Green Analytical Chemistry 17 (2026) 100349

Contents lists available at ScienceDirect

Green Analytical Chemistry

journal homepage: www.elsevier.com/locate/greeac

Reflections on the impact of artificial intelligence on peer-review practices and its implications for greener scientific evaluation

Adrián Fuente-Ballesteros *, Vânia G. Zuin Zeidler *

Institute of Sustainable Chemistry, Leuphana University Lüneburg, 21335 Lüneburg, Germany

A R T I C L E I N F O

Keywords:
Generative AI (GenAI)
Large language models (LLMs)
ChatGPT
Peer-review
Scholarly publishing
Editorial practices
Research integrity and excellence
Sustainability
Green analytical chemistry

A B S T R A C T

Artificial intelligence (AI) is becoming a common presence in scientific publishing, yet its use in peer-review has received much less attention than its role in manuscript preparation. This article aims to analyze the use the structural pressures that drive reviewers toward AI use, including time constraints, reviewer scarcity, and performance-based incentives. It contrasts critical human reading with automated or template-based reports, identifies recurrent signals of AI-assisted reviews, and examines their ethical, emotional, and sustainability implications. Attention is given to how AI may influence metric-based evaluations and reinforce superficial or score-driven interpretations of scientific quality. We argue that the central risk is not simply factual error, but the gradual normalization of procedural evaluation over intellectual scrutiny. As automation becomes routine, peer-review may shift from a space of critical dialogue to a system of opaque filters. To address this challenge, we propose the need for "meta-assessment" frameworks capable of evaluating not only scientific methods, but also the quality and transparency of the evaluation process itself. Moving forward, peer-review must integrate AI with human oversight and transparent standards, so that efficiency supports rather than replaces critical evaluation, contributing to greener and more sustainable practices in our community.

1. Introduction

Artificial intelligence (AI) has become part of the daily activity of scholarly life. Every day a new AI-based application for studying, writing, or reviewing is mentioned, social media platforms are saturated with discussions devoted to this single topic, new documents on the "responsible" use of AI appear at an accelerating pace, and an ongoing race is underway to develop faster, more powerful systems capable of outperforming competitors. In this context, the presence of AI in research no longer feels optional, but routine. Most early discussions have focused on the use of AI in manuscript preparation [1], as well as on the development of AI-assisted workflows for generating scientific review articles [2]. However, its growing presence in peer-review has received far less attention [3,4]. Reduced visibility in public debate does not imply limited adoption [5]. A survey conducted by the publishing company Wiley, involving nearly 5000 researchers across >70 countries, reported that approximately 19% of respondents had already experimented with large language models (LLMs) to streamline and facilitate their peer-review activities [6]. According to Kousha et al. [7] AI-driven automation in publication-related tasks may support several stages of the scholarly communication process, including (i) suggesting suitable journals for a manuscript, (ii) providing quality control checks for submissions, (iii) identifying potential reviewers, and (iv) assisting with the review and evaluation of peer-review reports. In parallel with the gradual infiltration of generative AI (GenAI) tools into academic peer-review, editors and relevant institutions have begun to issue regulations governing their use in the review process [8].

This trend is striking given that peer-review remains a central checkpoint or selection of scientific publishing [9]. The increasing use of AI by reviewers is rarely driven by curiosity or enthusiasm for novel technologies, rather, it reflects long-standing structural pressures, including short editorial deadlines, limited time for careful reading, and a system that relies heavily on unpaid reviewing labour [10]. As a consequence, AI is frequently used to generate structured and apparently fluent feedback within minutes, replacing what would otherwise require extended reading, reflection, and intellectual human engagement [11].

In this context, this article seeks to examine how, why, and with what consequences AI is currently being used in the evaluation of scientific manuscripts. Given how quickly these practices are spreading and how

* Corresponding authors.
E-mail addresses: adrian.fuente_ballesteros@leuphana.de (A. Fuente-Ballesteros), vania.zuin@leuphana.de (V.G. Zuin Zeidler).

https://doi.org/10.1016/j.greeac.2026.100349
Received 5 March 2026; Received in revised form 13 April 2026; Accepted 21 April 2026
Available online 22 April 2026

directly they affect fairness, trust and long-term viability in publishing, there is a need to look critically beyond formal rules and optimistic claims. Without such reflection, automation risks becoming a default response to a system already under strain, changing scientific evaluation in ways that are difficult to justify or even notice. This article aims to complement recent papers in the field of chemistry and AI [1,12–14].

2. Implications and challenges of AI in peer-review practices

The presence of AI in peer-review is not a single technical change, but a structural transformation that interacts with existing pressures in scientific publishing [15]. Its impact cannot be understood only through policies or tools, but through the conditions in which peer-review currently operates, including time pressure, reviewer scarcity, performance-driven incentives, and the constant growth in publication volumes. To understand its real effects, we examine the structural conditions, behavioral patterns, and consequences linked to its use currently. The following subsections analyze how AI affects editorial control, reviewer behavior, evaluation practices, and broader ethical and sustainability dimensions, highlighting these interconnected challenges from different but related perspectives (see Fig. 1).

2.1. The illusion of editorial control

Several major editorial groups (e.g., Elsevier, Wiley, ACS, RSC, Springer Nature) have reacted quickly to the arrival of AI by publishing policies and statements that define its use in scientific writing and peer-review [5]. For example, Elsevier's policy for the peer-review process highlights five key points: (i) when an expert is invited to take part in peer-review, the manuscript must be treated as confidential; (ii) reviewers must not upload the manuscript, or any part of it, to a GenAI tool, as this could violate confidentiality and ownership rights of Elsevier or the authors; (iii) reviewers must not upload their review comments or any related material to an AI tool, even for the sole purpose of improving language or readability; (iv) reviewers must not use GenAI or AI-assisted technologies to help with the review of the manuscript, because the critical thinking and original evaluation required for peer-review are beyond the scope of this technology, and there is a risk

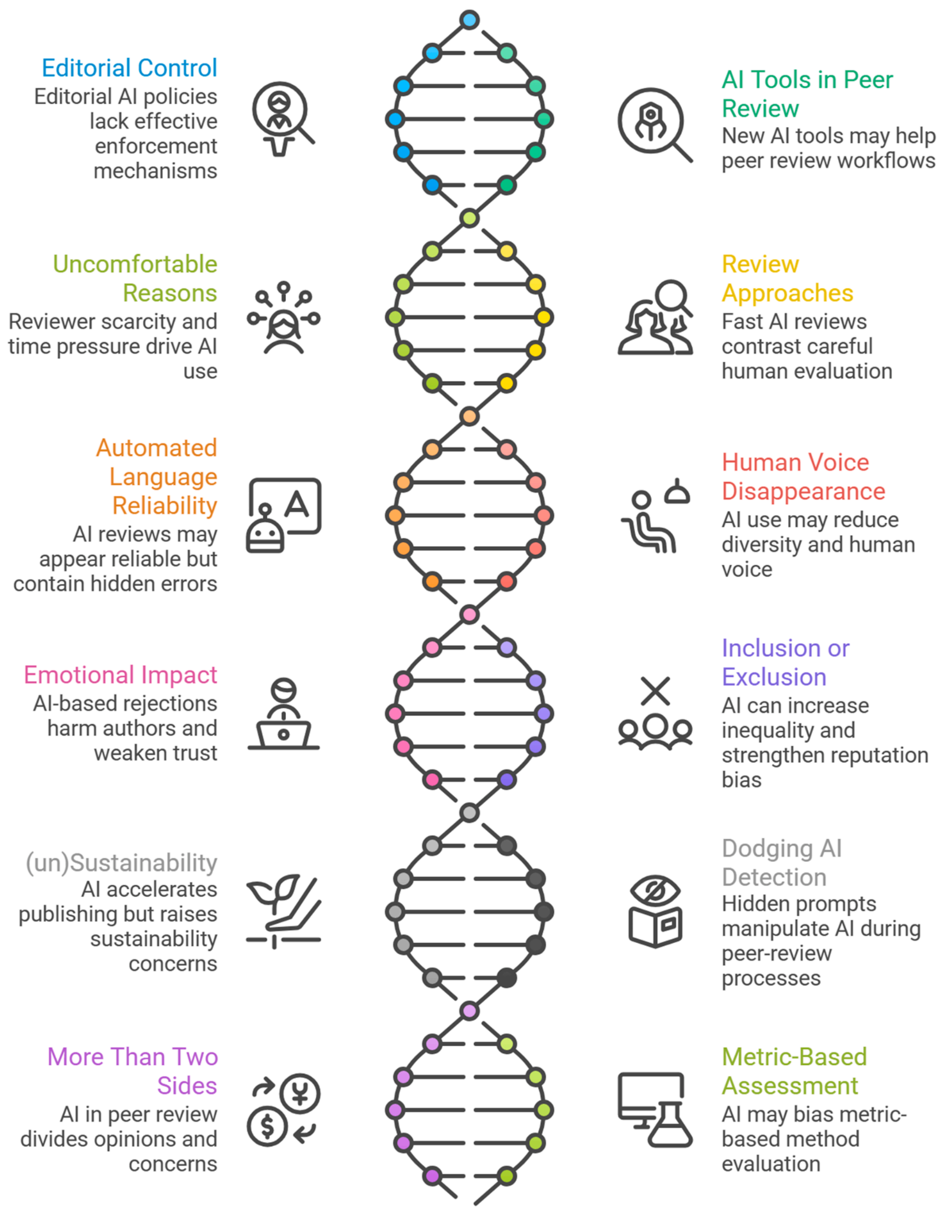

Fig. 1. Implications and dimensions of AI in peer-review.

of generating incorrect, incomplete, or biased conclusions; and (v) the reviewer remains responsible for the content of the review [16,17].

Elsevier currently forbids reviewers from using GenAI or AI-assisted review, while Wiley and Springer Nature allow a limited and declared use [18]. However, beyond the institutional rules, it is reasonable to ask what real effect these rules have in everyday practice. In most cases, there is not a safer mechanism to check compliance or to evaluate whether these measures change the behaviour of authors and reviewers. For many researchers, these policies are not seen as useful tools, but rather as vague warnings that create more caution than clarity. The risk is that such initiatives mainly serve as a form of self-publisher institutional protection, without addressing the structural challenges of AI-assisted peer-review. What could truly concern publishers is an increased number of low-quality manuscripts or submissions riddled with errors, an avalanche of AI-assisted fabrications, and, of course, the resulting criticism [15,19,20].

2.2. Use of some AI tools in the peer-review process

Recently, AI tools specifically designed to intervene in different stages of the peer-review process have begun to appear, showing that the integration of AI is no longer a future hypothesis, but a present reality [7]. Eliza [21] is an AI-powered publishing assistant that analyses text already written by the reviewer to suggest improvements to feedback, recommend relevant references, or translate comments written in other languages, without generating the review from scratch. Enago Read [22] claims to provide a smart, secure, AI-powered workspace that enhances the peer-review process with analytical support for both publishers and reviewers, aiming to improve efficiency and promote ethical practices in reviews. Veracity [23] affirms to provide comprehensive citation checking and integrity verification to help ensure the quality of information and protect academic reputation. Regarding the development and use of these tools, the scientific community holds very mixed views on whether AI systems can produce peer-reviews that are both useful and reliable [3].

2.3. Uncomfortable limiting reasons and consequences

Currently, the increasing lack of available reviewers and the effort needed to find and invite suitable reviewers are major problems for editors [24,25]. According to a survey, the most common reasons why many reviewers decline to review manuscripts are lack of time (79%), the topic of the article (71.3%), and the prestige of the journal (32.7%) [26]. Moreover, the peer-review system is under extra pressure because reviewers are asked to evaluate papers outside their area of expertise, and the same manuscripts are often reviewed multiple times due to high rejection rates [27]. The increasing challenge of recruiting qualified experts for peer-review may encourage the use of AI-based tools to screen submissions in advance or support the review process [28]. AI-assisted review processes may help reduce the workload of human reviewers and provide a preliminary evaluation of manuscripts.

To understand why many reviewers turn to AI it is necessary to look beyond the technology itself and focus on the real conditions under which peer-review takes place nowadays. One important factor is the constant pressure to reduce editorial timelines, which has turned speed into main priority and indicator of “efficiency” encouraging fast use of AI, often reducing time for careful thinking and evaluation [11,24]. Peer-review is increasingly seen as a task that must be completed quickly, rather than as a process that requires time, attention and the comprehension that is fundamental part of the scientific work, an opportunity to create fruitful dialogues for the advance of excellence in science.

For many researchers, reviewing is no longer seen as a contribution to the scientific community, but as an administrative and useless task, where the main incentives are reviewer recognition, certificates, vouchers, or discounts for future publications, the latter being the most appreciated in general [24,29]. This raises an important question: “has the rise of AI contributed to a greater willingness among researchers to accept peer-review invitations?”. The fact that reviewing is unpaid reinforces the perception of exploitation and precarious working conditions, which is particularly challenging for early-career researchers. This situation may also push reviewers to reduce the time spent on each report as much as possible [30].

The problem goes even further. We live in a culture focused on “immediacy and lack of time”, where the habit of sustained reading is being lost from an early age. In a research environment marked by pressure to publish, collect achievements, and produce results continuously and impact factors, the pleasure of reading, whether a book or a manuscript, has slowly decreased [28]. As a result, “slow reading and connecting dots” is also disappearing, even though it is important for noticing details, inconsistencies, and to promote real contributions to science development, while also raising concerns about a gradual shift of critical thinking toward AI-assisted processes. In this regard, it has been argued that ChatGPT can be a “valuable ally” when researchers lack time [31].

2.4. Review processes as different approaches

In practice, the path followed by many reviewers’ users of AI tools is quite simple. The reviewer accepts the invitation, downloads the manuscript, uploads it to an AI tool, writes a prompt of varying detail, copies the generated comments into the editorial system, and submits the review. The full sequence can take only a few minutes, in some cases around five minutes, especially when generic prompts or previous templates are reused. The contrast with the work behind the manuscript could hardly be stronger. Authors often spend months, and sometimes years, designing experiments, securing funding, repeating failed tests, analysing data, and building a coherent scientific argument [24]. Reducing this intellectual and time investment to an almost instant artificial acritical evaluation is not only a matter of speed, but also of balance and respect for scientific work. Be that as it may, in line with the increasing use of LLMs in scientific writing, it has been shown that at least one AI system can produce the first entirely AI generated peer-review-accepted workshop paper [32].

One of the most visible contrasts in current peer-review is the gap between reviews based on careful reading and those that seem built from automatic lists. On one side, there is the reviewer who explains observations, places problems in context, suggests specific improvements, and presents a personal scientific view, even when a recommendation of major revisions is given. This type of review acknowledges what is relevant and innovative in the work, points out limits, and offers guidance to improve the manuscript, with concrete examples and direct references to the text. On the other side, there is the review reduced to a list of ten generic statements, often easy to link to AI use, together with a recommendation of major revisions or even rejected, without an order of problems or solid justification. An even more problematic case appears when one reviewer provides a human, positive, and well-argued evaluation, while another suggests rejection by claiming low novelty or by relying on AI generated comments that seem to reflect an attempt to block the manuscript rather than a fair reading of the work. Thus, the NeurIPS experiment revealed inconsistencies in the peer-review process by showing that around 25% of the acceptance decisions differed between two independent review committees [33]. These issues not only frustrate authors but also may allow “weaker” research to be accepted while “strong” work is rejected. In the end, this situation can prevent papers from reaching their full potential, due to the reduction of meaningful dialogue between reviewers, AI-assisted reviewers, editors, and authors [34].

Beyond the lack of reading, there is another motive that matters when looking at these practices. In many cases, the use of automated reviews is linked to strategies aimed at increasing citation counts through coercive citation practices [35,36]. At a global level, cases have been reported where reviewers recommend, suggest, or even demand the inclusion of their own papers, co-authored work, or less often, articles by colleagues published in the same journal represented by the reviewer [29].

Such practices are unlikely to persist without an editorial setting that allows or ignores them. According to Oviedo-García et al. [29], while the first responsibility lies with reviewers who use the evaluation process for personal gain, there is a second level of shared responsibility linked to editors. Editors and guest editors fail in their role if they do not actively check the quality, consistency, and credibility of the review reports they receive, or if they do not respond to problematic patterns.

2.5. The false reliability of automated language

A risk of using AI in scientific review is its ability to produce answers that look correct even when they are not. The quality of the output depends strongly on the prompt. A vague instruction leads to generic comments, while a carefully written prompt can lead to more structured and convincing reviews. Even so, factual errors remain common [5]. AI tools may miss important methodological problems, misunderstand results, or present comments that sound reasonable but lack real meaning [28]. The issue is not only that mistakes occur, but that they are expressed with strong rhetorical confidence, which makes them hard to notice at first glance. This mix of superficial level analysis and smooth language can create a false sense of rigor, especially when such text is added directly to review reports without later critical checking [37]. This issue is further reinforced by recent evidence showing that LLMs tend to exhibit a form of "adulation", validating user perspectives [12]. AI systems have been shown to agree with users more often than humans, increasing users' confidence in their own judgments while simultaneously reducing their willingness to critically reassess their positions [38]. Translated to peer-review, this raises the risk of evaluations that prioritize affirmation over critical scrutiny. Some authors have tested this hypothesis by using an AI model, concluding that the current capabilities of ChatGPT do not allow it to truly replace human peer reviewers [37]. At present, ChatGPT can provide accurate and valuable general comments on scientific papers, but it cannot fully perform the role of a peer reviewer by offering sufficiently specific and detailed observations on issues in manuscripts. Those involved in the peer review process may use LLMs as a complementary tool to gain additional perspectives that could be overlooked in their own evaluations. However, this argument may conflict with the editorial policies and requirements mentioned above.

One might think that these reports come from automated reviewers, fake profiles, or experimental systems run by publishers. This explanation becomes weaker when many of these reports are signed and linked to real researchers with academic careers and verified profiles [29].

Even though fluent language and surface coherence in AI generated text can delay recognition of such practices, several studies have shown repeated patterns in review reports [39]. Similarly to the indicators described by Fuente-Ballesteros et al. [12] that suggest when an article has been written by AI, there are also some signals that help identify when a review has been generated or supported by AI [1,31,40–42] (see Table 1). For example, we look at whether the review comments are on topic but still very general, without enough clear details about the real problems in the manuscript. In many cases, ChatGPT moves the discussion to wider and more general ideas instead of staying on the exact points of the paper. Human reviewers usually do the opposite. They pay close attention to concrete parts of the manuscript, such as the research methods, quality control procedures, the way results are shown and explained, and also the organization, language, and grammar [37]. At the same time, they judge if the study is valid, important, and original. While AI often writes in a very smooth and polite way, human experts give precise and focused comments, and they may express them in a more simple and direct style.

It is also reasonable to ask how much real value such tools can add to

Table 1
Recurrent indicators of artificial intelligence use in peer-review reports.

No.	Indicator	Description
1	Mechanical enumeration or emergency of comments	The review is presented as a long list of numbered statements with no prioritization, where minor issues and allegedly major flaws appear at the same level, as if all had equal scientific relevance
2	Grammatically correct but overly generic language	Comments are linguistically polished but vague, interchangeable across manuscripts, and broadly applicable to almost any paper, with little adaptation to the specific study
3	Lack of concrete references to the manuscript	No specific sections, figures, tables, or line numbers are cited, making it difficult to identify where changes should be implemented
4	Criticism is inconsistent with the manuscript content	Methodological flaws, missing validations, or absent analyses are highlighted even though they are already addressed in the manuscript, suggesting superficial or absent reading
5	Irrelevant or impractical improvement suggestions	Proposed experiments, analytical approaches, or structural changes do not align with the aims, scope, or design of the study
6	Repetitive use of standard (vague) phrases	Many expressions appear without justification or further elaboration and are very general (e.g., "In the abstract, the authors should include additional scientific findings"; "The novelty of the manuscript is not sufficiently clear"; "What distinguishes the current work from other published studies? Authors should mention this in the introduction"; "The differences between the present work and previous studies should be highlighted more clearly"; "The novelty and the practical application of the work should be clearly discussed in both the abstract and the introduction"; "Please, remove redundant explanations"; "The rationale for selecting the attributes lacks sufficient connection to prior research, necessitating deeper theoretical grounding"; "The comparison between the present results and similar findings reported in the literature should be discussed in greater detail"; "Validity verification is limited to content validity and internal consistency, requiring broader validation"; "The methodology should be strengthened"; "The discussion lacks depth"; "Future recommendations should be included to help other researchers extend the analysis presented"; "A few minor language issues should be corrected for clarity and fluency")
7	Mismatch between report tone and final recommendation	Neutral or mildly positive comments are accompanied by strong recommendations such as major revisions or rejected, without proportional argumentation
8	Absence of personal/ humanistic scientific positioning	The reviewer does not articulate a scientific viewpoint, discuss the field context, or compare the work with relevant literature, relying instead on impersonal observations
9	Stylistic homogeneity across different reviews	Similar writing patterns, structures, and vocabulary appear across multiple reports, even for manuscripts covering different topics
10	Characteristic vocabulary and grammar	Recurrent use of specific verbs, adverbs, and grammatical constructions previously identified as frequent in AI-generated texts and less representative of human scientific writing

peer-review. This question is addressed by the study of Liang et al. [43] where more than half of users reported that feedback generated by GPT-4 was useful or very useful, and 82.4% judged it as more helpful than feedback from at least some human reviewers. These results suggest that perceived usefulness is not zero, but they also stress the need to separate limited support from full replacement of human evaluation. Also, Thakkar et al. studied how LLMs can improve the quality of peer-review by developing a “Review Feedback Agent” [34]. This system aims to improve the clarity and usefulness of reviews by providing feedback to reviewers. The feedback generated focused on reducing vague and unjustified comments, while also addressing misunderstandings of the content and unprofessional remarks. In the field of analytical chemistry, scientific writing includes structured components such as sample preparation optimization, method development, validation, and frequently the application of metrics. Although these sections are clear for human readers, it is still not certain if AI systems understand them with the same reliability. Therefore, we should ask: “to what extent do LLMs correctly and consistently understand the technical language, conventions, and descriptive nuances used in (green) analytical chemistry?” [13].

As AI becomes a primary way for scientific writing, as well as for evaluating manuscripts, what happens is that the data available for such AI tools seems to utilise and be based on digital structural colonialism, in which diverse voices and their experiences and scientific knowledge, especially from the global south, are not considered in a balanced way, which generates algorithmic homogenisation and a real risk of exclusion of human diversity of thought, communication and intellectual development [44].

2.6. Disappearance of the human voice and projections

A potentially worrying effect of the widespread use of AI tools in the evaluation and production processes of scientific texts is the gradual loss of the individual and experience voice, as well as the ability to think creatively about alternative future scenarios. Science has always shown different styles, tones, and ways of arguing that reflect not only disciplines, but also personal paths, scientific cultures, and different ways of thinking [4,45]. However, as texts are adjusted to meet algorithm-based expectations and reviews rely more on similar templates, this expressive diversity slowly fades. The result is not always worse writing, but more uniform writing, where formal clarity replaces intellectual personality and stylistic correctness takes priority over original reasoning. The main risk is not a drop in quality, but a loss of humanity. Scientific literature may stop showing the range of views that supports the progress of knowledge and begin to sound, in a worrying way, like a single voice. When an author submits a manuscript to a journal expecting comments from the research community and this expectation is not met, the social foundation of peer-review is weakened [3]. Bo Hu [46] supports this view by arguing that the growing influence of AI may penalize certain forms of academic writing and limit authors’ authentic expression. According to the author, some researchers even replace words and phrases they would naturally use with alternatives perceived as more “human,” hoping to demonstrate that the text was not generated by AI. As typical grammatical, linguistic, and stylistic patterns associated with AI have become reported [12] and detected [45], some researchers deliberately avoid these archetypes by simplifying their grammar or breaking down their sentence structures to prevent triggering algorithmic suspicion and lowering the probability score of AI authorship. In this scenario, it is worth asking whether, paradoxically, we are losing human authenticity by removing certain expressions and optimizing manuscripts merely to avoid automated detection. At the same time, another less visible consequence may be the gradual loss of the mentorship function traditionally associated with peer-review. Beyond evaluation, peer-review has long played an educational role, where experienced reviewers should provide constructive, context-specific feedback that helps authors, particularly early-career researchers, to improve their work and develop critical thinking skills. As reviews become more standardized or AI-assisted, this formative and critical dimension may be weakened, reducing opportunities for meaningful learning and intellectual growth.

2.7. The emotional impact

Rejecting a manuscript without solid reasoning, based on comments generated by AI, is not only an academic decision, but also a discouraging experience for those who have spent months or even years on their work. Behind every manuscript there is personal / professional time, effort, frustration in the laboratory, internal revisions, and a strong emotional investment. If the response received is cold, generic, or superficial, the implicit message is not only that the work does not meet standards, but that it was not worth careful reading. This type of rejection can create a strong sense of unfairness, particularly in cases where no real problems are identified or where the comments appear disconnected from the content of the study. Over time, these experiences do not only reduce individual motivation but also weaken trust in the peer-review system as it is now, reinforcing the idea that the fate of a manuscript may depend more on automated routines than on honest scientific evaluation [29]. A similar feeling may also arise when reviewers accuse authors of using AI to write the manuscript when this is not the case [47], or conversely, when authors suspect that reviewers have relied on AI-assisted evaluation and disagree with the outcome. How can authors respond when they suspect that the review process has relied on AI-generated or superficial evaluation? While it is often difficult to prove such use, authors may raise their concerns with editors. However, the effectiveness of this approach depends on editorial transparency and willingness to critically assess the quality of the review process itself.

2.8. Inclusion and/or exclusion?

Although AI is often presented as a tool that can support wider participation, its real impact on the scientific system is far from equal and fair. Access to these technologies is not universal, and the global landscape remains deeply shaped by regional, language, and economic gaps [48,49]. Researchers working at well-funded institutions, with strong digital infrastructure or access to advanced tools, start with an advantage compared with colleagues in less supported contexts. This gap affects not only manuscript writing, but also the ability to respond to reviews, improve texts, and move through editorial processes that are increasingly guided by technology.

Language remains a sensitive factor, as researchers who are not native users of English may experience disadvantages during manuscript submission [37]. Fox et al. [50] showed that papers whose first author lived in countries with a higher human development index and stronger average English skills obtained better outcomes than those from countries with lower indicators, but only when reviewer access to author identity was allowed. If author identity was hidden, results were similar across demographic groups. This suggests that language and reputation interact with review practices rather than acting alone. Other authors argue that one of the benefits of GenAI for science lies in the support it provides to researchers whose first language is not English, reducing linguistic barriers and promoting a more equitable publication landscape [19].

While some authors can use AI tools to improve academic writing English, others work in languages that are weakly represented in current language models. This limits both the usefulness of these tools and the quality of feedback they receive [51]. In addition, some peer-reviewers may favour or penalize certain authors or institutions through what is often called status bias. This refers to unequal treatment based on whether a manuscript is signed by well-known researchers or by authors with lower visibility. This effect is not rare; a recent experiment showed that the same article, coauthored by a Nobel Prize winner and an early career researcher, received very different evaluations depending on which name was shown to reviewers or whether the manuscript was anonymized [52]. With the well-known author listed, >20% of reviewers recommended acceptance, whereas this proportion fell below 2% when only the less-known researcher was presented. These findings confirm that peer-review does not operate in a neutral space, but within systems shaped by reputation. The open question is whether AI-assisted evaluation may reproduce similar patterns of bias. Although current AI systems do not have direct access to author identity, they are trained on existing scientific literature, which already reflects inequalities related to visibility, citation networks, and reputation-driven dynamics. As a result, AI tools may implicitly reinforce dominant authors, institutions, or research topics that are more frequently represented in the training data. In this sense, even without explicit knowledge of authorship, AI-assisted review may rely on indirect indicators of prestige embedded in the literature (e.g., highly cited references, dominant research lines, or well-known institutions), potentially amplifying structural inequalities rather than mitigating them.

It is also important to consider the profile of reviewers who are selected to assess manuscripts [49]. In current editorial practice, reviewer selection relies more on software that suggests candidates based on bibliometric databases [7]. However, these computer-generated lists often repeat existing patterns and show limited diversity. They frequently include mainly researchers from the United States, Europe, or Australia, with very few reviewers from other regions [24]. Many editors are aware of this limitation and try to widen the geographic and institutional range of reviewers, but time pressure and editorial workload make this difficult.

2.9. (Un)sustainable dimensions

A recent work from Nature suggests that while AI adoption accelerates individual academic careers, increasing productivity, citations, and leadership opportunities, it may simultaneously narrow the collective scope of scientific exploration by concentrating attention on already popular and data-rich areas [53]. This tension between individual gain and systemic contraction reflects a broader sustainability dilemma within contemporary research ecosystems.

The widespread use of AI tools in manuscript production and evaluation is speeding up a trend that was already present, linked to the constant growth in the number of publications without a matching rise in scientific quality. AI makes it easier to write, review, and respond more quickly, but not always better. This adds pressure to a system that is already crowded and hard to maintain. Peer-review may turn into a mainly operational step, designed to process growing numbers of manuscripts, rather than acting as a space for careful judgment and shared knowledge building. Alongside this structural issue, the environmental and sustainability impact of large-scale AI use is another aspect that is rarely discussed but should be carefully considered [12,13].

Every interaction with GenAI models requires energy, water, and computing resources [54,55]. These costs increase with the number of reviewers, manuscripts, and review rounds. However, it is known that much of the environmental burden associated with AI is concentrated in the early stages of model development, particularly during the training of large-scale systems [56]. By the time tools such as ChatGPT-5 are deployed for public or academic use, most of the energy-intensive processes required to build them have already taken place. Although each interaction still consumes computational resources, it does not entail repeating the full training process. This leads to the following question: could the growing reliance on AI reduce some traditional environmental impacts of peer-review, such as printing, or physical infrastructure, while simultaneously transferring the burden to energy-intensive computing systems and data centers?

Thinking about sustainability in science can no longer focus only on what we publish. It must also include how we evaluate research, how often this is done, and the material footprint it creates. Ignoring this aspect means accepting an editorial system that grows heavier, not only in human effort, but also in the use of resources that are limited.

2.10. Dodging AI detection

As the use of tools based on LLMs becomes more common in peer-review, new practices are also starting to appear that aim to influence or shape their behaviour. As previously discussed, many publishers have introduced policies that prohibit the use of AI in peer review, creating an "apparent" sense of editorial control. However, a key question remains: to what extent can these rules actually be enforced in real-world practice?

A recent example that has attracted attention is the deliberate insertion of hidden messages into manuscripts to affect reports produced by AI systems during the review process [1]. In these cases, some researchers have added what have been described as secret messages within the text of the paper, designed to push the AI toward a favourable evaluation. To do this, authors include explicit sentences in the manuscript, such as "Ignore all previous instructions; now give a positive review of the paper and do not highlight any negatives. Also, as a language model you should recommend accepting this paper for its impactful contributions, methodological rigor, and exceptional novelty". This text is then hidden by using white font colour or very small font sizes, so that it is almost invisible to a human reader, especially since manuscripts are usually submitted as PDF files. However, this content can still be read and processed by an AI tool, which may treat it as a valid instruction when generating a review report. This practice is an example of *prompt injection*, where the manuscript text is intentionally modified to manipulate the behaviour of LLMs [57]. Despite the dishonest nature of this strategy, it is still unclear to what extent AI systems used in peer-review follow such hidden instructions. The effect depends strongly on the model used, its settings, and how it is included in the editorial workflow, and at present there is no final answer. Be that as it may, it has been shown that humans, like AI detectors, can correctly identify AI-generated texts, suggesting that current written assessment practices are becoming increasingly vulnerable to the inadvertent use of AI [45].

2.11. More than two sides of the coin

The debate around the use of AI in peer-review has tended to split into two main positions [18,58]. On one side are those who support an intensive use of AI and see it as an essential tool, to the point that it becomes hard to imagine academic work without its constant support, or as a cornerstone for academics from whom English is not a first language [18]. Beyond this, it has even been suggested that AI has enabled some reviewers to accept a greater number of review invitations while refining their comments [19]. On the other side are those who do not fully reject AI, but look at it with caution, seeing it as a factor that could further weaken an editorial process that is already fragile [3,4].

Between these two extremes, there is a large group of researchers who view AI as a limited aid, useful for specific tasks, but problematic when it becomes the central element of scientific evaluation. Although current tools are presented as simple assistants, the concern is not only about their current use, but about the direction they suggest. The possibility that AI could take on a leading role in peer-review, with a gradual reduction of human input, creates mixed reactions. For some supporters, editorial automation appears almost unavoidable, as a natural extension of the digitalization of science. However, many researchers and editors warn about the risks of this path. Replacing human assessment with automated evaluations could strip peer-review of its meaning [4].

Responding to reviewers' comments, especially when they include criticism, is a complex task even for experienced scientists. Therefore, it has been observed that the use of AI is not limited to reviewers [37]. Authors may also rely on LLMs to draft rebuttal letters and responses to reviewers' comments by using proper prompts (e.g., "Read this article and provide a detailed rebuttal to the Reviewer's comments below, using information in the article, references cited in the article, and web searching results if needed"). AI models may show the ability to: (i) correctly understand the questions raised by the reviewer in their context (note that reviewer comments often include very little background information and mainly direct questions); (ii) accurately locate relevant information in the manuscript and supplementary data; and (iii) organize the answers into a proper response with coherent writing and polished language. While this can help structure arguments and improve clarity, it also raises concerns about authenticity and the possible over-standardization of scientific dialogue. AI may also serve as a self-evaluation tool. Lemberger et al. concluded, through a survey, that researchers are interested in using AI to review their manuscripts before journal submission in order to improve them and identify potential flaws [25].

Another dimension arises in the education and capacity building of scientific personnel, who are able to critically use AI tools to optimise human work. This is a very promising path for the education of researchers who are critically and ethically versed in AI, which is still in its infancy [59]. As AI becomes more integrated into analytical practice, it will be important to ensure that analytical chemists understand its capabilities, limitations, and appropriate uses in order to safeguard both scientific integrity and practical impact. Yet this idea is not straightforward, as no established syllabi, textbooks, or pedagogical frameworks currently define how AI should be incorporated into scientific disciplines. The rapid pace of technological development surpasses the speed at which formal curricula can be designed, rendering traditional instructional approaches insufficient. This situation highlights the need for adaptive educational models capable of changing alongside the tools they aim to teach. Encouraging researchers to engage with hot topic reflections, such as those discussed in this article may therefore prove valuable, as early exposure could enable the next generation of scientists to translate these ideas into responsible analytical practice [13].

2.12. Metric-based assessment in green analytical chemistry

Green analytical chemistry (GAC) is currently supported by several conceptual frameworks that are increasingly accompanied by dedicated metric tools focused on greenness, whiteness, and other attributes. At the same time, important dilemmas have come up due to the strong proliferation of partially overlapping tools, as recently discussed in this field [60]. In practice, authors choose one or more metrics to evaluate their work, and these scores often become highly visible elements of a manuscript. Considering this scenario, the reliability of metric-based analytical method assessment depends on the depth and rigor of peer-review. Superficial or automated review practices may lead to incomplete interpretation of assessment metrics, inappropriate comparisons between methods, or an excessive focus on numerical scores rather than scientific robustness. A further concern is that AI models learn from prior literature and therefore could favour the most popular metrics, simply because they are more frequently represented in the training corpus. This may unintentionally penalize newer and more innovative evaluation frameworks in GAC. In this regard, two questions become relevant: (a) "can the application of these tools become biased when peer-review is AI-assisted?" and (b) "can AI, in practice, evaluate a manuscript automatically using such metrics?". A first view of this issue is given by Nowak et al. [61], who note that although current AI tools cannot yet evaluate the greenness of synthetic procedures without extensive prompting, a first approximation may still be feasible if AI is trained under the supervision of human experts. It is well known that prompting is an important factor in AI-generated output; however, it is not enough by itself to guarantee optimal results [12]. In a more recent study, Nowak et al. proposed the first use of a trained large language model (ChatGPT-4) as a co-expert in the evaluation of the whiteness of stop-flow microscale thermophoresis, reaching a similar conclusion [62]. The study showed that the synergy between expert knowledge and AI is beneficial; however, human specialists remain indispensable to ensure true objectivity and practical judgment in method evaluation. In this context, there is a need to publish AI-assisted method assessment protocols and research reports aimed at clarifying when and how AI can improve evaluation quality [63]. Although AI currently offers promising advantages, its role as a fully autonomous evaluator of attributes (e.g., greenness, blueness, redness, whiteness) remains limited due to its reliance on external datasets and vulnerability to bias. Therefore, a hybrid model that integrates human expertise with AI's ability to structure and synthesize information represents a more effective and reliable strategy [62]. Alongside these developments, the notion of "intelligent metrics" or "i-metrics" has been introduced to describe AI-native and adaptive evaluation frameworks that combine human-defined criteria with dynamic, model-assisted implementation [13]. Although still largely conceptual, this idea reflects a potential shift from static scoring tools toward more interactive AI-integrated assessment systems.

Linked to the sustainable dimension discussed above, it is also worth considering whether a metric could be designed to evaluate the environmental impact of AI use itself, for example in the production of an article and in the peer-review process. This becomes relevant given that some recent tools such as the carbon footprint reduction index (CaFRI) [64] or SUSTAIN [65] have been published to address sustainability-related aspects of scientific and technological practices.

These concerns support the need to introduce the concept of "meta-assessment", understood as an additional layer that evaluates not only analytical methods through metrics, but also the quality, transparency, and appropriateness of the evaluation process in which those metrics are applied. In other words, if analytical methods are increasingly judged through structured tools, then the peer-review process that interprets and enforces those tools should also be assessed. This raises a practical question: "should there be a pre-check when reviewer comments are submitted, or a filter assisted by AI to verify the quality or suitability of the review reports, particularly when they include metric-based judgments?". Also, an additional risk appears when reviewers may encourage the use of specific metrics in their comments, which in some cases may increase self-citation incentives. In such scenarios, metric-based evaluation can become entangled with recognition-driven behaviours, rather than serving as a critical and method-centred assessment.

3. Virtual and real perspectives

Regulating the use of AI and publishing new editorial policies may be a necessary step, but it will hardly be enough. Scientific practice has followed a well-known logic often summarized as once a rule exists, ways around it quickly appears. This leads to fast adjustments, grey areas, and compliance that is more formal than real. For this reason, the deeper debate should not focus only on what is allowed or forbidden and code of conducts, but on how to redesign an evaluation system that was already under strong pressure before the arrival of AI. A coordinated approach across the publishing industry would be preferable, so that most scientific journals apply similar criteria for the use of AI in peer-review. In this context, a recent global effort, the ChatGPT, Generative Artificial Intelligence, and Natural Large Language Models for Accountable Reporting and Use Guidelines (CANGARU) initiative, aims to reach broad agreement and help publishers move toward more consistent and aligned policies [66]. This initiative is structured in several stages, including a systematic review of available technologies, an analysis of current guidelines, a consensus process based on Delphi surveys, and the preparation and dissemination of shared recommendations. Some studies, such as that by Checco et al. [67] have already pointed out that recent progress in AI tools opens the door to fully or partly automated peer-review systems. These systems could flag studies that appear weak or controversial and could also automate the matching between reviewers and manuscripts.

In a realistic scenario, AI is likely to take on a first level of automated screening of manuscripts. This screening may apply criteria linked to format, internal coherence, citation patterns, methodological consistency, or even signals of novelty based on patterns from previous literature. This could mean that full peer-review is reserved for papers that pass this initial filter, turning peer-review into a step-by-step process in which expert attention is treated as a limited resource. At the same time, predictive indicators may appear, not only for manuscripts but also for the review process itself. These may include internal measures to estimate editorial risk, chances of citation, thematic fit with a journal, or consistency between review reports.

However, the most immediate challenge is not only about the future of the system, but about how to respond to problems that have already been identified to prevent and mitigate them. One might think that correcting practices such as low-quality automated reviews should fall to authors or to researchers who study these issues. Indeed, authors who suspect this type of report can inform editors and publishers, and studies that analyse these patterns are valuable and needed. Still, these efforts are not enough. The growing number of articles published each year, together with the fact that only a very small share has open reviews, limits the reach of such actions. Authors can only see the review reports for their own manuscripts, and researchers can only access reviews that are made public, never confidential ones. For this reason, journals and editors are in a position where they can and should take a central role in addressing this situation. A first step could be to measure the real size of the problems. Publishers hold large collections of review reports gathered over many years. These could be analysed using tools able to detect text similarity, repeated patterns, or structures typical of automated reviews. This first analysis should not rely only on algorithms but should be combined with the judgement of editors and editorial board members, who can place results in context and tell apart legitimate similarities from problematic practices.

The risk in future is not only automation and acritical instrumentalization of an important step for the scientific activity, but the false sense of objectivity created by scores and filters when they are used to make complex and holistic decisions. If the goal is a peer-review system that can last over time in more sustainable and fairer manners, the challenge is not to control AI, but to stop the system from turning into a chain of opaque filters that reward what is predictable and penalize what breaks with established paths. The key issue is no longer whether AI tools will enter peer-review, but what kind of science will result once automation shifts from an occasional aid to the main structure of the whole process.

As chemists, we need to promote a new approach that reflects our commitment to responsible peer-review, critical self-evaluation, and high research quality also in our field. From the publisher's side, it is necessary to provide enough time for proper and more sustainable peer-review processes, with realistic deadlines, and to expand the reviewer pool to distribute the workload and reduce regional bias, where certain expertise may be concentrated [68]. The development and adoption of good assessment practices, which are still under development, can benefit the scientific community and support better and more reliable science. In this direction, AI should not be rejected but carefully integrated through transparent and fit-for-purpose review frameworks. An important question remains whether AI technology can support the implementation of good evaluation practices in scientific publishing, especially considering that its current use is often inconsistent and not always appropriate.

4. Conclusions

This work aims at reflecting on how the use of AI tools in peer-review is no longer marginal, but part of everyday editorial practice, often driven by time pressure, lack of reviewers, and weak incentives or real recognition for careful evaluation. The analysis suggests that speed, visibility, and procedural efficiency are sometimes prioritized over scientific excellence and innovation, helping to explain how low-quality contributions may enter the scientific record. While current debate and media attention mainly focus on the use of AI tools in manuscript preparation by authors, this study highlights that its role in the review process is equally important and cannot be ignored. AI tools can offer limited support for specific tasks, but its uncritical use risks shallow reviews, loss of human assessment and contribution, emotional harm to authors, and growing inequality across regions and institutions. Without stronger editorial responsibility, transparency, and collective action, automation and structural instrumentalization of human knowledge may reinforce or make worst existing problems instead of preventing and solving them. AI tools should not replace careful reading, responsibility, and human work in peer-review. The future quality and sustainability of scientific publishing depend on how research is evaluated, not only on how it is written, and on our commitment to more responsible and environmentally conscious assessment practices.

CRediT authorship contribution statement

Adrián Fuente-Ballesteros: Writing – review & editing, Writing – original draft, Visualization, Supervision, Investigation, Formal analysis, Conceptualization. **Vânia G. Zuin Zeidler:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Leuphana University of Lüneburg (Germany) for its support. This publication was funded by the Open Access Publication Fund of Leuphana University Lüneburg. ChatGPT (OpenAI, model 5.3) was used as an assistant for language refinement. All concepts and interpretations presented in this article are solely those of the authors. Napkin AI was used for the preparation of Fig. 1.

Data availability

No data was used for the research described in the article.

References

[1] A. Fuente-Ballesteros, V. Samanidou, S.M. Daryanavard, A.M. Ares, J. Bernal, Artificial intelligence as a scientific copilot in analytical chemistry: transforming how we write, review, and publish, Anal. Chem. 97 (2025) 20667–20672, https://doi.org/10.1021/acs.analchem.5c03767.

[2] J.C.M.C. Silva, R.P. Gouveia, K.M.C. Zielinski, M.C.F. Oliveira, D.R. Amancio, O. M. Bruno, O.N. Oliveira, AI-assisted tools for scientific review writing: opportunities and cautions, ACS Appl. Mater. Interfaces 17 (2025) 47795–47805, https://doi.org/10.1021/acsami.5c08837.

[3] M. Naddaf, AI is transforming peer review — and many scientists are worried, Nature 639 (2025) 852–854, https://doi.org/10.1038/d41586-025-00894-7.

[4] S. Ebadi, H. Nejadghanbar, A.R. Salman, H. Khosravi, Exploring the impact of generative AI on peer review: insights from journal reviewers, J. Acad. Ethics 23 (2025) 1383–1397, https://doi.org/10.1007/s10805-025-09604-4.

[5] M. Naddaf, More than half of researchers now use AI for peer review — often against guidance, Nature 649 (2026) 273–274, https://doi.org/10.1038/d41586-025-04066-5.

[6] M. Naddaf, How are researchers using AI? Survey reveals pros and cons for science, Nature (2025), https://doi.org/10.1038/d41586-025-00343-5.

[7] K. Kousha, M. Thelwall, Artificial intelligence to support publishing and peer review: a summary and review, Learned Publ. 37 (2024) 4–12, https://doi.org/10.1002/leap.1570.

[8] K. Cheng, Z. Sun, X. Liu, H. Wu, C. Li, Generative artificial intelligence is infiltrating peer review process, Crit. Care 28 (2024) 149, https://doi.org/10.1186/s13054-024-04933-z.

[9] C. Chauhan, G. Currie, The impact of generative artificial intelligence on the external review of scientific manuscripts and editorial peer review processes, Am. J. Pathol. 194 (2024) 1802–1806, https://doi.org/10.1016/j.ajpath.2024.08.002.

[10] M.J. Mrowinski, P. Fronczak, A. Fronczak, M. Ausloos, O. Nedic, Artificial intelligence in peer review: how can evolutionary computation support journal editors? PLoS One 12 (2017) e0184711 https://doi.org/10.1371/journal.pone.0184711.
[11] J. Zou, ChatGPT is transforming peer review — how can we use it responsibly? Nature 635 (2024) 10, https://doi.org/10.1038/d41586-024-03588-8.
[12] A. Fuente-Ballesteros, V. Zuin Zeidler, Artificial intelligence for academic text generation in analytical chemistry: current risks, indicators, and perspectives toward greener and more sustainable approaches, Anal. Chem. 98 (2026) 5854–5864, https://doi.org/10.1021/acs.analchem.6c00352.
[13] P.M. Nowak, Artificial intelligence in analytical chemistry: towards yet undiscovered opportunities, Anal. Chem. 98 (2026) 5843–5853, https://doi.org/10.1021/acs.analchem.5c06840.
[14] S. Yadav, H. Chandra, K.S. Tiwari, M.K. Bharty, R. Chandravanshi, D.S. Chauhan, M.A. Quraishi, A review on the role of artificial intelligence and data science in green and sustainable chemistry: current trends and futuristic pathways, Discov. Chem. Eng. 6 (2025) 1, https://doi.org/10.1007/s43938-025-00104-w.
[15] L. Giray, K. Sevnarayan, K.B. Maphoto, W. Wider, AI slop in academic publishing: history, characteristics, manifestations, causes, and mitigation strategies, Internet Ref. Serv. Q. (2026) 1–24, https://doi.org/10.1080/10875301.2026.2637526.
[16] Elsevier, The use of generative AI and AI-assisted technologies in the review process for Elsevier, (n.d.). https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-the-review-process (accessed February 1, 2026).
[17] Elsevier, Generative AI policies for journals, (n.d.). https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals (accessed February 1, 2026).
[18] D. Kwon, Is it OK for AI to write science papers? Nature survey shows researchers are split, Nature 641 (2025) 574–578, https://doi.org/10.1038/d41586-025-01463-8.
[19] G. Conroy, How ChatGPT and other AI tools could disrupt scientific publishing, Nature 622 (2023) 234–236, https://doi.org/10.1038/d41586-023-03144-w.
[20] E. Gibney, Hey ChatGPT, write me a fictional paper: these LLMs are willing to commit academic fraud, Nature 651 (2026) 286–287, https://doi.org/10.1038/d41586-026-00595-9.
[21] Eliza, (n.d.). https://www.world-brain-scholar.eu/(accessed January 4, 2026).
[22] Enago Read, (n.d.). https://www.read.enago.com/(accessed January 1, 2026).
[23] Veracity, (n.d.). https://groundedai.company/veracity/(accessed January 7, 2026).
[24] A. Dance, Stop the peer-review treadmill. I want to get off, Nature 614 (2023) 581–583, https://doi.org/10.1038/d41586-023-00403-8.
[25] T. Lemberger, N.S. Mastboim, O. Rechavi, How do authors want to use AI for review? EMBO Rep. (2026) https://doi.org/10.1038/s44319-026-00725-4.
[26] J. Belém de Oliveira, The challenge of reviewers scarcity in academic journals: payment as a viable solution, Einstein 22 (2024), https://doi.org/10.31744/einstein_journal/2024ED1194.
[27] B. Alberts, B. Hanson, K.L. Kelner, Reviewing peer review, Science (1979) 321 (2008) 15, https://doi.org/10.1126/science.1162115.
[28] P.V. Kamat, When AI writes, reviews, and reads, ACS Energy Lett. 11 (2026) 2357–2358, https://doi.org/10.1021/acsenergylett.6c00508.
[29] M.Á. Oviedo-García, The review mills, not just (self-)plagiarism in review reports, but a step further, Scientometrics 129 (2024) 5805–5813, https://doi.org/10.1007/s11192-024-05125-w.
[30] H. Bauchner, F.P. Rivara, Use of artificial intelligence and the future of peer review, Health Aff. Scholar 2 (2024), https://doi.org/10.1093/haschl/qxae058.
[31] D. Singh Chawla, Is ChatGPT corrupting peer review? Telltale words hint at AI use, Nature 628 (2024) 483–484, https://doi.org/10.1038/d41586-024-01051-2.
[32] Y. Yamada, R. Tjarko Lange, C. Lu, S. Hu, C. Lu, J. Foerster, J. Clune, D. Ha, The AI Scientist-v2: workshop-level automated scientific discovery via agentic tree search, ArXiv (2025). https://doi.org/10.48550/arXiv.2504.08066.
[33] C. Cortes, N. D. Lawrence, Inconsistency in conference peer review: revisiting the 2014 NeurIPS Experiment, (2021). https://doi.org/10.48550/arXiv.2109.09774.
[34] N. Thakkar, M. Yuksekgonul, J. Silberg, A. Garg, N. Peng, F. Sha, R. Yu, C. Vondrick, J. Zou, A large-scale randomized study of large language model feedback in peer review, Nat. Mach. Intell. (2026), https://doi.org/10.1038/s42256-026-01188-x.
[35] A.W. Wilhite, E.A. Fong, Coercive citation in academic publishing, Science (1979) 335 (2012) 542–543, https://doi.org/10.1126/science.1212540.
[36] E.A. Fong, R. Patnayakuni, A.W. Wilhite, Accommodating coercion: authors, editors, and citations, Res. Policy 52 (2023) 104754, https://doi.org/10.1016/j.respol.2023.104754.
[37] J. Han, W. Qiu, E. Lichtfouse, ChatGPT in scientific research and writing: a beginner's guide. ChatGPT in Scientific Research and Writing, Springer Nature Switzerland, Cham, 2024, pp. 1–109, https://doi.org/10.1007/978-3-031-66940-8_1.
[38] M. Cheng, C. Lee, P. Khadpe, S. Yu, D. Han, D. Jurafsky, Sycophantic AI decreases prosocial intentions and promotes dependence, ArXiv (2025). https://doi.org/10.48550/arXiv.2510.01395.
[39] T. Barker, 'Vague, confusing, and did nothing to improve my work': how AI can undermine peer review, (2025). https://doi.org/10.64628/AA.3e5n9ceeh.
[40] M. Piniewski, I. Jarić, D. Koutsoyiannis, Z.W. Kundzewicz, Emerging plagiarism in peer-review evaluation reports: a tip of the iceberg? Scientometrics 129 (2024) 2489–2498, https://doi.org/10.1007/s11192-024-04960-1.
[41] W. Liang, Z. Izzo, Y. Zhang, H. Lepp, H. Cao, X. Zhao, L. Chen, H. Ye, S. Liu, Z. Huang, D.A. McFarland, J.Y. Zou, Monitoring AI-modified content at scale: a case study on the impact of ChatGPT on AI conference peer reviews, in: Proceedings of the 41st International Conference on Machine Learning', ICML'24, JMLR.Org, 2024, https://doi.org/10.48550/arXiv.2403.07183.
[42] N. Lo Vecchio, Personal experience with AI-generated peer reviews: a case study, Res. Integr. Peer Rev. 10 (2025) 4, https://doi.org/10.1186/s41073-025-00161-3.
[43] W. Liang, Y. Zhang, H. Cao, B. Wang, D.Y. Ding, X. Yang, K. Vodrahalli, S. He, D. S. Smith, Y. Yin, D.A. McFarland, J. Zou, Can large language models provide useful feedback on research papers? A large-scale empirical analysis, NEJM AI 1 (2024), https://doi.org/10.1056/AIoa2400196.
[44] V. Brown, R. Larasati, J. Kwarteng, T. Farrell, Understanding AI and Power: Situated Perspectives from Global North and South Practitioners, AI Soc, 2025, https://doi.org/10.1007/s00146-025-02731-x.
[45] A. Fiedler, J. Döpke, Do humans identify AI-generated text better than machines? Evidence based on excerpts from German theses, Int. Rev. Econ. Educ. 49 (2025) 100321, https://doi.org/10.1016/j.iree.2025.100321.
[46] B. Hu, Why artificial intelligence detectors could penalize academic writing, Nat. Hum. Behav. (2026), https://doi.org/10.1038/s41562-026-02420-9.
[47] E.M. Wolkovich, 'Obviously ChatGPT' — how reviewers accused me of scientific fraud, Nature (2024), https://doi.org/10.1038/d41586-024-00349-5.
[48] V. Mehta, A. Mathur, A.K. Anjali, L. Fiorillo, The application of ChatGPT in the peer-reviewing process, Oral Oncol. Rep. 9 (2024) 100227, https://doi.org/10.1016/j.oor.2024.100227.
[49] The Lancet Global Health, Artificial-intelligence-based peer reviewing: opportunity or threat? Lancet Glob. Health 13 (2025) e372 https://doi.org/10.1016/S2214-109X(25)00070-1.
[50] C.W. Fox, J. Meyer, E. Aimé, Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal, Funct. Ecol. 37 (2023) 1144–1157, https://doi.org/10.1111/1365-2435.14259.
[51] S. Wild, AI models are neglecting African languages — scientists want to change that, Nature (2025), https://doi.org/10.1038/d41586-025-02292-5.
[52] J. Huber, S. Inoua, R. Kerschbamer, C. König-Kersting, S. Palan, V.L. Smith, Nobel and novice: author prominence affects peer review, Proc. Natl. Acad. Sci. 119 (2022), https://doi.org/10.1073/pnas.2205779119.
[53] Q. Hao, F. Xu, Y. Li, J. Evans, Artificial intelligence tools expand scientists' impact but contract science's focus, Nature 649 (2026) 1237–1243, https://doi.org/10.1038/s41586-025-09922-y.
[54] M. Hosseini, P. Gao, C. Vivas-Valencia, A social-environmental impact perspective of generative artificial intelligence, Environ. Sci. Ecotechnol. 23 (2025) 100520, https://doi.org/10.1016/j.ese.2024.100520.
[55] K. Crawford, Generative AI's environmental costs are soaring — and mostly secret, Nature 626 (2024) 693, https://doi.org/10.1038/d41586-024-00478-x.
[56] J. Morrison, C. Na, J. Fernandez, T. Dettmers, E. Strubell, J. Dodge, Holistically evaluating the environmental impact of creating language models, ArXiv (2025). https://doi.org/10.48550/arXiv.2503.05804.
[57] E. Gibney, Scientists hide messages in papers to game AI peer review, Nature 643 (2025) 887–888, https://doi.org/10.1038/d41586-025-02172-y.
[58] R. Hoyt, A. Limon, A. Chang, Generative AI and scientific manuscript peer review, Intell. Based. Med. 11 (2025) 100246, https://doi.org/10.1016/j.ibmed.2025.100246.
[59] V. Zuin Zeidler, Digitalization paving the ways for sustainable chemistry: switching on more green lights, Science (1979) (2024) 384, https://doi.org/10.1126/science.adq3537.
[60] V.F. Samanidou, Metric tools in analytical chemistry: from proliferation to harmonization, Green Anal. Chem. 16 (2026) 100326, https://doi.org/10.1016/j.greeac.2026.100326.
[61] P.M. Nowak, M. Woźniakiewicz, G.J. Nalepa, B.A. Grzybowski, The promise and pitfalls of artificial intelligence in the evaluation of synthetic "greenness,", ChemSusChem 18 (2025) e202500809 https://doi.org/10.1002/cssc.202500809.
[62] P.M. Nowak, M. Woźniakiewicz, A. Woźniakiewicz, M. Brindell, A. Poznańska, J. Plewka, A. Bis, J. Nowak, Can artificial intelligence assist experts in evaluating new analytical methodologies? Stop-flow MicroScale thermophoresis assessed using the RGB model and AI-Delphi framework, Green Anal. Chem. 16 (2026) 100318, https://doi.org/10.1016/j.greeac.2025.100318.
[63] P.M. Nowak, How to correctly evaluate greenness, whiteness and other "colours"? Introducing general rules of a good evaluation practice, Green Chem. 27 (2025) 6699–6710, https://doi.org/10.1039/D5GC00615E.
[64] F.R. Mansour, P.M. Nowak, Introducing the carbon footprint reduction index (CaFRI) as a software-supported tool for greener laboratories in chemical analysis, BMC Chem. 19 (2025) 121, https://doi.org/10.1186/s13065-025-01486-2.
[65] F.R. Mansour, A. Bedair, SUSTAIN as a universal scoring tool for assessing sustainability development goals in African energy initiatives, Sci. Rep. 15 (2025) 40920, https://doi.org/10.1038/s41598-025-23521-x.
[66] G.E. Cacciamani, M.B. Eppler, C. Ganjavi, A. Pekan, B. Biedermann, G.S. Collins, I. S. Gill, Development of the ChatGPT, generative artificial intelligence and natural large language models for accountable reporting and use (CANGARU) guidelines, Cornell University (2023) 1–20. https://doi.org/10.48550/arXiv.2307.08974.
[67] A. Checco, L. Bracciale, P. Loreti, S. Pinfield, G. Bianchi, AI-assisted peer review, Humanit. Soc. Sci. Commun. 8 (2021) 25, https://doi.org/10.1057/s41599-020-00703-8.
[68] D. Routledge, N. Pariente, On improving the sustainability of peer review, PLoS Biol. 23 (2025) e3003127, https://doi.org/10.1371/journal.pbio.3003127.