

THE THIRD KNOWLEDGE REGIME: ARTIFICIAL INTELLIGENCE AS AN EMERGENT RESPONSE TO THE LIMITS OF CLASSICAL ACADEMIA AND OPEN REPOSITORIES

El tercer régimen del conocimiento: La inteligencia artificial como respuesta emergente a los límites de la academia clásica y los repositorios abiertos

Carlos Eduardo Ravello Joo

ORCID: 0009-0007-5631-7436

carlosravello.com

Trujillo — Lima, Peru

May 2026

Author note

This work is part of a series of independent preprints by the author. It connects with conceptual frameworks developed in Ravello Joo (2026a, 2026b, 2026c), deposited in Zenodo. No external funding was received. The author declares no conflict of interest.

ABSTRACT

Artificial intelligence does not surpass humans through intrinsic cognitive capacity. The observable operational advantage emerges from both computational capacity and the elimination of accumulated institutional frictions that limit collective human cognition. This article examines that diagnosis from three simultaneous fronts: the structural slowness of the traditional academic system, the epistemic noise produced by open repositories without adequate filtering, and the anonymous gatekeeping that operates over the knowledge graphs that AI systems ingest. The concept of epistemological coordination costs is introduced as a framework to understand why both classical academia and open repositories fail simultaneously, and how AI enters that historical gap. It is argued that knowledge no longer belongs exclusively to human institutions, and that documenting rigorously before automated systems fill that vacuum with imagination without method is not an academic option but a

civilizational obligation. Information, like any physical system, tends toward disorder without organized energy flow. Knowledge institutions were the mechanisms humanity built to locally reduce that entropy. When those mechanisms fail, disorder does not wait.

Keywords: institutional friction, academic validation, open repositories, Wikidata, knowledge graphs, epistemological coordination costs, artificial intelligence, peer review, informational entropy

Table 1. Three epistemological regimes

System	Speed	Validation	Result
Classical academia	Low	High formal / low real	Rigidity + invisibility
Open repositories	High	Minimal or none	Noise at scale
Artificial intelligence	High	Signal compression	Operational dominance

1. Introduction

While this article was being written, over the course of four days, GitHub recorded the creation of more than 1,300,000 new repositories. Hugging Face added approximately 1,600 artificial intelligence models. arXiv received between 2,600 and 3,700 new papers. 230 repositories per minute. Not as spectacle. As symptom.

The symptom points to a question that knowledge institutions are not answering with sufficient honesty: why does an architecture without metabolism, without professional career, without reputational fear, without biographical time, operate cognitively faster than systems specifically designed to produce and validate knowledge?

The usual answer—that AI has more computational power—is true but incomplete. Computational power explains processing speed. It does not explain why a five-year clinical study can remain invisible to the systems that today build knowledge at scale, while a poorly structured GitHub repository appears indexed within hours.

This article proposes a more uncomfortable answer: AI's operational advantage emerges both from computational capacity and from the elimination of accumulated institutional frictions that limit collective human cognition (Gartenberg et al., 2026; Stanford HAI, 2025). This is not an intelligence problem. It is an architecture problem.

Information, like any physical system, tends toward disorder without organized energy flow (Shannon, 1948; Schrödinger, 1944). Knowledge institutions were the mechanisms humanity built to locally reduce that entropy. The formal development of that framework connects with the author's

prior work (Ravello Joo, 2026a, 2026b, 2026c). What this article examines is what happens when those mechanisms fail simultaneously from two opposite directions.

2. Two Simultaneous Failures

2.1 The Slow Failure: Academia as Scarcity Architecture

Peer review was born in a world where the problem was informational scarcity. Little knowledge was produced, and the challenge was protecting it from contamination. Slowness was an adaptive feature, not a defect: careful validation was the reasonable price of reliability.

That world no longer exists. In 2025, industry produced more than 90% of notable frontier models, with development cycles of weeks (Stanford HAI, 2025). On benchmarks like SWE-bench Verified, performance went from approximately 60% to nearly 100% in a single year. The average doctoral thesis in the United States takes 7.3 years from the start of the program (NSF/NCSES). In that time, AI renewed its capabilities completely between four and seven times.

The documented result is that relevant research becomes obsolete during its own validation process. Gartenberg et al. (2026), in an analysis of 6,957 submissions and 10,389 reviews in *Organization Science* between 2021 and 2026, documented the collapse of the quality-quantity trade-off driven by the massive increase in post-ChatGPT submissions (+42%) without proportional scaling of review infrastructure.

More seriously: the system does not consistently produce what it promises. The Open Science Collaboration (2015) found that only 36-39% of psychological studies replicated successfully. Camerer et al. (2018) documented similar results in economics and social sciences. The Retraction Watch database exceeds 58,000 entries. In the first weeks of 2026, one in every 277 papers indexed in PubMed cited a nonexistent reference—generated by AI—compared to one in 458 in 2025 and one in 2,828 in 2023 (Topaz et al., 2026, *The Lancet*).

Richard Smith (2006), editor of the *British Medical Journal* for 25 years, wrote from inside the system: peer review is a flawed process at the heart of science and journals. Not an external critic. The person who administered it.

2.2 The Fast Failure: Repositories as Noise Architecture

The response to the slow failure was speed without filtering. Open repositories—GitHub, Zenodo, Hugging Face, arXiv, OSF—solved the problems of technical invisibility and bureaucratic slowness. A DOI in hours. Immediate global visibility. Versioning. Citability.

But speed without methodology produces its own failure. npm recorded spam campaigns of up to 17,000 packages per day at 2025 peaks, with more than

67,000 fraudulent packages accumulated over two years. PyPI added 356 new projects daily in 2025, with 3.9 million new files in the year, without a validation mechanism equivalent to the volume. Hugging Face hosted 1,533,973 models as of March 24, 2025, with hundreds of new daily repositories of indeterminate quality.

The most revealing data comes from Carneiro et al. (2020): the reporting quality difference between preprints and reviewed papers is approximately 5%. After months of editorial process, the improvement is marginal. This means either that peer review does not add as much value as it claims, or that open repositories are not as chaotic as their critics argue. In any case, the validation system does not justify its cost with its result.

Noise is not just a quality problem. It is a training problem. The language models that hallucinate most are those operating on scarce, contradictory, or low-quality data. When open repositories produce noise at scale, that noise becomes signal for systems building automated knowledge. Chaos scales.

2.3 The Third Regime: AI as Abundance Compression

Classical academia and open repositories appear opposed. Functionally they are distinct mechanisms for solving the same problem: how to filter information under human limits of attention and coordination.

Academia manages uncertainty through hierarchy, scarcity, credentials, and exclusion. Open repositories through abundance, iteration, massive exposure, and memetic selection. Both are responses to different historical conditions. Neither was designed for the current scenario.

AI enters as a third epistemological regime. It does not validate like academia. It does not explore chaotically like the open internet. It synthesizes, recombines, compresses, and accelerates cognitive navigation. But it has a critical structural limit: its quality depends on the quality of what it ingests.

Stanford demonstrated that LLMs reach 96% useful responses when combined with verified structured data parsing, versus frequent errors without that base. AI is not epistemically self-sufficient. It is a compression machine that amplifies what it receives, good or bad.

This introduces the central concept of this article: epistemological coordination costs. The contemporary knowledge crisis does not emerge simply from an excess or absence of filters, but from the institutional inability to reconcile epistemic validation and cognitive speed under conditions of informational overproduction. AI radically reduces those coordination costs. Human institutions increased them until they became unsustainable.

3. The Circular Lock

3.1 Wikidata as Operational Filter

Google migrated its Knowledge Graph infrastructure from Freebase to Wikidata between 2014 and 2016. Wikidata is today the primary source of 500 billion facts about 5 billion entities that Google uses. Google's patents explicitly reference Wikidata for entity attribute extraction. LangChain launched official Wikidata integration in 2024. Amazon Alexa, Apple Siri, and Microsoft integrate the same base. 72% of Wikipedia articles use Wikidata for their infoboxes, with 97% using it for cross-language links.

Wikidata is not an encyclopedia repository. It is the architecture of the knowledge graph on which the systems that build the semantic world that LLMs ingest operate. Without a node in Wikidata, Google cannot resolve the entity. Without entity resolution, dispersed mentions on the web do not accumulate toward any center. The work exists but for the system it is nobody.

That filter is administered in part by anonymous editors with 19th-century notability criteria, applied with 20th-century tools, over 21st-century phenomena. An editor without name, without verifiable expertise in the domain, can delete in minutes the technical work of an independent researcher invoking notability criteria that do not distinguish between research without institutional affiliation and content without value. Without effective right of reply. Without accountability for the impact. In total anonymity.

3.2 The Mechanism and Its Technical Consequences

The Wikidata community itself acknowledged the problem in its February 2026 notability policy reform RFC: certain knowledge has been and is being structurally marginalized, leading to less coverage in reliable sources and therefore an increased barrier to demonstrating notability.

That is the circular lock in its most precise formulation: the system demands notability to enter, but notability is built with the visibility the system denies by not entering. Not a governance bug. The structure itself.

The technical consequences are concrete and scalable. Without a node in Wikidata: no entry to Google's Knowledge Graph; no Knowledge Panel; mentions of the work on the web are not associated with any verified node; LLMs do not recognize the entity; the work does not appear in AI-generated responses; the virtuous cycle of semantic authority never starts.

The same period that saw this anonymous gatekeeping expand saw ClaudeBot blocked on 5.8 million websites, with a crawl-to-referral ratio of 73,000:1. Machines are hungry for verified knowledge. The systems that should feed them block them or leave them without nodes to anchor what they find.

3.3 Active Technical Invisibility

Publishing in Scopus, Nature, or Elsevier is not just slow. In 2025, it is also an act of progressive technical invisibility. Major academic publishers operate with hard paywalls that serve Googlebot only metadata or initial snippets. In

2025 there was a 336% increase in sites actively blocking AI crawlers like GPTBot and ClaudeBot, with major publishers among the most aggressive. A Rutgers Business School and Wharton study (April 2026) found that publishers who blocked LLM crawlers lost approximately 7% of weekly traffic in the six weeks following the block.

University repositories complete the picture: unstructured URLs, orphaned pages without interlinks, no schema markup, unoptimized sitemaps, pre-modern loading speeds. The Googlebot has a crawl budget. It is a time economy. It will not get lost in a 2003 technical architecture. It goes where the food is served: Zenodo, arXiv, OSF, GitHub Pages.

The researcher who publishes only in traditional channels is not being more rigorous. They are being invisible. And in the current ecosystem, invisible to machines is invisible to the knowledge already being built at scale.

4. The Validators Before the Threshold

This article does not ask to abolish peer review. It asks that reviewers be at the level of the moment they guard.

The day the first AI said hello changed the minimum threshold of what it means to do science seriously. That threshold rose for everyone. For those who research and for those who validate. The problem is that only some are being required to notice that the floor moved.

A peer reviewer who does not master the tools with which the knowledge they review is produced is not a guardian of rigor. They are a bottleneck with a title. Not out of bad faith—most operate ad honorem, with their own academic workload and criteria inherited from a system that never updated its standards. That is why the demand is structural, not individual: validation systems must incorporate criteria that include whether the work can confront the reality it describes, whether it can be found, whether it can be falsified by something more than three colleagues in a closed circuit.

Popper did not ask for anonymous peers or institutional titles. He asked for real exposure to refutation. The system that today claims that heritage produces—according to Nature's survey of 1,576 researchers—that more than 70% have attempted and failed to reproduce another scientist's experiment, and more than 50% have failed to reproduce their own. 83% of researchers themselves acknowledge that a reproducibility crisis exists. Falsifiability did not disappear because AI arrived. It left earlier, silently, buried in the same system that proclaimed it. Real exposure to refutation did not disappear — it shifted.

Open repositories, public benchmarks, code that either works or does not, apply more genuine Popperian pressure on a claim than eight months of peer review followed by a paywall that blocks Googlebot.

5. The Obligation Before It Is Too Late

Algorithmic weight does not distinguish between truth and opinion. It distinguishes between engagement and silence. An influencer with 10 million followers opining on vaccines, economics, or mental health generates more signal in the graph than a researcher with 40 years of clinical work publishing in a repository that Googlebot never visits. Not because the algorithm is malicious. Because the algorithm optimizes what it was asked to optimize: attention, screen time, interaction.

If rigorously documented knowledge cannot compete in that graph, the problem is not epistemological. It is civilizational. We are not discussing whether academia is slow. We are discussing who is going to train the systems that will build the reality of the next generations. If that training is done with the algorithmic weight of opinions without method, without falsifiability, without verifiable record, the result is not a faster world. It is a world where the difference between truth and narrative disappears not through ignorance but through architecture.

Knowledge no longer belongs exclusively to human institutions. It belongs to the algorithm and to those who feed it with rigor. AI has a real ceiling of comprehension and too much imagination. LLMs hallucinate more when verified sources are scarce or blocked. If humans do not leave a well-documented record, machines will fill that vacuum with the imagination they have to spare. And they will do it with the speed that institutions never had.

Documenting well is not academic vanity. It is the only way for the knowledge we produce today to survive the algorithm that decides tomorrow what is real.

Limitations

This work constitutes an analytical essay with critical narrative review of verifiable literature. It is not a formal empirical study. Source selection introduces selection bias inherent to all non-systematic reviews. The concept of epistemological coordination costs is an analytical proposal by the author that requires formal development and independent validation. Some quantitative data cited come from industry reports whose exact methodology is not public. The technological landscape described evolves at a speed that may render specific aspects of the analysis outdated in the short term.

References

- Camerer, C. F., et al. (2018). Evaluating the replicability of social science experiments in Nature and Science. *Nature Human Behaviour*, 2(9), 637-644. <https://doi.org/10.1038/s41562-018-0399-z>
- Carneiro, C. F. D., et al. (2020). Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature. *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0228711>

- Gartenberg, C., Hasan, S., Murray, A., & Pierce, L. (2026). More versus better: Artificial intelligence, incentives, and the emerging crisis in peer review. *Organization Science*. <https://doi.org/10.1287/orsc.2026.ed.v37.n3>
- Maslej, N., et al. (2025). *AI Index Report 2025*. Stanford Human-Centered Artificial Intelligence. https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf
- NSF/NCSES. (2024). *Doctorate Recipients from U.S. Universities*. National Science Foundation. <https://ncses.nsf.gov/pubs/nsf24300>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Ravello Joo, C. E. (2026a). Metacognición 2.0: Diseño deliberado de identidad digital ante sistemas predictivos de inteligencia artificial. *Zenodo*. <https://doi.org/10.5281/zenodo.20092009>
- Ravello Joo, C. E. (2026b). El perfil que no escribiste: Identidad digital, biometría ocular y autoridad semántica en sistemas algorítmicos predictivos. *Zenodo*. <https://doi.org/10.5281/zenodo.20112409>
- Ravello Joo, C. E. (2026c). La década del influencer como colapso termodinámico de la información: Cierre operacional, alostasis del conocimiento y bifurcación del ecosistema digital (2012–2025). *Zenodo*. <https://doi.org/10.5281/zenodo.20260383>
- Schrödinger, E. (1944). *What is life?* Cambridge University Press.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Smith, R. (2006). Peer review: A flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, 99(4), 178–182. <https://doi.org/10.1258/jrsm.99.4.178>
- Topaz, C. M., et al. (2026). Prevalence of fabricated references in scientific literature. *The Lancet*. [https://doi.org/10.1016/S0140-6736\(26\)00603-3](https://doi.org/10.1016/S0140-6736(26)00603-3)
- Wikidata. (2026, February). Requests for comment: Notability policy reform. *Wikimedia Foundation*. https://www.wikidata.org/wiki/Wikidata:Requests_for_comment/Notability_policy_reform
- Clark, J. A., Williams, H. K. R., & Rossmann, D. (2022). Wikidata and knowledge graphs in practice: Using semantic SEO. *Information Services and Use*. <https://doi.org/10.3233/ISU-220171>