Who should still peer-review, for whom, and with what tools?

Stefano Mammola[1,2], Lorenzo Vilizzi[3], Paride Balzani[4,5], Danish A. Ahmed[6], Lucian Pârvulescu[7,8], Joana Garrido Nogueira[9,10], Manuel Lopes-Lima[9,10], Ronaldo Sousa[11], Mattias L. Johansson[12], Alessandro Nota[13], Carlos Cano-Barbacil[14], Ali Serhan Tarkan[3,15], Eva Záhorská[16], Anna Gavioli[17], Virginie Raybaud[18], Emili García-Berthou[19], Baran Yoğurtçuoğlu[20], Jingrui Sun[21,22,23], Ross N. Cuthbert[24], Elizabeta Briski[25], Ester M. Eckert[1], Barbora Števove[16], Teun Everts[26,27], Diego Fontaneto[1] , Stelios Katsanevakis[28], Riccardo Scalera[29], Michel Bariche[30], Valentini Stamatiadou[28], Marta Rodríguez-Rey[31] & Phillip J. Haubrock[5,32]

Acknowledgements

P.B. acknowledges funding by Humboldt Fellowship from the Alexander von Humboldt Foundation. E.G.-B. was funded by MICIU/AEI/10.13039/501100011033/ and FEDER, UE through project PID2023-146173NB-C21. T.E., S.K. and V.S. were supported by the European Union's Horizon Europe HORIZON-CL6-2024-BIODIV-01 project 'GuardIAS - Guarding European Waters from IAS', under grant agreement no. 101181413. P.J.H. was supported by the Marie Skłodowska-Curie Postdoctoral Fellowship HORIZON-MSCA-2024-PF-01 (Project DIRECT; Grant No. 101203662) within the European Union's Horizon 2024 research and innovation programme. C.C.-B. benefited from a postdoctoral contract (ref. JDC2023-052358-I), funded by the Spanish Ministry of Science, Innovation, and Universities (MICIU/AEI/10.13039/501100011033) and the FSE+. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or Horizon Europe. Neither the European Union nor the granting authority can be held responsible for them.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

This work started as a collective brainstorming session (ignited by a classic 'unpaid peer review' meme) in a dedicated group chat among carbon-based organisms – no AI chatbots involved, we believe. All authors contributed with ideas, suggestions, and comments. S.M. and P.H. led the writing.

Affiliations

[1.] *Water Research Institute, Molecular Ecology Group (MEG), National Research Council (CNR-IRSA), Verbania Pallanza, Italy*

[2.] *Laboratory for Integrative Biodiversity Research (LIBRe), Finnish Museum of Natural History (LUOMUS), University of Helsinki, Helsinki, Finland*

[3.] *University of Lodz, Faculty of Biology and Environmental Protection, Department of Ecology and Vertebrate Zoology, Lodz, Poland*

[4.] *GEOMAR Helmholtz Centre for Ocean Research Kiel, Wischhofstraße 1-3, 24 148, Kiel, Germany*

[5.] *Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, University of South Bohemia in České Budějovice, Zátiší 728/II, 389 01, Vodňany, Czech Republic*

[6.] *Department of Mathematics and Natural Sciences, Gulf University for Science and Technology, Hawally, Kuwait*

[7.] *West University of Timisoara, Timisoara, Romania*

[8.] *Crayfish Research Centre, West University of Timisoara, Timisoara, Romania*

[9.] *CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão, Portugal.*

[10.] *BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, Vairão, Portugal*

[11.] *CBMA – Centre for Molecular and Environmental Biology/ARNET-Aquatic Research Network & IB-S, Institute of Science and Innovation for Bio-Sustainability, Department of Biology, University of Minho, Campus Gualtar, 4710-057 Braga, Portugal.*

[12.] *Small Island Sustainability, University of The Bahamas, Freeport, Grand Bahama, The Bahamas*

[13.] *Department of Veterinary Sciences, University of Turin, Grugliasco, Italy*

[14.] *Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, Madrid 28006, Spain*

[15.] *Department of Basic Sciences, Faculty of Fisheries, Muğla Sıtkı Koçman University, Muğla, Türkiye*

[16.] *Department of Ecology, Faculty of Natural Sciences, Comenius University, Bratislava, Slovakia*

[17.] *Department of Environmental and Prevention Sciences, University of Ferrara, Ferrara, Italy*

[18.] *Université Côte d'Azur, CNRS, ECOSEAS, France*

[19.] *GRECO, Institute of Aquatic Ecology, University of Girona, 17003 Girona, Catalonia, Spain*

[20.] *Department of Biology, Hacettepe University, Beytepe campus, 06800, Çankaya, Ankara, Türkiy*e

[21.] *Yunnan Key Laboratory of International Rivers and Transboundary Eco Security, Institute of International Rivers and Eco Security, Yunnan University, Kunming 650091, Chin*a

[22.] *State Key Laboratory for Vegetation Structure, Function and Construction (VegLab), Kunming, China*

[23.] *Ministry of Education Key Laboratory for Transboundary Eco-Security of Southwest China, Yunnan University, Kunming, 650091, China*

[24.] *Institute for Global Food Security, School of Biological Sciences, Queen's University Belfast, United Kingdom*

[25.] *GEOMAR Helmholtz Institute for Ocean Research, Wischhofstraße 1-3, 24148 Kiel, Germany*

[26.] *Genetic Diversity, Research Institute for Nature and Forest (INBO), Geraardsbergen, Belgium*

[27.] *Biology Department, KU Leuven, Heverlee, Belgium*

[28.] *Department of Marine Sciences, University of the Aegean, 81100 Mytilene, Greece*

[29.] *Department of Science, Roma Tre University, 00146 Rome, Italy*

[30.] *Department of Biology, American University of Beirut, 1107 2020, Lebanon*

[31.] *Department of Life Sciences, University of Alcalá, 28805, Alcalá de Henares, Spain*

[32.] *School of Life and Environmental Sciences, Faculty of Health, Environment and Medical Sciences, Bournemouth University, Poole, Dorset, The United Kingdom*

Emails & ORCID

Stefano Mammola, stefano.mammola@cnr.it, ORCID: 0000-0002-4471-9055

Lorenzo Vilizzi, lorenzo.vilizzi@gmail.com, ORCID: 0000-0001-8103-885X

Paride Balzani, balzani@frov.jcu.cz, ORCID: 0000-0003-1549-7139

Danish A. Ahmed, Ahmed.D@gust.edu.kw, ORCID: 0000-0002-2490-1546

Lucian Pârvulescu: lucian.parvulescu@e-uvt.ro, ORCID: 0000-0002-1528-1429

Joana Garrido Nogueira joanafgnogueira93@gmail.com ORCID: 0000-0002-5576-3625

Manuel Lopes-Lima manuelpmlopeslima@gmail.com ORCID: 0000-0002-2761-7962

Ronaldo Sousa, ronaldo@bio.uminho.pt, ORCID: 0000-0002-5961-5515

Mattias L. Johansson, mattias.johansson@ub.edu.bs, ORCID: 0000-0003-3042-750X

Alessandro Nota alessandro.nota@unito.it, ORCID: 0000-0003-3383-9809

Carlos Cano-Barbacil, carlos.cano@mncn.csic.es, ORCID: 0000-0002-6482-5103

Ali Serhan Tarkan, ali.serhan.tarkan@biol.uni.lodz.pl, ORCID: 0000-0001-8628-0514

Eva Záhorská, eva.zahorska@uniba.sk, ORCID: 0000-0002-0851-7851

Anna Gavioli, gvlnna@unife.it, ORCID: 0000-0001-8222-9287

Virginie Raybaud, virginie.raybaud@univ-cotedazur.fr, ORCID: 0000-0001-6426-3326

Emili García-Berthou, emili.garcia@udg.edu, ORCID: 0000-0001-8412-741X

Baran Yoğurtçuoğlu, yokbaran@gmail.com, ORCID: 0000-0001-8144-6191

Jingrui Sun jingrui.sun@ynu.edu.cn, ORCID: 0000-0001-9046-448X

Ross N. Cuthbert r.cuthbert@qub.ac.uk, ORCID: 0000-0003-2770-254X

Elizabeta Briski ebriski@geomar.de, ORCID: 0000-0003-1896-3860

Ester M. Eckert, estermaria.eckert@cnr.it, ORCID: 0000-0002-7314-1715

Barbora Števove, barbora.stevove@uniba.sk, ORCID: 0000-0001-7838-3396

Teun Everts, teun.everts@inbo.be, ORCID: 0000-0001-7862-4209

Diego Fontaneto, diego.fontaneto@cnr.it, ORCID: 0000-0002-5770-0353

Stelios Katsanevakis, stelios@katsanevakis.com, ORCID: 0000-0002-5137-7540

Riccardo Scalera, scalera.riccardo@gmail.com, ORCID: 0000-0002-4100-3959

Michel Bariche, michel.bariche@aub.edu.lb, ORCID: 0000-0001-6831-4311

Valentini Stamatiadou, valentini.stamatiadou@gmail.com, ORCID: 0000-0001-8487-8168

Phillip J. Haubrock, Phaubrock@bournemouth.ac.uk, ORCID: 0000-0003-2154-4341

Abstract

In drafting *On the Origin of Species*, Charles Darwin developed his thoughts through extensive correspondence with trusted colleagues, seeking feedback on ideas, manuscripts, and proofs. This informal peer-review process, like many others of its time, was based on personal trust, academic respect, and reputation rather than guided by any institution, and it helped pave the way for modern peer review. Today, peer review has been largely corrupted by corporate greed: a handful of dominant mega-publishers exploit reviewers (and to an extent also editors) as an unpaid, external workforce expected to supply scientific judgement, quality control, and career signalling that journals package and sell but do not themselves produce. Amid an ever-increasing volume of review requests reaching researchers' inboxes daily, many respond by declining invitations outright, if they respond at all. Others accept but rely heavily on AI, since this saves time and a few more reviews will look good on their Web of Science profile. Thorough intellectual assessments are becoming increasingly rare. The current system is not sustainable, and change is urgently needed before expert judgement is fully displaced by speed, prestige signalling, and machine-written reports. Solutions include distributing unpaid peer-review labour more wisely, prioritizing non-commercial publishers and academia-friendly journals, exploring alternative business models, and redesigning peer review so that human expertise is concentrated on tasks requiring scientific judgement while low-order tasks are supported by AI. Above all, researchers, editors, funders, publishers, and societies should stop treating review as free certification labour and instead recognize it as a scarce scholarly resource that must be used where it contributes most to scientific quality.

Keywords: *academic labour; large language models*; *publish or perish*; *research assessment*; *scholarly publishing*

5

1. Introduction

Every academic knows the moment. You casually mention "peer review" at the coffee machine and within seconds you are flooded with anecdotes: a close colleague who pasted the latest review they received into an AI-detector and was told it was "100% machine written"; another who now reviews only for society-led journals, proudly declining all other invitations; or that grumpy professor who no longer reviews because "it's a waste of time and nobody pays me." Such strong opinions are unsurprising, as peer review and scientific publishing lie at the core of academic life, shaping not only our day-to-day work but also our career prospects. And the system is increasingly distorted by the pressure to publish, by speed, and, above all, by the greed of major publishing houses. But it has not always been this way.

Until the mid-twentieth century, scientific publishing was a limited, society-led service run by academics in support of scholarship. Peer review typically meant an editor or a board, all leading experts in their respective fields, outlining research agendas and deciding whose work to publish. After World War II, the rapid expansion of research transformed the scale and organisation of scientific publishing (Box 1): the number of journals indexed in Scopus rose exponentially, and the number of articles per journal, roughly constant until 1980, expanded dramatically (Thelwall & Sud, 2022). Published papers grew by about 47% between 2016 and 2022 (Hanson et al., 2024), reaching over 3.4 million new publications a year – one every 9.2 seconds. And volume is just one side of the equation: individual papers have become more methodologically complex and densely packed with information (Munafò et al., 2017; Cordero et al., 2016), less understandable by a wider audience (Plavén-Sigray et al., 2017; Bannet & Doubleday, 2020), and less disruptive (Park et al., 2023; Kanavagh, 2025), despite an increase in novelty-related language (Mammola, 2020; Ottaviani et al., 2025). With editors struggling to keep up with this volume and being asked to assess topics often far from their own expertise, sending manuscripts to external reviewers became the norm (Burnham, 1990; Baldwin, 2018).

The 'commercial journal industry', typified by Pergamon Press, was built on this expansion: it launched narrow specialist titles and sold library subscriptions to the whole bundle instead of circulating papers as a society service (Fyfe, 2022). A handful of publishers scaled the model, bought or displaced society lists and smaller journals, and now occupy the oligopoly of scholarly communication. This was made possible by building revenue streams around the unpaid labour of writing, editorial, and peer-review work: first as buyable print and library subscriptions, then through Open Access fees (Article Processing Charges, APCs) payable by authors, institutions, and, increasingly, countries via read-and-publish agreements such as DEAL in Germany, Jisc in the UK, and Bibsam in Sweden. Indeed, by signing Copyright Transfer Agreements, researchers (and, indirectly, labs, universities, or even countries standing

6

behind them) often legally transfer the exclusive rights on their own work to journal publishers, effectively abdicating control over how their research can be used, shared, and disseminated. Costs were further reduced by moving the remaining paid production work, such as copy-editing and typesetting, to in-house offices or third-party vendors in lower-cost labour countries, notably India and the Philippines (Sallaz, 2013). The result is a multi-billion dollar industry with profit margins of ≈32–38% (van Noorden, 2013; Beigel et al., 2025), largely financed by taxpayers, with little to no money reinvested into public libraries, academic societies (some of which have agreements with publishers), or community-based journals (Receveur et al., 2024; Galtier et al., 2026).

This commercial model has stretched reviewers from occasional editorial advisors into an unpaid, external workforce expected to supply scientific judgement, quality control, and career signalling that journals package and sell but do not themselves produce (Arboledas-Lérida, 2024). In 2020 alone, researchers contributed a conservatively estimated 100 million hours to journal reviewing worldwide; the estimated value of the time contributed by US-based reviewers alone exceeded USD 1.5 billion (Aczel et al., 2021). Typically, reviewers are asked to undertake up to four tasks at once: (1) decide if the manuscript belongs in that field and journal; (2) identify inconsistencies and flaws; (3) improve the manuscript; and (4) judge its likely citability for the journal's standing. These tasks are often performed under tight deadlines, with some journals allowing two weeks or less for review, and often just a few days for an editorial decision afterward because in the end, the quicker the turnaround time of a publishing process, the more a publisher can profit and promote its own efficiency metrics. And because journals already do most of the 'signalling' for careers and citations under the disguise of anonymous peer-review, what remains for the reviewer is error-checking and improvement – tasks that are often rushed, left incomplete, or dropped altogether.

Much (digital) ink has been spilled discussing how the peer-review system is intrinsically faulty (Nature Microbiology, 2018; Amaral, 2022; Haubrock et al., 2026) and resistant to change (Ferreira et al., 2016), how increasingly difficult it is to find reviewers (Albert et al., 2016; Fox et al., 2017; Lamberts et al., 2025; Montagna, 2026), how the quality of reviews is deteriorating (Peterson, 2020, Oviedo-García, 2024), and how the emergence of Large Language Models (LLMs) and other Artificial Intelligence (AI) tools has the potential to revolutionise the entire publishing and peer review world within a few years (Saad et al., 2024; Ebadi et al., 2025; Miao et al., 2026; Naddaf & Van Noorden, 2026; Brainard, 2026). Left to individual incentives, the present arrangement is unlikely to end well. When confronting these imbalances, two extreme responses are individually rational: decline to review, or let an LLM assist or draft the report. Both follow the same cost–benefit logic: the work is unpaid (with a few exceptions) and the effort behind a report is invisible to the research community, so a model that produces usable text at almost no cost is a

7

tempting shortcut. But what is rational for individuals could be harmful to the collective. Piecemeal fixes (paying reviewers, opening reports, reviewing only for academia-friendly journals, or unifying review guidelines; reviewed in Ferreira et al., 2016) each touch only a sliver of the burden, leave the bottlenecks and incentives in place, and can undercut one another. We are not here to add another complaint to the pile, as the diagnosis has been made often enough. Rather, after briefly touching upon the flaws of the contemporary peer-review system, decisions when to accept a review request, and the do and don'ts of AI-use, we make the case for a single, coordinated position in five claims, which are discussed in detail in Section 5 and summarized in Box 2. While our focus is on peer-reviewed journal papers, several of these ideas may apply to other peer-review outputs, including books, book chapters, and grant proposals.

Box 1 | A brief history of peer review

Peer-review originated in the 1600s and was predominantly an internal process (Ferreira et al., 2016). External refereeing is much younger and more contingent than the origin myth that Henry Oldenburg invented peer review in 1665 when he launched *Philosophical Transactions of the Royal Society* suggests (Spier, 2002), which ran on editorial correspondence, not independent reports. Collective society vetting (Royal Society of Edinburgh from 1731; the Royal Society's Committee of Papers from 1752) was still an in-house filter, and independently written referee reports only became routine at the Royal Society after Whewell's 1831–32 reform.

Until the mid-twentieth century, there was more journal space than copy and scientific publishing remained a limited, society-led service run by academics in support of scholarship, so editors and assistant boards solicited papers rather than rejecting them. The burden of publishing and guarding research was primarily shouldered by scientific societies, while peer review typically meant an editor or a board, all leading experts in their respective fields, deciding whose work to publish. The scale of the publishing system was certainly constrained also by technology: manuscripts were handwritten unique objects that could not be circulated cheaply, and even after typewriters and carbon paper (1890s) produced three to five copies, Science and JAMA still did not send work to outside reviewers until after 1940, while Nature made external refereeing mandatory only in 1973.

Only the commercial Xerox photocopier (from 1959) made true externalisation affordable, just in time for postwar research volume finally outstripping pages, turning occasional editorial advice into the mass, unpaid filter known as peer review that is now treated as science's gold standard. Moreover, the same unpaid labour system was built into the U.S. National Institutes of Health after World War II and subsequently other funding agencies globally, so that public money is still allocated by a volunteer filter (though often grant reviewing is modestly paid), required to score tens of thousands of proposals a year (Mandel, 1996).

With the advent of the internet, however, the costs of producing and delivering printed issues suddenly disappeared, allowing an escalation of the system. Publishers do have real costs, including peer-review administration,

8

production, and online hosting (Van Noorden, 2013; Naim et al., 2021). Yet, much of that production is outsourced as low-cost typesetting and copy-editing, often in countries such as India and the Philippines, where highly skilled and advanced editing labour is cheaper than in other countries, and is now being automated where possible (Lundblad, 2025). Grossmann & Brembs (2021) estimated efficient publication costs at about US$200–1,000 per article (typically ≈US$565–723 for a conventional journal publishing 100–1,000 papers a year), against a commonly cited subscription-equivalent price per article of about US$4,000. Meanwhile, reviewers globally are not paid but sometimes 'rewarded' with, e.g. non-stackable ~15% reviewer APC discounts (≈€400 off a €2,000–3,000 fee) or discounts on scientific books, steering public money and future submissions straight to the same commercial houses bypassing institutions and researchers.

2. What failed and why

Many reviewers take on peer-review work out of a sense of duty to the system and in the hope of reciprocity, expecting that their own manuscript or application may meet a competent reviewer (the so-called 'golden rule of reviewing'; McPeek et al., 2009; Dobusch et al., 2026). The stakes are not trivial: the ultimate decision may affect not only a project or paper, but potentially whether someone else's lab, contract, or career continues (especially for early-career researchers; Haubrock et al., 2026). Although the possibility for reviewers to learn something new, gain inspiration for new research avenues, and keep up with the latest research is real, reviewing for journals does not always offer practical or professional advantages. Concomitantly, reviewing rarely contributes to a resume, leaving the same limited pool of reviewers to 'certify' a rising volume of papers and proposals while journals and funders capture the career credit and related prestige (Fleming and Harley, 2024). What is being lost are the essential incentives to review, often culminating in review fatigue (Breuning et al., 2015; Johnson et al., 2018), and the faith in the process as transparent and fair (Ho et al., 2013). Combined with peer-review invitations too often mismatched with reviewers' expertise (Publons, 2018; Meyerson et al., 2025), peer-review may contribute to declining peer-review quality (Else, 2024) and increasing turnaround times (Björk et al., 2013). Evidence is therefore pointing towards the review system reaching a breaking point.

These days, editors spend as much time searching for reviewers as reading manuscripts (e.g. >10 invitations may be needed to secure reviewers where two or three once sufficed; Lamberts et al., 2025) and have to rely on editorial board members to frequently step into reviewer roles, slowing editorial decisions and increasing desk rejections (Kaltenbrunner et al., 2022). Most reliable reviewers are often early-career

9

researchers known to be 'more available' and in need of recognition (Meyerson et al., 2025), resulting in an increasing burden on them (Haubrock et al., 2026). Difficulties in recruiting reviewers may also lead editors to broaden their search beyond the most directly relevant experts, potentially creating a trade-off between reviewer expertise and availability. Reviewer reports may then become overdue, often rushed and of lower quality, and in some cases partly (or entirely) AI-generated (Russo et al., 2025). Those distortions are amplified by misuse of the review itself: review reports that are used to insert the reviewer's own citations, or to kill a paper by a non-colleague or rival rather than to judge the work fairly. It is unsurprising then, that peer review, once the hallmark of scientific quality, is now dubbed "slow, inefficient, costly, haphazard, biased, subject to abuse, and inaccurate" (Aczel et al., 2025: p. 1).

Meanwhile, to ensure manuscript processing and publishing speed, journals often impose very short turnaround times for reviews and editorial decisions, and often revert to undue rejection of otherwise potentially valuable contributions instead of sending manuscripts out to additional reviewers, disincentivizing further submissions by authors and lowering willingness to review (Seghier, 2025). And since timely manuscript handling is crucial for researchers, often journals make a "rejection but resubmission encouraged" decision instead of "major revisions", so that the revised manuscript counts as a new submission and the processing time appears faster in the journal's metrics. Yet journals continue to frame peer review as a professional duty while commercialising its role as a mark of quality and pressuring editors to increase publication output; editors who prioritise rejecting weak work have been sidelined when their standards conflict with this growth model (Boon & Baxter, 2023). Publisher contracts now increasingly tie an editor-in-chief's renewal to a quota of accepted papers, and route rejected manuscripts into sister journals, often fully open-access, rather than out of the portfolio (Siebert, 2018), pushing editors to lower the bar so that volume, and revenue, can be kept up.

3. Who to review for

If reviewing is scarce, largely uncompensated scholarly labour, the first decision that matters is no longer whether to review, but for what journals to review for. In practice, many are more willing to review for prestigious journals or funders rather than for low-impact journals due to the signalling value. Ideally, however, one would prioritize academia-friendly journals run by researchers, academic societies, universities, or museums [e.g. the DAFNEE database (Galtier et al., 2026) or the PEEER network (https://peeer.net/)], or even publishing platforms led by public research funders, such as Open Research Europe (European Commission, 2026). Some of these venues are Diamond Open Access, free for readers and authors. Others charge APCs, but being backed by an academic entity, may recirculate part of their revenues back into the academic community (e.g. prizes for students, travel grants, best-paper awards)

10

(Receveur et al., 2024). While these features are not a guarantee of fair journal practices, they are useful guides to navigate amidst increasingly crowded journal portfolios. Conversely, journals that simultaneously charge authors high fees and make a business of mass-publishing treat peer review as 'raw input' and therefore disregard the consequences of unpaid labour. It is therefore entirely rational for reviewers to take such considerations into account when deciding where to contribute their labour, and the same logic would apply when choosing where to submit manuscripts (McGill et al., 2018; Receveur et al., 2024; Galtier et al., 2026).

However, no-fee journals are not free to run either and in a market dominated by relatively few publishing houses, their costs must be borne elsewhere – institutions, societies, libraries, funders, or through in-kind and an increasingly exhausted volunteer academic labor. Publishing preprints and especially community commenting on preprints can perform part of the dissemination and evaluation traditionally organized by journals, without a commercial publisher in the middle and could become part of a sustainable long-term solution where big publishers are not needed anymore. However, although preprints have greatly expanded access and dissemination (Fu & Hughey, 2019), they have not displaced established journals as the main source of certification and prestige in research assessment. For example, the launch of Peer-Community In (PCI; https://peercommunityin.org) in 2016, moving the peer-review activity to preprints regardless of the final journal where the manuscript will be published, has grown steadily but did not take off widely after 10 years (Class et al., 2026).

Choosing where and for whom to review is therefore a form of important rationing, not a substitute for a different publishing system. Publishing and reviewing only for non-commercial journals may also not be the default best option in all cases. Especially for early-career researchers facing hiring and funding panels, a lower journal rank, a missing 'top-journal', or fewer citations can become a penalty in many assessment systems (McKiernan et al., 2019; Lim et al., 2025), despite longstanding initiatives to reduce reliance on journal-level metrics (e.g. rankings, impact factors; Krauss et al., 2023; Bordignon et al., 2023). As these declarations are often ignored due to the real need to rank applicants using tangible metrics, authors remain forced to target journals whose titles carry prestige in assessment, even though journal-level metrics are poor proxies for the quality of individual papers (Brembs et al., 2013; Hicks et al., 2015) – after all, reading a paper and understanding the science behind it takes time. Publishers extend the brand to capture that behaviour, as in the proliferation of closely related journals or spin-offs from already large journals, which expand the inventory of APCs and read-and-publish slots, so volume itself becomes the margin (McGill, 2024; Hanson et al., 2024). Advising early-career researchers to use only academia-friendly journals, while leaving assessment unchanged, is therefore unlikely to be effective – although some 'high-status' journals do appear in such lists (Galtier et al., 2026). A sound paper in a community journal may

still be disadvantaged in assessments that treat journal titles as proxies for quality, whereas people on short contracts cannot treat that as a moral luxury.

4. Where AI fails

Reviewer judgement is often far from reproducible: a meta-analysis of 48 studies covering 19,443 manuscripts found low inter-rater reliability between independent reviewers (Bornmann et al., 2010). Such disagreement does not necessarily imply error, as competent reviewers may legitimately differ in their assessments of rigour, novelty, significance, and interpretation, often because of high specialisation (Teixeira da Silva, 2026). Still, this limited reproducibility has fuelled interest in AI as a more standardised and consistent evaluator. LLMs can indeed be more consistent than humans (Joachim et al., 2025; Li et al., 2026) and can identify many of the same issues raised by human reviewers (Liang et al., 2024a). However, repeatability is not quality: the same studies reporting high AI consistency also found systematic leniency, inflated quality assessments, and poorer discrimination between human- and AI-generated research (Schherbiak et al., 2024; Joachim et al., 2025; Li et al., 2026).

Consequently, most publishers accept considerate AI use by authors (supported by disclosure statements) while often restricting its use by reviewers. This is primarily due to confidentiality, intellectual-property, accountability, and data-governance concerns. Indeed, pasting a manuscript into a public AI tool by the reviewer may expose unpublished methods, data, and ideas to a third party, potentially breaching the confidentiality and intellectual-property obligations of peer review. Depending on the AI provider, contractual terms, and privacy settings, submitted material may also be retained, reused, or used for model training. A study of the 100 highest-ranked medical journals found that 78 had guidance concerning AI in peer review; among these, 46 explicitly prohibited its use and 32 permitted some restricted use (Li et al., 2024). And yet, AI is increasingly penetrating the peer-review system anyway (Liang et al., 2024b), with over 50% reviewers using AI to help with peer-reviewing papers and writing reports (Naddaf, 2026). Furthermore, AI-based peer-review automation tools (e.g. Refine, ReviewerZero, Stanford's Agentic Reviewer, Google's Paper Assistant Tool, and multi-agent pipelines such as those around aiXiv) are on the rise (Brainard, 2026; Jayaram et al., 2026), while governance remains fragmented and incomplete (Abulibdeh et al., 2026).

The asymmetry between acceptable AI use for authors vs reviewers also reflects an implicit (and we think correct) distinction between AI assisting in the production of a claim and AI adjudicating one. Authors submitting AI-assisted text still bear full responsibility for what they claim; a reviewer's role is precisely to provide their informed opinion on such claims, something that current AI models cannot yet

12

substitute for. Recent simulation-based studies suggest that current LLMs are insufficiently robust to replace human reviewers in scholarly peer review and they can be manipulated by authors by injecting deliberate content into manuscript; still they can support the process by checking a manuscript for several types of errors and inconsistencies (Ye et al., 2024; Torrielli et al., 2026). LLM-based peer review has also been shown to carry biases in favour of senior authors from highly ranked institutions and authors with strong prior publication records (Ye et al., 2024; Vasu et al., 2026) – likely something intrinsic to the training data of LLMs as similar biases affect human reviewers as well (Severin et al., 2020) – or being overly positive and providing generic criticism (Fichtl et al., 2026).

What the current generation of models cannot substitute for is human judgment, though this argument is likely stronger for very disruptive papers rather than for the bulk of 'ordinary' manuscripts. Human review earns its keep on the odd expert point: the objection a model trained on existing literature would not have generated. While a machine check can find patterns and test consistency, the review that matters most is often a leap, and a machine trained on the *status quo* of human knowledge cannot make the leap (Zahavy, 2026) – at least yet (Bubeck et al., 2023; Hochberg & Thrall, 2026). For example, in a study of 139 research proposals, Machado (2026) found that greater novelty increased disagreement between GPT-4o and human evaluators, with humans more likely to favour highly novel proposals. This is not to say that humans are perfect either at evaluating disruptive novel work. Lynn Margulis's first paper discussing how eukaryotic cells were the product of bacterial symbiosis (endosymbiosis) was rejected 15 times before finally finding a home in *Journal of Theoretical Biology* (Knoll, 2012). George Akerlof's paper The Market for "Lemons", which laid the foundations for his later Nobel Prize-winning work on information asymmetry, was rejected by three leading economics journals before being published in the *Quarterly Journal of Economics* (Gans & Shepherd, 1994). Similarly, rejections involved the discoveries of the vitamin B5, the Dane particles (hepatitis B virus), the radioimmunoassay, CRISPR sequences, and many others (Lock, 1993; Gans & Shepherd, 1994; Campanario, 2009; Lander, 2016). Notwithstanding these extreme cases, the human leap remains indispensable, and systems built around AI screening should not erode the space for it. If AI has to be used in peer review, it should function as a secure, disclosed, task-specific, and auditable human–AI collaboration, not as a standalone peer-review machine (Bharti et al., 2024; Akella et al., 2025). The AI could either perform a first, mechanical screen on every manuscript or act as a second (or third) reader, while human reviewers independently evaluate the core scientific claims, verify rather than merely accept AI-generated criticisms, provide disciplinary and contextual judgement, and remain accountable for their recommendations. In other words, after AI runs, a human still has to add something a model trained on existing papers would not have said, and an editor has to notice if they did not. Without human control, AI review would push manuscripts toward the average, resulting in a hivemind

effect where safe framing, familiar methods, and incremental findings prevail (Baumann et al., 2026), all packaged in the standard 'Academese' (Pinker, 2015).

5. Suggestions for healthy and ethical peer review

5.1 Commercial titles last, and only with a limit

Public money pays for research in the name of the public good, so the unpaid labour that certifies it should serve science and society – not the same commercial firms that sell that certification back to the taxpayer. Reviewers should give their unpaid time first to community-owned and no-fee journals (McGill et al., 2018; Peterson et al. 2019; Receveur et al., 2024; Galtier et al., 2026), or even to commenting and giving feedback on preprints. Commercial titles should come second, and only if time remains or the manuscript itself is of genuine scientific interest. For example, requests from commercial or high-volume journals could be accepted when the article's topic falls within the unique expertise or interest of the reviewer, addresses a unique audience, or is a one-off community duty. This may be justified, but such cases should have a clear reason and a yearly limit.

Importantly, redirecting efforts towards academia-friendly journals will be sustainable in the long run only if paralleled by a change in the evaluation of scientific profiles and careers (Section 5.2). Indeed, while tenured staff may have the privilege of flexibility in their journal choices – publishing what they want, where they want – academics relying on grants and early-career researchers still need high-status publications in the current system. Nor do tenured staff operate in a vacuum: they face a dilemma between following their own ethical radar and boosting the careers of their co-authors, who may depend on those very high-status venues.

5.2 Change how we evaluate, not just where we review and publish

As reiterated repeatedly, funders and hiring panels should refrain from using journal ranks and metrics as proxies for the quality of individual papers (Scully & Lodge, 2005; Hicks et al., 2015; Cuellar et al., 2023; Lim et al., 2025), while reviewing and editing work should be weighed more in hiring and tenure decisions (Veríssimo & Roberts, 2013; Moher et al., 2020). Assessment of a few selected papers written by the applicant and judging the quality of the science therein provides a defensible basis for evaluating research than journal-level proxies, complemented by approaches such as narrative CVs (Bordignon et al., 2023) and bibliometric indices that go beyond blunt citation counts (Hicks et al., 2015; Cardoso et al., 2022). Meanwhile, metrics such as the fraction of publications in community-owned journals could be explicitly reported in CVs (e.g. the Author Usage Index in the DAFNEE database: https://dafnee.isem-evolution.fr/).

14

Making peer review work count more on CVs is somewhat trickier. Although there are a few indices aimed at quantifying peer-review work [e.g. a correction factor to the *h*-index (Veríssimo & Roberts, 2013); the Review Quality Instrument (van Rooyen et al., 2019); the R-index (Cantor & Gero, 2015)], none has been widely adopted. The paucity of metrics and their scarce adoption probably reflects the lack of any centralised, standardised repository of peer-review outputs in the first place, preventing science-of-science experts from exploring the space of indices suitable for the task. For the time being, we strongly encourage reviewers and editors to curate their own reviewer profile (currently the most used being Web of Science, which has replaced Publons, or ORCID), collecting there all their peer reviews and editorial contributions. The link to the reviewer profile could then be reported in applications and CVs, making these otherwise largely invisible scholarly contributions visible to assessment panels, as advocated by the Hong Kong Principles (Moher et al., 2020). This strategy is not perfect, as it emphasises quantity over quality, but if widely adopted it could nudge more research and development of bibliometric indicators and validated measures of review quality.

Without such systemic change, any advice to 'just use better journals' is viable mainly for people who already have tenure, while the rest of the community continues to donate unpaid labour to the titles that assessment rewards.

5.3 Pay editors first, then maybe reviewers

What the post-World War II volume boom replaced was an expert editor, paid to read and evaluate, who called on outside reviewers only when they lacked the relevant expertise. Remunerating editors, as some journals already do, is therefore a first necessity. A properly resourced editor is better able to carefully read and evaluate submissions, take the time required to handle multiple manuscripts, find merit in a submitted manuscript, differentiate a unique report from a generic or machine-written one, refuse to treat speed as the main performance metric, and decide whether a weak review should trigger another reviewer rather than a quiet rejection or acceptance. Such an editor can also properly moderate the review procedure itself – highlighting substantive reviewer comments that authors are strictly required to address, resolving conflicting suggestions by different reviewers, and removing *ad hominem* attacks from a report before it reaches the authors.

As for reviewers, payment can be an option, but only under the right conditions (Naddaf, 2026; Pagliaro et al., 2026; Teixeira da Silva, 2026). Any remuneration for a peer-review report turns reviewing into a target for gaming (Goodhart, 1975), so paying ordinary reviewers by the paper risks inviting fast, thin, recycled, or AI-written reports, unless assignments are capped and the editor scores the report before

15

money moves (Squazzoni et al., 2013). Payment also raises equity problems that have nothing to do with review quality: in several countries, especially in the Global South, public employees, including researchers, cannot legally receive outside income without cumbersome administrative procedures. In nations sanctioned by Western or NATO-aligned countries (e.g. Iran, Syria, North Korea, Russia), financial contracts with scientists are prohibited or heavily restricted by international banking and trade law. In the same vein, researchers employed by US federal agencies are generally barred from accepting outside payment for work related to their official duties, including reviewing. Any payment scheme should therefore start small (e.g. on a limited pool of reviewers a journal actually employs), rather than as a blanket policy that may exclude reviewers in some regions.

Grant evaluation already shows reviewer payment is, in principle, feasible. In European Commission-run funding schemes (e.g. Horizon Europe, COST, LIFE), experts are paid a €450 daily fee, typically amounting to €180–450 per proposal depending on time spent. The workflow is still independent reports plus a consensus step, and the non-monetary return (networks, insight into how proposals are judged) matters as much as the fee. Importantly, payment does not by itself equalise who can serve and accept extra income on top of a public salary, so the geographic bias persists. Furthermore, mediocre report quality still persists under paid grant review (Jayasinghe et al., 2003; Pier et al., 2018), so payment on its own is no cure.

Evidence from journals, where trials are scarcer, is more encouraging on the quality question specifically. In China, reviewer payment is already common practice, with journals offering a small honorarium for completed reviews (Chang & Yuan, 2020; Hu et al., 2023). An experiment with 1,500 reviewers found that payment accelerated reviews without reducing quality (Chetty et al., 2014), and a recent trial found that a payment of US$250 increased completion and modestly shortened turnaround with no detectable effect on review quality (Cotton et al., 2025). The fully open-access journal *Biology Open* published by the not-for-profit Company of Biologists, is testing a related model through its Fast & Fair peer-review initiative of pre-contracting and compensating reviewers, with payment conditional on meeting both quality and timing requirements. In its initial pilot, all 20 manuscripts received a first decision within seven business days, with review quality reported as maintained on the basis of editorial assessments (Gorelick & Clark, 2025). The initiative has since expanded, with 2025 data showing substantially shorter decision times, higher reviewer acceptance and completion rates, and no detectable deterioration in editor-rated review quality (Gorelick & Clark, 2026).

All of this ultimately suggests that unless it is done carefully, paying reviewers can be worse than the problem it is meant to solve. Hence, a multi-year trial of paid reviewing under strict limits could be

16

implemented to test whether paid review can work (Cheah & Piasecki, 2022). But none of this touches the deeper problem this paper is about: publisher profit. Paying editors and, eventually, some reviewers is worth doing regardless, because neither would meaningfully dent publisher margins, and because the current system already extracts this labour for free. But payment is not a substitute for the other pillars of this paper: it does not redirect effort towards community-based journals, it does not change how reviewer works is evaluated on CVs, and, done carelessly, it risks rewarding fast and thin reports and the abuse of AI.

5.4 AI as a tool under human authority

With LLMs now permeating society at every level, "Did you use AI for…?" is becoming a nearly pointless question and a blanket prohibiting its use in peer review is increasingly unrealistic. However, AI use must be strongly regulated by publishers, as quick language polishing of a reviewer report for clarity versus a substantive generative AI involvement in the assessment of a paper requires very different levels of transparency and scrutiny, not to mention that pasting an unpublished manuscript into AI can store that text on a third-party server and breach confidentiality. Such distinctions may be especially relevant for non-native English speakers, where language-only AI assistance could help reduce this additional burden without replacing the reviewer's scientific judgement (Thakkar et al., 2026).

AI in peer review should, however, stay as a disclosed aid under the reviewer's authority, not as a substitute for thoughtful human engagement. Used with confidentiality and accountability, it can flag inconsistencies across text, tables, figures, references and supplementary materials, and can support reporting checks, though performance varies by task, tool and prompt design (Eckmann et al., 2026; Erturk & Durmaz, 2026). Beyond this mechanical screening, LLMs have been proposed as useful copilots throughout the review pipeline in a different sense: as checklist assistants, aids to reviewer selection, generators of structured feedback on key evaluation criteria, and detectors of bias (Sun, 2025). They may also offer preliminary comments on conceptual gaps and missing critical literature (Liang et al., 2024a), granted that sources and claims are verified independently (Walters & Wilder, 2023). Free of these lower-order tasks, the human reviewer is left more time to exercise judgement on the core science behind a given paper and remains solely responsible for the final report – yet a breach of reviewer confidentiality remains. Meanwhile, the editor's role becomes assessing whether a report is more than a rewrite of the machine check, and whether that reviewer is asked again and eventually paid (see Section 5.3).

A key condition to make AI-assisted review sustainable lies in the prompt itself. LLM outputs can change substantially depending on how a task is formulated, which criteria are emphasised, and how strict the instructions are, which can alter both the content and the severity of a review (Erturk & Durmaz, 2026).

17

If AI checks are incorporated into peer review, journals should record the prompts, model, version, and outputs in a secure system. Once this check is standardised to a fixed model and prompt, there is little reason to leave it to each reviewer to run individually and inconsistently, as currently happens. Journals could perform the same automated check on every manuscript, and hand over its output to authors, reviewers and editors for their consideration. For example, this is already done by Frontiers publisher through their proprietary AIRA suite, which is directly integrated into the reviewer forum. Similarly, Springer Nature launched an in-house AI-driven integrity/quality screening system to flag problems before external peer review (Springer Nature, 2025). Importantly, this proposition sits in tension with the core argument of this paper that commercial publishers extract unpaid reviewer labour and sell scientific judgement they do not themselves produce. In transferring labour away from unpaid reviewers, commercial publishers could potentially monetise AI-screening as well (e.g. 'AI-checked' submission fees), particularly against the backdrop of an over-stretched reviewer pool with resultant compromises on review quality and efficiency. If centralised AI screening is to be worth adopting at scale, then it should be built and governed in ways that keep it from becoming another source of commercial capture (e.g. community-run or society-held infrastructure), and the review and the editorial decision must remain human.

6. Conclusions

Most critiques of the current peer review system ultimately root back to a single core problem: a publishing industry whose commercial incentives reward volume and profit over the quality of scientific assessment. And yet, however cynical or pessimistic we may feel, until a better system of quality control is devised and adopted, peer review remains the backbone of science and what distinguishes it from other forms of knowledge. While there is no simple, single solution to the peer-review crisis (e.g. pay reviewers, open reports, ban AI) given the historical resistance of peer-review to change from its main structural features relative to changes in the scientific and publishing environment (Ferreira et al., 2016), we believe that a community-based approach, encapsulated in five main priorities (Box 2), has the potential to slowly move the needle in the right direction. The core of these solutions can be distilled into answers to the three questions posed in our title. "Who should still peer-review?" has, in our view, a very simple answer: the scholarly community should continue to provide this fundamental service, as it is at the hearth of science as a unique knowledge system and the growth of the people who do research. But it is probably the second question, "For whom should we review?", that truly defines the contours of this effort: peer-review should be carefully rationed, making sure that we devote the most effort toward those actors that are friendly to, and aligned with, the long-term goals of the collective scientific enterprise, rather than primarily driven by commercial incentives. The third question, "with what tools?", may soon become less useful as AI evolves at a speed that makes predictions obsolete within months. Yet our current answer is concrete and transcends

18

the specific capabilities of ever-evoliving tools: AI should act as a disclosed support, run centrally and consistently by journals rather than improvised by individual reviewers, with archived prompts and models, and only on infrastructure that stays out of commercial capture. What it should not do is replace the judgement of a human reviewer, accountable to the authors and the field. Together, these three questions sketch a peer review system that puts the scientific community rather than shareholders back at the center of the process, and that is assisted, not replaced, by AI tools.

Box 2. | Five practical priorities for a more sustainable peer review process.

1. Editors of commercial journals who perform substantive editorial work should be paid, as a matter of basic fairness; reviewer payment should be piloted only under strict conditions (capped assignments, quality scored before money moves, contracted rather than flat-fee); and no version of payment excuses journals or funders from also fixing assessment and access.
2. Unpaid review goes first to community-owned and no-fee journals, and to serious commentary on preprints; commercial or high-volume titles come second, only with a reason and a yearly limit.
3. Funders and hiring panels should refrain from using journal rank as a proxy for paper quality; reviewing and editorial activities should count favourably on CVs; laboratories, departments, and societies should report the proportion of their publishing and reviewing that goes to Academia-friendly journals, so that effort becomes visible in assessment.
4. AI use beyond basic proofreading must be disclosed; unpublished manuscripts stay off public or otherwise insufficiently protected AI services; any machine check runs on journal-approved or owned systems with confidentiality and data-protection safeguards; prompts should be archived and never steered toward a predetermined verdict (e.g. “find reasons to reject…”, “what are the main flaws…?”); a review generated solely by an LLM should not qualify as peer review.
5. If a centralised AI screen is adopted, it should sit in community- or society-held infrastructure, not become another commercial fee on top of unpaid judgement.

References

Abulibdeh, R., Arslan, J., Ordóñez, S. A. C., Celi, L. A., Lunde, T. M., Ramos, M. (2026). From policing authors to stewarding science: Governing artificial intelligence across the scientific publishing pipeline in medical journals. *MIT Science Policy Review, 7*, 5-20. https://doi.org/10.38105/spr.cnmdfm0xb7

19

Aczel, B., Barwich, A. S., Diekman, A. B., Fishbach, A., Goldstone, R. L., Gomez, P., Gundersen, O. E., Von Hippel, P. T., Holcombe, A. O., Lewandowsky, S., Nozari, N., Pestilli, F., Ioannidis, J. P. A. (2025). The present and future of peer review: Ideas, interventions, and evidence. *Proceedings of the National Academy of Sciences*, *122*(5), e2401232121. https://doi.org/10.1073/pnas.2401232121

Akella, A. P., Siravuri, H. V., Rohatgi, S. (2025). Pre-review to peer review: Pitfalls of automating reviews using large language models. *arXiv preprint arXiv:2512.22145*

Albert, A. Y., Gow, J. L., Cobra, A., Vines, T. H. (2016). Is it becoming harder to secure reviewers for peer review? A test with data from five ecology journals. *Research integrity and peer review, 1*(1), 14. https://doi.org/10.1186/s41073-016-0022-7

Amaral, O. B. (2022). To fix peer review, break it into stages. *Nature*, *611*, 637. https://doi.org/10.1038/d41586-022-03791-5

Arboledas-Lérida, L. (2024). The 'unpaid' labour of peer reviewers and the accumulation of capital in the industry of academic publishing. *Critique, 52*(2–3), 317–338. https://doi.org/10.1080/03017605.2024.2364454

Baldwin, M. (2018). Scientific autonomy, public accountability, and the rise of "peer review" in the Cold War United States. *Isis*, *109*(3), 538-558. https://doi.org/10.1086/700070

Barnett, A., Doubleday, Z. (2020). The growth of acronyms in the scientific literature. *eLife*, *9*, e60080. https://doi.org/10.7554/eLife.60080

Baumann, J., Pei, J., Koyejo, S., Hovy, D. (2026). Stop automating peer review without rigorous evaluation. *arXiv preprint arXiv:2605.03202*

Becerra-Rodero, E., Ortega, J. L. (2026). The role of editorial boards in shaping scholarly communication: A large-scale mapping of their composition and structure. *Journal of the Association for Information Science and Technology*, *77*(10), 1334-1353. https://doi.org/10.1002/asi.70108

Beigel, F., Brockington, D., Crosetto, P., Derrick, G., Fyfe, A., Barreiro, P. G., Hanson, M. A., Haustein, S., Larivière, V., Noe, Ch., Pinfield, S., Wilsdon, J. (2025). The drain of scientific publishing. *arXiv preprint arXiv:2511.04820*

Bharti, P. K., Ghosal, T., Agarwal, M., Ekbal A. (2024). PEERRec: An AI-based approach to automatically generate recommendations and predict decisions in peer review. *International Journal on Digital Libraries*, *25*, 55–72. https://doi.org/10.1007/s00799-023-00375-0

Björk, B.-C., Solomon, D. (2013). "The Publishing Delay in Scholarly Peer-Reviewed Journals." *Journal of Informetrics*, *7*(4): 914–923. https://doi.org/10.1016/j.joi.2013.09.001

Boon, P. J., Baxter, J. M. (2023, November 20). A letter from the Chief Editors of Aquatic Conservation. Aquatic Conservation: Marine and Freshwater Ecosystems. https://www.jack-greenhalgh.com/a-letter-from-the-chief-editors-of-aquatic-conservation

Bordignon, F., Chaignon, L., Egret, D. (2023). Promoting narrative CVs to improve research evaluation? A review of opinion pieces and experiments. *Research Evaluation*, *32*(2), 313-320. https://doi.org/10.1093/reseval/rvad013

Bornmann, L., Mutz, R., Daniel, H. D. (2010). A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PLoS One*, *5*(12), e14331. https://doi.org/10.1371/journal.pone.0014331

Brainard, J. (2026). Ready for the robot reviewers? *Science*, *393*(6815), 963. https://doi.org/0.1126/science.z6vx92p

Brembs, B., Button, K., Munafò, M. (2013). Deep impact: unintended consequences of journal rank. *Frontiers in Human Neuroscience*, *7*, 291. https://doi.org/ 10.3389/fnhum.2013.00291

Breuning, M., Backstrom, J., Brannon, J., Gross, B. I., Widmeier, M. (2015). Reviewer fatigue? Why scholars decline to review their peers' work. *PS: Political Science & Politics*, *48*(4), 595-600. https://doi.org/10.1017/S1049096515000827

Bubeck, S., Chadrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., Zhang, Y. (2023, March). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Burnham, J. C. (1990). The evolution of editorial peer review. *Jama*, *263*(10), 1323-1329.

Campanario, J. (2009). Rejecting and resisting Nobel class discoveries: accounts by Nobel Laureates. *Scientometrics*, *81*(2), 549-565. https://doi.org/10.1007/s11192-008-2141-5

Cantor, M., Gero, S. (2015). The missing metric: quantifying contributions of reviewers. *Royal Society Open Science*, *2*(2), 140540. http://dx.doi.org/10.1098/rsos.140540

Cardoso, P., Fukushima, C. S., Mammola, S. (2022). Quantifying the internationalization and representativeness in research. *Trends in ecology & evolution*, *37*(9), 725-728. https://doi.org/10.1016/j.tree.2022.06.011

Cheah, P., Piasecki, J. (2022). Should peer reviewers be paid to review academic papers? *The Lancet*, *399*(10335), 1601. https://doi.org/10.1016/S0140-6736(21)02804-X

Chang, W., Yuan, J. (2020). Progress of peer review in international scholarly publishing. *Chinese Journal of Scientific and Technical Periodicals*, *31*(10), 1181–1192.

Chetty, R., Saez, E., Sándor, L. (2014). What policies increase prosocial behavior? An experiment with referees at the Journal of Public Economics. *Journal of Economic Perspectives*, *28*(3), 169–188. https://doi.org/10.1257/jep.28.3.169

Class, B., Bourguet, D., Guillemaud, T. (2026). Peer community in and peer community journal: a two-step diamond OA process giving research communities back control of publishing. *Insights*, *39*, 11. https://doi.org/10.1629/uksg.744

Cordero, R. J. B., de León-Rodriguez, C. M., Alvarado-Torres, J. K., Rodriguez, A. R., Casadevall, A. (2016). Life science's average publishable unit (APU) has increased over the past two decades. *PLoS One, 11*, e0156983. https://doi.org/10.1371/journal.pone.0156983

Cotton, C. S., Alam, A., Tosta, S., Buchman, T. G., Maslove, D. M. (2025). Effect of monetary incentives on peer review acceptance and completion: A quasi-randomized interventional trial. *Critical Care Medicine*, *53*(6), e1181–e1189. https://doi.org/10.1097/CCM.0000000000006637

21

Cuellar, M. J., Takeda, H., Truex III, D. T. (2024). Are ranked journal lists effective determiners of article quality? *Journal of Computer Information Systems*, *64*(5), 622-635. https://doi.org/10.1080/08874417.2023.2240725

Dobusch, L., Plotnikof, M., Wenzel, M. (2026). Reviewing is caring! Revaluing a critical, but invisibilized, underappreciated, and exploited academic practice. *Organization*, *33*(3), 458–474. https://doi.org/10.1177/13505084251343672 journals.sagepub.com

Ebadi, S., Nejadghanbar, H., Salman, A. R., Khosravi, H. (2025). Exploring the impact of generative AI on peer review: Insights from journal reviewers. *Journal of Academic Ethics*, *23*(3), 1383-1397. https://doi.org/10.1007/s10805-025-09604-4

Eckmann, P., Barnett, A., Bannach-Brown, A., Atria, E. P. B., Cabanac, G., Owen Franzen, L. D., Gazda, M. A., Hair, K., Howison, J., Kilicoglu, H., Labbe, C., McCann, S., Nachev, V., Roelandse, M., Salholz-Hillel, M., Schulz, R., Riet, G. T., Vorland, C., Bandrowski, A., Weissgerber, T. (2026). Use as directed? A comparison of software tools intended to check rigor and transparency of published work. *PloS One*, *21*(2), e0342225. https://doi.org/ 10.1371/journal.pone.0342225

Else, H. (2024). Biomed retractions have quadrupled in 20 years—Why? *Nature*, *630*(8016), 280-1. https://doi.org/10.1038/d41586-024-01609-0

Erturk, S. M., Durmaz, M. (2026). Large language models as peer reviewers: Prompt sensitivity and model-dependent reproducibility. *Academic Radiology*, *33*(9), 3642-3650. https://doi.org/10.1016/j.acra.2026.04.046. PMID: 42624573.

European Commission, (2026). A new era for Open Research Europe. Directorate-General for Research and Innovation, 26 March 2026. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/new-era-open-research-europe-2026-03-26_en

Ferreira, C., Bastille-Rousseau, G., Bennett, A. M., Ellington, E. H., Terwissen, C., Austin, C., Borlestean, A., Boudreau, M. R., Chan, K., Forsythe, A., Hossie, T. J., Landolt, K., Longhi, J., Otis, J.-A., Peers, M. J. L., Rae, J., Seguin, J., Watt, C., Wehtje, M., Murray, D. L. (2016). The evolution of peer review as a basis for scientific publication: Directional selection towards a robust discipline? *Biological Reviews*, *91*(3), 597–610. https://doi.org/10.1111/brv.12185

Fichtl, A. M., Ellinger, L., Kelber, J., Olík, K., Groh, G. (2026). AI-assisted peer review across research communities: From reviewer AI policies to LLM review quality. *arXiv preprint arXiv:2608.03581*

Fleming, P., Harley, B. (2024). Collegiality as control? How uncounted work gets done in the neoliberal business school. *Academy of Management Learning & Education*, *23*(1), 176–190. https://doi.org/10.5465/amle.2022.0486

Fox, C. W. (2017). Difficulty of recruiting reviewers predicts review scores and editorial decisions at six journals of ecology and evolution. *Scientometrics*, *113*(1), 465-477. https://doi.org/10.1007/s11192-017-2489-5

Fox, C. W., Meyer, J., Aimé, E. (2023). Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal. *Functional Ecology*, *37*, 1144–1157. https://doi.org/10.1111/1365-2435.14259

Fu, D. Y., Hughey, J. J. (2019). Releasing a preprint is associated with more attention and citations for the peer-reviewed article. *eLife*, *8*, e52646. https://doi.org/10.7554/eLife.52646

Fyfe, A. (2022). Self-help for learned journals: Scientific societies and the commerce of publishing in the 1950s. *History of Science*, *60*, 255-279. https://doi.org/10.1177/0073275321999901

Galtier, N., Belkhir, K., Antoine, P. O., Bibal, C., Boëte, C., Delsuc, F., Huchard, E., Kéfi, S., Maeder, Ch., Paradis, L., Puechmaille, S. J., Queffelec, A., Scornavacca, C., Smadja, C. M. (2026). Time to publish responsibly: DAFNEE, a database of academia-friendly journals in ecology and evolutionary biology. *Journal of Evolutionary Biology*, *39*(2), 190-199. https://doi.org/10.1093/jeb/voaf143

Gans, J. S., Shepherd, G. B. (1994). How are the mighty fallen: Rejected classic articles by leading economists. *Journal of Economic Perspectives*, *8*(1), 165-179. https://doi.org/10.1257/jep.8.1.165

Goodhart, C. A. E. (1975). Problems of monetary management: The UK experience. In: Papers in Monetary Economics, Palgrave, London. pp. 91-121. https://doi.org/10.1007/978-1-349-17295-5_4

Gorelick, D. A., Clark, A. (2025). Fast & Fair peer review: a bold experiment in scientific publishing. *Biology Open*, *14*(3), bio061982. https://doi.org/10.1242/bio.061982

Gorelick, D. A., Clark, A. (2026). Expanded implementation of fast & fair paid peer review reduces time to first decision without reducing review quality in a biology journal. *bioRxiv*, 2026-06. https://doi.org/10.64898/2026.06.02.729548

Grossmann, A., Brembs, B. (2021). Current market rates for scholarly publishing services. *F1000Research*, *10*, 20. https://doi.org/10.12688/f1000research.27468.2

Hanson, M. A., Barreiro, P. G., Crosetto, P., Brockington, D. (2024). The strain on scientific publishing. *Quantitative Science Studies*, *5*(4), 823–843. https://doi.org/10.1162/qss_a_00327

Haubrock, P. J., Kouba, A., Soto, I., Abreo, N. A. S., Vilizzi, L., Błońska, D., Serhan Tarkan, A., De Santis, V., Sousa, R., Nogueira, A. B., Heeren, S., Garrido Nogueira, J., Cano-Barbacil, C., Cremella, B., Strubbe, D., da Silva, J. P., Poloni, R., Deschepper, P., Thoré, E. S. J., Misteli, B., Sun, J., Panisi, M., Dudgeon, D., Stamatiadou, V., Galib, S. M., Maes, C., Baso, N. C., Beukema, W., Števove, B., Záhorská, E., Kurtul, I., Kaur, D., Huber, A. F., Halabowski, D., Guilherme Padilha, de.Assis, J., Balzani, P., Kaya, C., Bayçelebi, E., Mugnai, M., Ahmed, D. A., Saç, G., Aksu, S., Malik Kaya, B., Ciftcioglu, M., Jiang, X., Rodríguez-Rey, M., Martín-Forés, I., Gonçalves Leles, S., Pârvulescu, L., Pergl, J., Rossetto, F., Wei, H., Kirichenko, N. I., Briski, E., Gallitelli, L., Britton, R., Everts, T. (2026). How contemporary academic structures constrain scientific creativity and hold back Early Career Researchers. *Nature Human Behavior*, https://doi.org/10.1038/s41562-026-02575-5

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429-431. https://doi.org/10.1038/520429a

Humphreys, H. (2022). Payment and progress in peer review. *Lancet*, *400*(10347), 159. https://doi.org/10.1016/s0140-6736(22)00921-7

23

Ho, R. C. M., Mak, K. K., Tao, R., Lu, Y., Day, J. R., Pan, F. (2013). Views on the peer review system of biomedical journals: an online survey of academics from high-ranking universities. *BMC Medical Research Methodology*, *13*(1), 74. https://doi.org/10.1186/1471-2288-13-74

Jayaram, R., Tyler, D., Woodruff, D., Cortes, C., Matias, Y., Mirrokni, V., Cohen-Addad, V. (2026). Towards Automating Scientific Review with Google's Paper Assistant Tool. *arXiv preprint arXiv:2606.28277*

Jayasinghe, U. W., Marsh, H. W., Bond, N. (2003). A multilevel cross classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, *166*(3), 279–300. https://doi.org/10.1111/1467-985X.00278

Joachim, M. V., Dodson, T. B., Laviv, A. (2025). How artificial intelligence differs from humans in peer review. *Journal of Oral and Maxillofacial Surgery*, *83*, 1040–1050. https://doi.org/10.1016/j.joms.2025.03.015

Johnson, R., Watkinson, A., Mabe, M. (2018). The STM Report, 5th edition: An overview of scientific and scholarly publishing. Netherlands. https://coilink.org/20.500.12592/7qcc41

Kaltenbrunner, W., Birch, K., Amuchastegui, M. (2022). Editorial work and the peer review economy of STS journals. *Science, Technology, & Human Values*, *47*(4), 670–697. https://doi.org/10.1177/01622439211068798

Kavanagh, K. (2025). Want to do disruptive science? Include more rookie researchers. *Nature*, *646*(8084), 269-270. https://doi.org/10.1038/d41586-025-03117-1

Knoll, A.H. (2012). Lynn Margulis, 1938–2011. *Proceedings of the National Academy of Sciences of the USA, 109*(4), 1022-1022. https://doi.org/10.1073/pnas.1120472109

Krauss, A., Danús, L., Sales-Pardo, M. (2023). Early-career factors largely determine the future impact of prominent researchers: evidence across eight scientific fields. *Scientific Reports*, *13*(1), 18794. https://doi.org/10.1038/s41598-023-46050-x

Lamberts, R. P., Noordhof, D. A., Sandbakk, Ø., Mujika, I., Chamari, K., Roelands, B., Skorski, S., Haugen, T., Boullosa, D., Cortis, C., Janse de Jonge, X. A. K., McGawley, K., Loturco, I., Stevens, C. J., Pyne, D. B., Foster, C., Beneke, R., de Koning, J. J., & Langerak, N. G. (2025). The Silent Workforce of Academic Publishing: Time to Give Credit to Peer Reviewers. *International Journal of Sports Physiology and Performance*, *20*(9), 1181-1183. https://doi.org/10.1123/ijspp.2025-0303

Lander, E. S. (2016). The heroes of CRISPR. *Cell*, *164*(1), 18-28. https://doi.org/10.1016/j.cell.2015.12.041

Li, Z. Q., Xu, H. L., Cao, H. J., Liu, Z. L., Fei, Y. T., Liu, J. P. (2024). Use of artificial intelligence in peer review among top 100 medical journals. *JAMA Network Open*, *7*(12), e2448609. https://doi.org/10.1001/jamanetworkopen.2024.48609

Li, W., Zhao, R., Sun, P. T., Lee, J., Lu, O., Yuan, S., Sureshkumar, P., Chen, Y., Xiao, Z., Tang, Y. J. (2026). Large language models for peer review in biotechnology. *New Biotechnology*, *93*, 218–225. https://doi.org/10.1016/j.nbt.2026.03.007

Liang, W., Zhang, Y., Cao, H., Wang, B., Ding, D. Y., Yang, X., Kailas Vodrahalli, M. S., He, S., Scott Smith, D., Yin, Y., McFarland, D. A., Zou, J. (2024a). Can large language models provide useful feedback on research

papers? A large-scale empirical analysis. *NEJM AI*, *1*(8), AIoa2400196. https://doi.org/10.1056/AIoa2400196

Liang, W., Izzo, Z., Zhang, Y., Lepp, H., Cao, H., Zhao, X., Chen, L., Ye, H., Liu, S., Huang, Z., McFarland, D. A., Zou, J. Y. (2024b). Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. *arXiv preprint arXiv:2403.07183*

Lim, B. H., D'ippoliti, C., Dominik, M., Hernández-Mondragón, A. C., Vermeir, K., Chong, K. K., Hussein, H., Morales-Salgado, V. S., Cloete, K. J., Kimengsi, J. N., Balboa, L., Mondello, S., dela Cruz, T. E., Lopez-Verges, S., Sidi Zakari, I., Simonyan, A., Palomo, I., Režek Jambrak, A., Germo Nzweundji, J., Molnar, A., Saktiawati, A. M. I., Elagroudy, S., Kumar, P., Enany, S., Narita, V., Backes, M., Siciliano, V., Egamberdieva, D. Flores Bueso, Y. (2025). Regional and institutional trends in assessment for academic promotion. *Nature*, *638*(8050), 459-468. https://doi.org/10.1038/s41586-024-08422-9

Lock, S. (1993). The grossest failures of peer review. *BMJ: British Medical Journal*, *307*(6900), 382. https://doi.org/10.1136/bmj.307.6900.382

Lundblad, K. (2025). Proof of Concept: Typesetting of research in the academic publishing industry. *Logos*, *35*(4), 43-58. https://doi.org/10.1163/18784712-20240032

Machado, D. (2026). Generative AI bias against scientific novelty: a cautionary tale from a small-sample evaluation of research proposals. *Scientometrics*, *131*, 7479–7497. https://doi.org/10.1007/s11192-026-05792-x.

Mammola, S. (2020). On deepest caves, extreme habitats, and ecological superlatives. *Trends in Ecology & Evolution*, *35*(6), 469-472. https://doi.org/10.1016/j.tree.2020.02.011

Mandel, R. (1996). A half century of peer review: 1946-1996. Division of Research Grants, National Institutes of Health, Bethesda, MD. 340pp. https://archive.org/details/halfcenturyofpee00mand/page/340/mode/2up

McGill, B., Araújo, M., Franklin, J., Linder, H. P., Dawson, M. N. (2018). Writing the future of biogeography. *Frontiers of Biogeography*, *10*(3-4), e41964. https://doi.org/10.21425/F5FBG41964

McGill, B. (2024). The state of academic publishing in 3 graphs, 6 trends, and 4 thoughts. *Dynamic Ecology*. https://dynamicecology.wordpress.com/2024/04/29/the-state-of-academic-publishing-in-3-graphs-5-trends-and-4-thoughts/

McKiernan, E. C., Schimanski, L. A., Muñoz Nieves, C., Matthias, L., Niles, M. T., Alperin, J. P. (2019). Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. *eLife*, *8*, e47338. https://doi.org/10.7554/eLife.47338

McNutt, R. A., Evans, A. T., Fletcher, R. H., Fletcher, S. W. (1990). The effects of blinding on the quality of peer review: a randomized trial. *Jama*, *263*(10), 1371-1376. https://doi.org/10.1001/jama.1990.03440100079012

McPeek, M. A., DeAngelis, D. L., Shaw, R. G., Moore, A. J., Rausher, M. D., Strong, D. R., Ellison, A. M., Barrett, L., Rieseberg, L., Breed, M. D., Sullivan, J., Osenberg, C. V., Holyoak, M., Elgar, M. A. (2009). The golden rule of reviewing. *The American Naturalist*, *173*(5), E155-E158. https://doi.org/10.1086/598847

Meyerson, L.A., Suzzi-Simmons, A., Simberloff, D. (2025). Quantifying reviewer declines in scientific publishing: twenty-one years of data from biological invasions 2002–2024. *Biological Invasions*, *27*, 223. https://doi.org/10.1007/s10530-025-03679-1

25

Miao, J., Davis, J. R., Zhang, Y., Pritchard, J. K., Zou, J. (2026). Reimagining research papers as interactive and reliable AI agents. *Nature*. https://doi.org/10.1038/s41586-026-11044-y

Moher, D., Bouter, L., Kleinert, S., Glasziou, P., Sham, M.H., Barbour, V., Coriat, A.M., Foeger, N., Dirnagl, U. (2020). The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology*, *18*(7), e3000737. https://doi.org/10.1371/journal.pbio.3000737

Montagna, P. A. (2026). Estuaries and coasts data explains why peer review is slow and there is a reviewing crisis. *Estuaries and Coasts*, *49*(5), 155. https://doi.org/10.1007/s12237-026-01782-7

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021. https://doi.org/10.1038/s41562-016-0021

Naddaf, M., Van Noorden, R. (2026). AI-detection tools have made huge leaps forward — how good are they? *Nature*, *656*(8129), 808-811. https://doi.org/10.1038/d41586-026-02569-3

Naddaf, M. (2026). Paid to peer review: firms are paying researchers to assess manuscripts for journals. *Nature*, *658*, 306. https://doi.org/10.1038/d41586-026-02277-y

Nature Microbiology, (2018). Peer review is not broken. *Nature Microbiology*, *3*, 1. https://doi.org/10.1038/s41564-017-0093-3

Okike, K., Hug, K. T., Kocher, M. S., Leopold, S. S. (2016). Single-blind vs double-blind peer review in the setting of author prestige. *Jama*, *316*(12), 1315-1316. https://doi.org/10.1001/jama.2016.11014

Ottaviani, G., Martínez, A., Petit Bon, M., Mammola, S. (2025). On the quest for novelty in ecology. *Peer Community Journal*, *5*, e59. https://doi.org/10.24072/pcjournal.567

Oviedo-García, M. Á. (2024). The review mills, not just (self-) plagiarism in review reports, but a step further. *Scientometrics*, *129*(9), 5805-5813. https://doi.org/10.1007/s11192-024-05125-w

Pagliaro, M., Della Pina, C., Ciriminna, R. (2026). Ending the peer review crisis: Sharing revenues with reviewers. *European Review*. https://doi.org/10.1017/S1062798726100829

Park, M., Leahey, E., Funk, R.J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, *613*, 138–144. https://doi.org/10.1038/s41586-022-05543-x

Peterson, A. T., Glor, R. E., Soberón, J. (2019). More on the future of publishing in biogeography. *Frontiers of Biogeography*, *11*(1), e42880. https://doi.org/10.21425/F5FBG42880

Peterson, D. A. (2020). Dear reviewer 2: Go f' yourself. *Social Science Quarterly*, *101*(4), 1648-1652. https://doi.org/10.1111/ssqu.12824

Pier, E. L., Brauer, M., Filut, A., Kaatz, A., Raclaw, J., Nathan, M. J., Ford, C. E., Carnes, M. (2018). Low agreement among reviewers evaluating the same NIH grant applications. *Proceedings of the National Academy of Sciences*, *115*(12), 2952–2957. https://doi.org/10.1073/pnas.1714379115

Pinker, S. (2015). Why academics stink at writing. The *Chronical of Higher Education*, *61*, 1–18. https://grad.ncsu.edu/wp-content/uploads/2016/06/Why-Academics-Stink-at-Writing-1-2.pdf

Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C., Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *eLife*, *6*, e27725. https://doi.org/10.7554/eLife.27725

Receveur, A., Bonfanti, J., d'Agata, S., Helmstetter, A. J., Moore, N. A., Oliveira, B. F., Petit-Cailleux, C., Rievrs Borges, E., Schultz, M., Sexton, A. N., Veytia, D. (2024). David versus Goliath: Early career researchers in an unethical publishing system. *Ecology Letters*, *27*(3), e14395. https://doi.org/10.1111/ele.14395

Russo, G., Horta Ribeiro, M., Davidson, T. R., Veselovsky, V., West, R. (2025). The AI review lottery: widespread AI-assisted peer reviews boost paper scores and acceptance rates. *Proceedings of the ACM on Human-Computer Interaction*, *9*(7), CSCW486, 1 – 28. https://doi.org/10.1145/3757667

Saad, A., Jenko, N., Ariyaratne, S., Birch, N., Iyengar, K. P., Davies, A. M., Vaishya, R., Botchu, R. (2024). Exploring the potential of ChatGPT in the peer review process: an observational study. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *18*(2), 102946. https://doi.org/10.1016/j.dsx.2024.102946

Sallaz, J. J. (2013). Your paper has just been outsourced. Global Dialogue 3(4). https://globaldialogue.isa-sociology.org/articles/your-paper-has-just-been-outsourced

Scully, C., Lodge, H. (2005). Impact factors and their significance; overrated or misused? *British Dental Journal*, *198*, 391–393. https://doi.org/10.1038/sj.bdj.4812185

Shcherbiak, A., Habibnia, H., Böhm, R., Fiedler, S. (2024). Evaluating science: A comparison of human and AI reviewers. *Judgment and Decision Making*, *19*, e21. https://doi.org/10.1017/jdm.2024.24

Seghier, M. L. (2025). Paying reviewers and regulating the number of papers may help fix the peer-review process. *F1000Research*, *13*, 439. https://doi.org/10.12688/f1000research.148985.4

Severin, A., Martins, J., Heyard, R., Delavy, F., Jorstad, A., Egger, M. (2020). Gender and other potential biases in peer review: cross-sectional analysis of 38 250 external peer review reports. *BMJ open*, *10*(8), e035058. https://doi.org/10.1136/bmjopen-2019-035058

Siebert, S. (2018). The "champagne tower" of science publishing. *International Higher Education*, *94*, 21-22. https://doi.org/10.6017/ihe.2018.0.10564

Squazzoni, F., Bravo, G., Takács, K. (2013). Does incentive provision increase the quality of peer review? An experimental study. *Research Policy*, *42*(1), 287-294. https://doi.org/10.1016/j.respol.2012.04.014

Springer Nature, (2025, January 7). Springer Nature: Authors, editors and peer reviewers supported with launch of new AI tool [Press release]. https://group.springernature.com/gp/group/media/press-releases/ai-tool-to-help-streamline-integrity-and-ethics-checks/27730892

Spier, R. (2002). The history of the peer-review process. *Trends in Biotechnology*, *20*(8), 357-358. https://doi.org/10.1016/S0167-7799(02)01985-6

Sun, Z. (2025). Large language models in peer review: challenges and opportunities. *Scientometrics*, *130*, 5503–5546. https://doi.org/10.1007/s11192-025-05440-w

Teixeira da Silva, J.A. (2026). Paying peer reviewers: benefits, risks, and challenges. *Naunyn-Schmiedeberg's Archives of Pharmacology*, *399*, 9453–9470. https://doi.org/10.1007/s00210-025-04969-0

Thakkar, N., Yuksekgonul, M., Silberg, J., Garg, A., Peng, N., Sha, F., Yu, R., Vondrick, C., Zou, J. (2026). A large-scale randomized study of large language model feedback in peer review. *Nature Machine Intelligence*, *8*(3), 326-336. https://doi.org/10.1038/s42256-026-01188-x

27

Thelwall, M., Sud, P. (2022). Scopus 1900–2020: Growth in articles, abstracts, countries, fields, and journals. *Quantitative Science Studies*, *3*(1), 37-50. https://doi.org/10.1162/qss_a_00177

Torrielli, F., Locci, S., Rapp, A., Di Caro, L. (2026). Exploiting large language models in peer review: indirect prompt injection attacks and integrity probes. *Scientometrics*, *131*, 4889–4951. https://doi.org/10.1007/s11192-026-05695-x

Van Noorden, R. (2013). Open access: The true cost of science publishing. *Nature*, *495*, 426–429. https://doi.org/10.1038/495426a

van Rooyen, S., Black, N., Godlee, F. (1999). Development of the Review Quality Instrument (RQI) for assessing peer reviews of manuscripts. *Journal of Clinical Epidemiology*, *52*(7), 625–629. https://doi.org/10.1016/S0895-4356(99)00047-5

Vasu, S. S. M., Sheth, I., Wang, H. P., Binkyte, R., Fritz, M. (2026). Justice in judgment: Unveiling (hidden) bias in LLM-assisted peer reviews. In Findings of the Association for Computational Linguistics: ACL 2026, pp. 307-330. https://doi.org/10.18653/v1/2026.findings-acl.14

Veríssimo, D., Roberts, D. L. (2013). The academic welfare state: making peer-review count. *Trends in Ecology & Evolution*, *28*(11), 623–624. https://doi.org/10.1016/j.tree.2013.07.003

Walters, W. H., Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific reports*, *13*(1), 14045. https://doi.org/10.1038/s41598-023-41032-5

Ye, R., Pang, X., Chai, J., Chen, J., Yin, Z., Xiang, Z., Dong, X., Shao, J., Chen, S. (2024). Are we there yet? Revealing the risks of utilizing large language models in scholarly peer review. *arXiv preprint arXiv:2412.01708*

Zahavy, T. (2026). Position: LLMs can't jump. In Forty-third International Conference on Machine Learning Position Paper Track. https://openreview.net/attachment?id=klU4737opt&name=originally_submitted_PDF

28