

Scientific research and misuses of generative AIs: lightning in blue sky?

Pesquisa científica e uso abusivo de IAs Generativas: relâmpago em céu azul?

La investigación científica y el uso indebido de la IA generativa: ¿un rayo en cielo azul?

Edgar Lyra

Pontifícia Universidade do Rio de Janeiro (PUC-Rio), Rio de Janeiro/RJ, Brazil

<https://orcid.org/0000-0003-3664-3537>

edlyra@puc-rio.br

Abstract

The academic community is currently grappling with the misuse of generative artificial intelligence in academic production and scientific research. There is a growing number of complaints and retractions of texts recently published in specialized journals, as well as the circulation of publications improperly produced by AIs. The reasons for this state of affairs need to be discussed, among other things, because they call into question the issue of epistemic authority. The hypothesis to be explored in this article is that the conditions of possibility for this phenomenon go back to times before generative AIs, that is, to an economy of academic publication based on productivism and questionable practices of textual reference and validation, in other words, on a merely indexical textuality, in which the examination of sources and their real content has been lost. In short, the aim is to reflect on the naturalization of these practices and, in this context, on the current admissibility of generative black boxes.

Keywords: artificial intelligence; large language models; scientific research; epistemology; rhetoric.

Resumo

A comunidade acadêmica se vê hoje às voltas com o uso indevido de inteligências artificiais generativas na produção acadêmica e na pesquisa científica. Sobre o número de denúncias e retratações de textos recentemente publicados em revistas especializadas, ou de circulação de publicações indevidamente produzidas por inteligências artificiais (IAs). As razões desse estado de coisas precisam ser discutidas, entre outras coisas, por colocar em questão a

autoridade epistêmica. A hipótese a explorar neste artigo é que as condições de possibilidade do fenômeno retroagem a tempos anteriores às IAs generativas, isto é, a uma economia da publicação acadêmica pautada pelo produtivismo e por práticas de referência e validação textual questionáveis, em outros termos, por uma textualidade meramente indexical, na qual se perde de vista o exame das fontes e de seu real conteúdo. Pretende-se, em síntese, refletir sobre a naturalização dessas práticas e, nessa chave, sobre a atual admissibilidade das caixas-pretas generativas.

Palavras-chave: inteligências artificiais; grandes modelos de linguagem; pesquisa científica; epistemologia; retórica.

Resumen

La comunidad académica se enfrenta actualmente al uso indebido de la inteligencia artificial generativa en la producción académica y la investigación científica. Aumenta el número de informes y retractaciones relacionados con textos publicados recientemente en revistas especializadas, así como la circulación de publicaciones elaboradas de manera inapropiada mediante IA. Las causas de esta situación merecen ser debatidas, sobre todo porque ponen en entredicho la autoridad epistémica. La hipótesis que se explora en este artículo es que las condiciones que propician este fenómeno son anteriores a la IA generativa; derivan de una economía de la publicación académica impulsada por el «productivismo» y por prácticas cuestionables en materia de citación y validación textual —o, dicho de otro modo, por una forma de textualidad meramente indicial que pierde de vista el examen real de las fuentes y su verdadero contenido. En resumen, se busca reflexionar sobre la normalización de estas prácticas y, en este contexto, sobre la aceptabilidad actual de las «cajas negras» generativas.

Palabras-clave: inteligencias artificiales; grandes modelos de lenguaje; investigación científica; epistemología; retórica.

1 Introduction

Since 2022, with the public launch of ChatGPT, a notable hype around new computational possibilities can be observed. Generative artificial intelligences (gen-AIs) have been making themselves present in all areas of individual and social life, scientific research being no exception. The availability and prodigious practicality of these resources have made not only their use irrefutable but also their abuse predictable.

Unlike symbolic computation, with its if-then-else chains, which allowed John McCarthy and colleagues to speak, in a different sense from the current one, of “artificial intelligence” (AI) as early as 1955 (McCarthy *et al.*, 1995), the processes that generate today’s formidable results are not auditable or even entirely understandable in their architecture. Current computational power is stochastically applied to immeasurably large masses of data, producing increasingly complex and variegated associations and chains of meaning. Additionally, except when “hallucinations” particularly stand out, outputs of these machines are, in general, uncritically received as palatable and usable for the vast majority of purposes.

Everything happened very quickly. There was no time, for example, for a sober and in-depth debate on the “epistemological status” of the results thus produced. In other words — and this is precisely what needs to be detailed and investigated —, these results resemble “inductively legitimized opinions” with “statistical parameters” that, despite reasonable for most uses, end up thoughtlessly and indulgently invested with epistemic authority. Rather, they end up receiving, for all intents and purposes, the status of “knowledge”.

Without neglecting the effects of the ubiquitous use of these tools in everyday life, the issue becomes especially acute when we look at their presence in scientific research. Given the logic of profit and competitiveness that structurally moves the business model centered on big techs, scientific institutes and research universities constitute the ultimate framework in which the discussion of these effects can be carried out independently and responsibly. Additionally, it is difficult to say what will happen if even these reflective niches are decisively captured by the methodological premises they are supposed to problematize.

The issue is that these institutions also find themselves grappling with academic productivism¹ or, worse, with pervasive productivity demands from industry and its imposingly utilitarian applications. The hypothesis here is that the productivism of the last decades fertilized the soil for current methodological licenses, with a consequent delegation of authority and responsibility to processes which are, in truth, opaque even to the most specialized users. However, in a scenario in which it is necessary to publish in order not to perish, and, furthermore, in which it is necessary to index and capitalize on what is produced to survive, the use of the current smart machinery becomes irresistible.

It would be surprising and disheartening if this were not the case, but reports of spurious uses of these technologies began to appear — arising from haste, expediency, naivety or simple dishonesty. An example with reasonable visibility is the 2023 article, by Richard van Noorden, which relates the withdrawal of more than ten thousand articles published in specialized journals that year (Van Noorden, 2023)². Although the text shows occurrences located predominantly in special issues, in conference papers, and in countries such as Saudi Arabia, Pakistan and Russia, the diagnosis is worthy of attention.

¹ Academic productivism is, in a general sense, the subordination of knowledge production to a logic and rhythms dictated by competitive practices necessary for professional survival, in a system increasingly demanding of results and dependent on funding. While we will not retrace the history of this productive pressure, identifiable by anyone in the knowledge professions, reflections on the problem can be found, for example, in Rezende *et al.* (2023) or in Zuin and Bianchetti (2015).

² In addition to Noorden’s article, a tracking source for public article withdrawals is the RETRACTION WATCH website.

A similar article was published by Swedish professors Jutta Haider, Kristofer Söderström, Björn Ekström and Malte Rödl (Haider *et al.*, 2024). The authors express strong concerns with the proliferation of spurious texts embedded in Google Scholar, especially in sensitive areas such as health, environmental studies, and computing, possibly producing misinformation, tampering with meta-analyses, and hacking public policy support bases — texts, to make matters worse, which are difficult to effectively withdraw from circulation.

These findings certainly give rise to regulatory efforts, ranging from general legislation such as the AI Act (European Union, 2024) to academic manuals such as the “*Living guidelines on the responsible use of generative AI in research*”, the second edition of which was released by the European Commission in 2025 with the result of an analysis of twenty-one similar guides, published by universities, governmental, and non-governmental institutions. The problem is that, although welcome and necessary, the advent of these norms does not account for the roots of the phenomenon, nor is it capable of effectively regulating or mitigating its consequences³.

In other words, in addition to the side effects caused by the spurious or reckless use of these new technologies, it is necessary to discuss what systemic changes their widespread adoption has been causing to scientific research, more general academic production, and scientific credibility, all while faced with a society exposed to a flood of information.

The understanding of what is most radically at stake defines, in turn, a pressing demand for critical reflection on academic productivism itself, which involves the rediscovery of the *raison d'être* of scientific production and, even more seriously, of what, at this point, we still classify as “scientific knowledge”. As these demands transcend the scope of an essay like this, we shall be content to shed light on the need to topically shift the current discussion towards the above-mentioned priorities.

Accepting this as what is within our reach, the following section draws on the literature available to discuss the reasons why even zealous researchers today resort so eagerly to these new forms of “knowledge production”. After listing some of the more tangible risks of this eagerness, we suggest that the current indulgence with “generative black boxes” has roots in methodological licenses⁴ and rhetorics hitherto little discussed, although they have

³ This insufficiency was addressed, for example, by Lyra and De Souza (2025), in “Less Visible Risks of AI-based Technologies: The Civilizational Threat of Discursive and Epistemic Dysfunction”.

⁴ In other words, throughout this essay we draw attention to tolerances, conveniences, and methodological naturalizations.

been in force for much longer in the world of publish or perish⁵. Finally, we indicate how practices of denaturalization of these behaviors may, perhaps, among other efforts, contribute to the most common attempts to regulate the use of AI through written and institutionally approved laws and standards.

2 Reasons for general adherence to the use of generative AIs in scientific research

We start with Lisa Messeri and Molly Crockett's 2024 text, which discusses the illusions of scientific understanding in the age of research which has been turbocharged by generative artificial intelligences. According to the authors:

Scientists must contend with various research constraints, including finite time, fixed budgets and limited cognitive capabilities. AI tools are viewed as solutions to these barriers, enabling scientists to be more productive (in terms of generating more scientific work) and more objective (free from bias and subjectivity). (Messeri; Crockett, 2024, p. 50).

The authors identify four categories of AI use, uses that aim to circumvent restrictions to greater productivity. These categories are recovered to support broad considerations on the subject, as the article itself allows us to glimpse that the choices and interweaving of these uses by scientists depend on research type and criteria, which can be more or less flexible, according to convenience and occasion. The four categories are: Oracle, Surrogate, Quant, and Arbiter.

The first category accounts for the use of AIs for curating and streamlining systematic literature reviews. Given the enormous volume of material produced worldwide and the need to not repeat what has already been said, resorting to machines that cover the field becomes necessary. Of course, dealing with this methodological need can vary greatly, but it is possible to point out some snags that are difficult to get around. The first is that machines will more easily detect what is stochastically more present in the bases covered, which inevitably introduces a kind of remissive endogeny in research, leaving the question of whether the use of AI is really capable of eliminating bibliographic biases or if, on the contrary, it produces conveniently highlighted methodological recursions. The suspicion is that, increasingly, AI curators exclusively select articles that, having made "oracular use" of generative AIs, maximize their chances of integrating the databases covered by such curators in the future. Authors such as Ilia Shumailov and colleagues from the United Kingdom and

⁵ It is difficult to pinpoint the source of this expression, as its use today is almost cynically naturalized. We suggest some reading in this regard in note 1 above.

Canada even discuss whether this recursion could “collapse the models” of genAIs that produce them (Shumailov *et al.*, 2024)⁶.

The expansion of the field in question and the flexibility of the parameters, in turn, possibly adopted to mitigate responsive narrowing of curator AIs, may generate an increasing number of titles to be examined. Faced with the impossibility of doing it the old way, reading article by article, book by book, the solution is again to delegate to the machine the task of making summaries and organizing lists of more relevant articles, and these summaries and hierarchies will also need more or less judicious, or biased, prompts and validations, accompanied by at least one very serious risk, which is that of “dealing with sources in a merely indexical way”, particularly serious in fields such as the Humanities and Social Sciences. We will return to this in the fourth section of this article. In any case, this possibility is linked with the title of the article by Messeri and Crockett (2024): one migrates from impotence in front of a huge and unknown ocean to a wisdom which is mechanically anabolic and illusory in its real understanding of what it purports to explain.

The second category — Surrogate — relates to the prior or abbreviated testing of hypotheses, aiming to reduce, or even circumvent, the performance of laborious field research or expensive laboratory tests. In other words, it involves simulating the behavior of analyzed entities and phenomena, aiming to skip steps and restrict the field of final experimentation. There is also a lot of room for choices and criteria definition, with questions ranging from the representativeness of the simulating algorithms of the tests, to the discernment of what must imperatively be sought in the real tests.

The category of Quant seems to us closely related with that of delegation, without overlapping it. It seems to us to be more focused on tidying up the research field and fermenting initial ideas than on the prior testing of human generated hypotheses. The rapid scrutiny of large masses of data in search of patterns, regularities, deviations, and anomalies can, in fact, quickly point out issues worthy of investigation, though it remains to be discussed the extent to which it does not offer a subterfuge for conducting bureaucratic or artificial research, without genuine hypotheses or real background questions. It is clear that today there are domains into which the human mind cannot peer in finite time, which, therefore, demand the use of generative AIs. However, it is worth asking whether this demand or methodological license does not risk pushing research, even in niches in which such scanning is not strictly necessary, to the mere identification and calculation of

⁶ It should be noted that the careful analysis, in its methodology, of the trends indicated by the authors is not within my momentary reach. The concern is still worthy of record.

regularities and deviations, discouraging more complex heuristics of interpretation and understanding of the world.

In any case, together with the two previous categories, such a use of AI seems to account for a quantitative and artificial increase in productivity, especially narrowed in the direction of what Messeri and Crockett (2024) call the “monoculture” of scientific production. If the path that guarantees survival is the most agile and technologically traversed path, the dictates of this agility understandably come to figure as standard procedures.

The fourth category is Arbiter. Its imaginable use is diverse. The validity of the use of AIs to evaluate academic submissions is discussed, raising the question of how far this outsourcing can go, as is also questioned in the judiciary system, to what extent is it appropriate, without risk of bias and injustice, to delegate heavy judicial work to AIs. In the case of peer review, there is a considerable snowball effect to consider: the use of AIs in research increases the production of articles and the number of submissions to be evaluated, delegating at least part of the work to AIs. Add to this the practical temptation to use these tools in situations in which the reviewer is not strictly familiar with the topic and the return period is short (Donker, 2023).

It is also reasonable to conjecture that, in order to maximize their chances of acceptance by journal A or B, researchers “with a more pragmatic spirit” use generative AIs to previously review their articles and adapt them to the respective reviewing standards, taking care, of course, to erase these traces before finalizing the submission. To be fair, the introduction of appropriate scientific-intellectual mannerisms in texts is nothing new in the world of publish or perish, recalling the Sokal case⁷, when an article full of bibliographic references and occasional jargon was accepted by a qualified journal (*Social Text*), even though, as its author later admitted, it only contained theoretical nonsense and phrases without real meaning (Sokal, 1996).

3 Concerns

The typology proposed by Messeri and Crockett (2024) is part of an attempt to draw attention to some risks caused by the use of AIs in the pipeline of scientific production, risks insufficiently discussed even by the community of its most careful users. The article begins by reporting the flirtation with the full automation of the use of AI in scientific

⁷ The entry “*Sokal Affair*”, from Wikipedia, offers a great introduction to the topic, including good bibliographic sources. For those who wish to go deeper into the imbroglio, it recommends reading the book written by Sokal and Bricmont (1998), entitled “intellectual impostures”, with the caveat that it advances a rather dubious diatribe against a certain postmodern intelligentsia of the time.

research (Messerli; Crockett, 2024, p. 49). It then lists more commonly recognized threats to science and society: algorithmic bias, environmental cost, general misunderstanding of the real capabilities of AI, and, finally, the implicit exploitation of human labor. The authors make timely mention of Kate Crawford's excellent book (2021), *Atlas of AI*, which evidences how much disproportionately low-paid labor needs to be mobilized, and how much infrastructure needs to be installed for Alexa to turn on the light in our home using a simple voice command.

Considering that the recognition of these risks is usually accompanied by the hope that they can be addressed in a technical way⁸, the authors shift the focus to other snags even more insufficiently discussed and difficult to mitigate, the most important of which is indicated in the title of the article: the illusion of growing understanding of the world, provided by the introduction of generative AIs. Once again, we freely explore these contributions, clarifying them and adding other effects of the systematic use of gen-AIs in scientific research, sometimes even more serious than those actually explored in the article.

The question that opens their final section, "Why do scientists believe in AI tools?" (Messerli; Crockett, 2024, p. 52) is both "rhetorical" and relegated to a limbo of irrelevance and inability to cope. It is rhetorical in a "practical" point of view, and the quadruple categorization proposed by the authors can factually answer it. Either one "believes in" and incorporates these tools into research, or one falls behind and "perishes". However, theoretical, and especially philosophical, reflection on the subject is avoided by ignoring questions about the epistemological status of what is produced by generative AIs. How is "knowledge" here understood? Is it true that nothing more important is lost when we take for knowledge information that is stochastically validated?

Although they offer a framework with good questions in this latter direction, the authors turn their focus to the illusory understanding hypertrophy that new practices and methods end up feeding. We end up having a kind of reissue of the old scientific pedantry⁹, now backed by artificial intelligence. In fact, belonging to scientific communities and access to what is discussed in them can feed the illusion of knowing more than is actually the case. It is common and noticeable, for example, that graduate students participating in these communities, proud of this belonging, talk about the results and formulations of senior researchers as if they had full understanding of these syntheses, not considering, in the present scope, the moral issue of stating or omitting credit. The most relevant question here

⁸ We conjecture that the authors are referring to technical-bureaucratic instances or restrictive legal regulations.

⁹ We use the term "pedantry" in this essay in the sense of posing as knowing that which is not actually known—posturing not necessarily accompanied by arrogance or affectation.

is epistemological: the fact that certain researchers do not infrequently stumble when asked to explain the syntheses that they opportunely use. Of course, not only postgraduate students are given to such pedantism, so it is worth insisting that our interest in this habit is more structural than moral.

The fact is that, with generative artificial intelligence entering the scene, things have become quite more complicated. Often treated as “research colleagues”, these devices allow one to speak with remarkable authority about subjects that have not been investigated or even carefully examined. This “symbiosis” not only authorizes discourse on these topics but also allows information that is not properly problematized and validated to be rapidly used in scientific production and dissemination. The illusion of having a broad view not only of the contours of the field, but of its more complex injunctions, is doubly deleterious, discursively and cognitively, producing “textual-discursive garbage” that can be incorporated into databases from which new “information” will be collected (Naddaf; Quill, 2025)¹⁰.

In a slightly different note, the authors of our guiding text use the agricultural metaphor of “monoculture” and its “susceptibility to pests and diseases” to refer to the risk generated by the comprehensive illusion now at stake¹¹. The idea of monoculture is justified by the tendency to prioritize what AIs are capable of processing and, *pari passu*, the epistemic and discursive implications that are part of them. Radicalizing the threat, we may foresee the emergence of the *homo prompter*¹² in our civilizational horizon, one who not only confines their cognitive abilities to the computational way of operating, but who begins to speak as if they were a machine.

However, by prioritizing the structural aspects of the growing tendency to use AI in the production of “knowledge”, other complications can and should be noted and discussed, even in a more speculative way. Investing in the idea of *homo prompter*, what new “electronic brains” seem to do is to raise to hitherto unknown powers a capacity of the human spirit particularly amenable to computational mimesis, which is the associative-inductive capacity. The “discourse” thus produced proves acceptable for operating with

¹⁰ The authors suggest that tens of thousands of references found in articles published in 2025 are invalid. See also Roe, Furze and Perkins (2025), an article that compares the productive profusion of generative AIs to a kind of “digital plastic” – created by man, ubiquitous, malleable and persistent in information ecosystems.

¹¹ Pests and diseases (*pests and diseases*), metaphors used by Messeri and Crockett (2024, p. 53), refer to vulnerabilities to errors of understanding and bias, as well as a discouragement regarding innovation. The authors mention several works which are attentive to the value of epistemic diversity and the dangers of monocultures.

¹² A similar term—*homo promptus*—has been proposed by Linden, Hachem and Kondyli (2025) in the article “*Homo Promptus*: Predicting the Impact of generative AI on human memory and creativity”. The conceptual construction is very interesting and worthy of attention, though not exactly congruent to the one presented here. Further work on possible complementarities in a future article would be worthwhile.

unimaginably large databases and peering into them at breakneck speeds. The human being, narcissistically fascinated by this partial and prodigious mirror, easily forgets the other faculties of the spirit that allow him to heuristically elaborate reasonable or convincing discourses, that is, without having the myriad speeds and bases of consultation that machines use.

Since the issue of this “ostracism” is important, we evoke, for now, lived and remembered experience, the layers of meaning production that mobilize emotions, stories and reputations, intuitions and contextualization, doubt, interrogation, and the search for meaning in different levels of amplitude and depth being fundamental in this construction, having as a counterpoint the human measure of what is possible and reasonable. Co-opted by an artificial way of being and intrinsically deprived of access to the formidable operational power of generative AIs, the discursivity of the *homo prompter* loses plasticity and becomes a kind of frayed fabric, rife with holes and gaps, which makes it more and more dependent on the devices to which it addresses its prompts.

There is a necessary caveat: a strongly speculative hypothesis like the one formulated above, filled with images and contextual illusions, could hardly be produced in a stochastic-inductive way, even granting machines everything that their apologists proclaim. Admitting that this conjecture needs to be properly tested, we are sure that no one who has read this text thus far will be horrified by it.

Finally, it is not possible to make an exhaustive list of the human capacities ostracized by the advent of *homo prompter*, so we offer only one more illustration of the current state of affairs, the one that we consider perhaps the most imperceptible of all. It regards our ability to live longer with what we do not know, to keep questions in suspense and reorganize our certainties and cognitive fields from these suspensions. Until a few decades ago, when a question assailed us, we had to hope that we could address it to whomever could answer it, be it a person, a book, or nature. Even simple questions, such as what is Namibia’s capital, had to wait for an atlas to come within reach. Until then, we were left thinking that it could not be Luanda or Nairobi. The question that remains is what happens when this epistemic and temporal experience gives way to the immediacy of smartphones and the AIs installed in them.

Things tend to get complicated when entering the field of current scientific practices. By definition, genuine research, which goes beyond the search for information available elsewhere, addresses something that is not known and needs to be investigated, needing to use methods which are valid and appropriate to the problem, as well as patience

and inventiveness. Here we point out and emphasize the need to continue detailing how the availability of oracles, behavioral simulators, pattern detectors and response enhancers interferes in these investigative processes.

4 Lightning striking in a blue sky?

Circling back to the beginning of this essay, the frenzy associated with the introduction of generative artificial intelligences was impressive when, at the end of 2022, OpenAI launched ChatGPT-3.5 for public use. That first version answered questions on the most varied subjects, generating a mixture of enthusiasm and restlessness in various areas of human activity.

In academia, discussions centered on the detection of plagiarism, the effects of the tool on student and teacher activities, and the impacts of this invention on research, development, and publication of scientific articles (Van Dis *et al.*, 2023). Everything seemed to boil down to a question of anticipating and curbing misuse, of preventing damage and repairing damage, as if we were not aware of the chthonic modifications underway in the manner we deal with knowledge, in the constitution of our subjectivities, in short, in our general insertion in the world.

The truth is that little consideration has been given to the relationship between certain implicit beliefs about education and research, already widespread at the time, and the fear triggered by the advent of gen-IA. By giving us something that resembled quick knowledge, supported both technically and scientifically, gen-AIs mimicked, with several advantages, what was being done in schools and universities, which is being transformed in the wake of multiple and uncomfortable productivity requirements, although little examined and discussed (Glass, 2018)¹³.

It certainly wasn't lightning in a blue sky. In the best-case scenario, which is riskily bet upon here, this disruption could open an unusual window of opportunity to discuss very important issues which have been, thus far, closed off by productivist inertia.

It may be hard to admit, but in more or less sophisticated ways, it has been the usual in the academic establishment to search for information and combine it competently, coherently, justifiably, tasks frighteningly facilitated by quick and summarized access to

¹³ The author worked for many years as a *Writing Studies* teacher in the USA and followed the changes in students' academic writing related to the contribution of new computational technologies. His doctoral thesis, with a strong freirean reference, was entitled "Software of the Oppressed: reprogramming the invisible discipline".

digital bibliographic references, meta-analyses, and data mining, in addition to the simulation, quantification, and review functions discussed above.

A caricature — though unfortunately not particularly cartoonish — provides a way to better understand the phenomenon. It has become common in academic textuality in recent decades, certainly more so in some areas than in others, for a researcher to seek to validate a proposition by making it be followed by the formula: (X, 2025; Y, 2014; Z, 1999; W, 2018). Even with the “early” search engines, note-taking and traditional direct reading were already being supplanted by the availability and use of indexers, quick summaries, and content associations increasingly driven by “distant reading” — assuming, that is, that these new digital hermeneutics can even be linked to “reading” in any palatable sense¹⁴.

Just imagine the discomfort, in a doctoral thesis defense, of someone asking in what sense the author of the work examined thinks that Y, in his 2014 publication, would be saying essentially the same thing as he did in his text. Or that X (2025), Y (2014), Z (1999) and W (2018) may be saying the same thing about the topic under validation? However, since it is rare for anyone to ask this type of question, it remains to be determined why a certain scientific community has allowed such a form of “epistemic validation” to have transit.

It is worth asking, rhetorically and epistemologically, why pseudo-inductive devices like these are allowed to function as forms of discursive validation. Does it imply that the reader should know all the sources provided? Rather, who would bother to gather them beforehand, and stop reading at every mention to reach for the source and check if it says what the author says it says? Or should they, *a priori*, believe in these references as today they believe in gen-IAs? Why, after all, today, in countless academic fields, is the requirement waived, except for literal citations, that the “references” be accompanied by the pages where what is alleged had been said? What about occasions when, as is common in the humanities, voluminous classic books are indexed, mixed with diverse articles, short and not so short? Is it suggested that there is, in classic texts, a monolithic idea, devoid of tensions and confirming the postulation of the one who uses it? And that the reader knows the nuances of the sedimentations made in the bibliographic reference?

Of course, there are situations in which such practices are understandable, for example, when it is intended to show that many researchers have focused on a theme or issue, but it is not to these exceptions that we draw attention to here. Nor do we bother to illustrate and map these occurrences, believing that the simple call to attention will allow the

¹⁴ A preliminary version of this discussion can be found in Lyra (2024): “ChatGPT as a Digital Iceberg — educational considerations”.

reader to identify them here and there without difficulty. It is no exaggeration to point out the ill-fated joke played by Alan Sokal on the journal *Social Text* is itself an indication of the acceptance of undue and supposedly authoritative indexical validations. The article displayed, overall, more than 200 titles of diverse breaths in its starry bibliography.

Beyond the perhaps hyperbolic and generalizing nature of this discussion, generative AIs are far superior to any human being in the art of examining data, of assigning values to possible associations, of identifying what is usable as a function of prompts or set goals, and to combine the selection in the form of a “smart” response. With human feedback and supervision, driven by the increase in computational capacity perhaps at quantum levels, these machines will become more and more “knowledgeable”, and it is even difficult to foresee how this interactivity will evolve.

As a necessary parenthesis, in other occasions we dealt with an even harder and longer core of this historical construction, left aside due to the focus of this essay on research turned to scientific publication. Put succinctly, at the end of the first decade of the previous century, John Dewey already drew attention to deleterious educational practices, especially in basic science education, guided by the preferential presentation of the arrival points of the processes of knowledge production, that is, of stabilized models at the end of formidable historical and methodological sagas, without these sagas being minimally evidenced. We can certainly conjecture that this school-validation of scientific black boxes would be corrected in the higher segments of education, especially among researchers. But, judging by the state of affairs we are dealing with here, this is definitely not the case¹⁵.

In summary, if our standards of education and knowledge construction were already so conditioned by the whip of productivity and epistemic tolerance, to the point of accepting “heuristics” such as those described above, the admissibility of the black boxes characteristic of stochastic calculations should not cause any astonishment. If this was all knowledge is about, perhaps it’s better to let generative AIs simply take the helm, as they manage “stochastic mining” of X, Y, Z and W better than any of us.

5 Final remarks

The considerations made so far point to several very serious risks of the hegemonic and mandatory use of generative AIs in life in general, particularly in the production of

¹⁵ A more detailed presentation of this formulation can be found in Lyra (2025): “for a philosophy of opinion” (p. 21 *et. seq.*). The works of John Dewey evoked are: *Science as Subject-Matter and as Method* (1910) and *Method in Science Teaching* (1916). The example that paradigmatically served me was that of the solar system, commonly presented to students as if it did not have a formidable history of debates and validation of hypotheses.

knowledge, in scientific research and, even more incisively, in theoretical reflection on these habits and their consequences.

It is not relevant to list them all here, if only because they intertwine with one other in ways that are difficult to foresee in their myriad possibilities. These final considerations divide themselves, therefore, around only three main points.

The first point is the emphasis on the discursive-epistemic bankruptcy that these habits help perpetrate and accelerate. This corrosion threatens to liquidate real possibilities of theorizing about the effects of the hegemony of the AIs and, with it, the reasoned discussion about possible care and containment of the current technicist furor. In this case, on the horizon we see the problem folding on itself, which risks making the scientific community and its various stakeholders, paraphrasing the philosopher Socrates, no longer able to realize that they do not realize what they are doing or saying. In this scenario, we live the situation that Kant said should be avoided at all costs, the one in which “one milks the goat while the other holds the sieve” (1994, p. 83).

The second point to recapitulate and emphasize is the idea that the ground for the current admissibility of computational black boxes was being prepared for a long time by a culture of results and productivity that, historically and pervasively, intruded into the heart of scientific production. This point is particularly important because it paves the way for some modest, albeit feasible, alternatives of resistance to the current state of affairs, alternatives addressed above all to researchers still capable of metacognitive questions, no matter what their original area. Transdisciplinarity, with all its difficulties, is certainly a very important element for the success of any epistemological alternative.

The third and final point is an illustration of some positions and practices of guidance, research, and dissemination of results, which, not meant to be exhaustive, allow similar ideas to come to light. It is, above all, a matter of decidedly abandoning merely indexical references or, in other terms, of treating authors and their works as persons, not as authoritative indexes or mere providers of discursive validation. This means having to contextualize and justify the referrals, investing in discursivities that are more focused on sharing arguments than on the bureaucratic search for legitimacy. Would it be too much to want, after all, to encourage the new generations to practices that are more interrogative, less pedantic, more honestly hesitant, less defensive, above all more generous and desirous of real dialogue?

My rhetorical hope is that the denaturalization of forms of validation such as those described above can contribute to the reestablishment of new levels of dialogical and

epistemic care, in which everyone talks about what they know, as far as they know; that they strive to make themselves understood, to themselves as well; that they have the courage to say that they did not understand (and continue investigating) what they did not understand; finally, that they start talking and writing as if they were people, not machines producing indexable articles.

References

- CRAWFORD, K. *Atlas of AI: power, politics, and the planetary costs of artificial intelligence*. New Haven: Yale University Press, 2021.
- DEWEY, J. Science as Subject-Matter and as Method, *Science*, Washington, v. 31, n. 787, p. 121-127, Jan. 1910.
- DEWEY, J. Method in Science Teaching, *General Science Quarterly*, New York, v.1, n.1, p. 3-9, Nov. 1916.
- DONKER, T. The dangers of using large language models for peer review. *The Lancet Infectious Diseases*, London, v. 23, n. 7, p. 781, Jul. 2023.
- EUROPEAN COMMISSION. *Living guidelines on the responsible use of generative AI in research*. Brussels: RTD Publications, 2025.
- EUROPEAN UNION. *Artificial Intelligence Act* (Regulation (EU) 2024/1689), Brussels. Available at: <https://artificialintelligenceact.eu/the-act/>. Accessed: 16 May 2026.
- GLASS, E. R. *Software of the Oppressed: reprogramming the invisible discipline*. 2018. Thesis (PhD in Philosophy) — Cuny University, Nova York, 2018.
- HAIDER, J.; SÖDERSTRÖM, K. R.; EKSTRÖM, B.; RÖDL, M. GPT-fabricated scientific papers on Google Scholar: key features, spread, and implications for preempting evidence manipulation. *Harvard Kennedy School Misinformation Review*, Boston, v. 5, n. 5, p. 1-12, Set. 2024.
- KANT, I. *Crítica da Razão Pura*. Lisbon: Gulbenkian, 1994.
- LINDEN, K.; HACHEM, H-H.; KONDYLI, V. Homo Promptus - predicting the impact of generative AI on human memory and creativity. *Memory, Mind & Media*, Cambridge, v. 4, n. 15, p.1–20, Set. 2025.
- LYRA, E. O ChatGPT como Iceberg Digital — considerações educacionais. *O que nos faz pensar*, Rio de Janeiro, v.32, n.54, p.105-121, Jan.-Jun. 2024.
- LYRA, E. Por uma Filosofia da Opinião. Rio de Janeiro/Lisbon: NAU/Documenta, 2025.
- LYRA, E.; DE SOUZA, C. Less Visible Risks of AI-based Technologies: the Civilizational Threat of Discursive and Epistemic Dysfunction. In: TARANTOLA, A. M. (org.). *Artificial Intelligence and Care of Our Common Home: a Focus on Industries, Finance, Education and Communication*. Milan: Vita e Pensiero, 2025. p. 63-76.

MC CARTHY, J.; MINSKY, M.; ROCHESTER, N.; SHANNON, C. *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence [online]*, ago. 1995. Available at: <https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>. Accessed: 12 Apr. 2026.

MESSERI, L.; CROCKETT, M. J. Artificial intelligence and illusions of understanding in scientific research. *Nature*, London, v. 627, p. 49-58, Mar. 2024.

MESSERI, L.; CROCKETT, M. J. AI surrogates and illusions of generalizability in cognitive science. *Trends in Cognitive Sciences*, London, v.30, n.3, p. 203-2015, Mar. 2026.

NADDAF, M.; QUILL, E. The growing problem of hallucinated citations. *Nature*, London, v. 652, p. 27-29, Apr. 2026.

RETRACTION WATCH: Tracking retractions as a window into the scientific process. Available at: <https://retractionwatch.com/>. Accessed: 16 May 2026.

REZENDE, C. H.; BONANNI, R. J. O.; AZEVEDO, H. J. C. C. Publish or perish: a cultura acadêmica em crise. *Revista Científica Multidisciplinar Núcleo do Conhecimento*, São Paulo, v. 06, p. 62-74, Jul. 2023.

ROE, J.; FURZE, L.; PERKINS, M. Digital plastic: a metaphorical framework for critical AI literacy in the multiliteracies era. *Pedagogies: an International Journal*, Singapore, v. 21, n. 2, p. 1-15, Set. 2025.

SHUMAILOV, I.; SHUMAYLOV, Z.; ZHAO, Y.; PAPERNOT, N.; ANDERSON, R.; GAL, Y. AI models collapse when trained on recursively generated data. *Nature*, London, v. 631, p. 755-759, Jul. 2024.

SOKAL, A. Transgressing the Boundaries: towards a transformative hermeneutics of quantum gravity. *Social Text*, Durham, n. 46/47, p. 217-252, Spring/Summer. 1996.

SOKAL, A.; BRICMONT, J. *Imposturas Intelectuais*. Rio de Janeiro: Record, 1998.

VAN DIS, E. A. M.; BOLLEN, J.; VAN ROOIJ, R.; ZUIDEMA, W.; BOCKTING, C. L. ChatGPT - five priorities for research. *Nature*, London, v. 614, p. 224-226, Feb. 2023.

VAN NOORDEN, R. More than 10,000 research papers were retracted in 2023: a new record. *Nature*, London, v. 624, p. 479-481, Dec. 2023.

WIKIPEDIA. *Sokal Affair*. Available at: https://en.wikipedia.org/wiki/Sokal_affair. Accessed: 16 May 2026.

ZUIN, A.A.S.; BIANCHETTI, L. O produtivismo na era do “público, apareça ou pereça”: um equilíbrio difícil e necessário. *Cadernos de Pesquisa da Fundação Carlos Chagas*, São Paulo, v. 45, n. 158, p. 726-750, Oct./Dec. 2015.

Translation: Ananda Missailidis

Submission date of original in Portuguese: 22/02/2026

Approval date of original in Portuguese: 18/05/2026