIZE & UPDATE

Agents are careless with your files unless you tell them otherwise.

Insert as tracked changes

Count references before & after

Confirm the file still opens

If anything mismatches — abort, don't save

## STAGE 5 · ORGANIZE & UPDATE Slide 38

Agents are careless with your files unless you tell them otherwise. Make the safety checks part of your instructions.

Insert these references into [file] in alphabetical order as tracked changes. Skip any that are already present. Count the references before and after and confirm the arithmetic. Validate that the file still opens. If anything does not match, abort and report rather than saving.

Write a handoff brief for the next session: what was done; what is staged but not yet merged; what is unresolved and why; the two highest-value next tasks; and the mistakes not to repeat.

### Example: Book Project

Your goal is to set me up to revise X for Y.

You will not be writing any text.

- 1) Read A, B & C [abstract, proposal, first version, ToC, notes etc.].
- 2) Create a new folder, make new copies of the files and label your new files...
- 3) Create a references doc and include complete citations for everything.
- 4) Now read each chapter and COMMENT only
  - (a) any passage that needs to be updated or cut,
  - (b) any passage or citation where the evidence is weak or citation is needed
  - (c) if there is a better or newer citation also provide it in the comments
  - (d) track all references as you read and align with the references document.
- 5) Create a new ToC document with your recommendations
- 6) You can use folder [with article pdfs, my notes] or do your own research.
- 7) Create a new folder on Google docs and upload copies

### Example: Book Project (More) Complete Prompt

Your goal is to set me up to revise X for Y. Act like a senior editor with deep experience in Z. You will not be writing any text. We have a word limit of # including citations. Here is the publisher style guide; keep that in mind as you consider cuts and updates.

1. Read A, B & C [abstract, proposal, first version, ToC, notes etc.].
2. Create a new folder D and make new copies

(new doc files) of the files in E. Label your new files Book F ch# etc. Keep the same formatting from the style sheet. Make new versions of references, introduction etc. as well.

3. Add a new file for G based on the ToC; use the other chapters for style but use the contents H.

4. Create a complete references doc and include complete citations for everything cited in the book in format I.

5. Now read each chapter and COMMENT only on (a) any passage that needs to be updated or cut; (b) any passage or citation where the evidence is weak — if a citation is needed, provide it in the comments as an inline text citation (per the style of the book) and add it to the references; (c) if there is a better or newer citation, also provide it in the comments; (d) keep track of the references as you read and compare them to the references document — add anything missing (with track changes on and a comment showing where it appears), and flag anything in the references that is NOT in the text with a highlight rather than deleting it.

6. When you are done, create a new ToC document with your annotations and recommendations for additional updates, chapters or sections.

7. Use folder J [with article PDFs, notes and your research] as needed, or do your own research. I am especially interested in K. My audience is L. Consider all of this as you think about where new chapters or sections might go, and put it into the ToC document v4\_annotations.

8. Create a new folder on Google Docs and upload all of this inside that folder with comments tracked.

## STAGE 6

## FEEDBACK

The readers you don't have in the room.

What would a hostile referee at your target journal say?

The reader who misunderstands you

A panel of scholars who disagree  
Each reads the others, then ranks changes  
A dashboard of what most agree to fix



## STAGE 6 · FEEDBACK — the readers you don't have in the room Slide 40

What might an average reader / college professor / IRS auditor find confusing, objectionable or exciting? Give me feedback from a reader who misunderstands my intentions. Create feedback that will challenge me, including feedback with inaccurate information.

Do multiple individual analyses of this text by a wide group of scholars with different views

of the field. Start with X and Y [pick your own list of scholars/rivals] and include recent work on Z. Include scholars from these competing theoretical areas [list] or the authors of [list of books].

Then have each one read all of the other reviews and compare what they think is good and what needs improvement. Create a dashboard chart that lists the recommended changes and how many and which advisors recommend each. Cluster the recommendations agreed by most advisors at the top. Write a brief report that gives me three solid suggestions for improvement and highlights what parts of this research are most contested. Provide citations for the work that most challenges this. Use this analysis to ask me thought-provoking questions about how I might improve my chances of acceptance at the Journal of X.

## REUSABLE SKILLS



Here are my special rules for how to do this task

### 1. Source-Check Pass: *Verify my draft's claims match what is cited.*

- Flag uncited assertions and overclaims
- Trace to the original, not the secondhand source
- *Your rule:* what counts as a credible source in your discipline

### 2. Case Chronology

*List all of the evidence for this case in chronological order*

- List the source of each time point
- Highlight all disagreements explicitly
- *Your rule:* Put what is firmly established and uncontested in bold.

first, grade last; coach by question, not correction.

Extract a text's thesis, premises and hidden assumptions first; steelman before critiquing.

Source-Check Pass: flag uncited assertions and overclaims; trace to the original, not the secondhand source.

Your rule: what counts as a credible source in your discipline.

Case Chronology: list all of the evidence for this case in chronological order; list the source of each time point; highlight all disagreements explicitly. Your rule: put what is firmly established and uncontested in bold.

► [The QR code on this slide — skills, templates and examples](#)

## SAVE REUSABLE SKILLS (slide 42)

When you have refined a set of rules for a task you will repeat, save it as a Skill so you never have to re-explain it. "Here are my special rules for how to do this task." The slide shows two; the pattern is the same each time — the task, the steps, and one rule that is yours and not mine.

Evaluate the credibility of sources in this way...

Apply my rubric criterion by criterion; feedback

## Research Update Skill “add these refs...”

When I share a new study or report: verify, integrate and sync.

### Step A — Verify before anything else

- Fetch the actual source.
- Confirm: authors, year, exact title...
- Quote figures only from the source.
- Check for retractions, corrections, and published critiques (search...)
- Classify the evidence tier and carry it into every note you write:
  - RCTs etc...

### Step B — Integrate into the manuscript

Add the APA entry to [references doc] as a tracked insertion.

Insert one comment/ch, tuned to that chapter's argument.

Label in comments: ADD / UPDATE / COUNTER-EVIDENCE etc

figures from the source, not from press coverage. Check for retractions, corrections and published critiques.

Classify the evidence tier and carry it into every note you write: RCTs and randomized field experiments; systematic reviews and meta-analyses; quasi-experiments and large-N administrative data; surveys and perception data (attitudes, never outcomes); preprints and working papers; practitioner and press sources (cite as practice, not evidence).

House rule: famous-but-fragile studies (contested preprints, corrected papers) are cited only as cultural phenomena, never as evidence. If a new source contradicts something already in the manuscript, that is the most valuable outcome — write it up as counter-evidence and say which existing claim needs qualifying. If verification fails or is ambiguous, report that and stop.

Step B — Integrate. Add the APA entry to the references document as a tracked insertion, alphabetically, deduped. Write one comment per target chapter, labelled ADD / UPDATE / CUT / COUNTER-EVIDENCE / EVIDENCE CHECK, with the finding, its key numbers, its evidence tier and where it fits.

The safety protocol — never skip it. For every file edit: count what is there before, run the edit, verify the count equals before + n, validate the file opens, and only then replace the original. If any check fails, abort that file and say so.

Step C — Slide citations. If the finding is one I might cite in a workshop, append the full citation to the relevant Slides and Citations document, matching that document's existing format.

Step D — Website. Preview the exact text in chat before saving, and get explicit approval before publishing.

Close with a compact summary: what was verified and its tier, which chapters got notes and why, what went into references, slides and website, what is synced versus pending, and any caveat.

## Example: a Research Update Skill

### Complete Skill

When I share a new study or report, your job is to verify it, integrate it into the manuscript apparatus, and keep all copies in sync. Accuracy is paramount: never add anything unverified, always distinguish evidence tiers, and hold positions only when the evidence supports them.

Step A — Verify before anything else. Fetch the actual source (the DOI, SSRN, arXiv or publisher page). Confirm authors, year, exact title, venue, and the specific numbers to be cited — quote

## Building a Skill



1. Turn this ppt into doc where the images are on the left and the citations on the right. Here is an example.
2. I revised your output. Review it and learn.
3. Try again with this new ppt.
4. Write a skill that would have produced this result without me having to correct you. Include the specific mistakes you made along the way and what fixed them..

## Building a Skill (slide 43)

You do not write a Skill from scratch. You do the task once with the AI, correct its output by hand, and then ask it to write down what it should have known. The correction is the training data: what it got wrong and what fixed it is the part worth saving.

### Complete Prompt

1. Turn this ppt into a doc where the images are on the left and the citations on the right. Here is an example.
2. I revised your output. Review it and learn.

3. Try again with this new ppt.

4. Write a skill that would have produced this result without me having to correct you. Include the specific mistakes you made along the way and what fixed them.

This document is the output of exactly that loop.

- ▶ [The QR code on this slide — skills, templates and examples](#)

## VIBE CODING, WEBSITES and VISUALIZATIONS

### Vibe Coding

Create and deploy an interactive superhero-themed game to teach the Bingham plastic model through visual simulation to college students in both English and Arabic.

Got it — I'll deliver the game in **Arabic + English**, packaged as a **standalone web app** you can open in your browser or run locally without installing anything complicated.



Create a simulation. The prompt: “Create and deploy an interactive superhero-themed game to teach the Bingham plastic model through visual simulation to college students in both English and Arabic.”

- ▶ [Play the Claude version](#)
- ▶ [Play the ChatGPT version](#)

### Schedule Tasks

Every Monday at 7am, review all of my course data in the LMS (chats, completion, scores, time-on-task etc.) and compare it to the baseline averages from previous weeks. Flag anything that's shifted more than 10% from baseline. Send me an email organized by course with a flag for any metric that has changed and include any possible explanation and remedies.

**Tell me how I can better help my students?**

Schedule a task, then ask it a question you cannot answer yourself: “Tell me how I can better help my students.”

### Interactive Database & Website



Prompt



#### STEP 1

Do an exhaustive search of academic resources about 19th century women composers of solo piano music.

#### STEP 2

Create a searchable global data base...

#### STEP 3

Create a **website**



#### STEP 4

Find sheet music not on my computer...

#### STEP 5

Order missing works from ILL... 

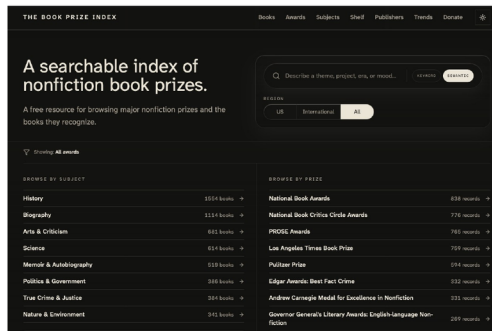
Example: Women at the Keys — a searchable global database of 19th-century women composers of solo piano music, built from an exhaustive search of archives, catalogues and IMSLP, then deployed as a public website.

- ▶ [The full prompt, workflow and the versions produced by different AI models](#)
- ▶ [The result — Women at the Keys](#)

## Interactive Database & Website



A searchable index of non-fiction book prizes  
By Benjamin Breen, UC Santa Cruz



Example: a searchable index of non-fiction book prizes, by Benjamin Breen, UC Santa Cruz.

► [Book Prize Index](#)

## Interactive Database & Website



Example: Piano Repertoire Explorer — over 250,000 published piano works. Fable and Sonnet collected publication records, catalogues and reviews into one data set, then deduplicated, cross-referenced, checked and tagged it at scale (Joe Verdi and G. Verdi are probably the same composer if associated with an Italian opera, but only solo piano arrangements should be counted). AI can scale things far beyond my human limits.

► [Piano Repertoire Explorer \(v1\)](#)

### Broad Street, 1854

You are Dr John Snow. Find the killer.



On a laptop  
[josebowen.github.io/  
BroadStreet1854/](http://josebowen.github.io/BroadStreet1854/)

### Three Mile Island, 1979

You can stop the meltdown.



On a laptop  
[josebowen.github.io/  
/TMI1979](http://josebowen.github.io/TMI1979)

► [Broad Street, 1854](#)  
► [Three Mile Island, 1979](#)

# EVERYONE IS NOW AN AI MANAGER

EVERYONE IS NOW AN

## AI Manager



Which tasks should you delegate?



AI needs direction, feedback & correction.

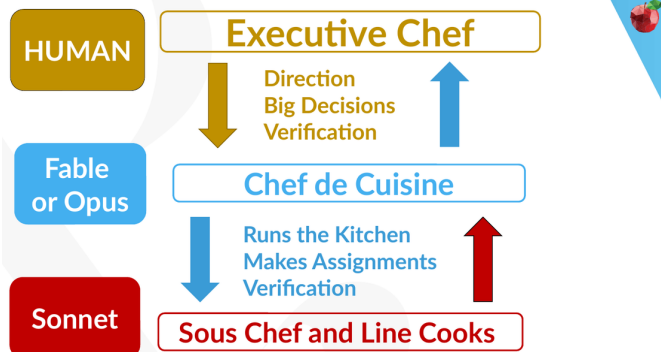


AI delegation is a skill, not a shortcut.



AI needs your judgment.

Using agentic AI is like being an executive chef. I make the big decisions and do the verification, but Fable is my Chef de Cuisine, who then creates assignments for Sonnet the Sous Chef. For the piano project, Fable wrote the instructions for Sonnet to do the data retrieval — often a dozen agents working at once in archives across Europe — and then checked the data.



## WILL AI MAKE YOUR RESEARCH MORE PRODUCTIVE?

### Will AI make your RESEARCH more productive?

“...productivity among GenAI users rose by 15% in 2023 relative to non-users and **further** increased to **36%** in 2024...”

The estimated **improvements in journal quality** are smaller but still positive, with mean impact factors rising by 1.3 percent in 2023 and 2.0 percent in 2024.”

Filimonović, D., Rutzer, C., & Wunsch, C. (2025, Oct). Can GenAI Improve Academic Performance? Evidence from the Social and Behavioral Sciences. <https://arxiv.org/abs/2510.02408> page 14

Filimonović, D., Rutzer, C., & Wunsch, C. (2025, October). Can GenAI improve academic performance? Evidence from the social and behavioral sciences. arXiv [working paper; see p. 14]. <https://arxiv.org/abs/2510.02408>

## “Pre-Read” Tools

- **Refine.ink** - identifies errors in reasoning, calculation, and references across parts of a paper
- **MANDATORY** audit and verification for publication in American Economic Association and the Econometric Society Journals.
- **Paperreview.ai**

## Stanford Agentic Reviewer

**How it works**

- Upload & Submit**  
Upload your paper PDF and enter your email address.
- Get Notified**  
Receive an email notification when the AI review is complete.
- View Review**  
Come back to see the AI review and suggestions for you.

**Paper PDF \***

Choose PDF file or drag and drop  
Max 10MB - First 10 pages analyzed

**Email Address \***  
your.email@university.edu

**Target Venue (Optional)**  
Select a conference/journal

**Submit for Review**

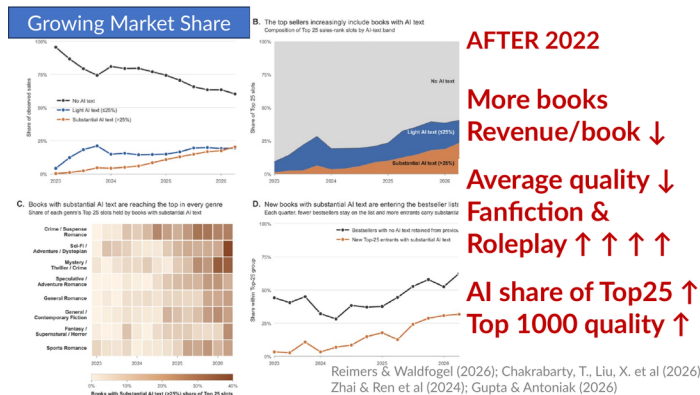
“Pre-read” tools for journals and conferences

► [Refine.ink](https://www.refine.ink)

► [Paperreview.ai](https://paperreview.ai)

Refine. (2026, August 6). Refine partners with the American Economic Association and the Econometric Society. [company blog] <https://www.refine.ink/blog/economics-journal-partnerships>

Reimers, I., & Wadfogel, J. (2026). AI and the quantity and quality of creative products: Have LLMs boosted creation of valuable books? NBER Working Paper 34777. [working paper] <https://doi.org/10.3386/w34777>



Chakrabarty, T., Liu, X., et al. (2026, July). Generative AI floods and dilutes the market for books. arXiv [preprint]. <https://arxiv.org/pdf/2607.20349>

The fanfiction and roleplay line comes from the WildChat corpus: in more than 500,000 consented, English-language ChatGPT conversations, over a third involve generating fiction, especially fanfiction and erotica which generate what the authors call a "solipsistic reader-writer," "who both generates and consumes fiction within a closed conversational loop."

Gupta, N., Antoniak, M., & Walsh, M. (2026, June 23). AI

fiction in the wild. arXiv [preprint; provisionally forthcoming in MFS: Modern Fiction Studies]. <https://arxiv.org/abs/2606.22748>  
Zhao, W., Ren, X., Hessel, J., Cardie, C., Choi, Y., & Deng, Y. (2024). WildChat: 1M ChatGPT interaction logs in the wild. ICLR 2024. [consented corpus of 1M conversations; the dataset behind Gupta et al.] <https://arxiv.org/abs/2405.01470>

## AI Is Changing Scientific Research

1. AI “accelerates manuscript output, reduces barriers for non-native English speakers, and diversifies the discovery of prior literatures.”
2. “However, traditional signals of scientific quality such as language complexity are becoming unreliable indicators of merit.”

Keigo Kusumegi et al. (2025) Scientific production in the era of large language models. *Science* 390, 1240-1243. DOI:10.1126/science.adw3000

Kusumegi, K., et al. (2025). Scientific production in the era of large language models. *Science*, 390, 1240–1243. <https://doi.org/10.1126/science.adw3000>

## New Work Flows?

### PRE-WRITING

- Ideas
- Drafts
- Outlining
- Sounding Board
- Questioning Assumptions
- Research Assistant



### EDITING

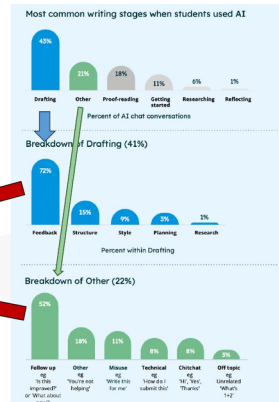
- Targeted Feedback
- Coherence
- Alternative Versions
- Style Enhancement
- Audience Testing
- Reflection

80% AI-generated text =  
**19.4%** in US  
 9.8% in UK

Most common student requests:

"Is my paper good?"  
 "Am I aligned with the rubric?"  
 "How can I improve?"

Turnitin Q2 2026 Learning Integrity Insights Reports



Turnitin. (2026). Q2 2026 learning integrity insights report. [vendor detection data]  
<https://www.turnitin.com/press/us-higher-education-students-outsourcing-writing-to-ai-at-twice-the-rate-of-uk-australian-peers>



AI Is Changing  
**AVERAGE**

Work with AI is  
**RAISING  
 STANDARDS**

#### Checks

Choose checks to run

- ☐ Grammar  
Fix spelling and grammar errors.
- ☐ Brevity  
Omit needless words
- ☐ Cliches  
Replace over-used phrases
- ☐ Readability  
Simplify convoluted sentences
- ☐ Passive Voice  
Convert passive voice to active voice
- ☐ Confidence  
Remove excessive hedging (I think, probably, etc)
- ☐ Citation  
Identify claims that need evidence
- ☐ Repetition  
Remove repeated words
- ☐ Custom  
What should Lex check for?

Writing with Lex.page

► [Lex](#)

- Brainstorm intro ideas
- Identify weak arguments
- Flag confusing parts
- Tell me what this means (Clarity)
- Give me better analogies
- Ideas for paragraph transitions
- Thesaurus with context

**AI doesn't reduce work,  
 it intensifies it.**

- **TASK EXPANSION**
  - You can do the work of others
- **BREAK PROMPTING**
  - Work for breaks, lunch or overnight
- **MULTITASKING & MANAGING AI**
  - Faster work pace

Ranganathan, A., & Ye, X.M. (2026, Feb 9) AI Doesn't Reduce Work—It Intensifies It. Harvard Business Review  
<https://hbr.org/2026/02/ai-doesnt-reduce-work-it-intensifies-it>

Ranganathan, A., & Ye, X. M. (2026, February 9). AI doesn't reduce work — it intensifies it. Harvard Business Review.  
<https://hbr.org/2026/02/ai-doesnt-reduce-work-it-intensifies-it>

## Where does AI belong in Research?



Understanding?  
Verification?  
Production?

Technology is neither good nor bad; nor is it neutral.

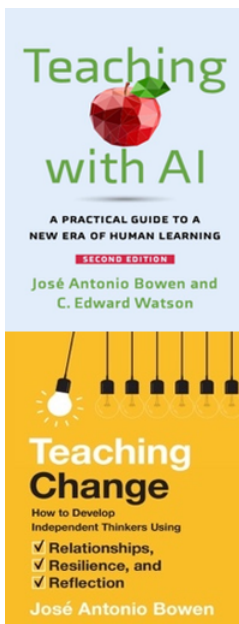
Melvin Kranzberg (1917-1995) Professor of History of Technology



Which tasks should you do?

Kranzberg, M. (1986). Technology and history: "Kranzberg's laws." *Technology and Culture*, 27(3), 544–560.

## CITATIONS & RESOURCES



- ▶ [WeTeachWithAI.com](https://www.weteachwithai.com) contains many many more resources, prompts, assignments and links. It is updated weekly if not daily.
- ▶ **Buy the book at 30% off!**
- ▶ [Teaching with AI at Johns Hopkins University Press — code HTAI25 for 30% off](#)
- ▶ [Teaching with AI on Amazon](#)
- ▶ If you don't want to use AI, but want students to want to do things the hard way, then read **Teaching Change: How to Develop Independent Thinkers Using Relationships, Resilience and Reflection**.
- ▶ [Teaching Change at Johns Hopkins University Press — code HTWN for 30% off](#)
- ▶ [Teaching Change on Amazon \(often 25% off\)](#)