Generative AI in Astronomical Research

How astronomers use LLMs?

Emerging results and ethical concerns from professional surveys

Morgan **Fouesneau**

Max Planck Institute for Astronomy – *Data Science Dept.*

fouesneau@mpia.de

This talk would not be complete without a "survey" of its own.

There will be four questions. These are fully anonymous

The **elephant** in the room

Let's start with two obvious questions

POLL #1

What is **your career stage**?

Master · PhD · Postdoc · Permanent · Other

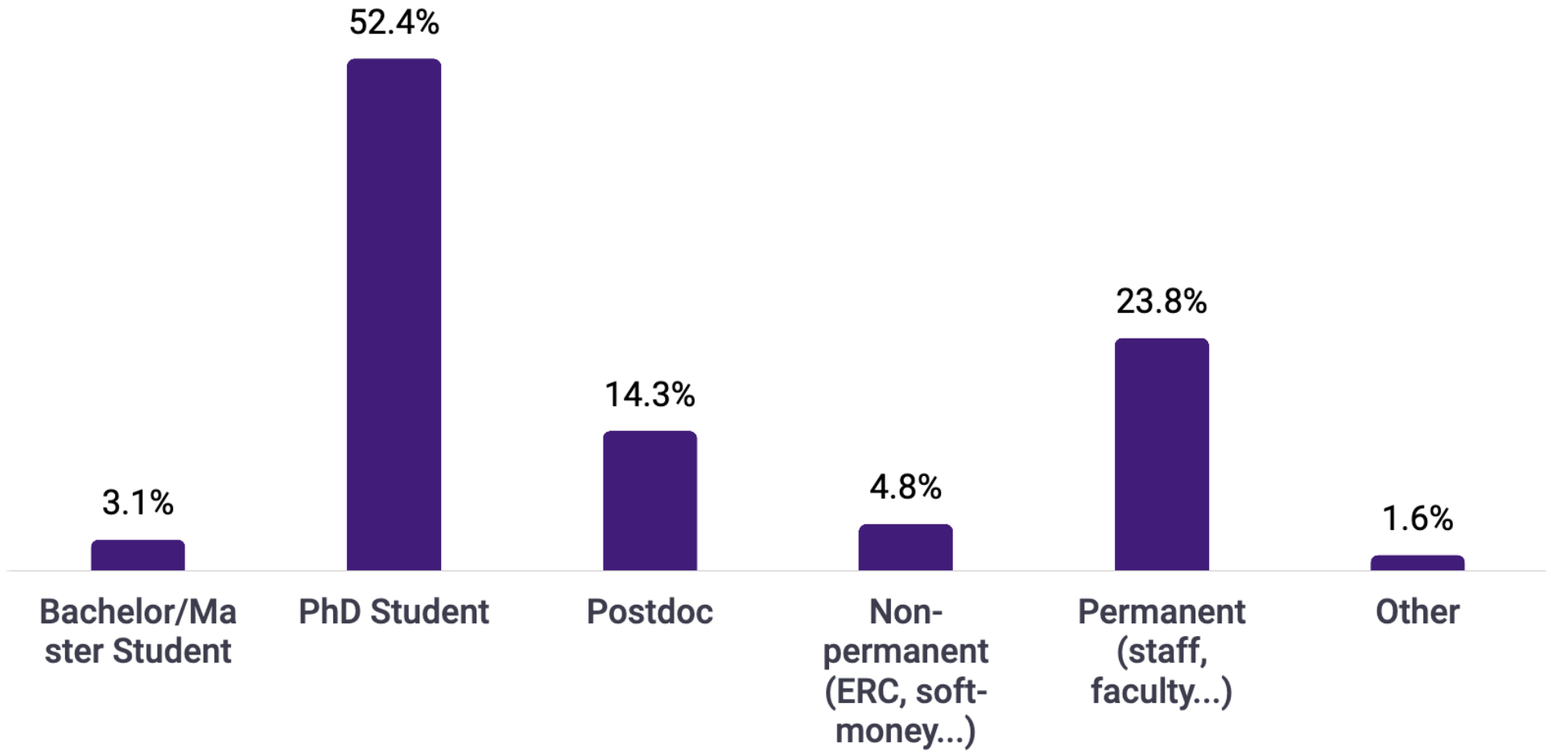
POLL #2

How many of you used an LLM **this week**?

The **elephant** in the room

Let's start with tw

What is your career stage?



73 total participants | 63 votes

The **elephant** in the room

Let's start wit

Did you use an LLM **this week**

48 total participants | **38** votes

You are not alone

The adoption numbers converge across every survey

81%

adoption rate cross-disciplinary
Liao et al. 2024 — 816 researchers

84%

at MPIA (n=98)
internal survey from July 2026

similar numbers for
20

countries survey of academics
Mohammadi et al. 2026

This is not a niche. It is the new normal.

Where the models actually deliver value

Raw Input / Friction	AI-Augmented Outcome
Manual boilerplate coding (FITS files, ADQL queries, parsers, bug finding)	**Code generation & Scaffolding** *Ziegler et al. 2024:* 55.8% faster completion with AI assistance
Blank page paralysis and unstructured ideas how to approach a problem, early drafts paralysis	**Manuscript drafting & Brainstorming** *Noy and Zhang 2023:* 40% reduction in writing time, 18% quality improvement (453 college-educated pros)
Overwhelming volume of daily publications astronomy and cross-disciplinary; knowledge silos	**Literature Search & Synthesis** Rising domain-specific searching tools e.g., Pathfinder *(Iyer et al. 2024)*, RAG bot *(Hyk et al. 2025)*, OpenScholar *(Asai et al. 2026)*
Structural English-language publication barriers astronomy and cross-disciplinary; knowledge silos	**Accessibility & communication** **massive equalizer** for non native speaking astronomers (Shyr et al. 2024)

How do **we** use AI?

POLL #3

What are the **primary tasks** you tackle with LLMs?

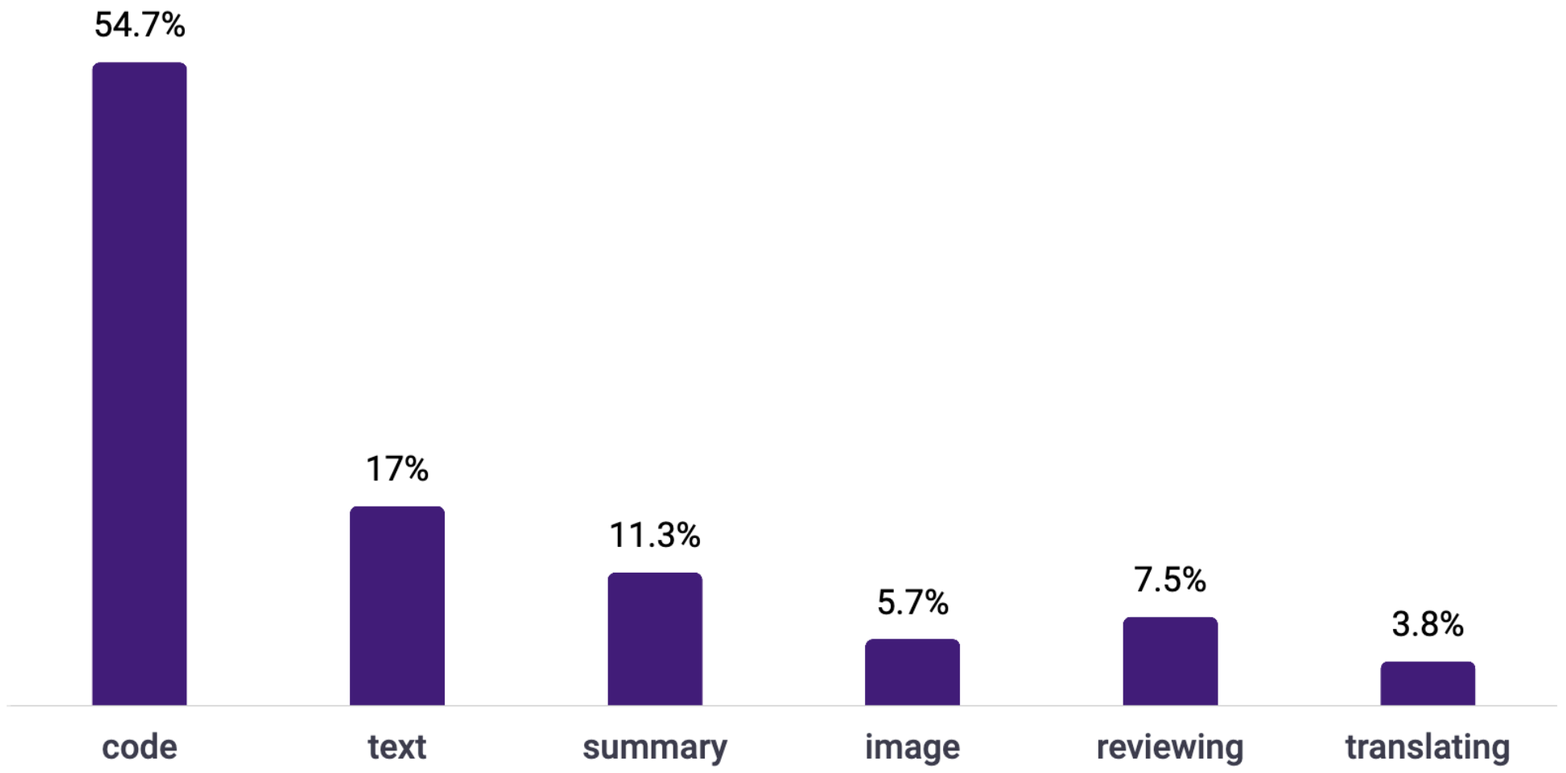
code · text · image · summary · other

pigeonhole.at

ML4ASTRO3BIRD

How do **we** use AI?

What are the primary tasks you tackle with LLMs? (you can pick more than one)



47 total participants | **53** votes

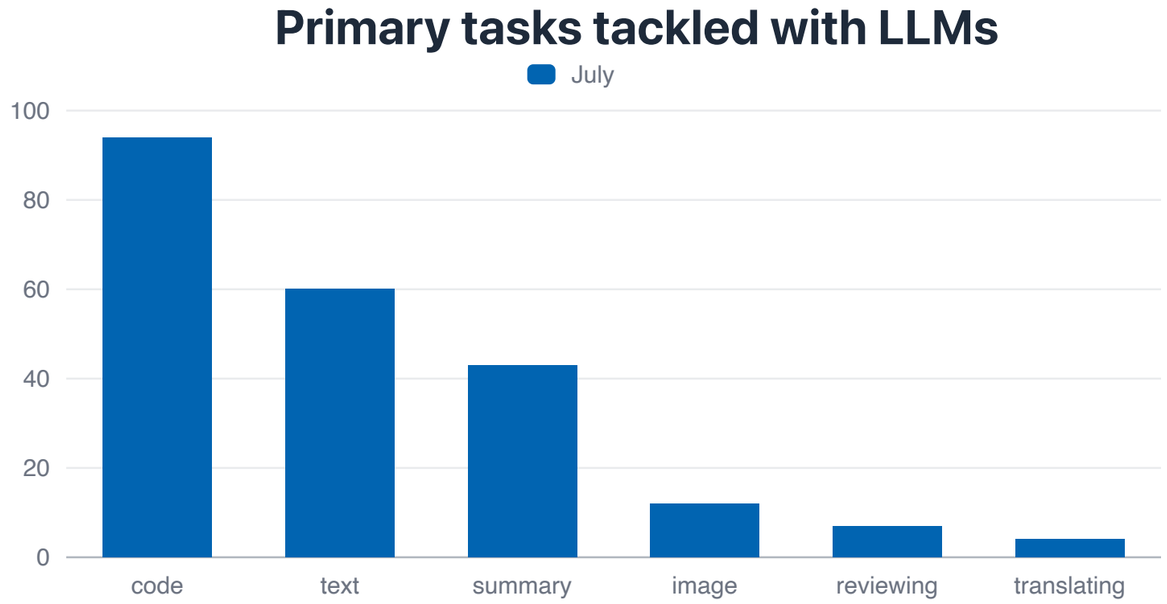
How do we use AI?

92%

use LLMs for coding

58%

use agentic interactions (vs. chat)



Survey of 100 users at MPIA in 2026

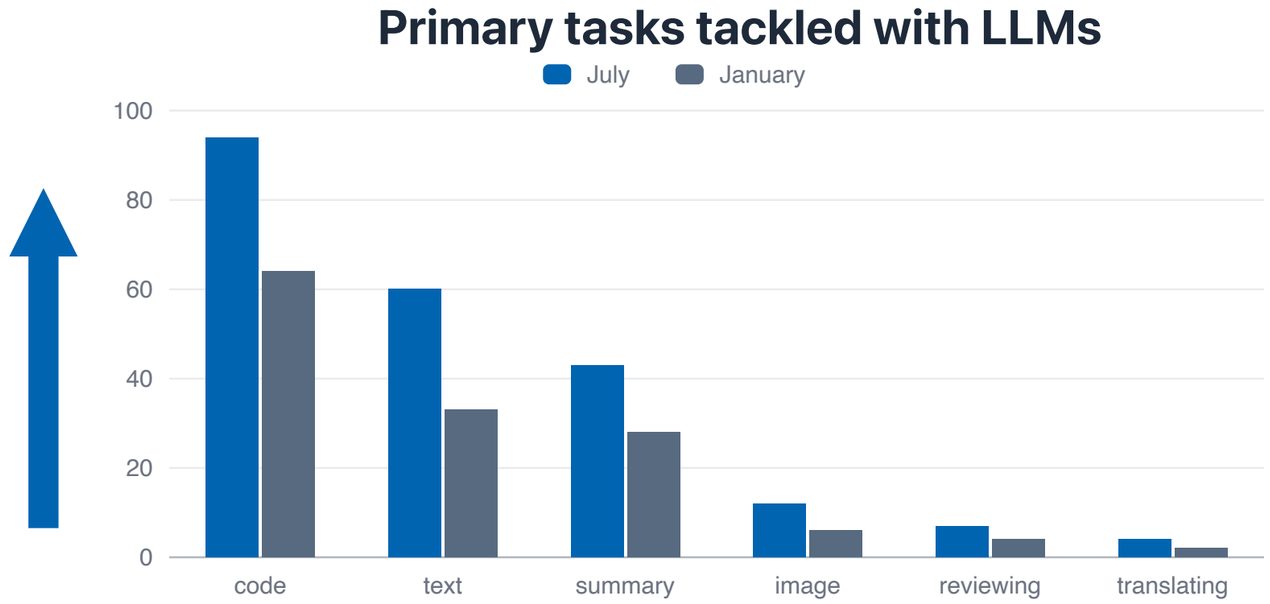
Adoption rate has vastly increased

92%

use LLMs for coding

58%

use agentic interactions (vs. chat)



Survey of 100 users at MPIA in 2026

6 months **massive evolution**

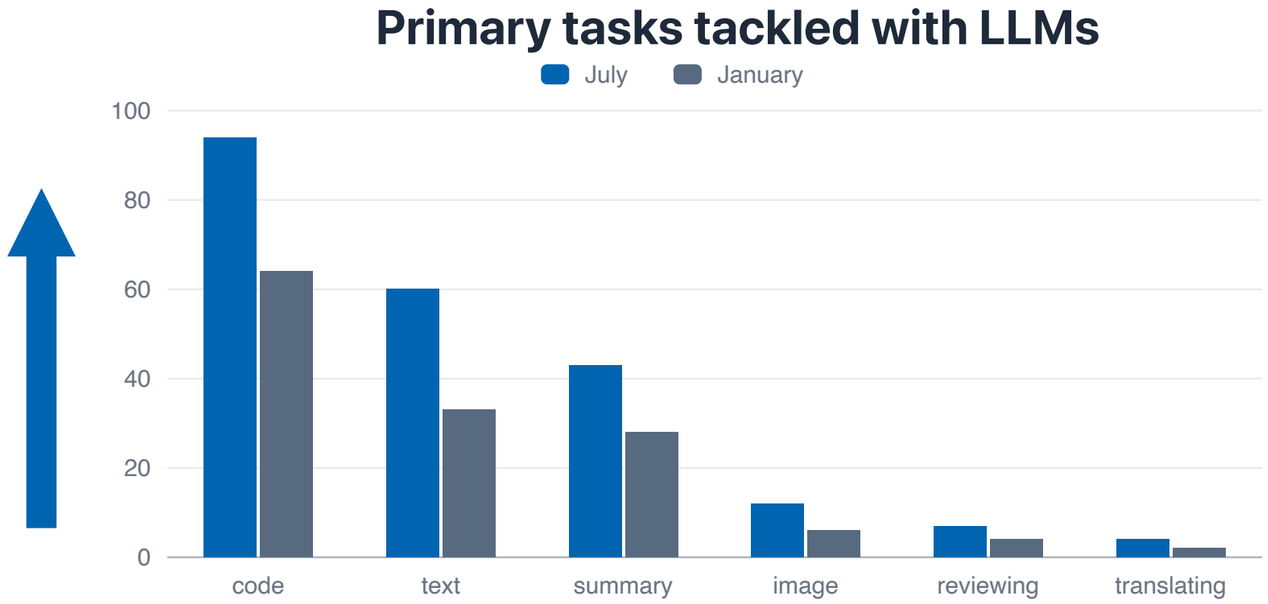
Adoption rate has vastly increased

92%

use LLMs for coding

58%

use agentic interactions (vs. chat)



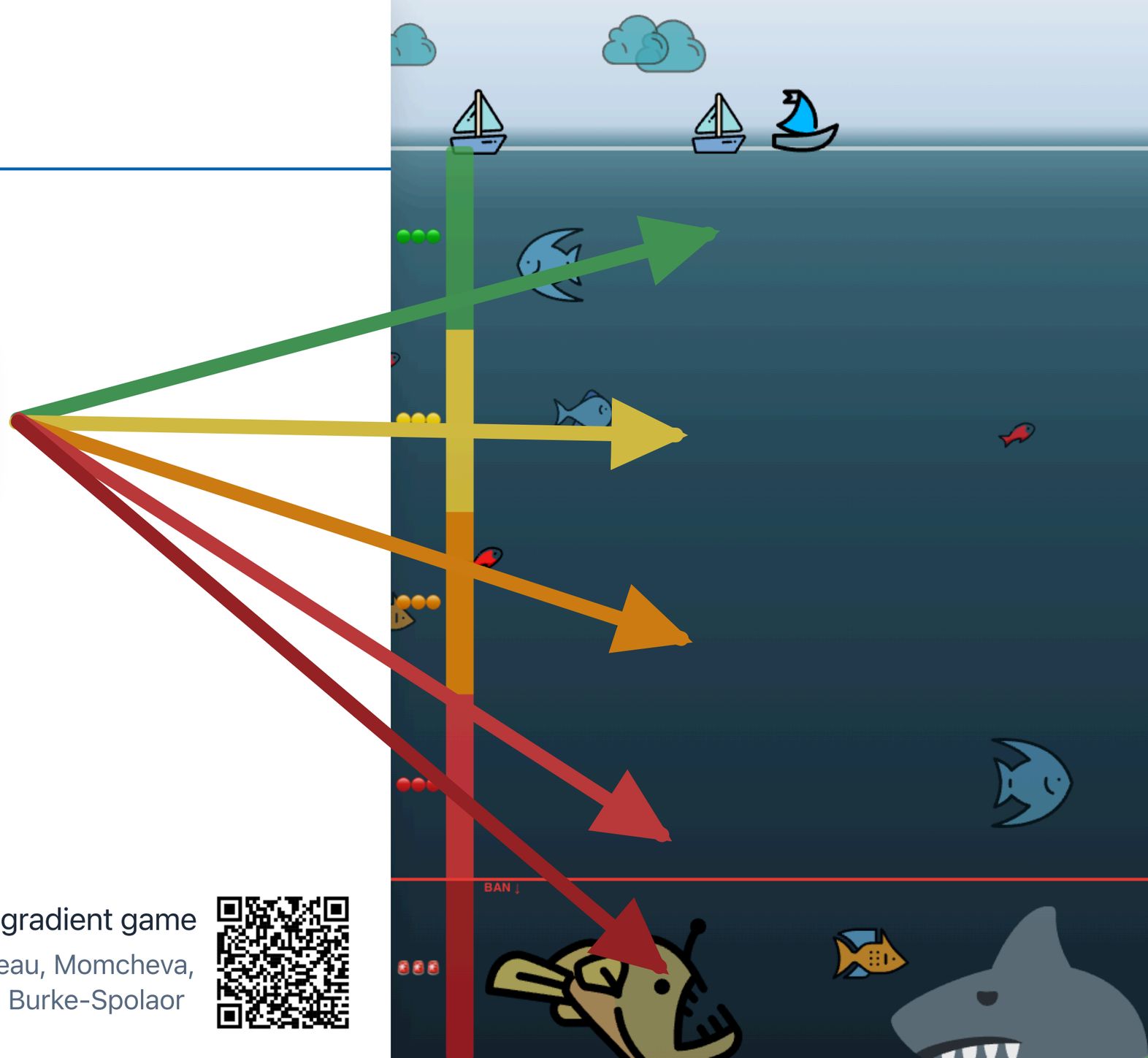
Survey of 100 users at MPIA in 2026

6 months **massive evolution**

Do we all agree on good practice?

Ethical gradient

little scenario
using LLMs/AI



Ethical gradient game

Fouesneau, Momcheva,
Salditt, Burke-Spolaor

Ethical gradient

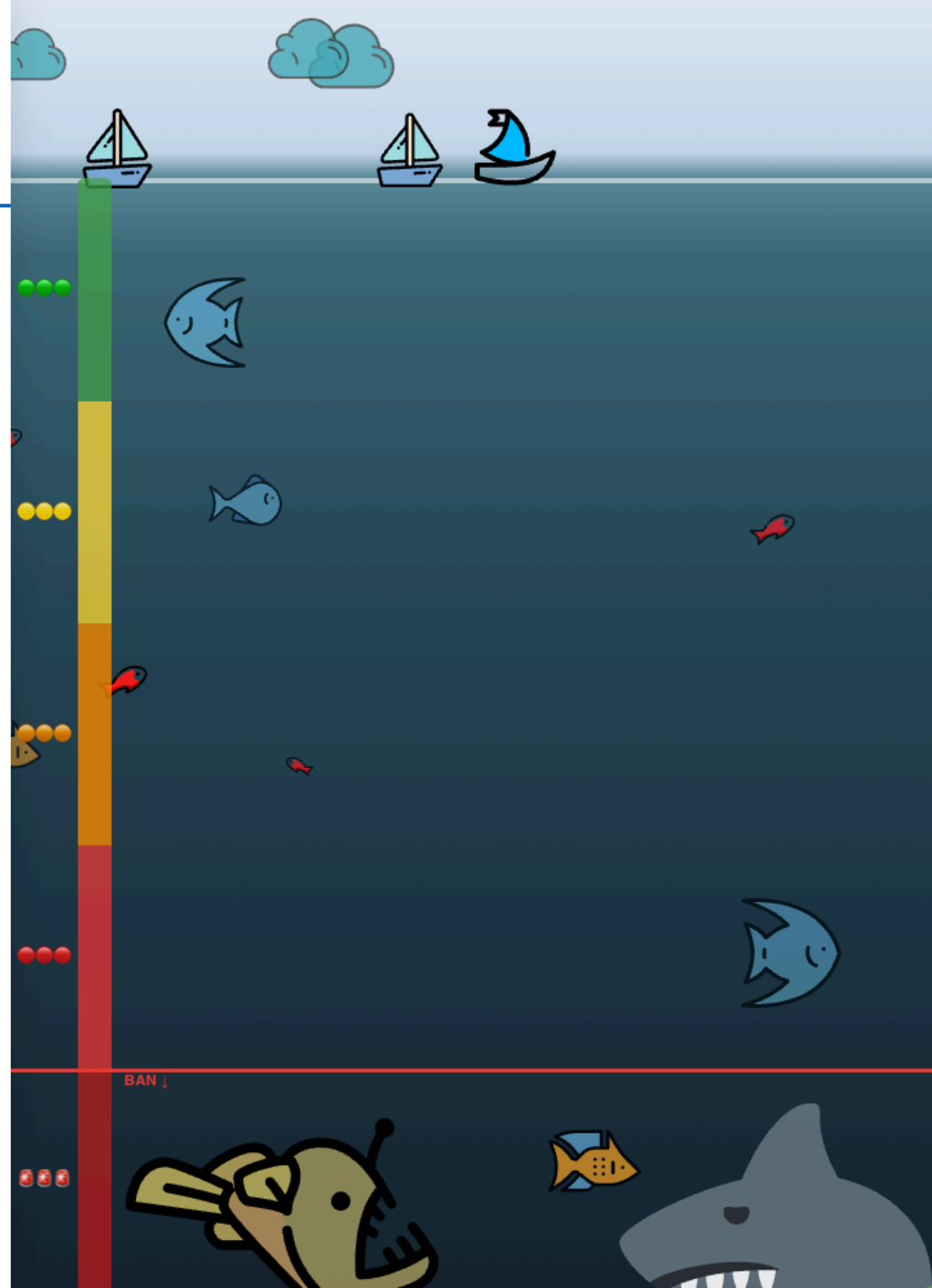
A committee uses an LLM to rank grant proposals by quality and novelty and uses the ranking directly.

A researcher uses an LLM as their main tool for finding relevant papers in an unfamiliar field.

A researcher trains an LLM on their own papers to generate first drafts of new manuscripts, which they edit carefully.

A reviewer uploads a confidential manuscript under peer review to a public LLM to draft a referee report.

A student asks an LLM to explain a mathematical concept.



Ethical gradient game

Fouesneau, Momcheva, Salditt, Burke-Spolaor

Ethical gradient

It is not always clear
and opinions may differ

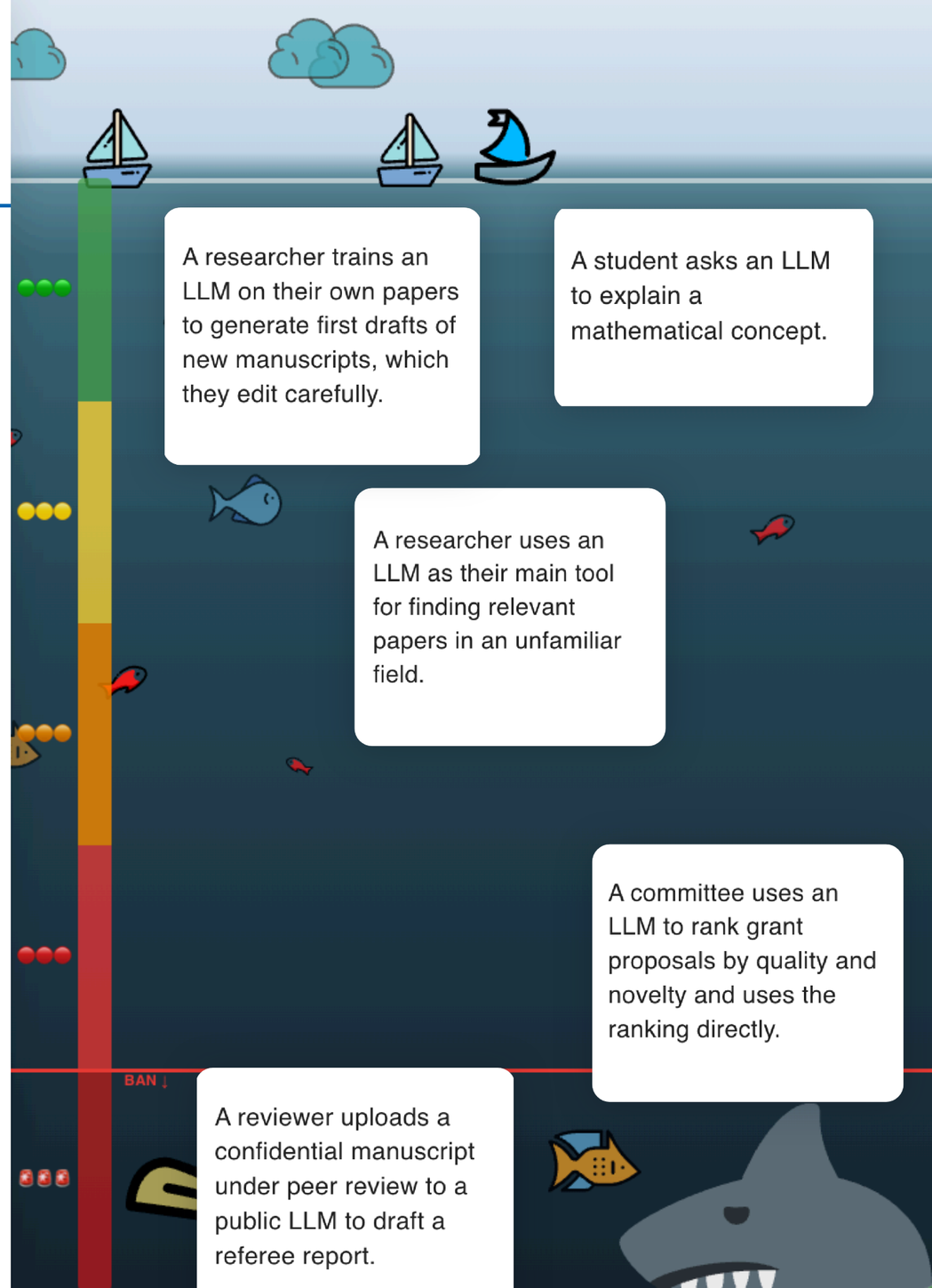
Community norms are lagging

Don't take this as my final rank

Many scenarios, have a go yourself - after this talk :)

Ethical gradient game

Fouesneau, Momcheva, Salditt, Burke-Spolaor

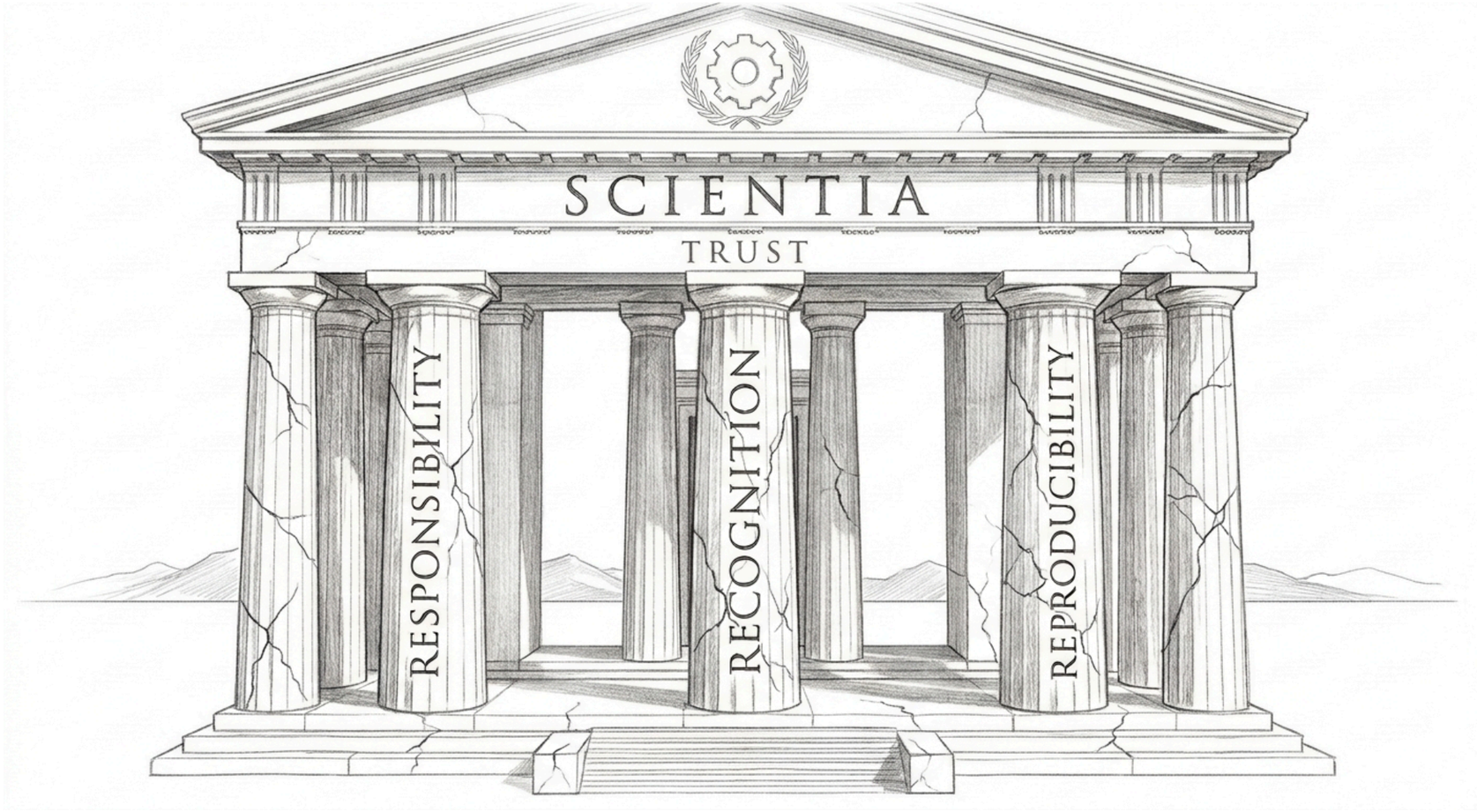


Adoption rate has vastely outpaced the development of community norms

The question is not whether to use them. It's what we value when everyone does.

What we value

Science relies on trust. Trust rests on three pillars.



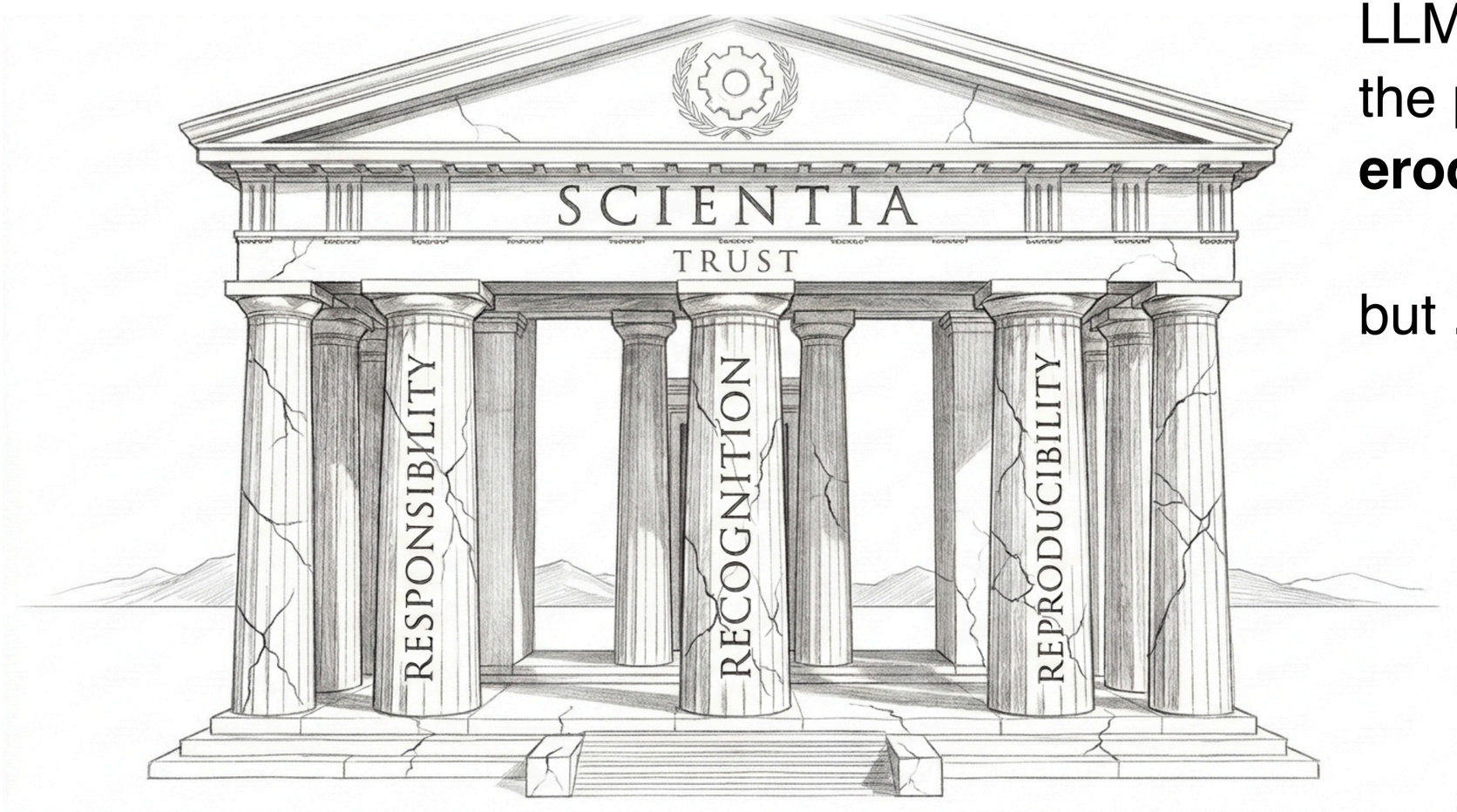
Who did the work?
Who stands behind it?

Who gets credit?
What counts as contribution?

Can others verify?
Does the code actually work?

image generated with Gemini
Inspiration credit Iva Momcheva

What we value



LLMs have
the **potential to deeply erode** all three.

but ...

Who did the work?
Who stands behind it?

Who gets credit?
What counts as contribution?

Can others verify?
Does the code actually work?

AI is an **amplifier** of **existing issues**

AI isn't breaking astronomy. **It's revealing what we've been "tolerating".**

- The paper count was never a quality metric
- Authorship was already questionable.
- Contribution metrics were already variable.
- Verification was already time consuming...

AI just made these cracks impossible to ignore.

Where trust breaks down the most

Hallucinations and fabrication

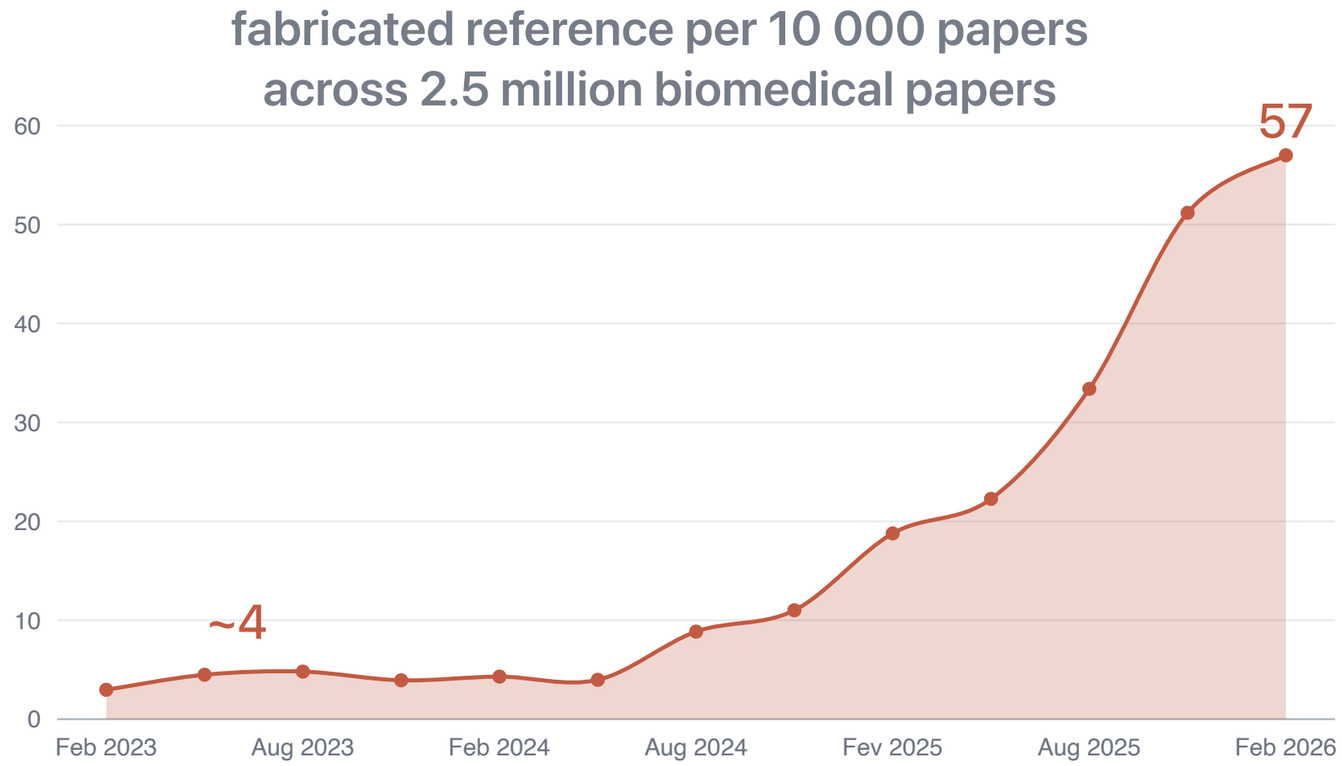
Erosion of responsibility

55%

ChatGPT references fabricated
Walters & wilder 2023

146K

hallucinated citations on
arXiv, bioRxiv, PubMed Central, and SSRN
Zhao et al. 2026

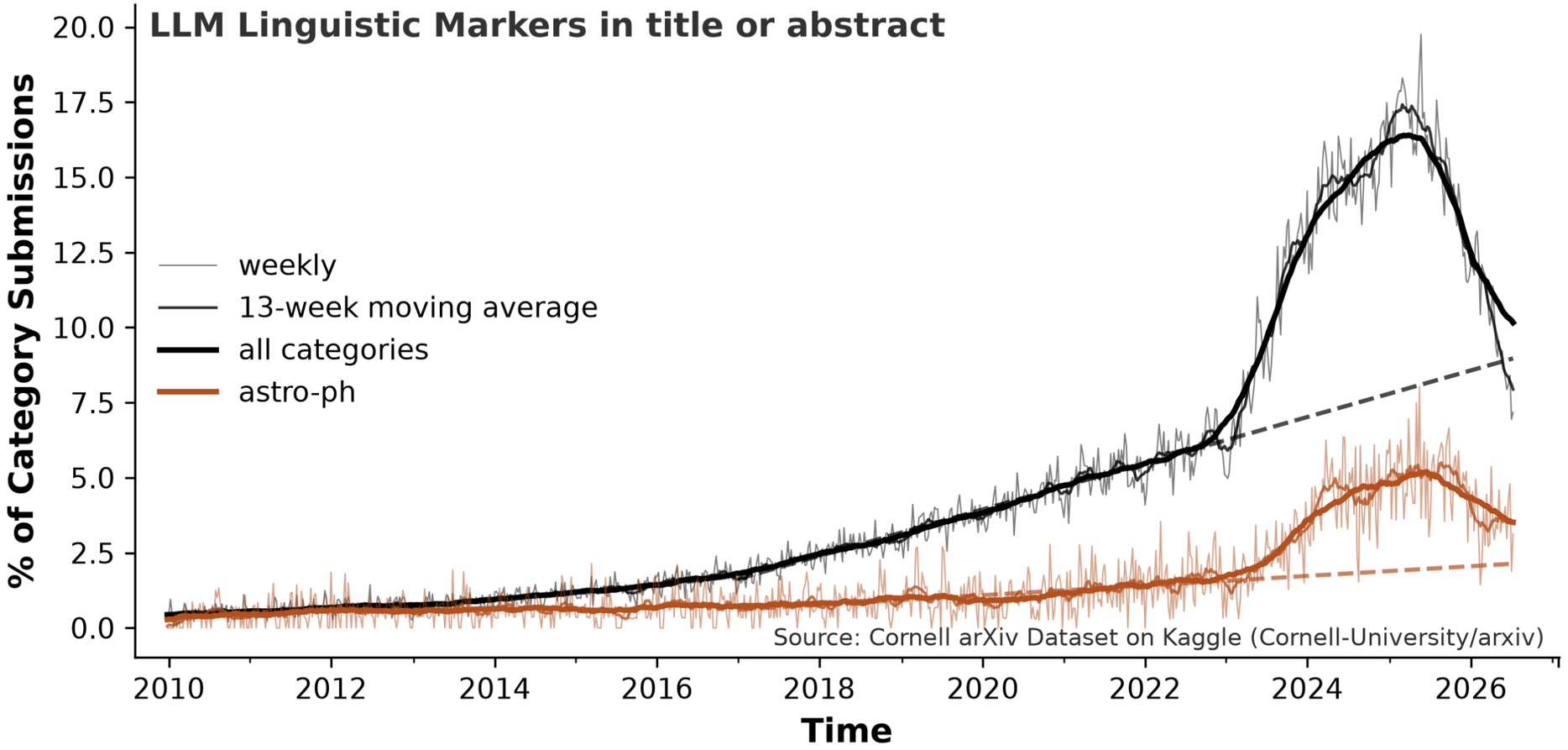


Topaz et al. 2026

One fake citation in a review propagates through every paper that cites it.

LLM linguistic markers in arXiv papers

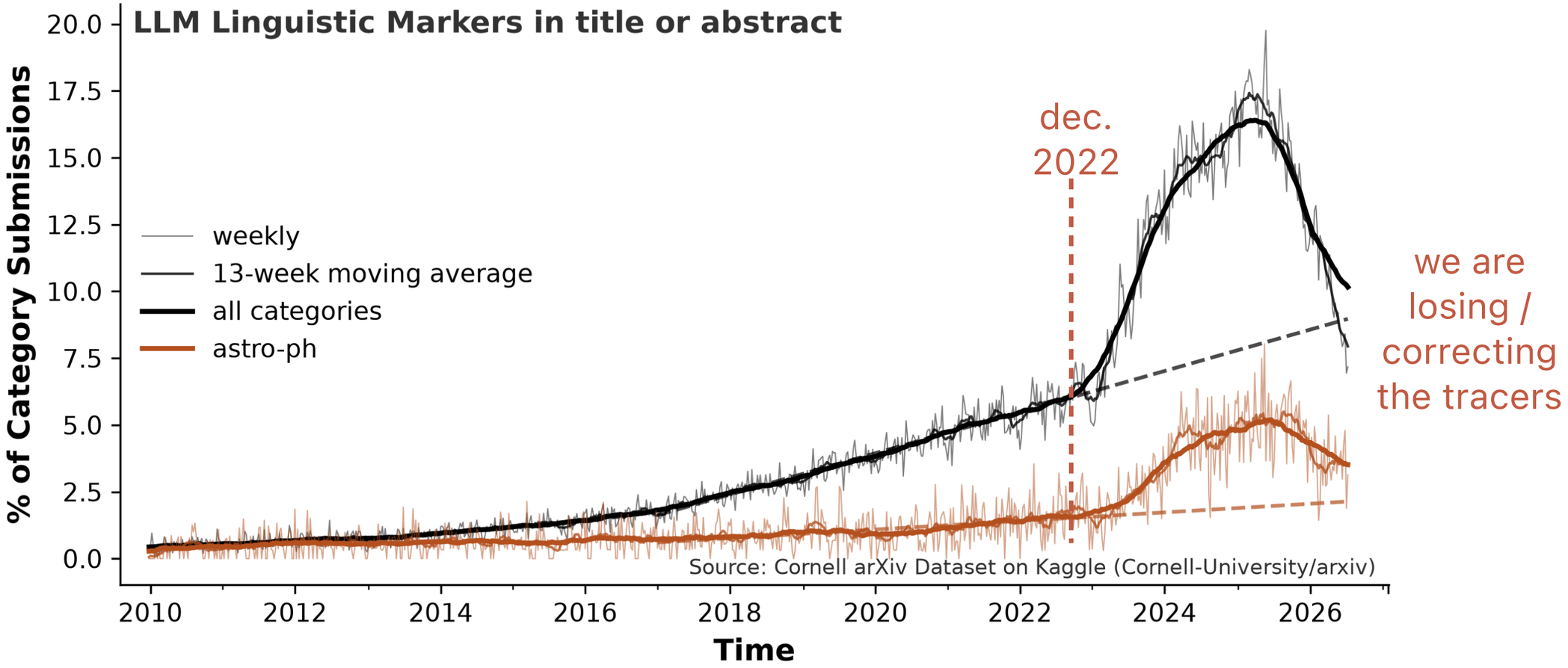
Erosion of signal



Our analysis of 22 markers selected from published LLM detection literature

LLM linguistic markers in arXiv papers

Erosion of signal



we are losing / correcting the tracers

Our analysis of 22 markers selected from published LLM detection literature
Strong correlation with ChatGPT coming out on November 30, 2022

Confident wrongness

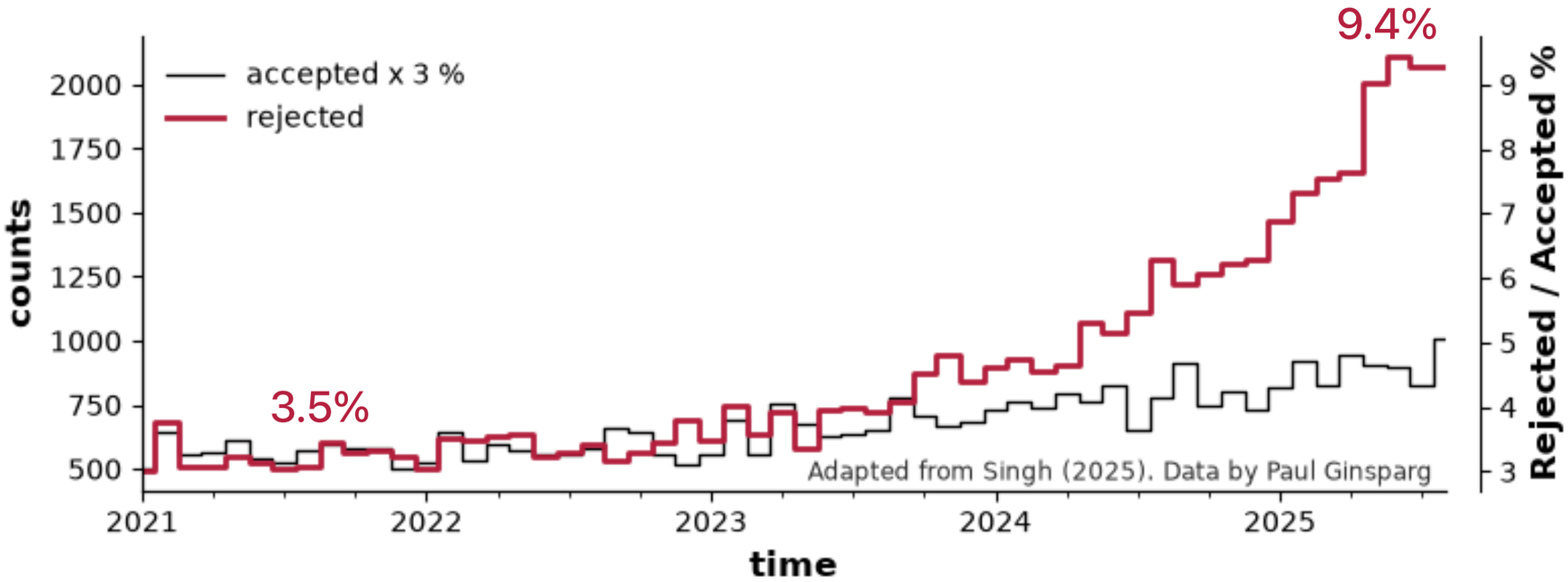
Erosion of attribution

Measurably more complex

Longer sentences, more syllables, harder to read, specific wordings

Astarita et al. 2024, Geng & Trotta 2024

arXiv rejection rejection rate increased from 3.5% (2021) → 9.4% (2025).



It sounds plausible & sophisticated — that's what makes it dangerous.

"All em-dashes in these slides were human-generated", Terence Tao (ICM, 2026)

Code that runs but isn't right

Erosion of reproducibility

more reworking

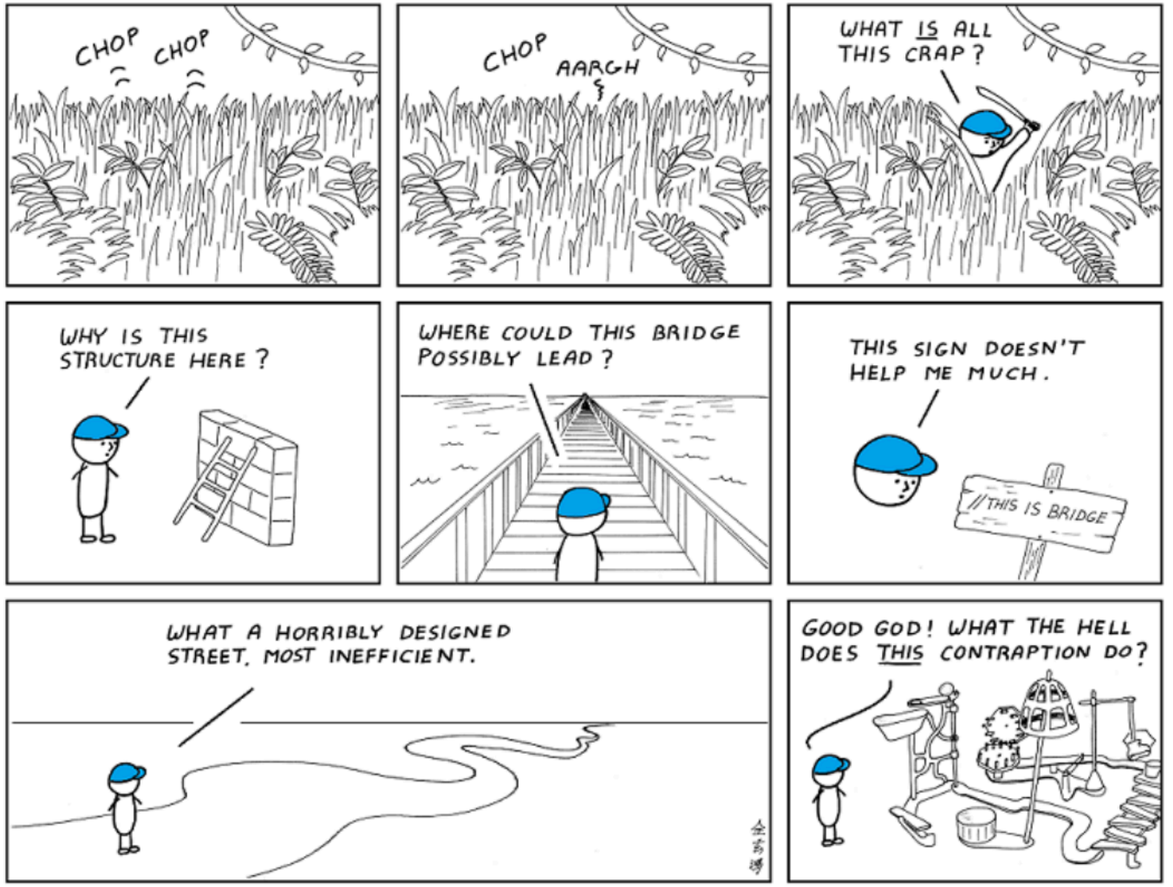
to ensure maintainability

Licorish et al. 2025

24%

of AI-generated code was plagiarized

Gupta et al. 2025



Abstruse Goose Comic 432

I hate reading other people's code.

The code runs — but does it do what you think?
Reproducibility requires more than execution.

The Asymmetry

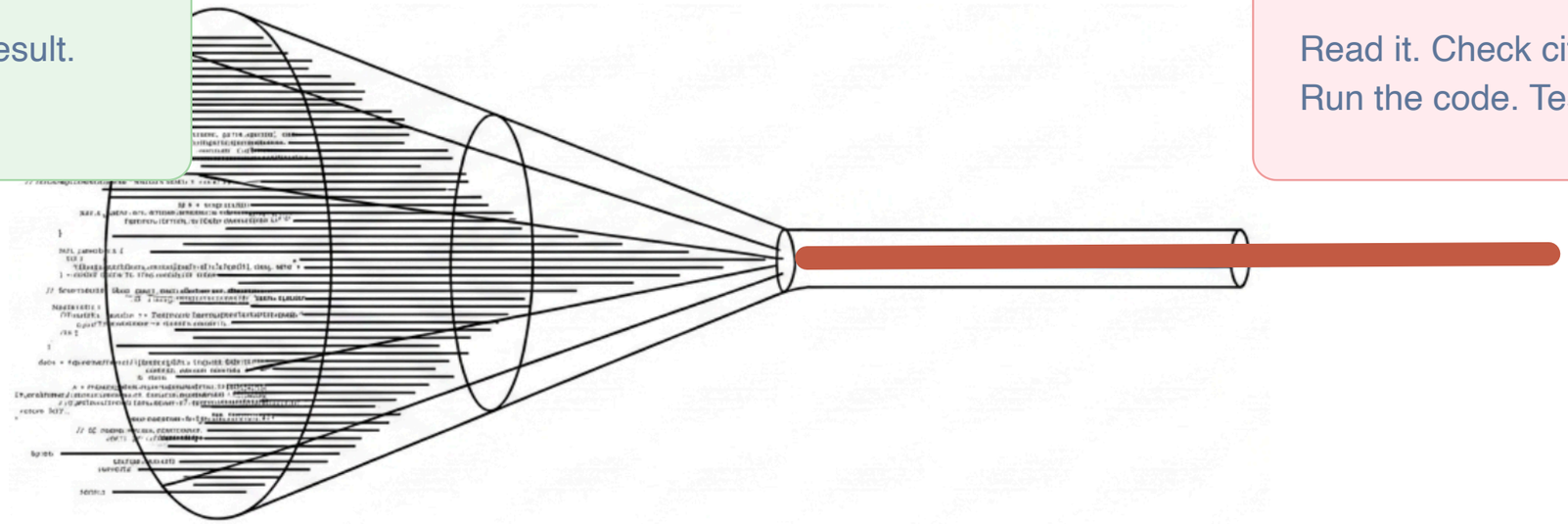
Generation is cheap. Verification is expensive.

The structural problem

GENERATION

seconds

Type a prompt. Get a result.
No expertise required.



VERIFICATION

hours

Read it. Check citations.
Run the code. Test edge cases.

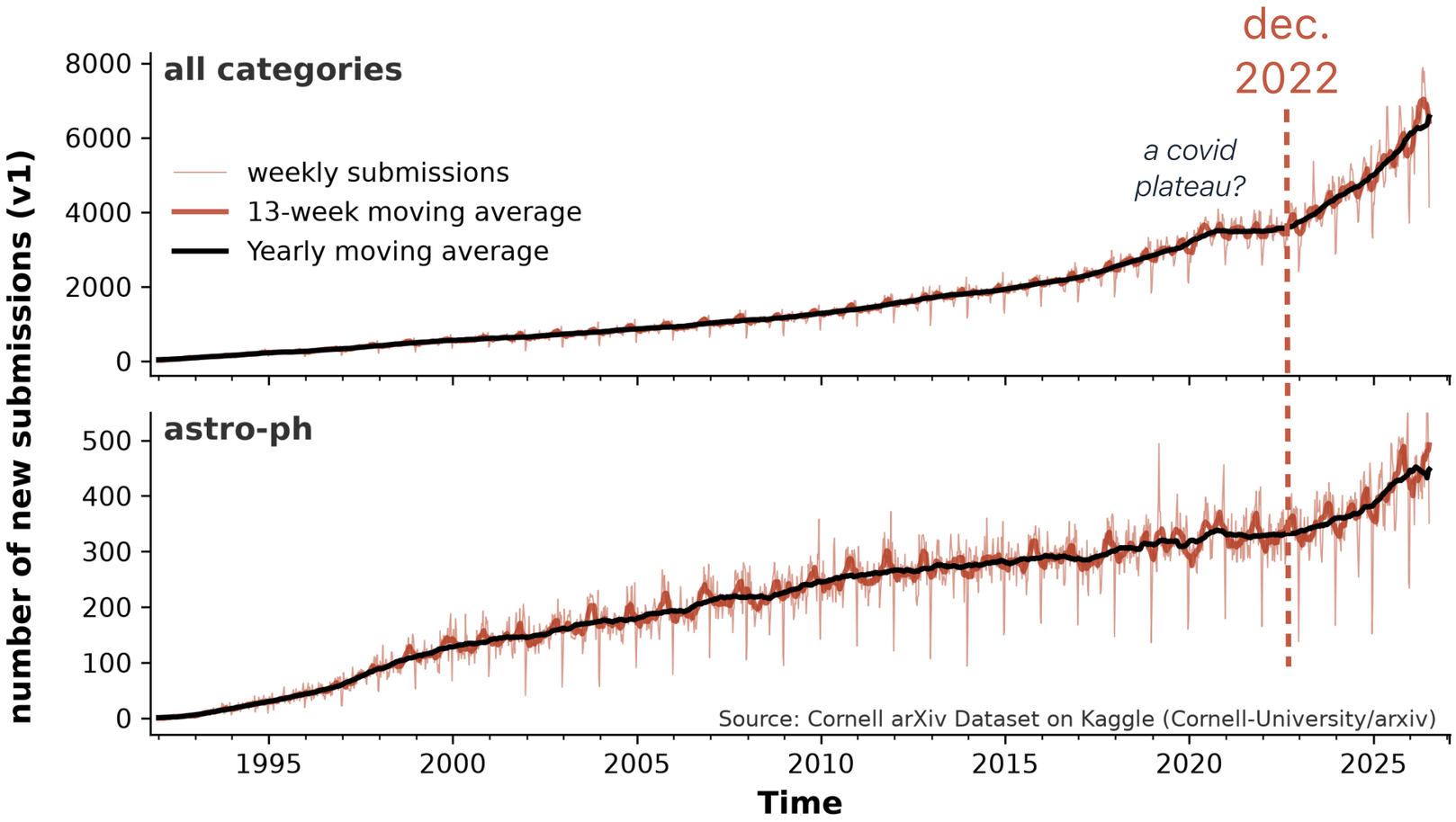
Our current economics favor volume over quality. That's structural, not personal.

and trust is then delegated as a task for the receiver (users, co-authors, readers, committees, ...).

arXiv submission volume keeps climbing

The structural problem

All categories: ~6,000/week. Astro-ph: ~400/week.



This isn't just about papers — it's about people

How do you find the pearl when everyone looks good now?

The gap between good and bad candidates, good and bad proposal cannot be assessed with old methods.
Publications? Everyone has them. Citations? AI inflates them. Code? AI writes it.

How do committees distinguish signal from noise without interviewing everyone or running trial periods?
(which academia cannot afford)

Use AI? AI loves itself, prone to errors, propagates biases

The assessment crisis is **the trust crisis,** personalized.

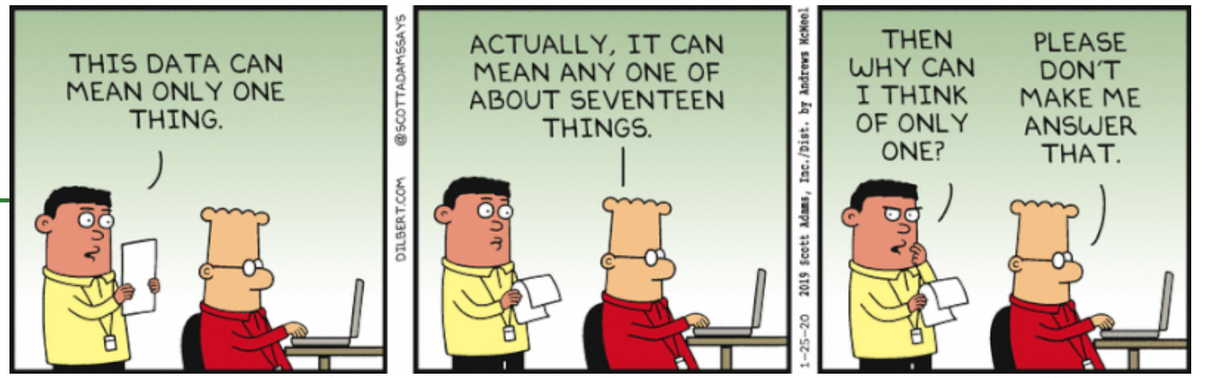
It's not just academics, it's also in education, industry, ...

What (can) we do about it

three levels of actions: individually, as a community, and institutions

What **you** can do **today**

Individual practice



Love Dilbert: Data Can Only Mean One Thing 1–25–20 Dilbert.com

Own the work	Can you defend the key decisions in the code, statistical methods? Can you explain why each step is necessary? If you can't explain it, you don't understand it, you should not submit
Verify Citations	Have you verified that every reference actually exists? Have you read the key papers your work builds on?
Run your code	Have you tested against known answers? Could you debug it if it broke?
Be transparent	Do you give enough for someone else (or you is 2 monts) to redo the work? Improve the reproducibility, publish prompts, assumptions, codes, and data

These cost time at *generation*. They preserve trust at *verification*.

What we decide **together tomorrow**

Community norms

Mandatory refereeing	Journals should implement 1 paper for first author. 2 for designated co-authors. Similar to ALMA proposals and NeurIPS journal (but lower load to start). — create noise, share the burden
Automated checks tools	Why don't we check plagiarism, citation, statistical sanity? — and before humans. JOSS and others already do that (ArXiv also)
Explicit contributions	Should we redefining “Scientific Contribution” in the Age of AI? Why not list the specific contributions of each co-author (code, papers, ...) like Nature does? — we cannot be all 1st author
Shared disclosure	Can we agree on one granular standard for all journals and codes? — Not optional

Create direct feedback loops between generation and verification

What universities and funders **must do then**

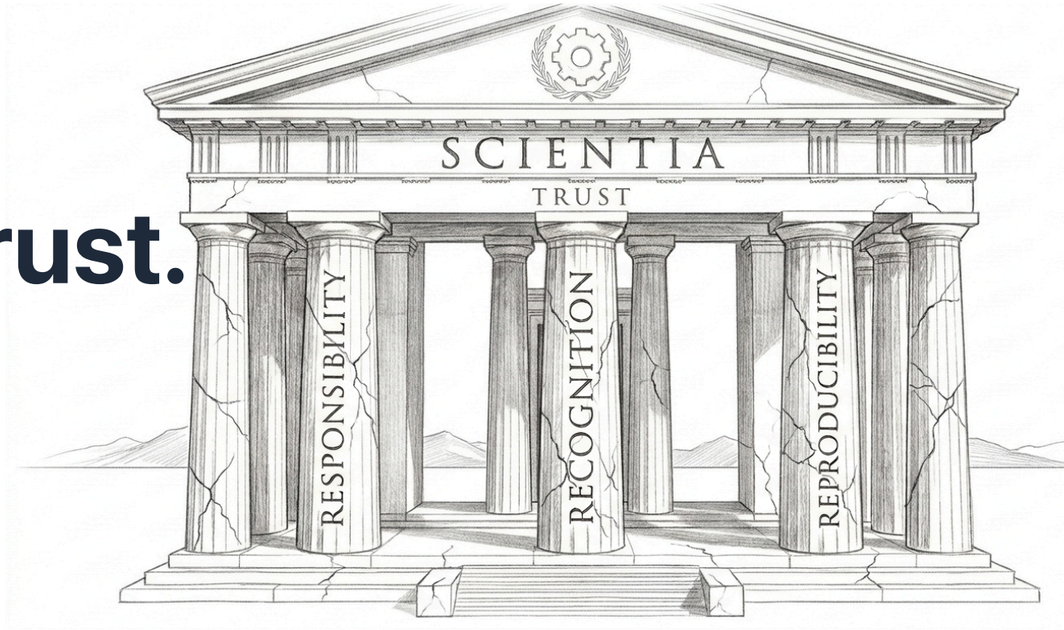
Institution norms

Update quality metrics	Invest to define/find alternative quality/impact signals? (expert curation, structured peer endorsements, reproducibility audits...)
Shift hiring criteria	Jobs should not reward volume but reward contribution Support the community to evaluate contribution
Invest in education	Teach critical AI literacy, not just prompt engineering Focus teaching on critical thinking rather than API reflexes
Provide a real stand	Actively set guiding examples instead of passively relying on good will. Provide a framework to compose with, debate and move forward

The system needs to create the structural support for the community and individuals.

the choice

We can let these tools erode trust. Or we can use this moment to reinforce their pillars.



Individual

Own. Verify. Run. Disclose.

Community

Referee. Contribute. Check. Standardize.

Institutions

Fund. Reward. Update. Educate.

What can we agree today to do tomorrow?

The tools are here, and won't go away.

POLL #4 (and last)

What is a norm we should adopt as a community **this year**?

pigeonhole.at
ML4ASTRO3BIRD

This question will remain open until the end of the day

What can we agree today to do tomorrow?

The tools are here, and won't go away.

Q: What is one norm we should adopt as a community **this year**, which one would it be?

11 VOTES — **Anonymous** • 10 minutes ago
Stating clearly in papers the contributions of each author and how AI was used.

10 VOTES — **Anonymous** • 9 minutes ago
A community-norm rubric that does not discourage but does clearly explain AI use

7 VOTES — **Anonymous** • 9 minutes ago
Some sort of (finite time period) ban from publishing if caught with citations that don’t exist.

4 VOTES — **Anonymous** • 9 minutes ago
AI statement, with tickable list of how it was used at every step, with space for the author to elaborate on each point. Standardised list. Mandatory for arxiv and all journals. Always accessible.

4 VOTES — **Anonymous** • 20 minutes ago
Explocit contributions (including AI) statement

1 VOTE — **Anonymous** • 5 minutes ago
Agreeing on the fact that LLMs are a tool. Misuse is the fault of the user. So we should use it for things like testing, text corrections, discussions etc. Not for outsourcing my complete job.

1 VOTE — **Anonymous** • 6 minutes ago
Explain significance and impact of papers rather than just a counting them and/or list only top 5 papers

1 VOTE — **Anonymous** • 7 minutes ago
Educate everyone (but specially students going through education in this era) on AI literacy and what is being sacrificed by outsourcing to agents

1 VOTE — **Anonymous** • 8 minutes ago
Publish code for reproducibility

0 VOTES — **Anonymous** • 4 minutes ago
Shifting the direction of AI from doing our science for us to do admin tasks (e.g. travel expense forms) so that we have more time for our science

The question is not whether to use LLMs/AI.
It's what we value when everyone does.

And what we do <u>now</u> to keep it.

Morgan **Fouesneau**

Max Planck Institute for Astronomy - *Data Science Dept.*

fouesneau@mpia.de