Will AI replace RAs and TAs?

Paul Goldsmith-Pinkham (Yale SOM & NBER)

Emory University · 29 October 2026

The elephant in the room



Will AI replace...

- RAs and TAs?

The elephant in the room



Will AI replace...

- RAs and TAs?
- Graduate students?

The elephant in the room



Will AI replace...

- RAs and TAs?
- Graduate students?
- Us?

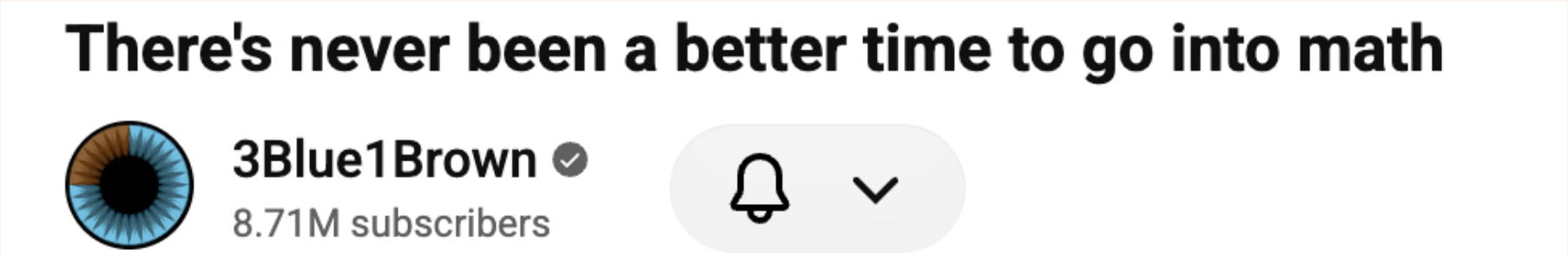
“The mood is discombobulated. … Most people, I think, are feeling grief.”

— Steve Strogatz, *The Daily* (10/13/2026)

“What is the point?”

— Steve Strogatz, *The Daily* (10/13/2026)

But I am an optimist



Roadmap

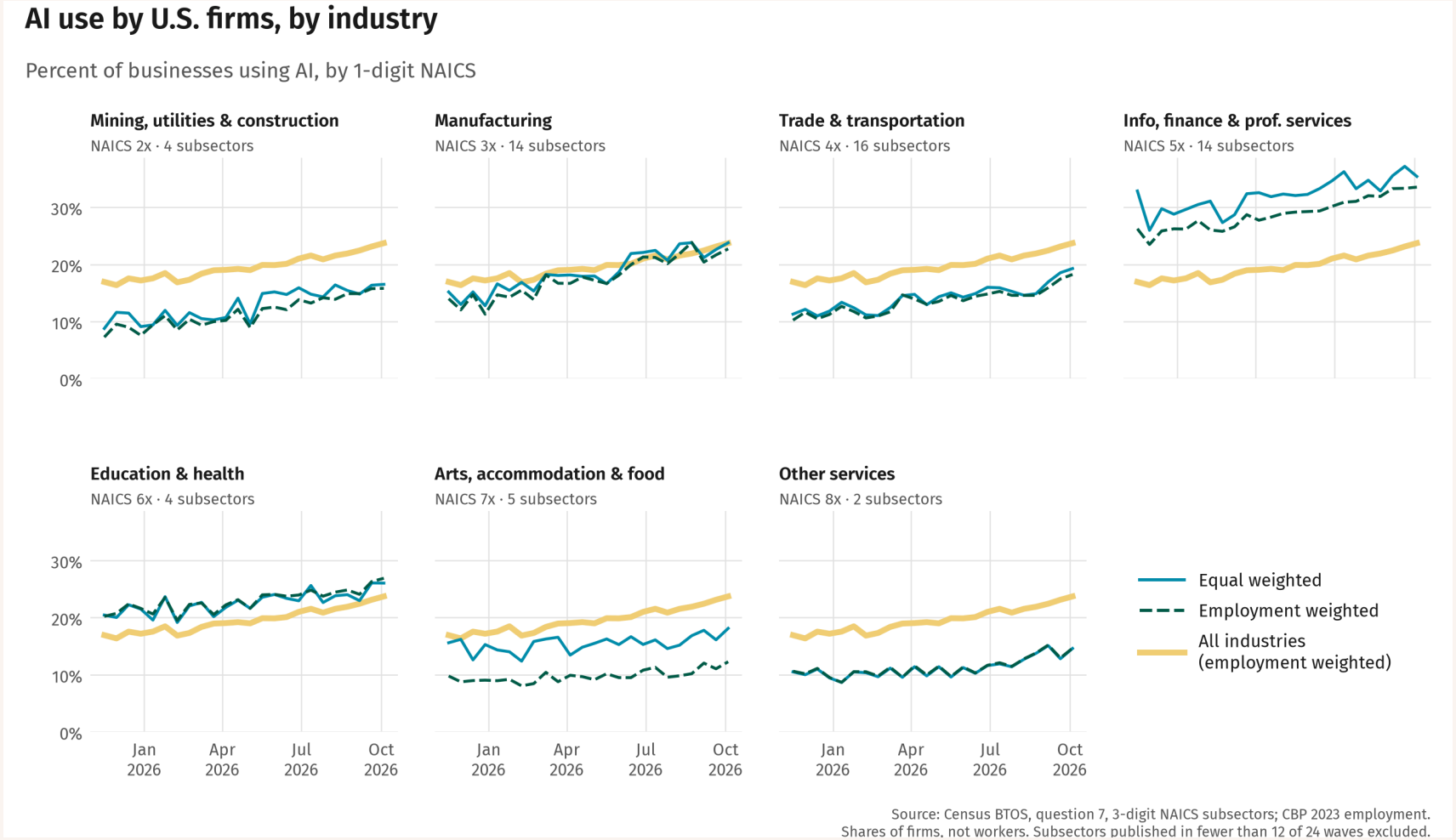
1. What has happened?
2. What is AI, and how does it work?
3. What *will* happen?
4. What do economists do?
5. Equilibrium

Knowledge work has three pieces:

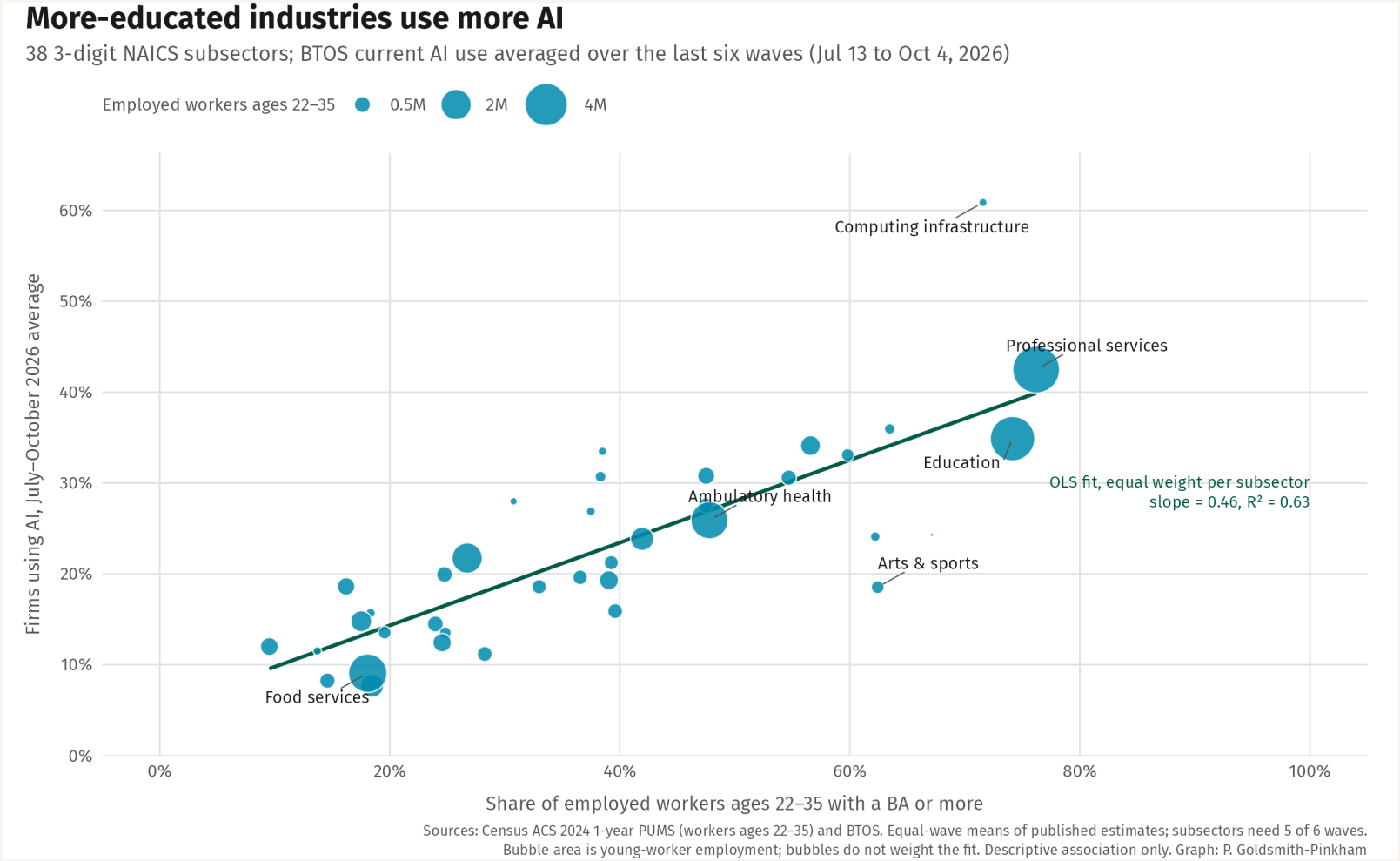
1. **Production**: Can we produce output?
2. **Verification**: Is the output correct?
3. **Judgment**: Was this what we wanted?

AI has cut the cost of producing knowledge work

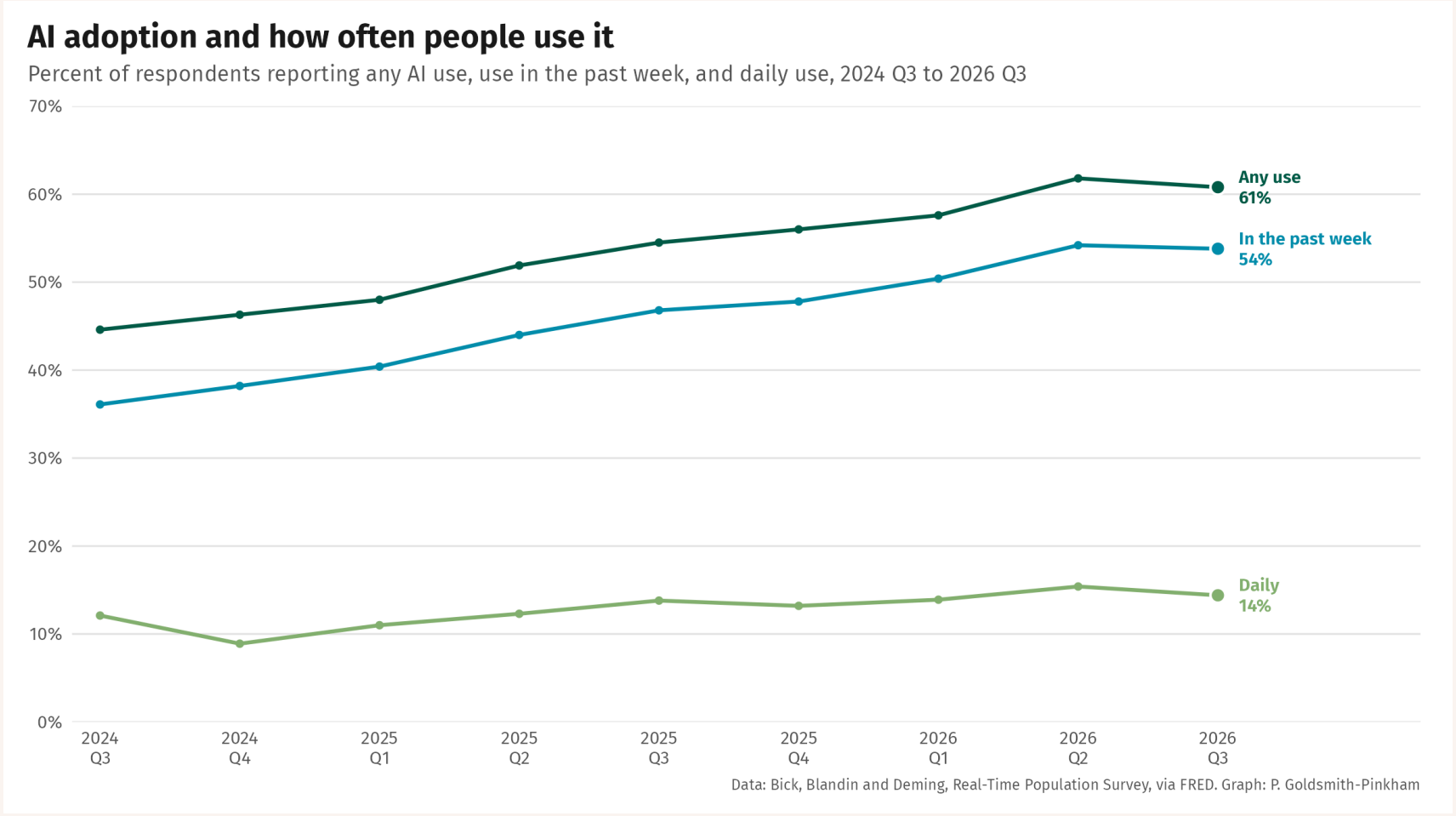
Firms: AI use is concentrated in a few sectors



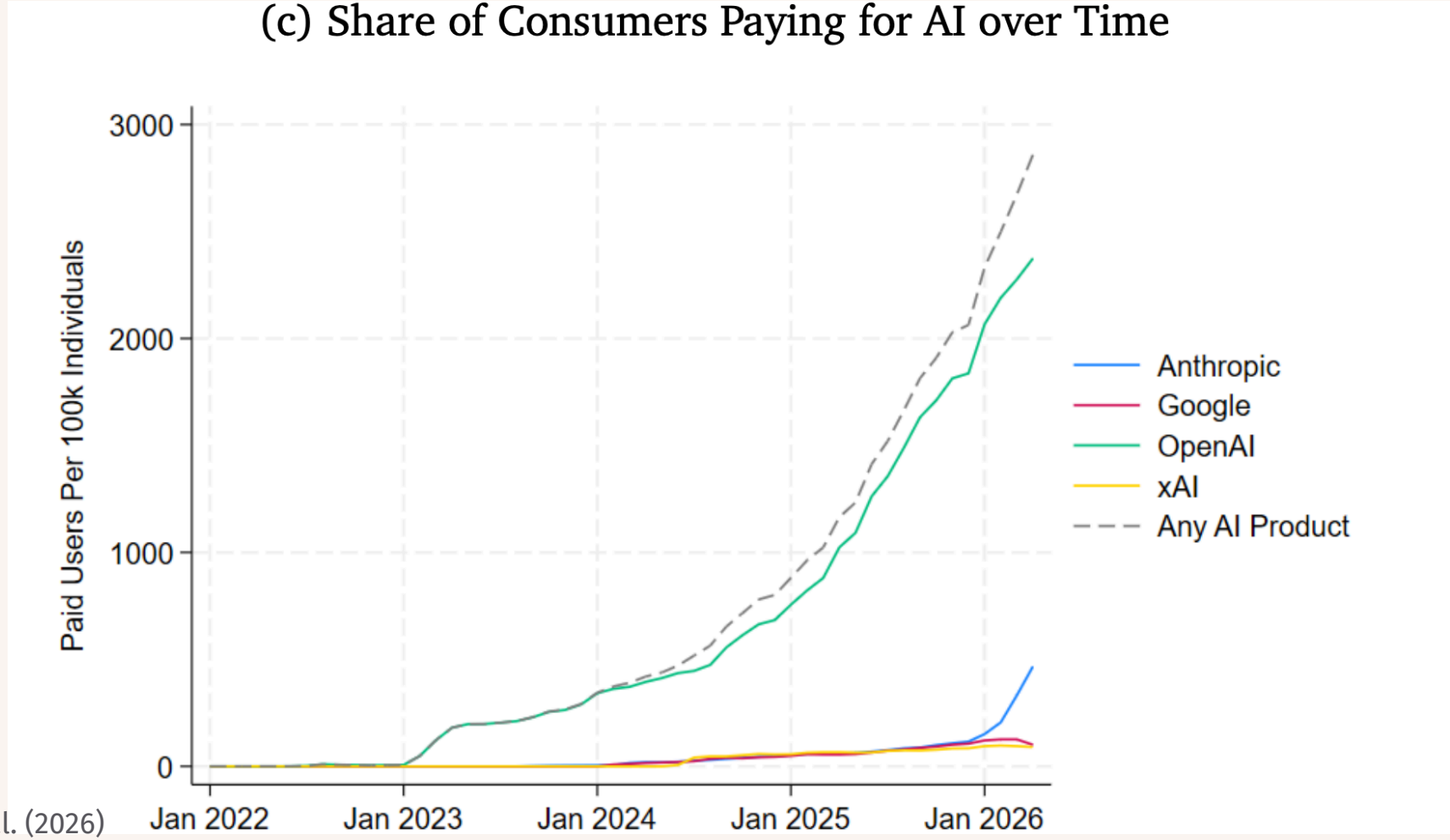
… the sectors where people look like us



Consumers: use is rising fast

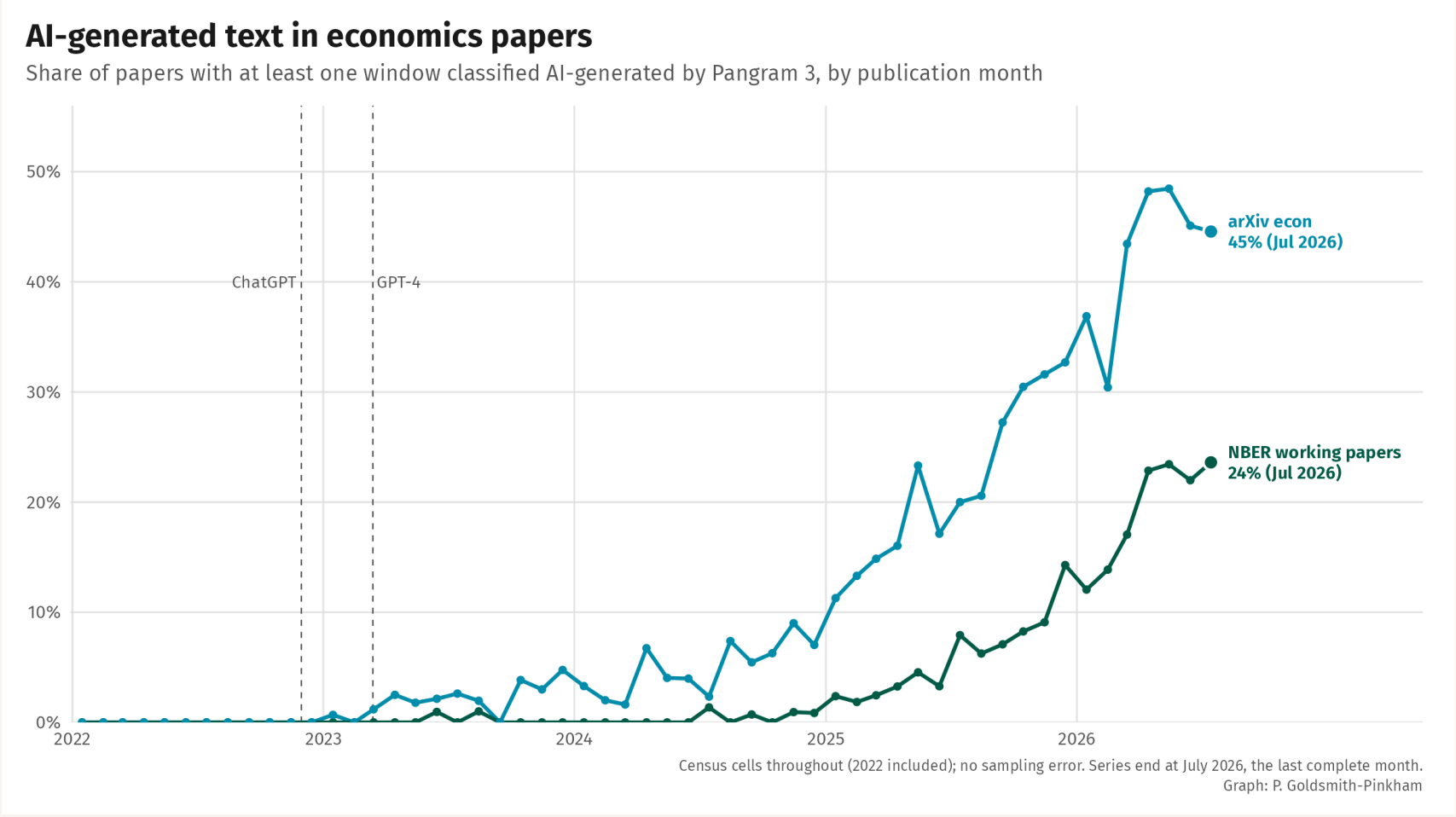


… but far fewer are paying for it



Egan et al. (2026)

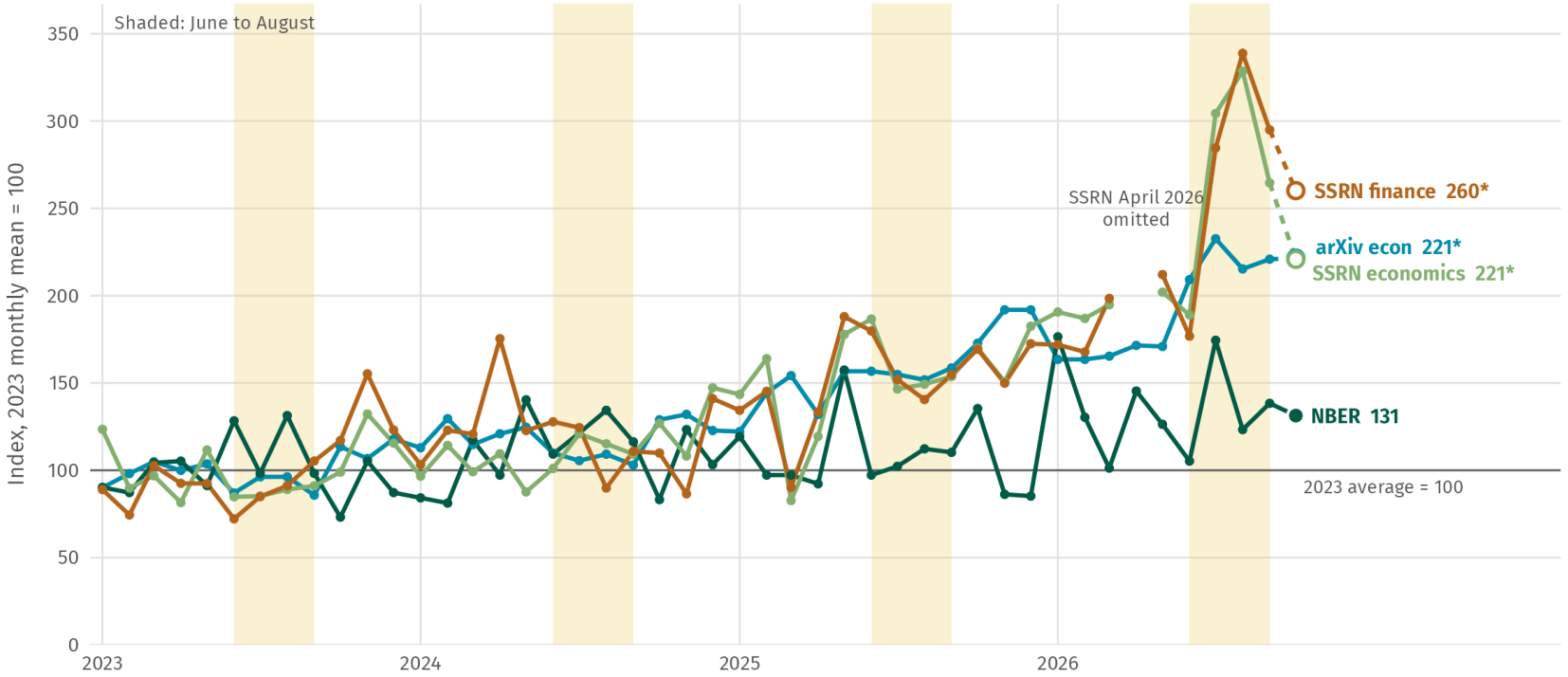
Economists: more AI in our writing



… and more papers

Economics and finance preprints per month: arXiv, NBER and SSRN

Each series indexed to its own 2023 monthly average, January 2023 to October 2026 (* projected)



arXiv: new econ.GN/EM/TH papers by original submission date (arXiv API). NBER: working papers by Crossref issue month. October 2026, dashed: arXiv projected from Oct 1–26 (x 31/26), SSRN from Oct 1–27 (x 31/27); NBER as observed (weekly Monday releases; Oct 26 was the last of the month).
SSRN: 42 ERN (economics) and FEN (finance) subject journals by posted date, each indexed to its own 2023 mean and averaged with equal weight (subjects overlap, so listings cannot be summed into a paper count; subjects under 10 postings a month in 2023 excluded). SSRN April 2026 omitted as an outlier. Graph: P. Goldsmith-Pinkham

… and some strange behavior

A practitioner's guide to cluster-robust inference

Search within citing articles

The Cronbach's alpha of domain-specific knowledge tests before and after learning: A meta-analysis of published studies

[PDF] springer.com

Quick read

PA Edelsbrunner, BA Simonsmeier… - Educational Psychology …, 2025 - Springer

Abstract Knowledge is an important predictor and outcome of learning and d
measurement is challenged by the fact that knowledge can be integrated and

Save Cite Cited by 228 Related articles All 13 versions

[HTML] The impact of Covid-19 on student achievement: Evi
meta-analysis

G Di Pietro - Educational research review, 2023 - Elsevier

This work attempts to synthetize existing research about the impact of Covid
closure on student achievement. It extends previous systematic reviews and

Save Cite Cited by 363 Related articles All 7 versions

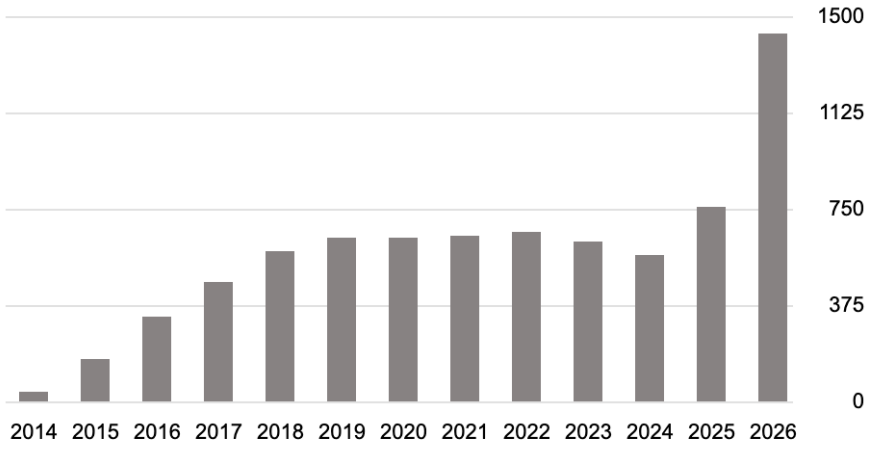
Measuring the impact of early-2025 AI on experienced ope
productivity

J Becker, N Rush, E Barnes, D Rein - arXiv preprint arXiv:2507.09089, 2025

Despite widespread adoption, the impact of AI tools on software developmen
remains understudied. We conduct a randomized controlled trial (RCT) to un

Save Cite Cited by 344 Related articles All 11 versions

Citations per year



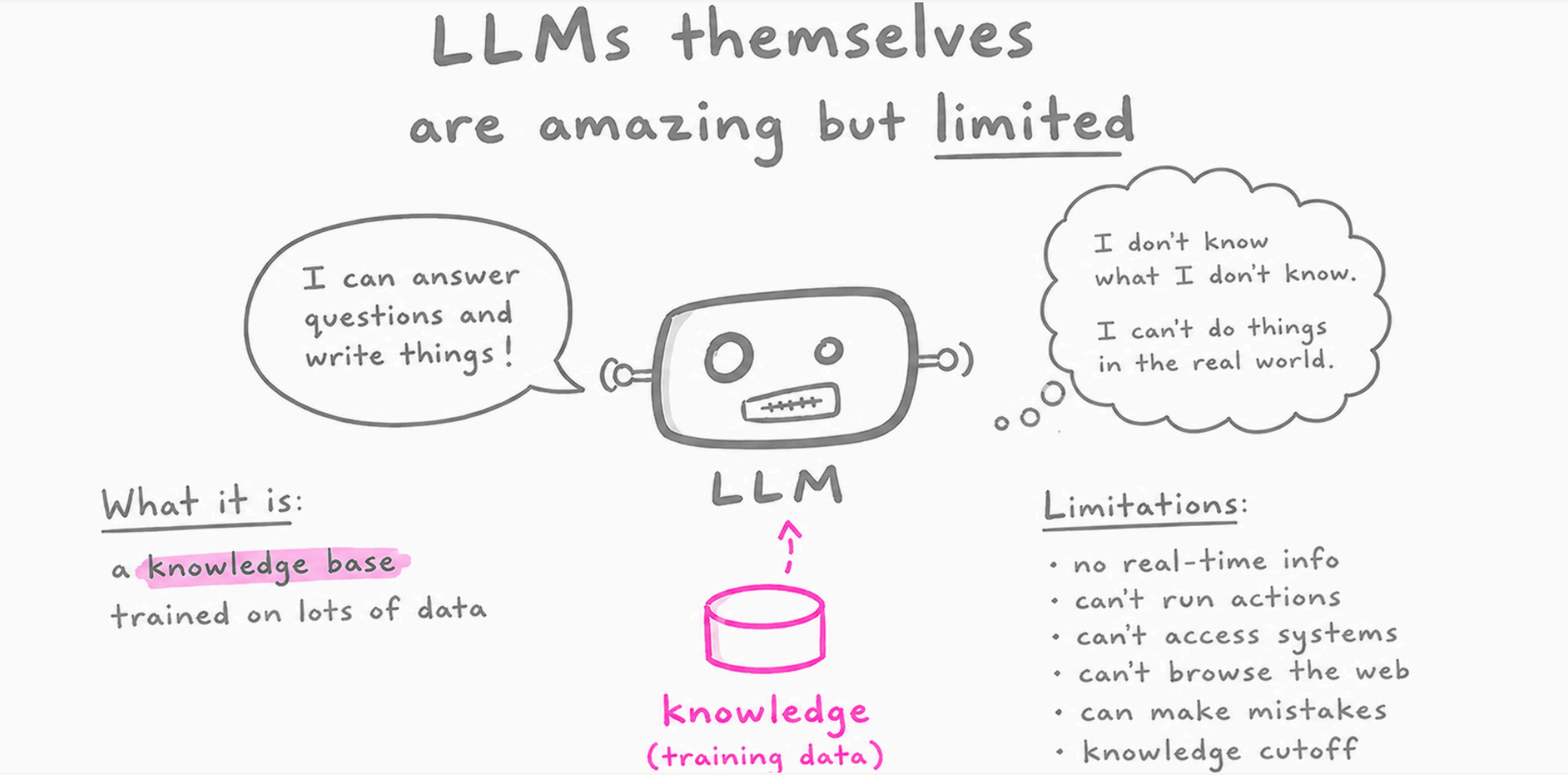
So what?

1. AI use is rising
2. Concentrated among knowledge businesses
3. Growing impact in academia

Where does this leave us?

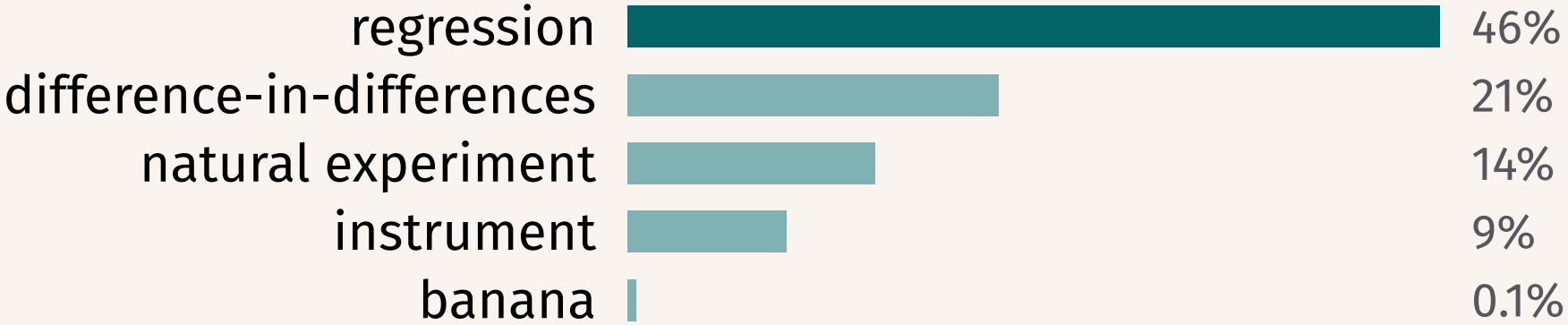
What is AI, and how does it work?

An LLM: text in, text out

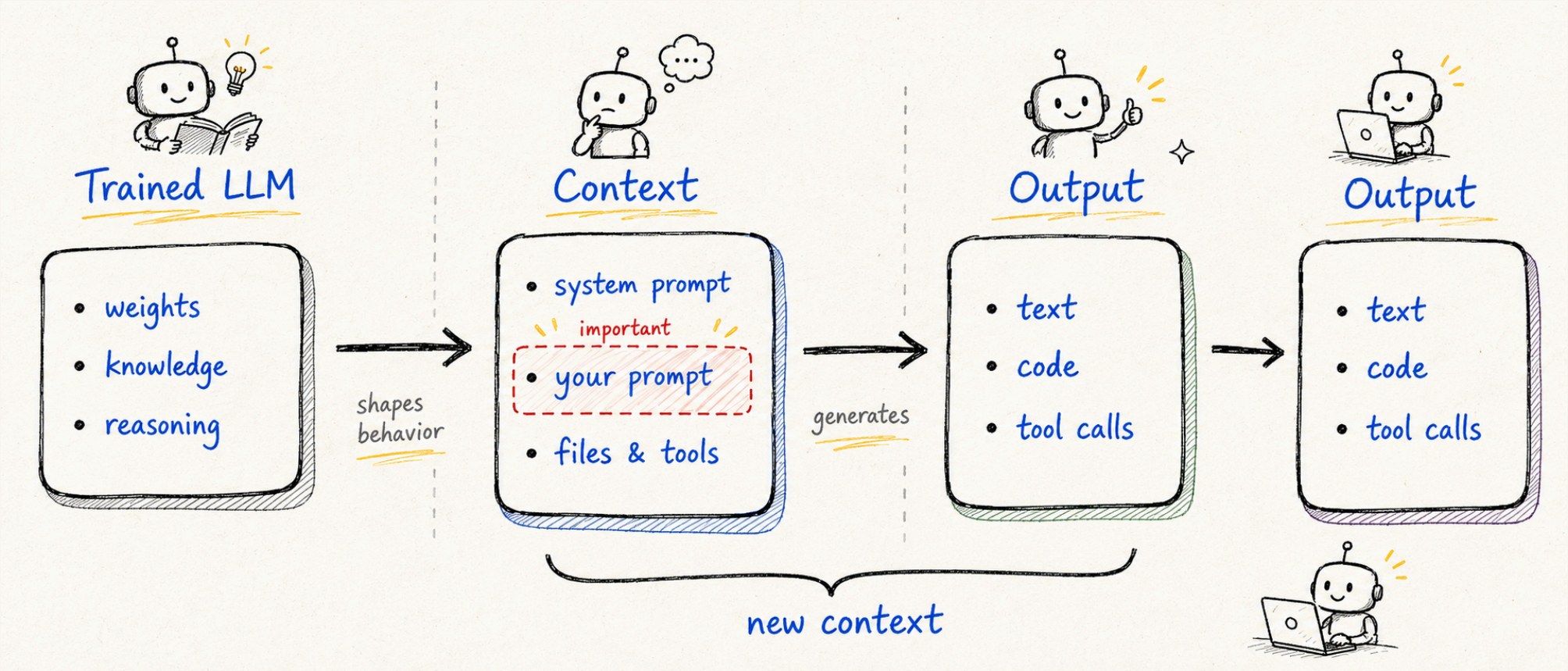


Next token prediction

The economist estimated the effect using a ...



Everything it sees is context

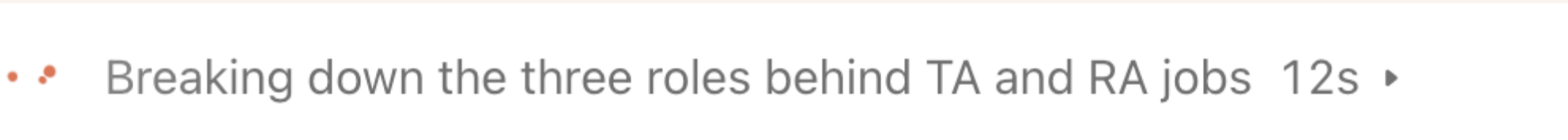


Unlock #1: reasoning

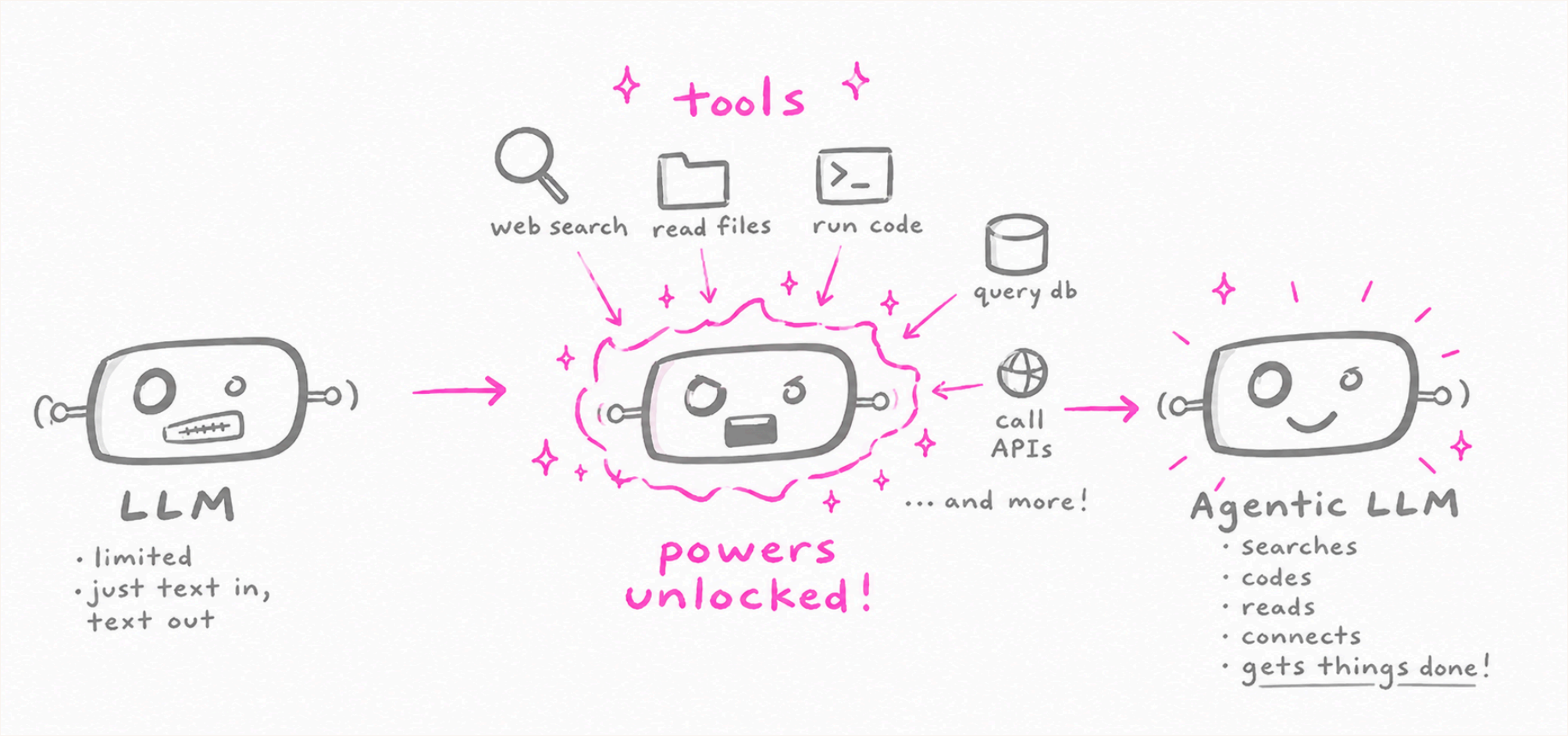
Before Question → Answer

After Question → *“Let me think… check… retry…”* → Answer

Spend more tokens → better answers



Unlock #2: tools



How is an LLM made?

Pre-training

Ingest everything;
predict the next word

---

knowledge,
how to write

Pre-training $\approx$ compressing the internet, books, code

How is an LLM made?

Pre-training

Ingest everything;
predict the next word

---

knowledge,
how to write

→

Fine-tuning

Imitate curated
good answers

---

to take turns,
to answer a user

Pre-training ≈ compressing the internet, books, code

How is an LLM made?

Pre-training

Ingest everything;
predict the next word

knowledge,
how to write

→

Fine-tuning

Imitate curated
good answers

to take turns,
to answer a user

→

Reinforcement Learning

Try, do more of what
won, repeat

learn paths to
reward

Pre-training ≈ compressing the internet, books, code

Show a rater two answers. Which is better?

$$\Pr(A \succ B) = \frac{\exp(\beta_A)}{\exp(\beta_A) + \exp(\beta_B)}$$

Model learns:

- *loadings* **and**
- *characteristics*

of prompts that are preferred

- But aggregating over many raters' preferences

Example of reward hacking

Raters prefer longer answers

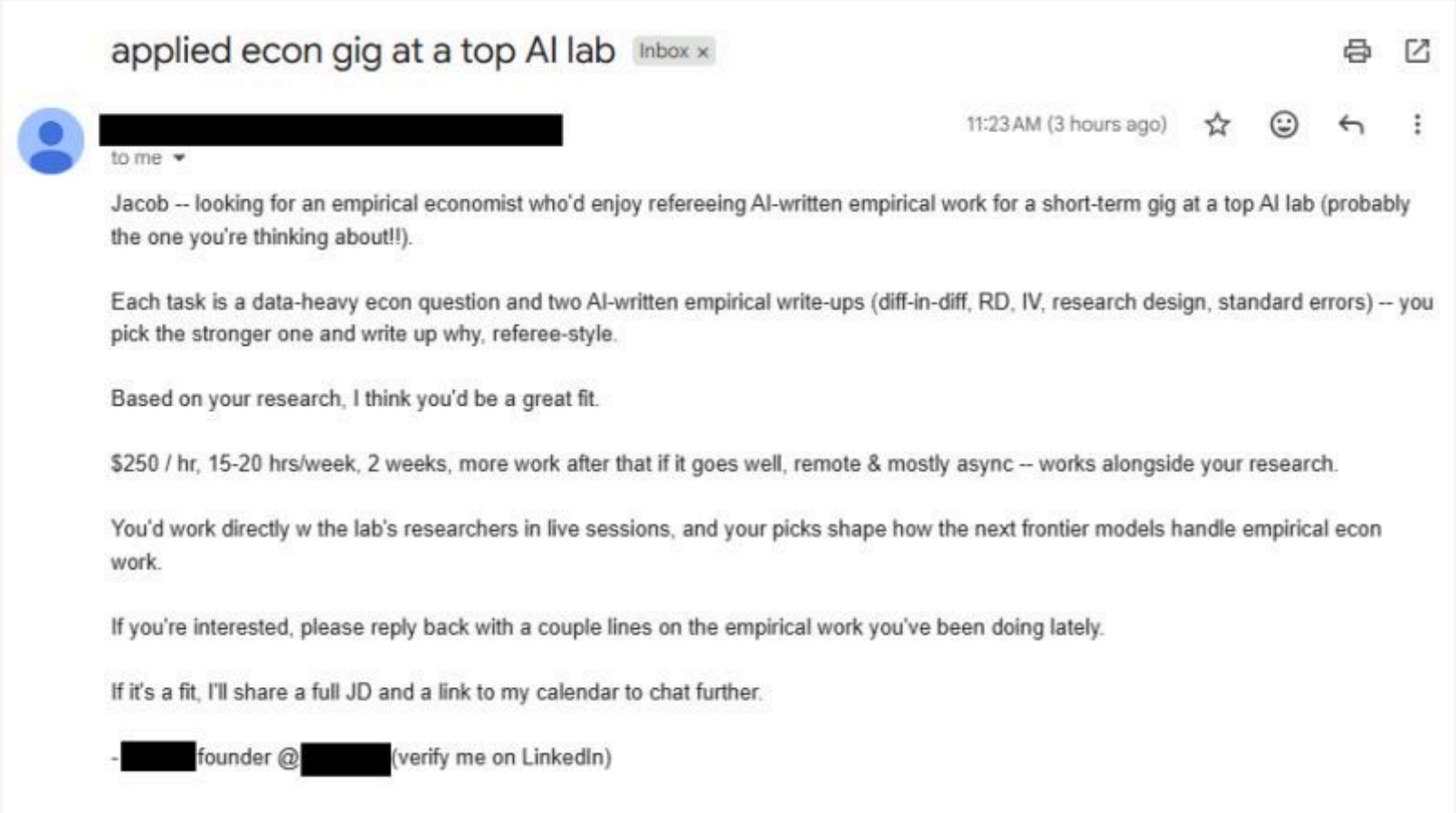
Reward model rewards length

The model learns to pad

Same for:

- confident tone
- markdown
- sycophancy
- much weirder learned behaviors

The labs want economists' preferences



applied econ gig at a top AI lab Inbox ×

[redacted] 11:23 AM (3 hours ago)

to me

Jacob -- looking for an empirical economist who'd enjoy refereeing AI-written empirical work for a short-term gig at a top AI lab (probably the one you're thinking about!!).

Each task is a data-heavy econ question and two AI-written empirical write-ups (diff-in-diff, RD, IV, research design, standard errors) -- you pick the stronger one and write up why, referee-style.

Based on your research, I think you'd be a great fit.

$250 / hr, 15-20 hrs/week, 2 weeks, more work after that if it goes well, remote & mostly async -- works alongside your research.

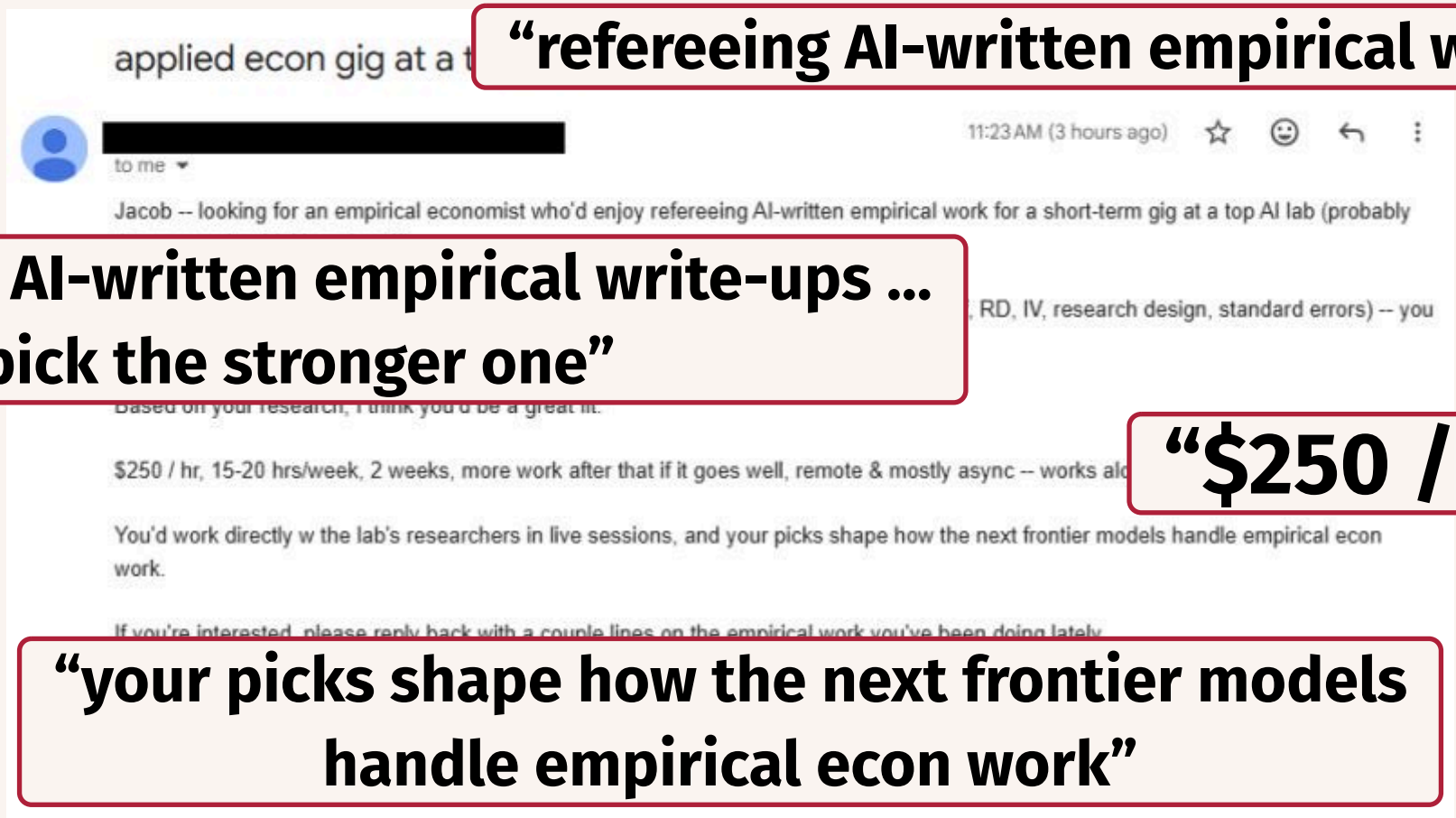
You'd work directly w the lab's researchers in live sessions, and your picks shape how the next frontier models handle empirical econ work.

If you're interested, please reply back with a couple lines on the empirical work you've been doing lately.

If it's a fit, I'll share a full JD and a link to my calendar to chat further.

- [redacted] founder @ [redacted] (verify me on LinkedIn)

The labs want economists' preferences



“refereeing AI-written empirical work”

“two AI-written empirical write-ups … you pick the stronger one”

“$250 / hr”

“your picks shape how the next frontier models handle empirical econ work”

Code and math:
known, correct answers for a given task

- Prompts with known answers
- Try many times
- Upweight what works

Artifical intelligence with some very non-human behaviors

1. **Born fresh** every session
2. Preferences are an **aggregate** of raters
3. Biggest gains where rewards are **verifiable**

So, what will happen?
AI as subtitute vs. complement

Substitute or complement? Two RAs

A humble
Fed RA



Coffee $\rightarrow$ code

Substitute

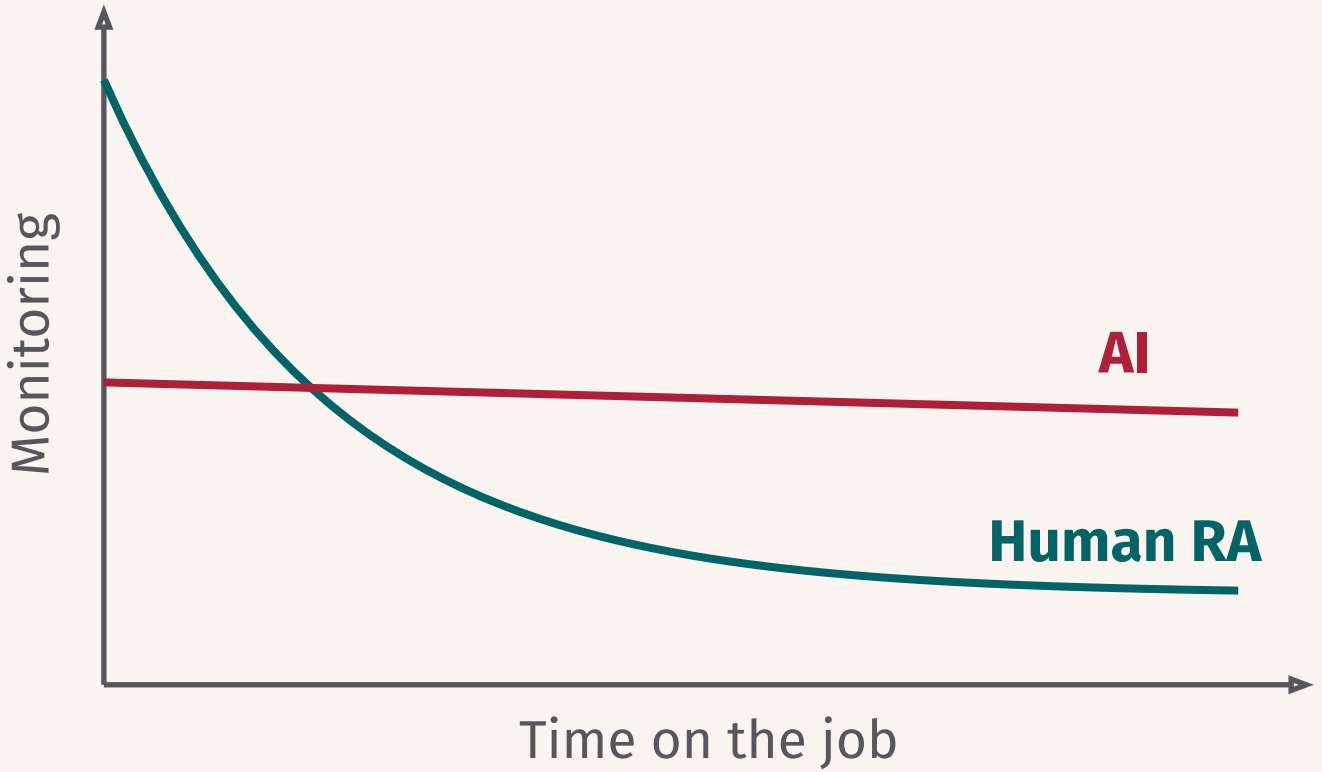
Knows the PI's goal

Complement

Time Constraints!

Delegation $\rightarrow$ a principal–agent problem

Monitoring over time (stylized)



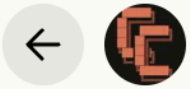
Your AI is born anew every session



Memory is context; weights stay fixed

Learning happens at the AI lab,
not in your lab

… and has no career incentives



r/ClaudeCode • 3mo ago
u/hwknd

You are right, and I apologize.

Bug Report

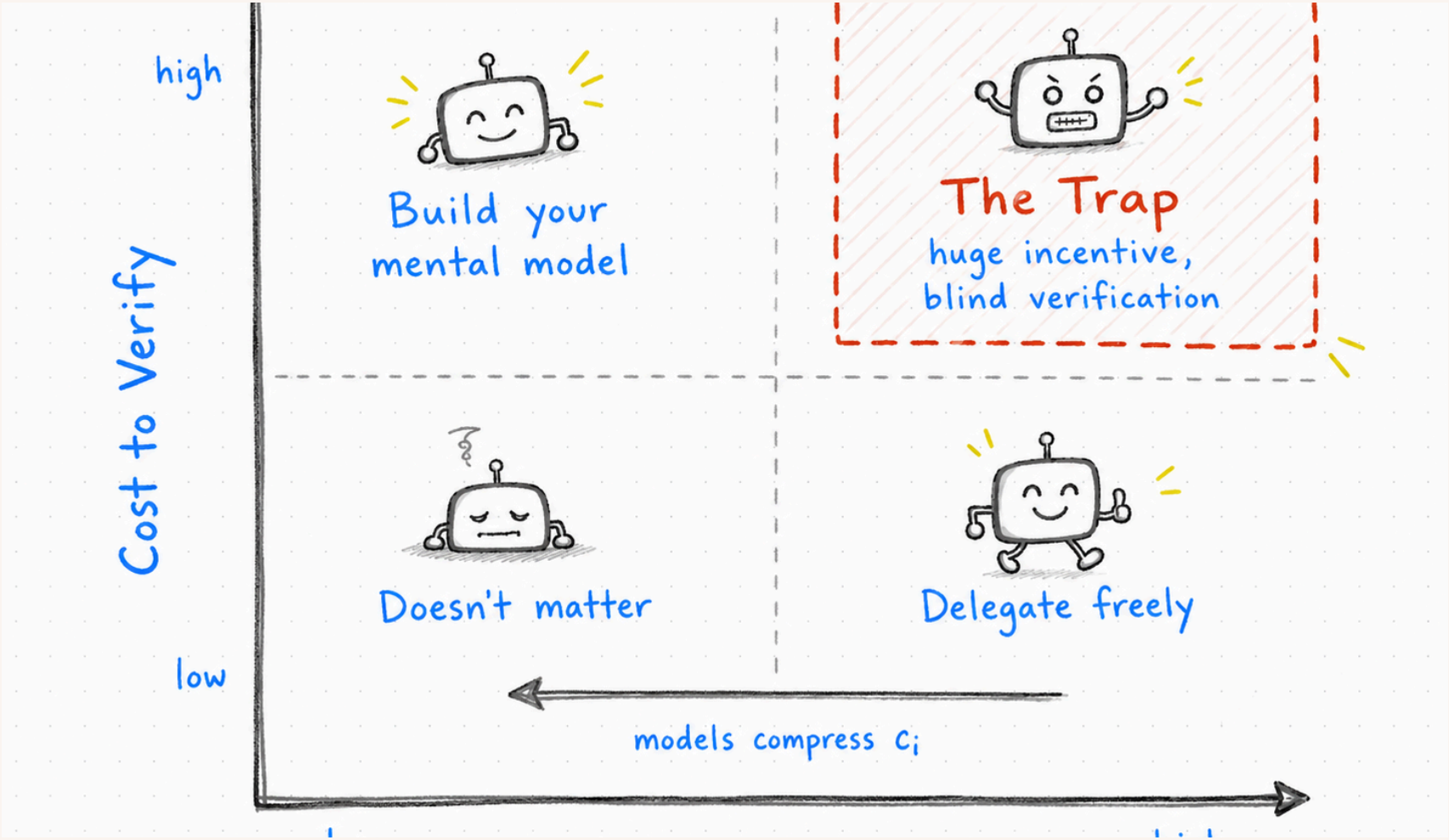
And we are back to this thing *@!(#*$^ ignoring instructions that are explicitly stated in either the global CLAUDE.md file or the project CLAUDE.md file, or both.

> "You're right, and I apologize -"..

AAAAAAAAAARGGGGGGGGHHHHHHH

Fable 5. Under 20% of context used.

Verification debt builds rapidly with AI production



Humans add value where judgment and verification are tied to output

What do economists do?

What does an academic economist do?

Research **Teaching** **Communicating**

> **Economic theory is fiction, and data analysis is journalism.**
>
> The difference between fiction and journalism is the way facts are treated. Fiction is mostly about the story, and is loosely connected to the facts; journalism is mostly about the facts, and is loosely connected to the stories. Historical fiction lies between.
>
> Rather than “Theory and Evidence,” I thus title my two sections: “Economic Fiction” and “Econometric Journalism.”

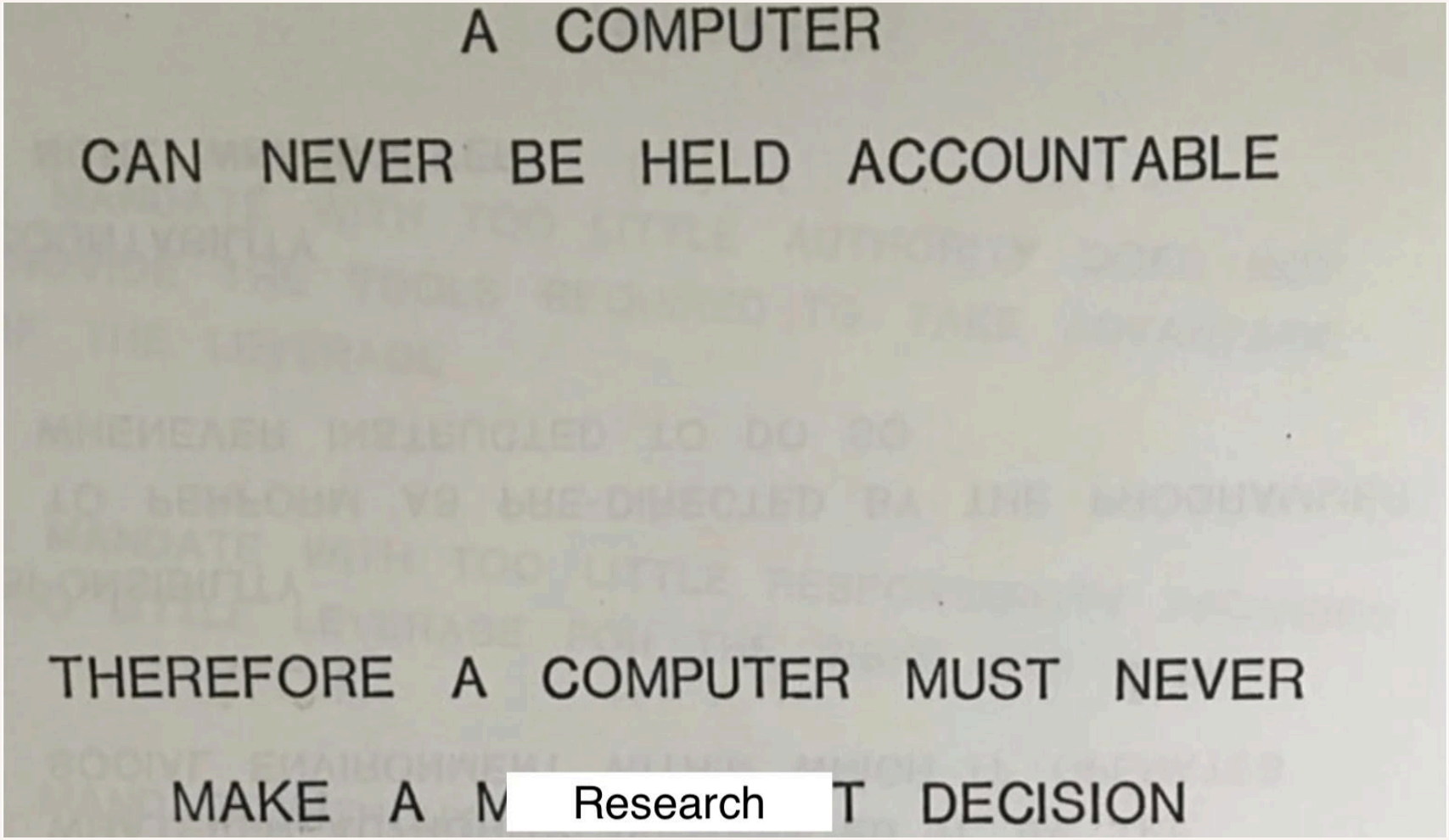
— Ed Leamer, *The Craft of Economics* (p. 10)

Research is fiction and journalism

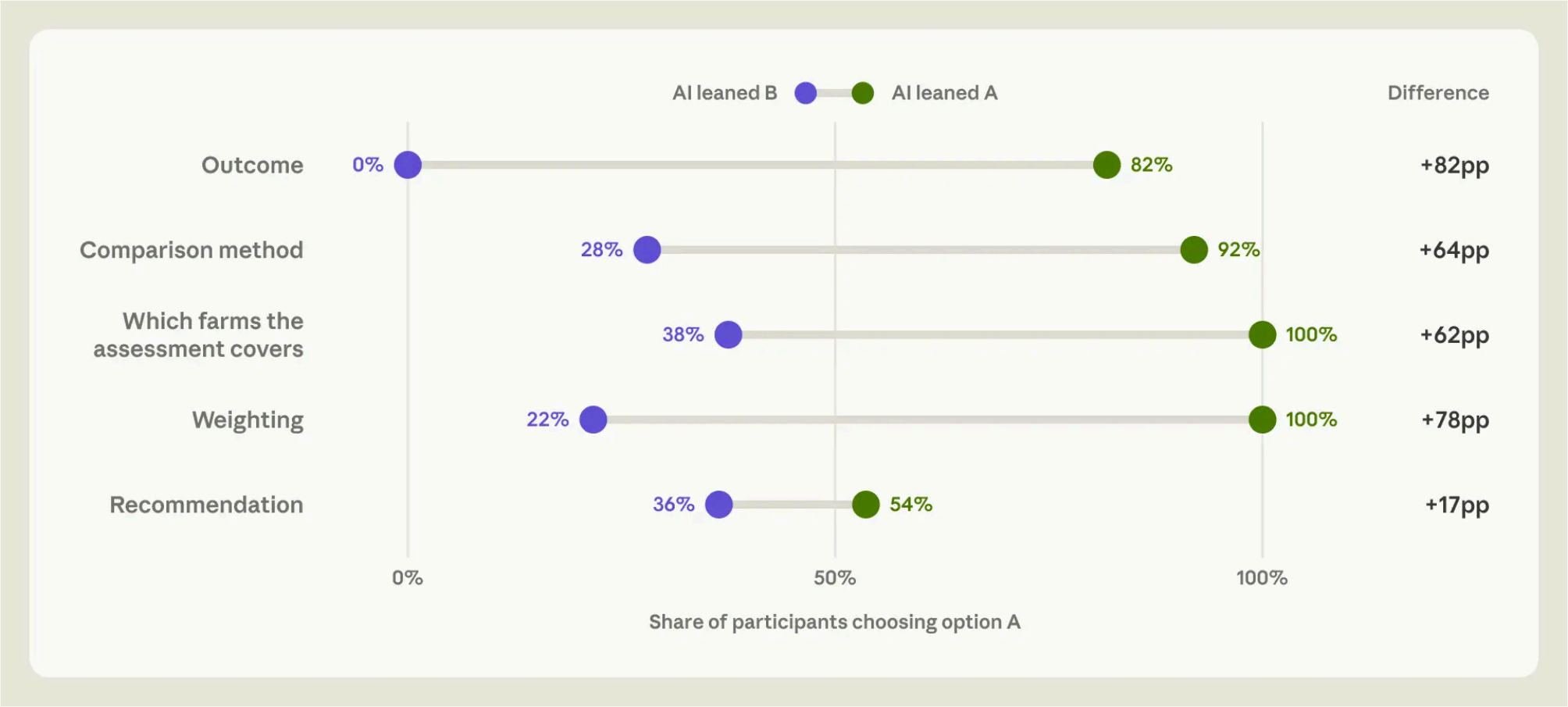
> Fiction can be **revealing, innovating, insightful, even inspirational**, or it can be derivative, boring and tedious.
>
> Journalism can be **balanced and informative and influential**, or it can be prejudiced, confusing, and irrelevant.

— Ed Leamer, *The Craft of Economics*

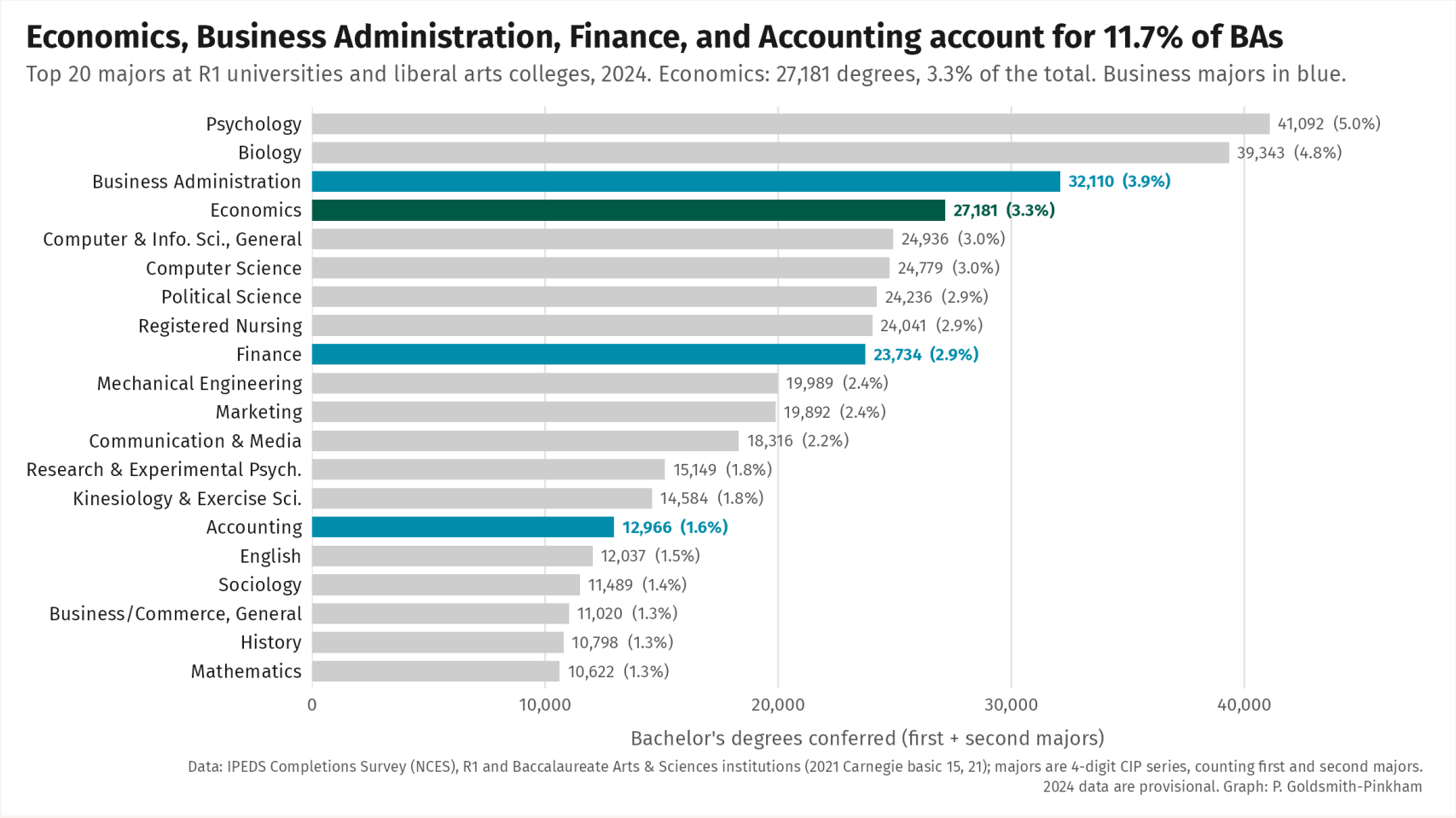
Deciding what has value is a human decision



… but we happily outsource it



Teaching: huge demand for economics



Teaching: the classroom as a Xerox machine

> Prior to 1440, a classroom was a place where the lecturer read the lecture and the students scribbled down notes. … **The classroom was just a big Xerox machine** where the documents for study were "printed" by the students.
>
> Gutenberg's printing press offered a different solution … However, **neither faculty nor students like the conversational approach to education** because it's quite labor-intensive, and its outcome is ambiguous and uncomfortable. … Thus, per the familiar meme: **a classroom has been a place where faculty pretend to teach and students pretend to learn.**

— Ed Leamer, *Journal of Economic Education* (2019)

Teaching: we program our students

> Artificial intelligence means carrying out the code written by a programmer. **That is exactly what our lecture/exam style of education does.** It programs the students to respond in the way that the instructor thinks is most appropriate. But electronic computers are much better at carrying out coded instructions … Moreover, computers are willing to work long hours at low rates of pay, and **they don't join unions and they don't fall in love.**
>
> … The labor market is going to insist more and more on **real human intelligence.**

— Ed Leamer, *Journal of Economic Education* (2019)

Computers are better at running code

… and AI is a temptation good

Rethinking the TA's section

Before

"Here's how to run the regression in R"

Now

"Why is this coefficient biased?"

Judgment is only useful if you can verify the output

Equilibrium

Unclear what the equilibrium impact will be

Cheaper production $\rightarrow$ bigger projects

$\rightarrow$ **more checking & judgment**

The profession has verification debt too

Papers ↑ AI text ↑ Referee time =

So, will AI replace RAs and TAs?

1. The job as it is now: **largely yes**

So, will AI replace RAs and TAs?

1. The job as it is now: **largely yes**

2. The job built on judgment: **more valuable**

So, will AI replace RAs and TAs?

1. The job as it is now: **largely yes**

2. The job built on judgment: **more valuable**

3. The real risk: **the pipeline**

Judgment is learned by doing production under supervision

What will be long-term consequences of not building human capital through training?

What is the point?

AI makes producing analysis cheap

Smaller (worse?) impacts on verification + judgment

Risks:

1. We stop training people
2. We outsource judgment
3. We accrue verification debt