Maxwell's Ghosts

Artificial Intelligence as Epistemic Infrastructure

Paul Friedl

06.11.2026 doi: 10.59704/30ae12a02c4af3ce

Almost exactly 35 years ago, a fisherman found a naked body floating 20 miles south-west of Gran Canaria. The body, it was quickly established, was of British business magnate Robert Maxwell, who only a couple of hours earlier had vanished from his private yacht, the Lady Ghislaine.

Maxwell had lived an altogether unlikely life. Born to a poor Orthodox-Jewish family as Ján Ludvík Hoch, he escaped Nazi-occupied Czechoslovakia to France as a teenager, became decorated as a captain in the British army and, after the war, built a publishing and media empire that turned him into one of the wealthiest and most flamboyant businessmen in the world.

The cornerstone of Maxwell's riches came from his undertakings in academic publishing, in particular owning and running the big publisher Pergamon Press. Maxwell is arguably the central inventor of modern academic publishing, who realised that a respectable-sounding journal would have scientists supply their findings free of charge, thereby transforming academic journals from "public goods [into] private commodities".

In doing so, he created an incredibly, some might say impudently, profitable business model that still regularly yields margins upwards of 30 or even 40 per cent. For universities and scholars, however, the new structure proved less rewarding. In a system where publicly funded science communities not only freely hand over their findings but also donate vast amounts of labour to review them on publishers' behalf just for university libraries then to pay arbitrary, increasingly prohibitive prices to access those same resources, academic communities have effectively become a captive audience, their labour easily exploited. Little surprise that most see the system as fundamentally broken.

Thirty-five years later, a new class of businesses is looking to profit from academic production: AI companies, from juggernauts like OpenAI to specialised firms like Consensus or Scite.AI, are energetically targeting universities and other higher education actors, while providing free accounts to students.

There can be little doubt that generative AI will fundamentally transform all dimensions of scientific and academic activity, if it hasn't done so already. This also means that generative AI will become a central new *epistemic infrastructure* for science communities, shaping the accessibility, organisation and presentation of scientific knowledge and other epistemic resources. How much is at stake with this transformation can already be seen in mathematics, where leading figures of the field, among them twenty-five Fields Medallists, see a "severe misalignment" between AI and academia that might have a "destructive effect" on the discipline and poses a "general threat to intellectual work".

Whether academic communities can prevent the crisis of academic production from deepening – and perhaps even gain greater control over its means – will depend not least on whether they find answers to three interconnected questions: What changes with the arrival of a new epistemic infrastructure? What risks does a commercial ordering of generative "AI for science" entail? And what can academic communities do to protect their epistemic resources and ensure that the transformation follows a scientific rather than a commercial logic?

Five Questions for Epistemic Infrastructures

To grasp what changes, it helps to look at generative AI as the latest in a long line of epistemic infrastructures. Just as with the introduction of other epistemic

infrastructures, such as publication indices and databases, preprint servers or *Google Scholar*, generative AI will not only transform the cost of epistemic resources but will also reshape their structure. These transformations can be categorised under several rubrics:

1. Corpus: Which knowledge feeds the information pools that underpin "science AI"? Based on what criteria, following what logic are selection decisions made? Which knowledge is seen or declared as (sufficiently) scientific?
2. Access: Who can access this knowledge and how? Which factors determine access (money or other exchange goods, location, technological facilities, …)?
3. Ordering: How is knowledge ordered? How is it indexed, ranked, contextualised and presented?
4. Transparency: What can scientists know about the infrastructure, e.g. its corpus, (the mechanisms of) its ranking, or the logic by which it compiles and synthesises knowledge? What can they know about specific claims, e.g. their provenance?
5. Feedback: How will AI itself alter beliefs, imaginaries and practices of scientific knowledge production? How will its epistemography affect aspects such as reception and citation standards, the (perceived) synthesisability of knowledge or what counts as scientific progress?

Early Warnings

There can be little doubt that generative AI in science will effect changes across all five of these rubrics. Some developments can already be observed, others are still pending.

Regarding corpora, the past two years have seen a surge in academic licensing deals, often worth tens of millions, through which academic publishers open up scientific literature to AI companies for AI training or as context for retrieval-augmented generation. Because most publishing contracts include a comprehensive grant of rights, publishers can usually strike such deals without seeking authors' consent, remunerating them or even informing them – and have for those reasons often drawn heavy criticism from scholarly communities. Meanwhile, publishers (and scientists) are also bringing lawsuits against AI companies, alleging that they have trained their systems on copyrighted material obtained, among other sources, from online shadow libraries.

Moreover, to protect their market position, large academic publishers have begun to sharply restrict free access to journal and book abstracts, undermining digital aggregation tools like bibliographic databases, search engines or library catalogues, which depend on them. At the same time, industry insiders worry that commercial, rather than scientific, demand will shape the corpus, with applied fields like medicine, law or computer science thriving, while less marketable research – e.g. in the humanities, from smaller languages or without the possibility of quantified benchmarks – gets left behind. Smaller publishers, lacking "attractive" catalogues and technological expertise, might also not survive the transition, not to speak of the large swaths of as yet undigitised materials.

Regarding access, universities and libraries are of course worried that access to epistemic resources will become even more expensive. In a highly oligarchic market with arcane pricing structures and strong bargaining power asymmetries, licensing dynamics are already seen as economically unsustainable. The arrival of another layer of intermediation may deepen this crisis. As things stand, older licences are mostly silent on AI use, while newer ones range from outright bans to carve-outs for university-run systems. Where licences do not cover AI use, libraries may soon have to pay twice: for publishers' catalogues and their use in AI tools. Nor does the AI market promise relief: providers are racing to lock in customers in a market that already shows signs of high concentration.

In terms of ordering, research communities are demanding that AI companies respect the central scientific principle of attribution. The complexity of AI training methods and murky practices around the use of user inputs, however, cast doubt on how far attribution can and will be guaranteed in practice. Citation-based indexing and ordering of scientific knowledge

may generally lose relevance. Commercial incentives may also shape ordering itself: with ChatGPT already showing (so far clearly labelled) ads and licensing deals privileging individual publishers' catalogues, AI providers might eventually sell premium placement for paywalled articles or other publisher products.

Regarding transparency, AI systems can often already give some account of the sources and limits of their corpora, but far less of how they select and rank information. This is especially problematic, considering that generative systems inescapably favour certain sources over others, while often creating deceptively confident "answer bubbles". Much will therefore depend on whether research AI tools build interfaces that show users what answers rest on, including corpus contents, selection logics and the sources behind specific claims. Some tools, such as Consensus, already point this way.

Finally, in terms of feedback, research and citation practices may change considerably – with the result, as already suggested above, that the networked production and ordering of knowledge, long shaped by snowballing (citation tracking) and the citation-based linking of publications in central digital indices such as Web of Science or Google Scholar, loses relevance. Some also worry that knowledge presented by AI might not only be mistaken for deep, comprehensive and neutral, but could also nudge research towards monocultures in which certain kinds of knowledge and knowledge production come to dominate while creativity suffers – concerns for which early empirical studies offer some, if still tentative, support. Ultimately, in the shadow of AI, most, if not all, disciplines will also need to rethink what they consider scientific progress, raising fundamental questions about disciplines' objectives and methods.

Countermeasures

So what can and should academic communities do to work towards a more accessible and transparent design of scientific generative AI as an epistemic infrastructure aligned with scientific principles?

First, to counter the enclosure of epistemic resources, existing scholar-led publishing initiatives, which move publishing out of the hands of commercial providers and into those of academic institutions, deserve far greater support and should be scaled up.

At the same time, universities and public libraries need to strengthen their bargaining power so as not to lose further ground in the financial tug-of-war with publishers and technology providers. As with many other measures, this is only likely to succeed if institutions pool their resources (financial, technical, intellectual) – and develop shared long-term strategies.

Most important of all, however, universities must take care not to end up, through careless infrastructure choices, in dependencies that would rob them of the freedom to decide over their own academic resources in future. Universities should therefore be particularly wary of the partnerships now being struck with large AI companies, which risk locking core components of research (data, methods, skills) so deeply into proprietary software environments that they later can no longer be extracted at reasonable cost – even if providers present alluring freebies and especially if universities are meant to pay with data instead of money.

At the same time, building home-grown academic language models seems a doubtful bet: despite falling costs, they are unlikely to catch up with frontier tools any time soon.

More persuasive are calls for model-agnostic middleware through which researchers can pair different language models with different bodies of knowledge and literature. Libraries – experts on questions of selection, neutrality, transparency and diversity – should have a central role in building, procuring and implementing such infrastructures.

Beyond infrastructure, however, academic communities also need to renegotiate their practices of research in general and of publishing in particular. In some disciplines, the sheer speed and volume of academic output already today mean that new findings can hardly be systematically tracked, assessed and used – whether in peer review or in one's own research – without the help of AI. Continued acceleration thus compounds the risk of structural dependence on AI technologies and their providers. Short of further

automating research and publishing, AI-based peer review included, academia perhaps has little choice but to slow down in concert. *Cambridge University Press*'s managing director, for one, sees no other way out of the crisis. Such a deceleration could draw on various levers. Taking a cue from the caps on submissions per author already common at conferences and preprint servers, hiring committees could similarly limit the number of publications applicants may submit, while journals could make submission conditional on a minimum number of prior citations.

ABOUT THE AUTHOR
Paul Friedl
Dr. Paul Friedl is a legal researcher with a focus on data and technology law. He is currently a postdoctoral researcher at Humboldt-University Berlin.

SUGGESTED CITATION
Friedl, Paul: Maxwell's Ghosts: Artificial Intelligence as Epistemic Infrastructure, Verfassungsblog, 06.11.2026, doi: 10.59704/30ae12a02c4af3ce.

Licensed under CC BY-SA 4.0.

Langzeitarchiviert bei IntRechtDok unter: doi.org/10.17176/20261106-133819-0.