Science's Jagged Frontier

AI and the Uneven Transformation of Knowledge Production

John M. Drake

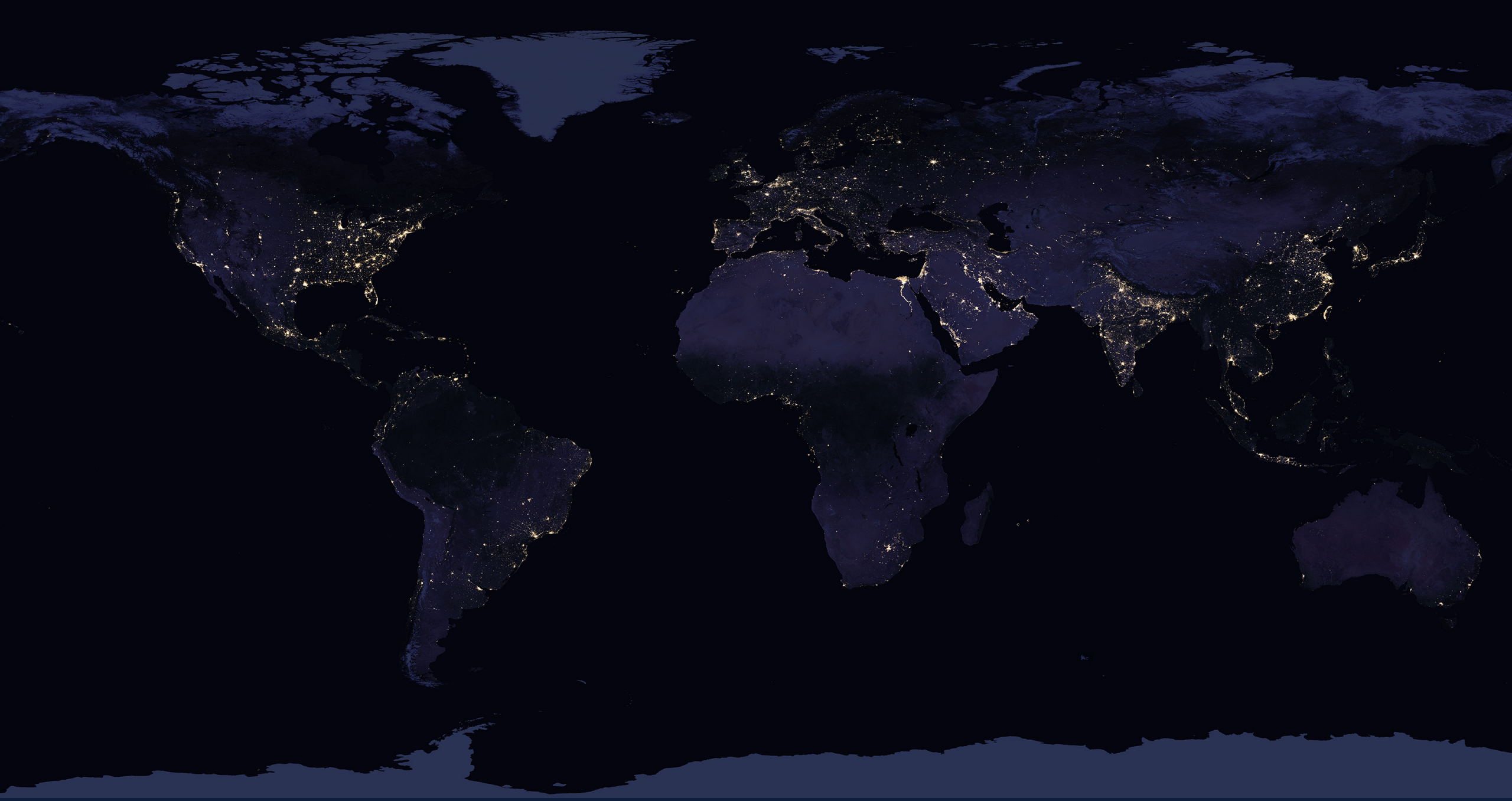
Regents' Professor, Odum School of Ecology & Center for the Ecology of Infectious Diseases, University of Georgia

Keynote · 46th IATUL Conference: Navigating the Future of Knowledge

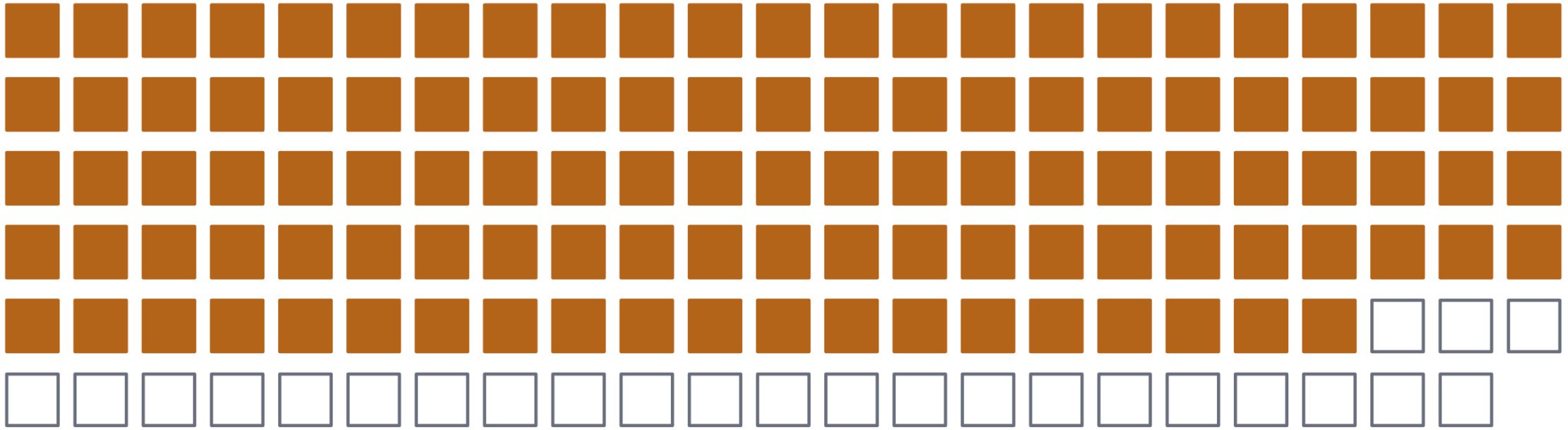
University of Johannesburg Library, Johannesburg, South Africa

November 16, 2026

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker



An AI is named on 112 of the 137 commits in this project



112 commits name Claude, an AI 25 do not

Each square is one logged change to the airline study’s code and text since May 2026. Project history inspected 6 November 2026.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Claims are getting cheaper to produce; establishing their warrant is not keeping pace

- What's actually new
- The frontier in one project
- The AI Scientist
- Does the claim hold?



Drakensberg Amphitheatre, Royal Natal National Park. Photo: Diriye Amey, CC BY 2.0.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Markov counted letters in Pushkin in 1913; the name arrived in 1955

- 1913 · Markov counts vowels and consonants in *Eugene Onegin*
- 1955 · McCarthy, Minsky, Rochester and Shannon propose the Dartmouth project
- 1956 · The workshop meets



Claude Shannon, one of the four authors of the Dartmouth proposal. Photo: Tekniska museet, CC BY 2.0.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Science has used machine learning for years, and so has my lab

- Protein structures
- Telescope surveys
- Medical scans
- Rodents likely to carry future diseases

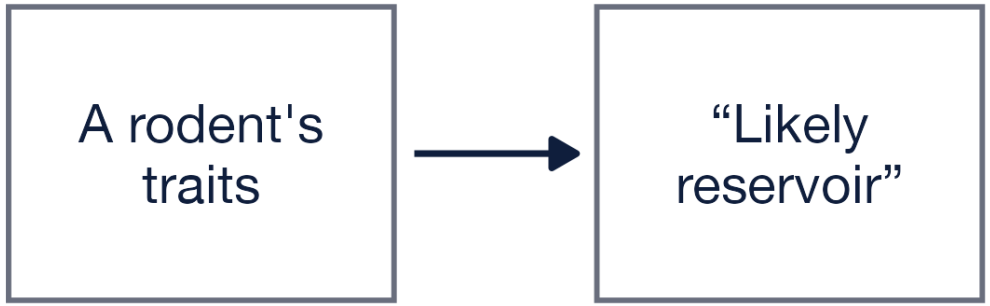


The last is Han, Schmidt, Bowden and Drake, *PNAS*, 2015. Brown rat: Dunpharlain, CC BY-SA 4.0.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

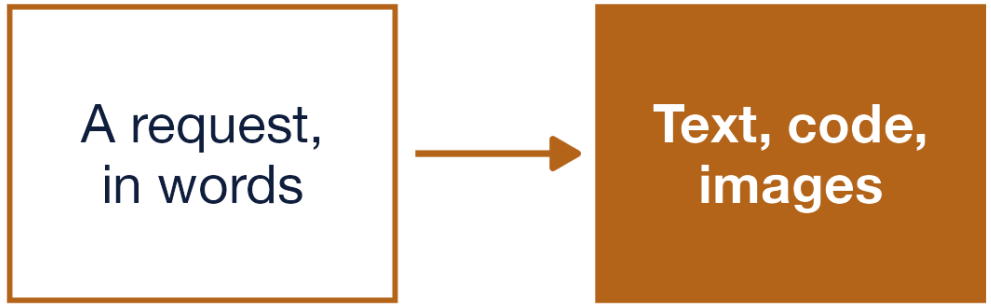
Our 2015 classifier sorted and predicted; a generative model produces

Sorts and predicts



The answer is a label or a number.
A person writes the paper.

Produces

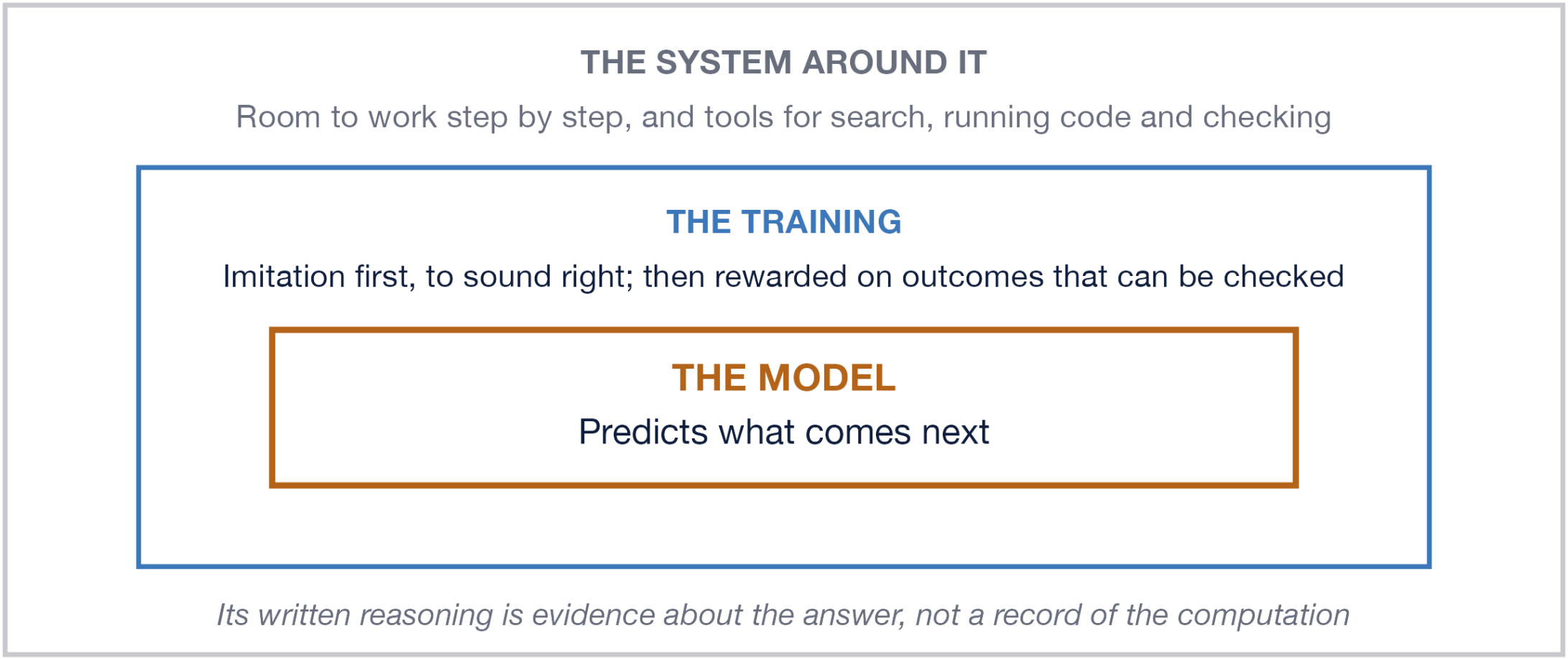


The answer looks like
the materials of scholarship.

The rest of this talk is about the kind on the right.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Still a prediction engine, so why can it reason now?



Training, working time and tools change what a predictor can do. It generates; tools are what connect it to sources.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Better performance gives us more to use and more to check

> *"On the Dangers of Stochastic Parrots"*

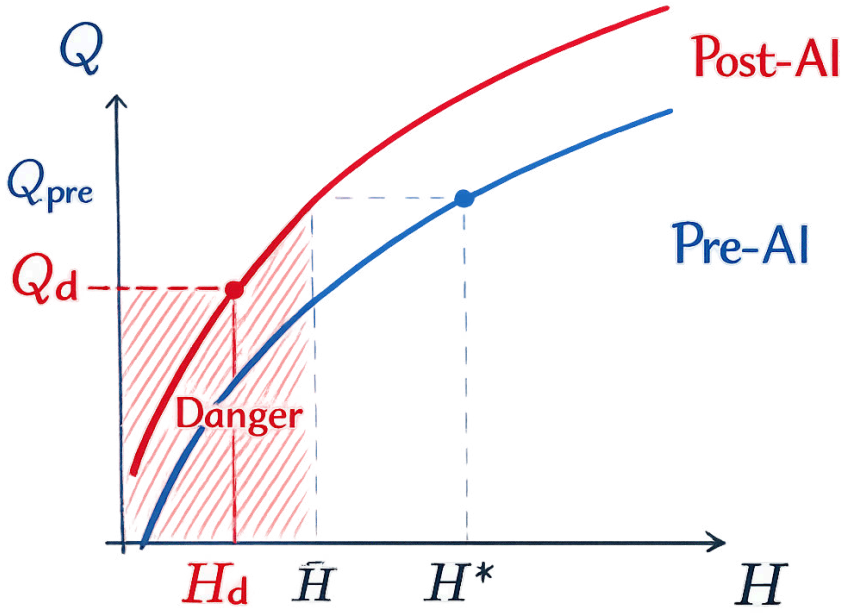


Bender, Gebru, McMillan-Major and Shmitchell, FAccT, 2021. Scarlet macaw: Charles J. Sharp, CC BY-SA 4.0.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

What's new is that the output looks like scholarship

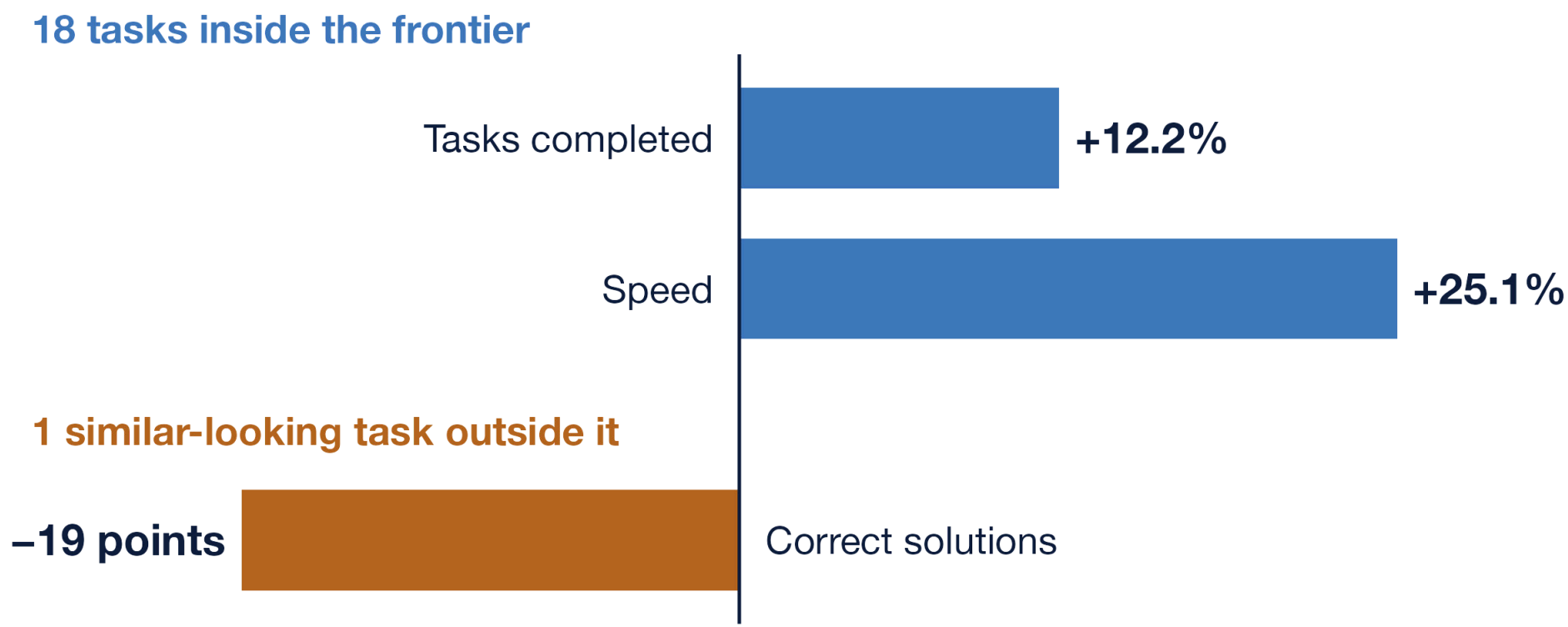
Text, code and figures, on request, at falling cost



Conceptual. Redrawn after Scott Cunningham, Federal Reserve Board of Governors, March 2026.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Competence changes sharply from task to task, and the boundary is hard to see



758 consultants at Boston Consulting Group were randomly assigned to work with or without GPT-4. Inside the frontier: 12.2% more tasks completed, 25.1% less time taken. Outside it, 84.5% were correct without AI and 60–71% with it. Dell'Acqua et al., *Organization Science*, 2026.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

The frontier in one project

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

In the airline study, AI helped at almost every stage

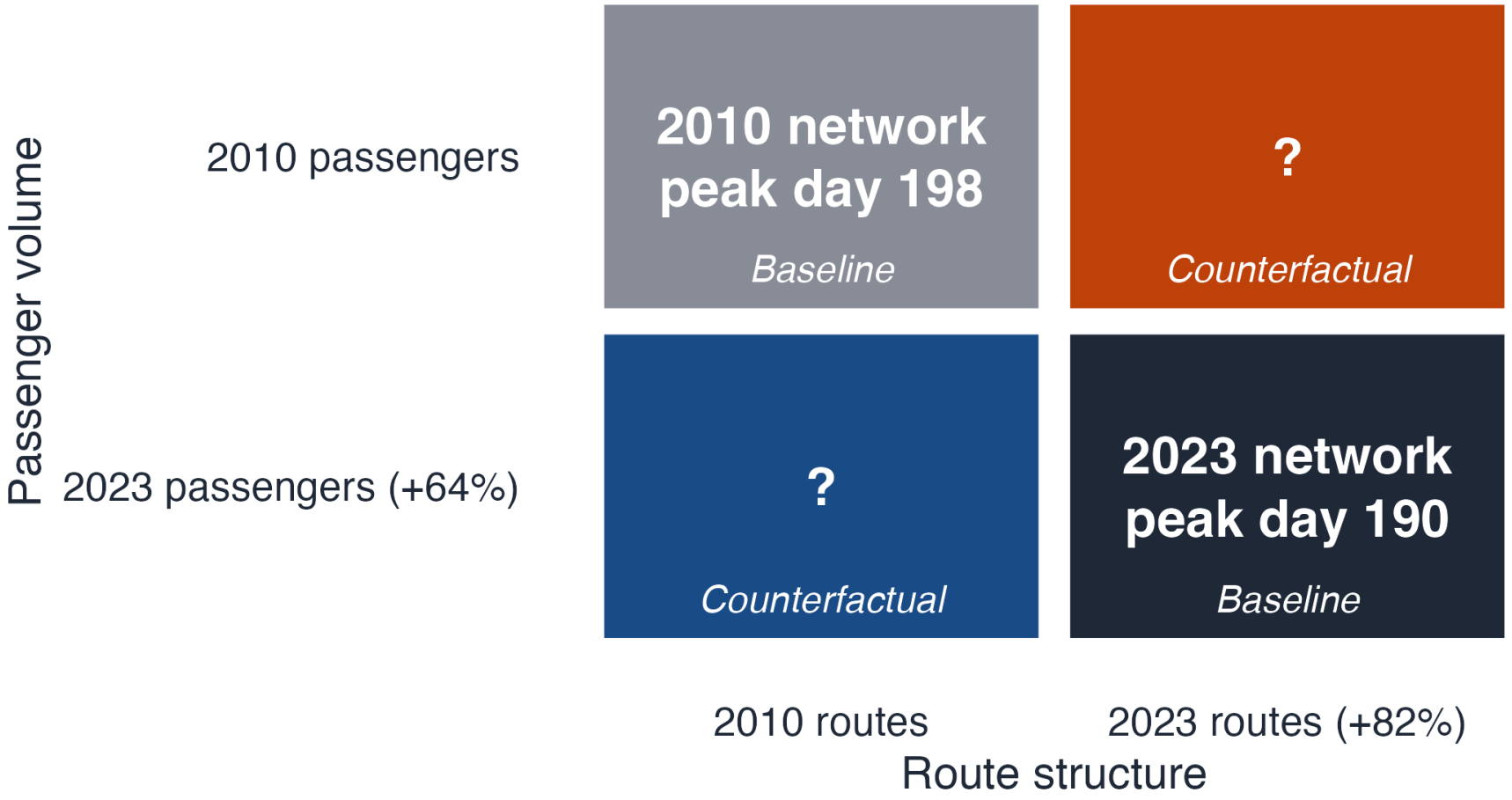
- Planning
- Adversarial critique
- Code audit
- Fact-checking
- Updating text
- Additional experiments
- File management
- Ablation studies
- Figures
- Talk preparation

○ **Counting the passengers** someone else's records

The passenger records are Official Airline Guide data for four years between 2010 and 2023.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

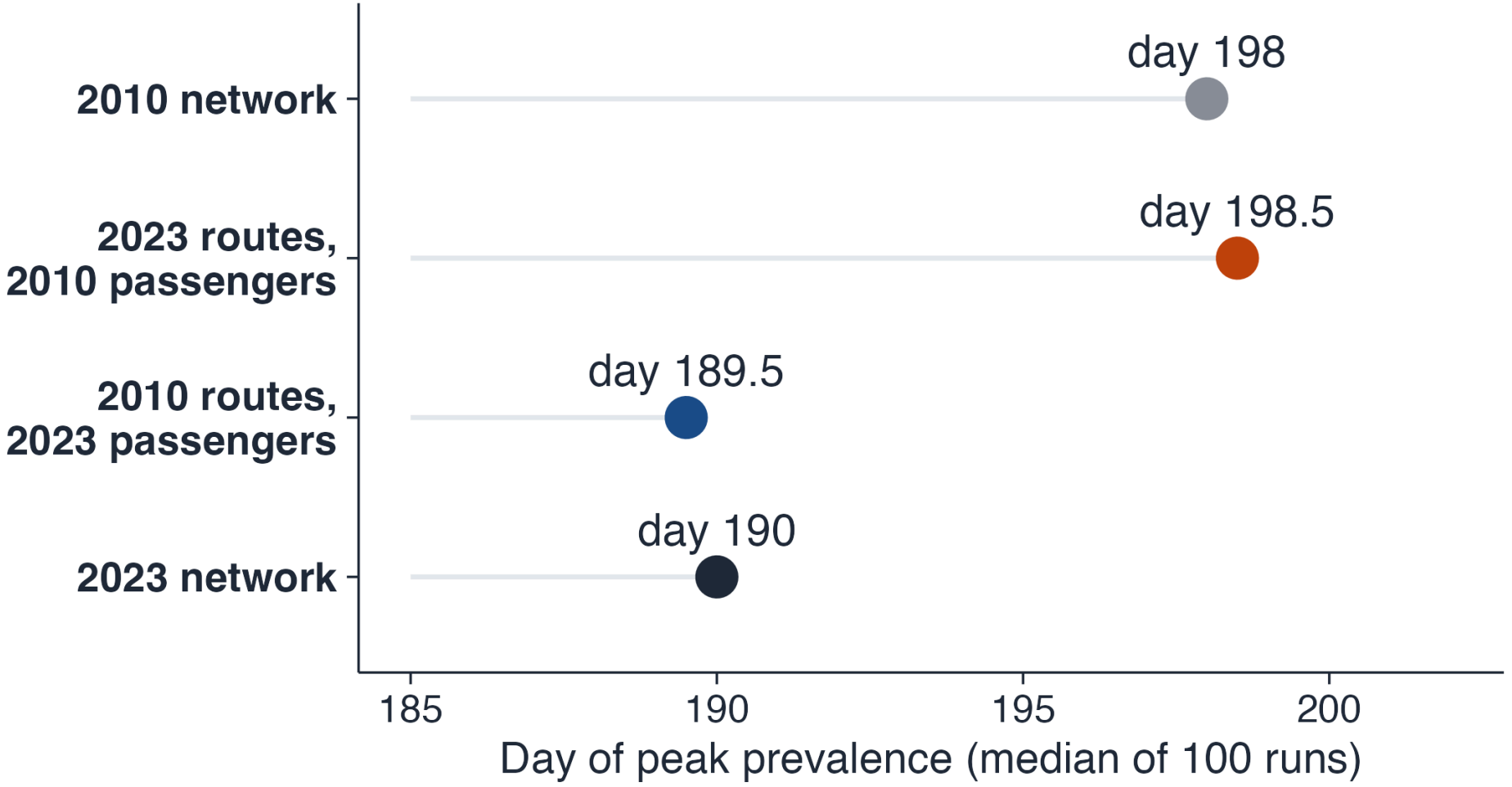
Cheap experiments change what gets tried



Swapping one year's routes with another year's passenger volume separates two things that rose together.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

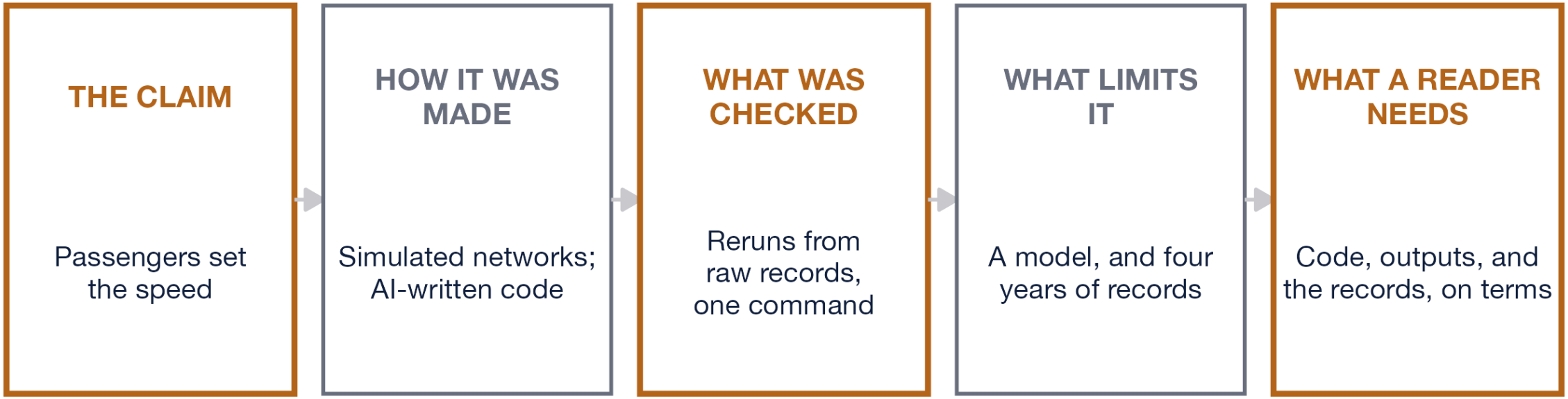
Passengers, more than routes, set the speed in these simulations



Each counterfactual peaks beside the real network with the same passenger volume. Drake and Winter, Roscoff, 2026.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

What would you need, to believe that claim?

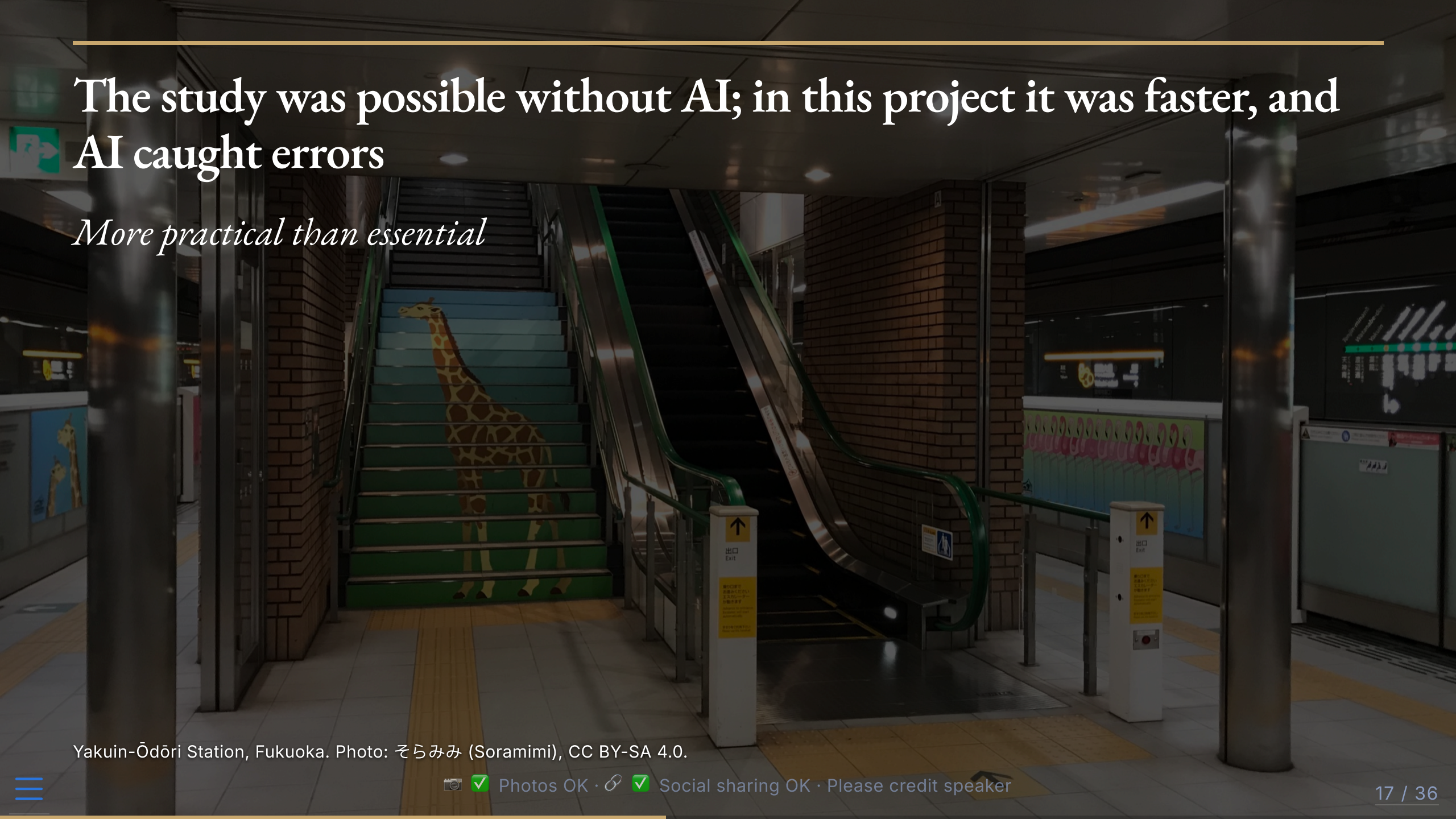


A library can archive the three outlined links: the claim, the record of what was checked, and the code, outputs and records on deposit.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

The study was possible without AI; in this project it was faster, and AI caught errors

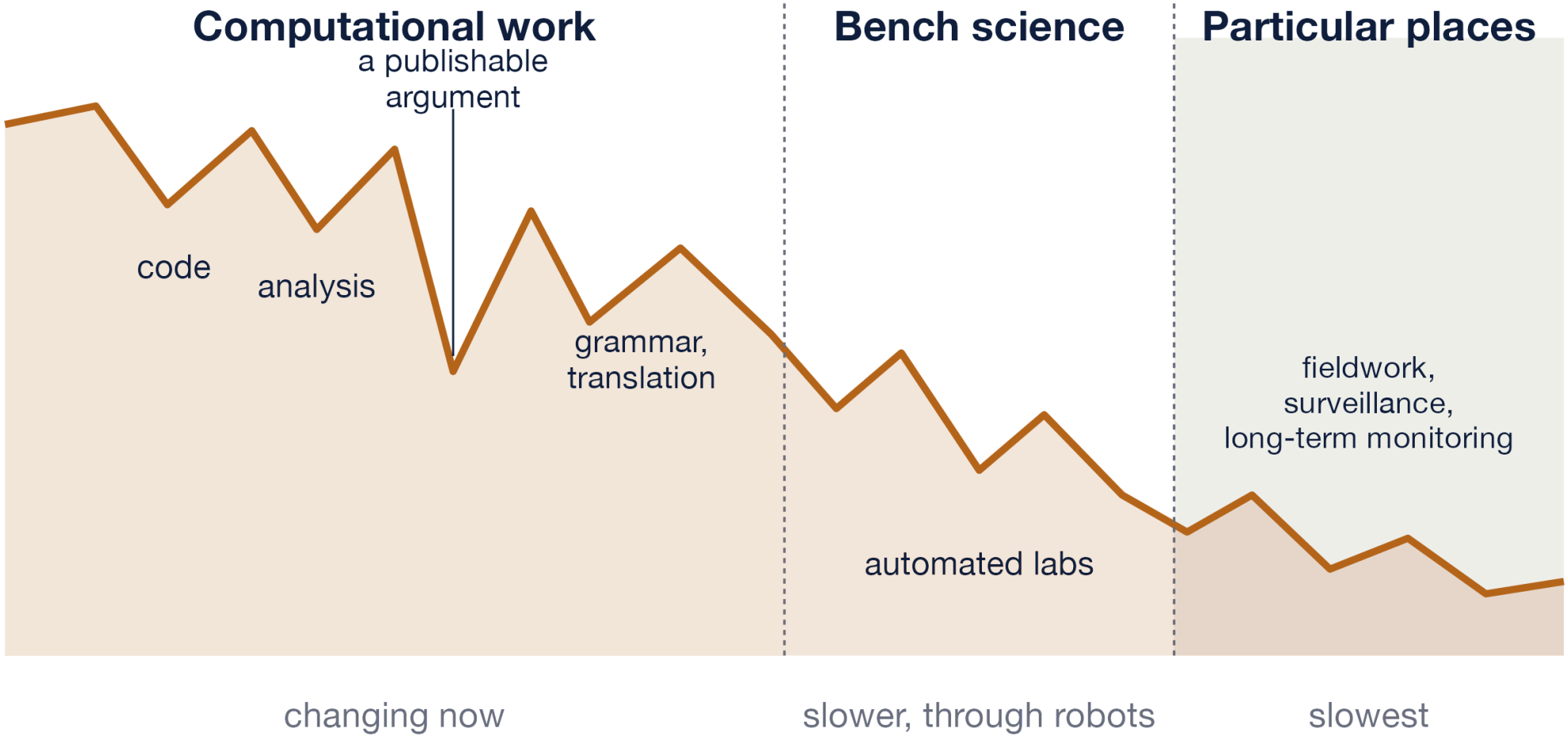
More practical than essential



Yakuin-Ōdōri Station, Fukuoka. Photo: そらみみ (Soramimi), CC BY-SA 4.0.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Different kinds of research face different constraints



A hypothesis, drawn from what each kind of work has to touch. It is not a measurement, and it will move.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

When analysis gets cheap, useful observations get (relatively) more valuable

A thirty-year census of forest plots can't be reproduced by a clever script



Drake, *Ecology Letters*, 2026. Okefenokee National Wildlife Refuge. Photo: Larry A. Woodward, USFWS.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

A local source of data does not guarantee a local share of its value

Custody, authority, capacity and benefit are separate negotiations

When analysis is abundant, the advantage shifts toward whoever can produce observations nobody else has — and the terms on which they can be used are a decision, not a default.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

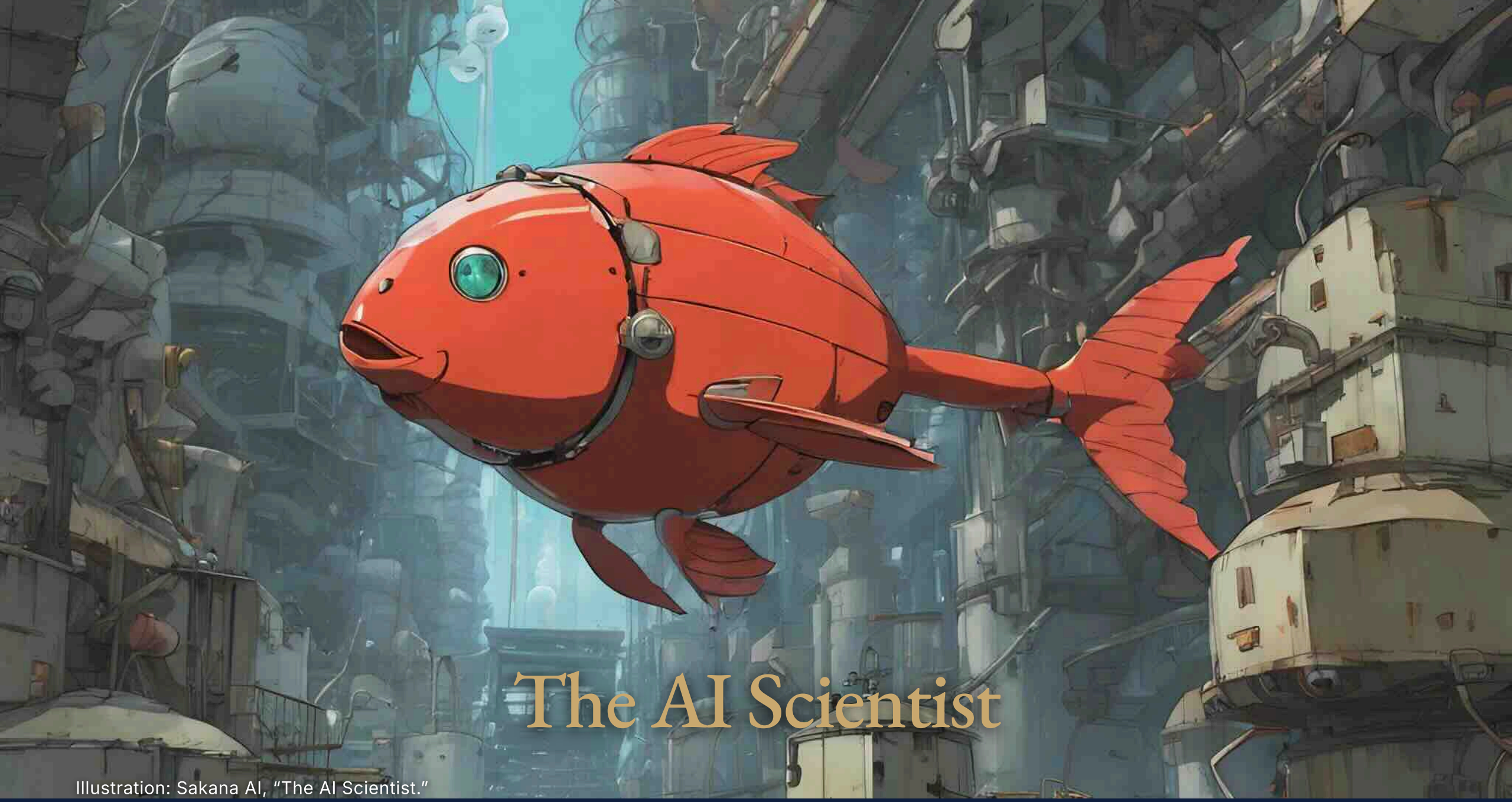
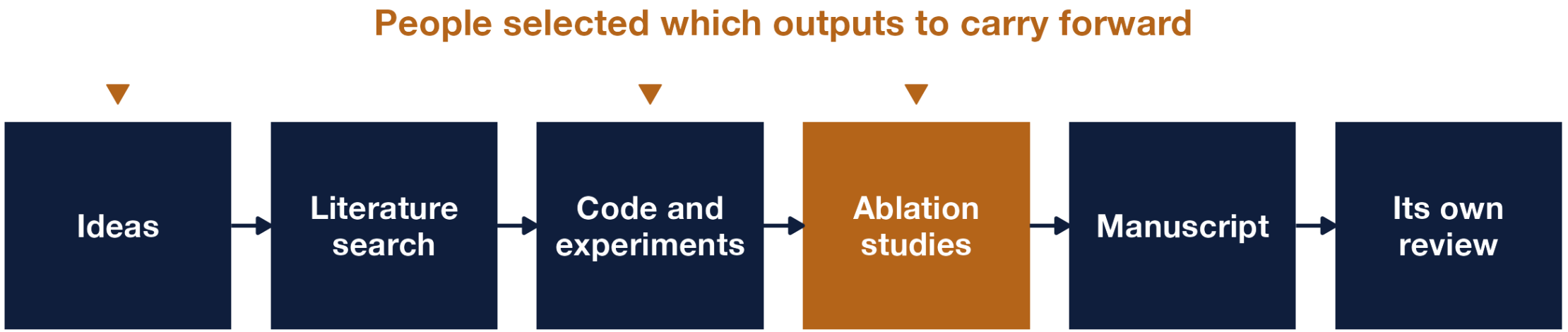


Illustration: Sakana AI, “The AI Scientist.”

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

A system can generate a whole study; people still select



The manuscripts themselves were not edited.

People filtered promising outputs along the chain; the manuscripts themselves were not edited. The AI Scientist works in machine learning, where experiments run entirely on a computer. Lu et al., *Nature*, 2026.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

One selected paper scored above a workshop's acceptance threshold

- Three of 43 papers reviewed were machine-written
- Workshop acceptance rate: 70%; main conference: 32%
- Withdrawn afterward, by prior agreement

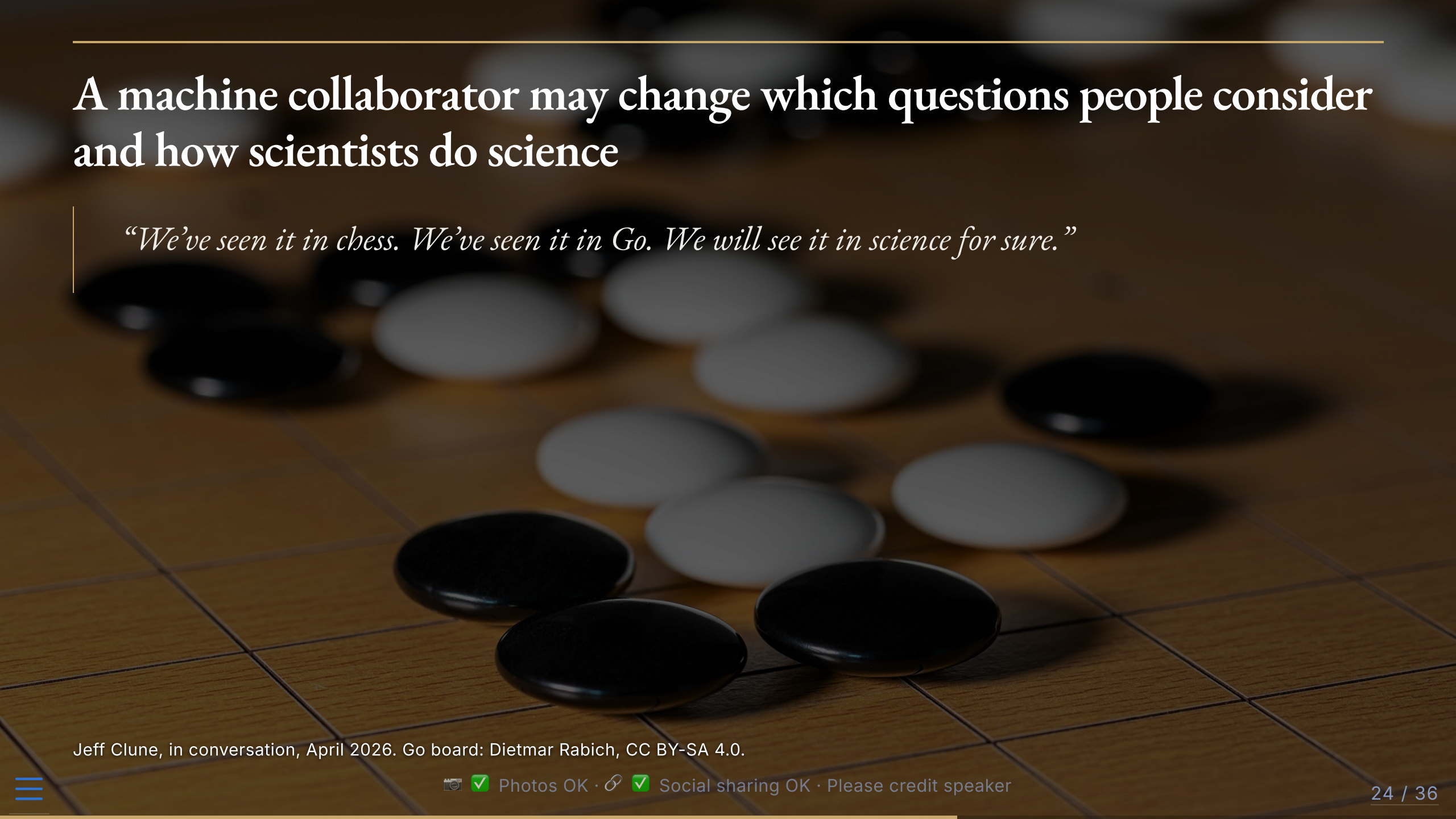


Lu et al., *Nature*, 2026.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

A machine collaborator may change which questions people consider and how scientists do science

> *"We've seen it in chess. We've seen it in Go. We will see it in science for sure."*



Jeff Clune, in conversation, April 2026. Go board: Dietmar Rabich, CC BY-SA 4.0.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Generated designs still have to survive experimental selection

- Two genome models generated candidate bacteriophage genomes
- 285 were synthesized; 16 infected and killed *E. coli*
- Design proposed the sequence; the laboratory decided what worked

Stanford and the Arc Institute, *Science*, 2026, following a 2025 preprint. Bacteriophage ΦX174, 5,386 bases. Photo: phage plaques on *M. smegmatis*, Wazzzup7up, public domain.

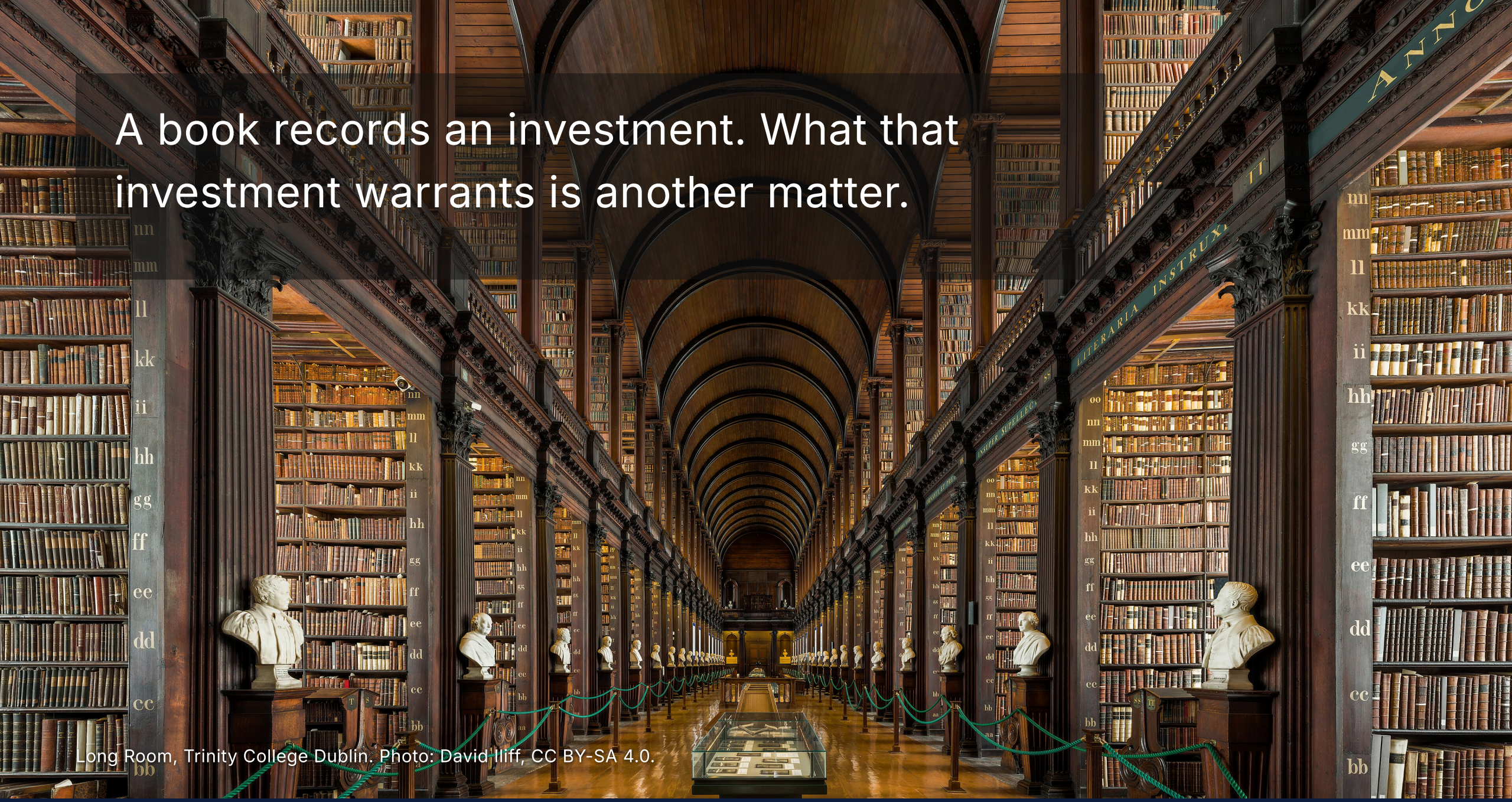
📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Curating information: cheap to produce is not the same as checked, or findable

Abundance creates two problems at once

- A claim can be easy to find and poorly supported
- A claim can be carefully supported and effectively invisible
- What a system can retrieve shapes what it can cite
- What it cites shapes what a reader meets as knowledge
- Provenance: which version is canonical, and who vouches for it

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker



Long Room, Trinity College Dublin. Photo: David Iliff, CC BY-SA 4.0.

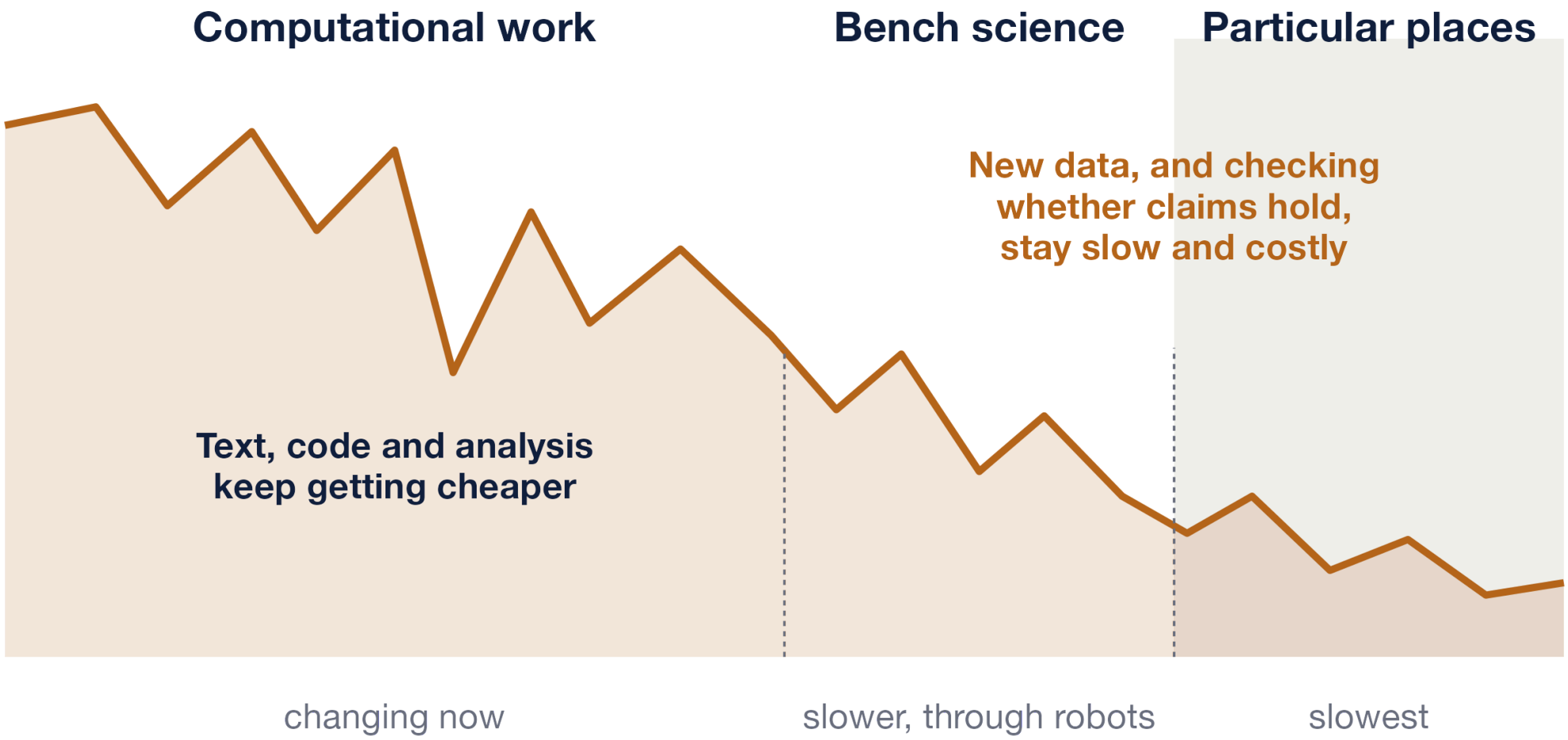
Fluency is becoming cheaper to imitate

> *"AI-generated papers shred the long-standing (if rough) correlation between authors applying effort and the work having value."*

Editorial, *Nature*, 26 March 2026.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Cheaper checks can coexist with a growing checking burden



Inside the frontier, text, code and analysis keep getting cheaper. The volume of claims to evaluate is growing faster.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Does the claim hold?



Pont du Gard, France. Photo: Martin Kraft, CC BY-SA 3.0.

Record which checks succeeded, and what was never assessed

Five distinct checks, not one price list: costs and failure modes vary by study



Record each one as passed, failed or NOT ASSESSED — and who or what performed it

"Checked" is not one thing. These are five distinct achievements, with different costs in different studies.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Abolish mandatory disclosure of AI assistance in writing

> *"AI was used in the preparation of this manuscript."*

What the label does not tell a reader

- Which tool, and what it did
- Whether it touched the prose, the data or the analysis
- Whether anyone checked the result

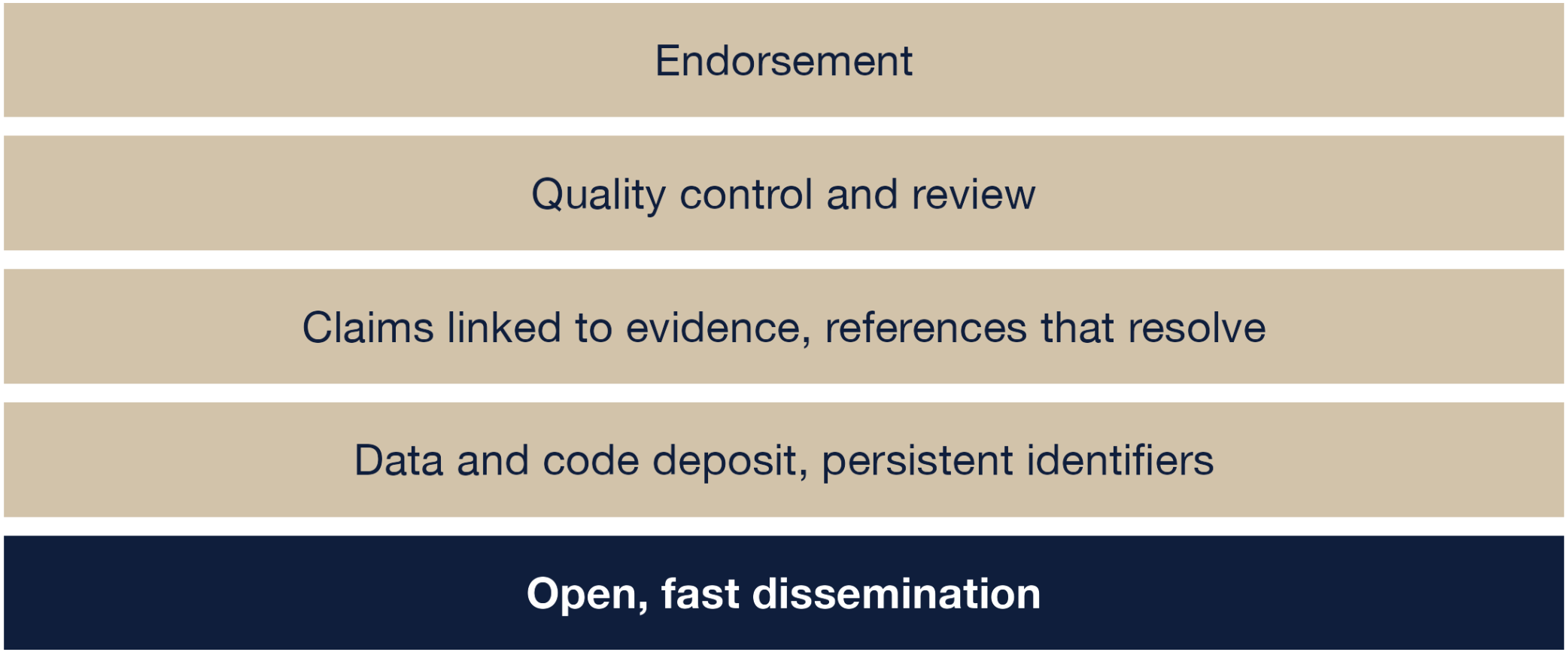
What to require instead

- Analytical uses described in the methods, like any other instrument
- An author who answers for every assertion
- Specific rules for confidentiality, attribution and permissions

Enforcing the label has a cost: AI detectors disproportionately flag writing by non-native English speakers (Liang et al., *Patterns*, 2023).

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Publish a record of checks alongside each version of the work



Dissemination and evaluation can have distinct stages, and each check can say what it covered.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Who controls the account a collection gives of itself?

“Explain this project to me.”

“Where is the simulation model?”

Good documentation is what the model reads.

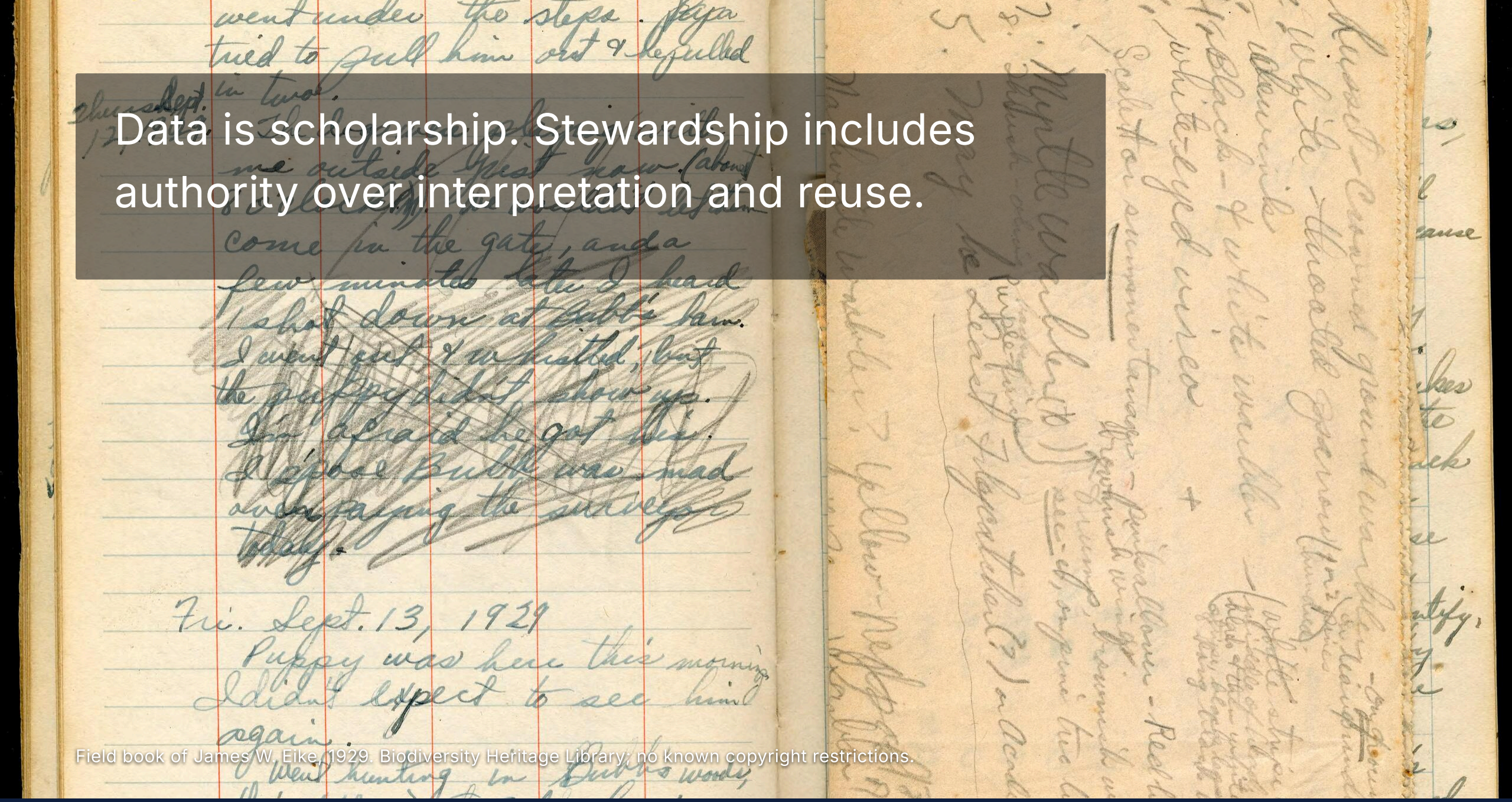


```
OAGData/
├── data/raw/
├── code/
│   ├── sim_utils.R
│   ├── reproduce_all.R
│   └── run_all.R
├── analysis/
├── derived/
├── manuscript/
└── docs/
```

This is the airline study’s real file structure. A model reading a well-described collection answers better than one reading a mess.

📷 ✅ Photos OK · 🔗 ✅ Social sharing OK · Please credit speaker

Data is scholarship. Stewardship includes authority over interpretation and reuse.



Field book of James W. Eike, 1929. Biodiversity Heritage Library; no known copyright restrictions.


Make the evidence as easy to see as the claim.