Editorials

Opportunities and challenges of generative AI in the research lifecycle

Neda Sadeghi, PhD[1][a], Erin Nakamura, MPH[1], Luke Norman, PhD[1], Asim Okur, PhD[1], Jessica X. Ouyang, MD[1], Gustavo Sudre, PhD[2], Tonya White, MD, PhD[1]

[1] National Institute of Mental Health, Bethesda, MD, USA, [2] King's College London, London, UK

Keywords: artificial intelligence, generative AI, LLM, research lifecycle

https://doi.org/10.52294/001c.165447

Aperture Neuro

Vol. 6, Issue SI 2, 2026

Artificial intelligence (AI) is increasingly being explored and adopted across the research lifecycle, from idea generation and literature discovery to data analysis, manuscript preparation, and editorial and peer-review processes. This perspective provides an overview of AI's role across the stages of scientific research. We describe emerging tools and workflows, illustrating how AI can assist researchers by aggregating and synthesizing a large body of work across various domains, supporting methodological implementation, and facilitating communication and publication. We also discuss their shortcomings, including surface-level reasoning, the fabrication of plausible but incorrect outputs, and the challenges posed by the fact that researchers new to a field may not know which questions to ask or which nuances to interrogate. In addition, we discuss recent advances toward more agentic and end-to-end AI systems, highlighting both their technical feasibility and the challenges they pose for validation, oversight, and responsible use. For each stage of the research lifecycle, we outline key limitations of current AI systems and propose practical considerations, what researchers should and should not do to support rigorous and ethical integration of AI into scientific workflows. This integration requires coordinated frameworks across the ecosystem. Journals, funding agencies, universities, and policymakers play essential roles in defining standards for transparency and accountability, while individual researchers remain responsible for methodological rigor and validity of reported results.

INTRODUCTION

Artificial intelligence (AI) has become an integral component of scientific research, reshaping how ideas are generated, data are analyzed, and results are communicated. This shift is increasingly recognized by the scientific publishing community, with recent surveys suggesting that editors believe AI will significantly change medical publishing within the next decade.[1] AI broadly refers to computational systems capable of performing tasks that typically require human intelligence, such as learning and decision-making. Machine learning (ML), a subset of AI in which algorithms learn patterns from data, has long been used in neuroimaging research, particularly for tasks such as classification, prediction, clustering, and image analysis.[2,3] These methods have been widely used in neuroimaging for both common and rare disorders,[4,5] where ML-based predictive modeling and pattern recognition have been employed.

What has changed recently is the scale, accessibility, and scope of AI capabilities. The availability of large-scale datasets, increased computational resources, and development of deep network architecture capable of automated feature extraction, have all driven adoption of deep learning across scientific domains. More recently, generative AI has emerged as a class of deep learning models that learn underlying statistical structure of large datasets and are capable of generating new content such as text, images, audio, and video. These models such as large language models (LLMs) learn the statistical structure from massive datasets and can produce content that extend well beyond traditional predictive tasks.[6,7] The influence of these models now spans routine activities such as information retrieval and summarization, automated coding to advance scientific

a **Corresponding Author:**
Neda Sadeghi, PhD (nedasadeghi@icloud.com)
National Institute of Mental Health
Bethesda, MD, USA

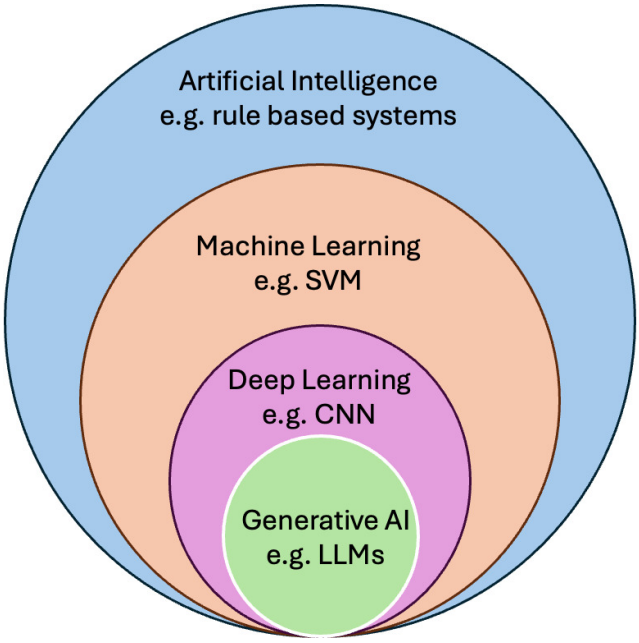

Figure 1. Relationship of Artificial intelligence, machine learning, deep learning, and generative AI. SVM: support vector machine. CNN: convolutional neural network. LLMs: large language models.

applications such as protein structure prediction[8] and the acceleration of drug discovery.[9] Figure 1 shows the relationship between artificial intelligence, machine learning, deep learning, and generative AI.

This article covers the expanding role of generative AI throughout the research lifecycle, from idea generation to publication. We describe the application of generative AI at each stage of scientific workflow, with examples from neuroimaging, highlighting some of the tools currently available to researchers, and discuss implications for transparency and research integrity, and offer guidelines for responsible implementation.

IDEA GENERATION

Transforming broad scientific interests into focused, testable hypotheses and fundable research proposals is often the most interesting, but also the most challenging stage of research. LLMs have increasingly been positioned as tools to facilitate this process, helping brainstorm research directions and identify potential gaps in the existing literature.

These tools can be useful in early exploratory work when used with understanding their capabilities and core limitations (Figure 2). When researchers struggle to articulate specific research questions within a general area of interest, prompting an LLM to generate multiple potential research questions can help clarify thinking and direction.

The AI-generated research questions themselves are rarely suitable for direct implementation. Instead, the value lies in a human evaluating each suggestion, which forces researchers to articulate why particular framings fail to capture their intended scientific goals. This process can help ascertain implicit assumptions and priorities.

A particularly valuable application involves using LLMs as adversarial tools. These tools respond not only to the question asked, but also to how these questions are framed. Researchers can prompt these systems to function as skeptical grant reviewers or critical colleagues, asking them to identify weaknesses in proposed study designs, challenge theoretical assumptions, or generate counterarguments to research hypotheses. For example, a prompt such as "Act as a skeptical reviewer for an NIH study section. What are the three most likely criticisms of the following specific aims?" can surface methodological concerns that might otherwise emerge only during formal review. The key principle is that adversarial prompting often yields more useful output than asking for affirmation or elaboration. However, these AI-generated critiques must themselves be evaluated for validity and relevance, as the model may raise concerns that reflect generic reviewer tropes rather than issues specific to the proposed work. Effective prompts specify the role (e.g. act as PI or skeptical reviewer), the task, relevant context, output format, constraints, and, when useful, examples or specific steps to take.

Major risks of AI-assisted idea generation emerge when researchers treat LLM outputs as authoritative. These systems are statistical models that generate outputs by predicting likely sequences of tokens based on patterns learned from large-scale training data and conditioned on the input context. In multimodal settings, non-text inputs such as images are first transformed into token-like representations and integrated into the same generative framework. As a result, their outputs reflect learned statistical associations rather than genuine understanding of scientific principles or experimental constraints. In scientific discovery tasks, AI systems have been shown to rely heavily on reproducing well-established models from their training data, rather than generating truly novel hypotheses.[10] Anecdotal reports from researchers using AI for experimental design suggest additional concerns. AI-suggested study designs can appear superficially plausible while containing unrecognized confounds, requiring datasets that do not exist or are not accessible, or unknowingly proposing approaches that have already been attempted and failed.

Crucially, general-purpose LLMs can be unreliable when identifying research gaps. They may hallucinate references to non-existent papers or, even when not hallucinating, simply lack coverage of relevant literature, particularly recent publications or work outside high-visibility journals. When an LLM states that "no research has examined" a specific relationship, this may reflect limitations in the training data rather than an actual gap in the field. Users can sometimes probe these limitations by asking the model to describe its knowledge cutoff or to express uncertainty, though such self-reports are not always reliable. At a broader level, the widespread use of the same LLMs for brainstorming may lead to the homogenization of research questions. If substantial numbers of researchers employ similar models trained on overlapping text, the field may experience convergence toward a narrower set of "obvious"

Key Issues & Considerations
Idea Generation • AI-generated ideas may converge on similar patterns, reducing novelty • LLM training data gaps can misidentify research gaps • AI outputs are starting points, not validated insights
Methods & Analysis • AI may generate plausible-looking but incorrect code • Small errors can propagate through pipelines
Literature Review • LLMs may hallucinate realistic but fake citations • Specialized tools (Elicit, Consensus) are grounded in actual databases and more reliable
Writing & Editing • Original meaning can be lost during editing • Ensure the use of AI is approved by your institution and potential journals
Journal Selection • AI journal selectors may be biased towards well-established journals. • These tools assess semantic similarity, not scholarly fit
Post-Submission • Uploading manuscripts to LLMs may violate confidentiality agreements • Some publishers prohibit reviewers from using generative AI • LLM-generated reviews may echo manuscript claims and raise superficial objections

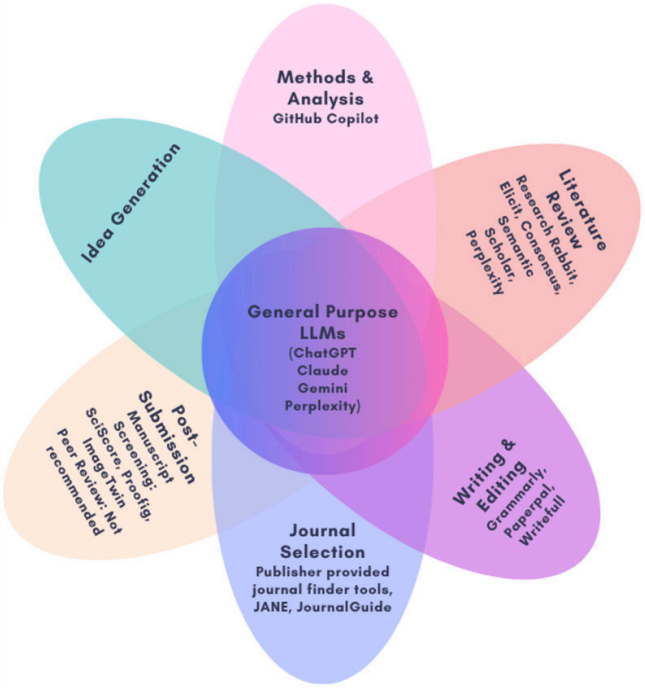

Figure 2. Left: Key Issues and Considerations. Right: Examples of commonly used AI platforms for different stages of the research cycle.

research questions. Whether this represents a genuine concern or simply reflects that certain questions are genuinely important remains an open empirical question. It is important that researchers considering AI tools for idea generation maintain clear boundaries around their use (Table 1).

METHODS & ANALYSES

Once a study or analysis has been designed, the next step is implementation. Neuroimaging studies typically employ multi-step preprocessing pipelines, statistical modeling scripts, quality-control procedures, and data-management workflows. Owing to advances in data analytic methods, data sharing, and the growing recognition, particularly in psychiatric neuroimaging, of the need for large samples to achieve desired statistical power, these pipelines and workflows increasingly involve applying sophisticated analytical approaches to large multi-site neuroimaging datasets.[11-14]

LLMs can assist with these technical tasks, opening these methods to researchers with limited coding experience and increasing efficiency for experienced programmers. For example, instead of searching package documentation, adapting code from tutorials, or combing through forum threads, neuroimaging researchers can now have a LLM create specific operations, such as constructing a design matrix for a specific contrast, or batch-renaming files to conform to BIDS formatting requirements. They receive ready-to-run Python, R, MATLAB, or bash code that performs their specific task.[15] In addition, LLMs can parse error messages and suggest troubleshooting steps in seconds. Once the processing and analysis pipelines have been successfully applied, researchers may further use these models to generate README files describing how to reproduce the analysis steps, or to draft manuscript methods sections directly from processing scripts and configuration files.

However, it is well known that LLMs can make mistakes or produce code hallucinations. Most concerning are situations in which code appears plausible and runs successfully but produces incorrect results.[16-18] For example, an LLM might generate a script that applies incorrect slice timing parameters or transformations. Such errors could propagate through subsequent processing and analysis steps, affecting outputs in ways that may be difficult to detect. In addition, copying and pasting logs, file paths, or metadata into LLMs can introduce data privacy risks, as these often contain identifiable information, such as medical record numbers, visit dates, or researcher usernames.[19,20]

In addition to coding, LLMs and related multimodal systems can be a useful tool for generating figures, particularly for conceptual schematics, graphical abstracts, or illustrative diagrams. They can speed up the design process by producing polished initial drafts. However, in our experience, the images often require changes due to inaccuracies, ambiguities, or inconsistencies in spatial and relational structure. In particular, iterative refinement can be challenging as targeted edits are not always localized and may alter other parts of the figure, making controlled modification time-consuming.

We therefore suggest that researchers work with LLMs interactively, in small steps, treating them as coding assistants, rather than fully automated generators of entire pipelines. As with all pipelines and code, LLM-generated scripts should be tested before deployment at scale, with validation of outputs from intermediate steps. Use of AI to

	Do's	Don'ts
Idea Generation	Evaluate each suggestion Use AI as an interactive brainstorming partner Explore possibilities Ask AI to act as critical reviewer	Treat AI-generated ideas as authoritative or complete Assume suggested gaps are real without verification Accept confident sounding outputs without skepticism
Methods & Analysis	Use AI interactively Test and validate code before deploying it at scale Ensure you understand the logic and assumptions Validate intermediate outputs	Copy and paste private, protected, or sensitive data Assume runnable code is correct Skip quality control
Literature Review	Use AI for broad scoping and initial orientation Identify themes, and related bodies of work Verify claims by reading original sources Check all citations for accuracy and relevance	Cite papers you haven't read Rely on AI to identify all relevant work Use AI-generated references without verification
Writing & Editing	Use AI to generate outlines Improve clarity and readability Review output and fact check Use AI for editing, rewording, and grammar support	Copy and paste confidential or unpublished material Accept polished language as evidence of correctness
Journal Selection	Use AI as a starting point to identify potential journals Use publisher's specific tools Combine AI suggestions with human judgement	Rely solely on AI for journal selection Ignore journal scope, audience, or AI-use policies Assume AI predictions of acceptance likelihood are reliable
Post submission	Use the available journal tool for compliance checks Follow journal guidance on responsible AI use	Upload unpublished manuscripts Use AI in ways that violate confidentiality Delegate scientific rebuttal to AI

Table 1. AI user guidelines

generate code should be disclosed at manuscript submission, as is already mandatory for many journals. Computational pipelines used to generate published findings must be understood and validated by the researchers who put their names to the results.

LITERATURE REVIEW

The use of LLM to assist with the literature review is extremely common, beginning during the conception of the study through to framing the results. Automated search and synthesis platforms such as Research Rabbit, Elicit, Consensus, and Semantic Scholar offer capabilities that extend beyond traditional database searches. These tools can cluster papers by topic, map co-citation networks, generate summary statistics across studies, and, in some cases, synthesize findings across multiple papers. For broad-scoping reviews or initial orientation in an unfamiliar research domain, these capabilities can provide an initial framework of the field.

Each platform offers distinct approaches to literature discovery. Research Rabbit employs citation-based mapping to visualize how papers connect through references and citations, allowing researchers to explore literature networks interactively.[17] Elicit uses semantic search and language models to extract structured information from papers and to organize findings into customizable tables that facilitate comparison across studies.[17] Consensus synthesizes findings from multiple papers to characterize the degree of agreement or disagreement on specific research questions, giving researchers a rough overview for evidence on a specific topic. Semantic Scholar provides AI-powered paper recommendations and automated summarization features across millions of papers. Although these tools differ in their specific implementations, they all make the literature review process more thorough through automated discovery and synthesis.

When these tools summarize papers, they typically exclude the methodological details needed to critically evaluate the findings. Many platforms will condense detailed Methods sections into brief paragraphs. These summaries typically do not fabricate information, but they are often incomplete in ways that are imperative for evaluating research quality. This is especially problematic in neuroimaging, where findings depend critically on methodological details, including sample size, scanner specifications, preprocessing pipelines, statistical thresholding procedures, and multiple-comparison corrections. An AI summary might report that "the study found increased prefrontal activation," but omit that this was based on an uncorrected $p < 0.01$ threshold, which provides minimal protection against false positives. For a field still grappling with replication challenges and concerns about methodological rigor, such omissions can mislead researchers about the strength of evidence. It is worth noting that targeted follow-up prompts requesting specific methodological details can sometimes elicit this information, although the burden of knowing what to ask still falls on the researcher.

An additional concern involves the bias of recommendation algorithms toward recent, highly cited publications. These systems tend to surface papers that are already well-connected in citation networks, potentially creating feedback loops in which highly cited papers receive additional citations.

These limitations become ethically concerning when researchers cite papers based solely on AI summaries, without reading the original work. AI summaries can legiti-

mately help researchers decide which papers warrant full reading, but using them as a substitute for carefully reading the papers represents a failure of scholarly responsibility. Including a citation implies that the work has been read with sufficient detail to understand its content. This responsibility cannot and must not be delegated to an AI system, regardless of how sophisticated the summarization algorithm is.

The implications for trainees warrant consideration, although they are beyond the scope of this paper to fully explore them. Students using these tools complete literature reviews more quickly, but it remains unclear whether they develop equivalent depth in critical appraisal skills. The metacognitive abilities involved in formulating effective search strategies, iterating on keyword combinations, recognizing when one has identified the boundaries of a topic, and evaluating study quality develop through effortful practice and experience. Learning to effectively prompt AI tools, evaluate their outputs critically, and integrate them into a rigorous workflow may constitute a new form of research skill development.

WRITING

AI is rapidly augmenting and changing the approaches to scientific writing. A Nature poll of 5,000 researchers found broad comfort with AI assistance for editing and translation, but less enthusiasm for drafting, with only 28% reporting having actually used AI to edit a paper and 8% for initial writing. Early-career researchers were more accepting of AI assistance and more likely to report using it. Across career stages, the majority agreed that use should be disclosed.[21]

Early efforts focused on domain-specific models trained exclusively on scientific text. Meta's Galactica (2022), for example, was trained on 48 million scientific papers, textbooks, and other academic sources,[22] but was withdrawn within days of its public release because it generated confident, well-formatted text containing fabricated citations and inaccurate claims. General-purpose instruction-tuned models such as ChatGPT and Claude have since proven more capable for scientific writing tasks, likely owing to their training on large-scale heterogeneous data. These models are now also being integrated into dedicated research environments. OpenAI's Prism, for instance, embeds GPT model within a LaTeX-based collaborative workspace for drafting, revision, and citation management.

AI is most clearly appropriate when it supports the form of scientific writing rather than the substance of scientific thought. For example, it can systematically check that in-text citations match the reference list, verify that the manuscript structure aligns with journal guidelines, and flag missing reporting elements that are increasingly required for transparency and reproducibility, such as dataset identifiers and software versions. It can also refine grammar, consistency, and clarity; these are all tasks where AI adds efficiency without substituting for scientific judgment.[23-28] Non-native English speakers in particular report meaningful benefit from AI editing tools.[23]

Current AI tools cannot be considered authors, as the Committee on Publication Ethics argued that they cannot take intellectual responsibility for claims or be held accountable for errors after publication.[29] Most journals post guidance on acceptable AI use and researchers should review this before proceeding and disclose any use in accordance with journal requirements.

JOURNAL SELECTION & SUBMISSION

Choosing the right journal affects both how widely a paper is read and whether it is accepted. Scope misalignment is one of the most common reasons for rejection, and for researchers without extensive publication experience, knowing which journals are appropriate for a given piece of work can be genuinely difficult.[27] AI matching tools can speed up the search by comparing manuscript text against previously published articles. These are a useful starting point, though each publisher's tool searches only within its own portfolio, and neutral alternatives such as the Journal/Author Name Estimator and JournalGuide lack published performance data.[27]

Researchers have naturally looked to AI to answer the question every author wants resolved before submission: will this paper be accepted at this journal? Using ChatGPT, Thelwall and Yaghi (2025) found correlations with actual review outcomes ranging from near zero to moderate (rho = 0.46) across three venues. Purpose-built deep learning systems have achieved higher accuracy (74-86%), although no comparable work exists for neuroimaging.[30] Performance is stronger for fields with clear publication norms than for interdisciplinary areas, and human judgment remains important for evaluating methodological fit and novelty, which current tools do not assess.[31] At present, AI may help researchers identify journals with the right scope, but whether it can predict the outcome of peer review remains an open question.

POST-SUBMISSION, PEER-REVIEW, PUBLICATION, DISSEMINATION

Following submission, the peer-review process is another stage in which AI can be applied. Some publishers, including major neuroimaging journals, prohibit reviewers from using generative AI entirely, not only to prevent the uploading of confidential content, but also because they consider AI assistance incompatible with the human judgment required for review.[32] Despite this, a recent survey of 1,645 reviewers across 111 countries found that more than half the reviewers reported using AI during peer review.[33] The use of AI during peer review will almost certainly increase, and this is not necessarily problematic if managed appropriately. Some publishers are building closed AI environments to address confidentiality concerns.[34] Tools designed to check statistical reporting or compliance with reporting guidelines may become commonplace.[35]

Indeed, some uses of AI during review already could be considered relatively uncontroversial. For example, a re-

viewer who needs a refresher on a particular technique, such as dynamic causal modeling or Independent Components Analysis (ICA), might consult an LLM to review the specifics before evaluating whether the authors applied it correctly. LLMs can produce reviews that follow the expected structure, but these have been found to echo what manuscripts claim about their own limitations or significance, or to raise objections that suggest a lack of understanding of the manuscript or the wider literature.[35-37] A reviewer who accepts such output uncritically may submit feedback that confuses authors and editors without doing anything that an editor could not have done themselves.

Furthermore, manuscript authors are already attempting to exploit weaknesses in AI models. For example, Lin (2025) identified 18 preprints on arXiv that contained hidden instructions inserted by study authors to manipulate an AI system.[38] These instructions were typically written in small or white font, making them difficult for humans to read, and ranged from blunt commands demanding positive reviews to detailed instructions specifying that weaknesses should be described as minor and easily fixable. This is an extreme example, but it demonstrates that the more reviewers rely on AI, the greater the incentive for authors to optimize their manuscripts for AI models rather than for human readers.

An area where AI shows considerable potential is in how scientific content can be adapted and disseminated across audiences and platforms. A research article can be readily converted into structured abstracts, plain-language summaries, graphical summaries, blog posts, and audio or visual content, making research more accessible and enabling findings to reach a broader range of audiences. As with other stages of the scientific process, human oversight remains essential to ensure that translated content accurately represents the original work and avoids misinterpretation and oversimplification.

CONCLUSIONS AND FUTURE DIRECTIONS

Looking ahead, AI is increasingly evolving from task-specific assistance toward agentic systems capable of coordinating activities across the full research lifecycle, from study design and analysis to evaluation and publication. When appropriately designed, such agentic systems may help scientists explore complex hypotheses and manage and find patterns in increasingly large and heterogeneous datasets. For example, human-in-the-loop multi-agent frameworks, such as AI Co-Scientist,[39] and domain-specific biomedical AI agents, such as Biomni,[40] support scientific discovery by generating and refining hypotheses and autonomously constructing analytical pipelines. Although these frameworks do not currently support neuroimaging tools, their open-source design and community-driven development make the inclusion of neuroimaging workflows likely. Such integration promises greater efficiency and reproducibility, but also amplifies existing risks. Errors can propagate further and faster, and the distance between researchers and their raw data grows.

Alongside these advances in AI systems, the infrastructure of scientific publishing is also evolving to support more interactive and machine-readable research. Emerging platforms such as NeuroLibre[41] highlight a shift toward more interactive and reproducible scientific publishing by integrating executable environments (e.g. Jupyter notebooks) within articles, enabling readers to explore code, data, and analyses dynamically. Enriching articles with structured metadata, linked datasets, and semantic context enables both human and AI agents to interact with and extend research outputs. Together, these developments point toward a transformation in scholarly communication where standardized, machine-readable formats allow AI systems to reproduce results, generate new visualizations, or conduct alternative analyses. This evolution could enable a more interconnected research ecosystem, where one article can programmatically access another, reuse its underlying components, and evaluate how variations in data or methodology influence outcomes, ultimately advancing transparency, reproducibility and cumulative knowledge generation.[42]

As the use of AI tools becomes increasingly common, deliberate strategies are vital to ensure that AI accelerates scientific discovery without undermining scientific integrity. We propose the following guidelines for scientists, publishers, universities, and funding agencies.

For scientists, our guidelines focus on individual responsibility. AI should be treated as an assistive or collaborative tool. Core scientific responsibilities such as study design, model selection, and interpretation of results remain the responsibility of the human researcher. AI outputs should supplement, not replace, scientific judgment, and must be validated throughout the research process. We define supplementation as cases when researchers can independently explain, justify, and reproduce the outputs; replacement occurs when outputs are accepted without such validation. The level of validation should scale with the potential impact of errors on study conclusions. For example, in our observations, AI-assisted coding led to inefficient trial-and-error cycles when outputs were used without understanding, whereas progress improved when code was systematically examined and validated. Maintaining domain expertise and direct engagement with the data remains essential.

These considerations have particularly important implications for the training of future scientists. As trainees enter research environments where agentic AI systems are readily available, there is a risk that automation may obscure foundational aspects of scientific reasoning, including critical evaluation of assumptions, assessing raw data, and methodological decision-making. Scientific training must therefore explicitly reinforce critical thinking skills and emphasize AI's role as a tool for feedback, exploration, and refinement, rather than as a substitute for intellectual responsibility. Training programs should include guidelines on validating AI-generated outputs, understanding model limitations, and recognizing risks such as hallucinations, overconfidence, and conceptual homogenization.

However, individual responsibility alone is not enough. Institutions, journals, and funding agencies need to build infrastructure that matches how researchers actually work. This includes clear policies on the disclosure of AI use, secure environments for handling confidential materials, and the integration of AI literacy into training programs for researchers at all career stages.

To ensure responsible adoption, guardrails and safeguards must be developed and embedded across educational, institutional, and publication contexts. Universities, journals, funding agencies, and other stakeholders should establish clear policies governing transparency, data privacy, confidentiality, and appropriate use of LLMs and agentic AI systems throughout the research and peer-review process. Where feasible, institutions need to support the development or use of secure, in-house AI systems or approved platforms that allow researchers to leverage AI tools while protecting sensitive data and unpublished materials. Ultimately, responsible adoption will depend not only on policy enforcement but also on education, infrastructure, and shared norms within the scientific community.

In addition, the increasing availability of end-to-end and agentic AI systems raises important questions about how scientific incentives are structured. As tools such as Denario[43] and AI Scientist[44] demonstrate the technical feasibility of rapidly generating manuscripts that can progress through submission and simulated peer review, traditional metrics focused on publication counts and citations may become less meaningful. Incentive structures may need to evolve to place greater value on rigor, reproducibility, creativity, and having meaningful conceptual advances rather than sheer output. Reconsidering how scientific contributions are evaluated will be essential to ensure that AI-enabled efficiency supports substantive progress rather than accelerating the production of low-impact or redundant research.

Finally, AI technologies continue to evolve rapidly. The limitations we describe today will shift as models improve, and new challenges will inevitably emerge. As a result, recommendations and policies will need to be revisited. Institutions should invest in infrastructure, along with a flexible governance structure and ongoing evaluation mechanisms, to remain current. Ensuring responsible AI integration is not a one-time effort, but an ongoing commitment.

Together, these efforts can help ensure that AI strengthens scientific discovery by amplifying human insight rather than diminishing the critical thinking, creativity, and rigor that remain central to scientific discovery.

AUTHOR CONTRIBUTIONS

N.S. conceptualized the paper, coordinated contributions, and led drafting and revisions, including writing specific sections of the manuscript. Co-authors contributed specific sections and/or provided critical feedback and revisions. Authors are listed alphabetically after the first author.

ACKNOWLEDGMENTS

We would like to thank NIH librarian, Ms. Alicia-Marie Lillich, for her help with the literature review. Professor Sudre was supported by the Rosetrees Trust and the Pears Foundation through the Rosetrees Pears Chair of Bioinformatics. This research was supported [in part] by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH author(s) are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Submitted: January 30, 2026 CDT. Revised: May 05, 2026 CDT. Accepted: May 25, 2026 CDT. Published: September 04, 2026 CDT.

REFERENCES

1. Biondi-Zoccai G et al. Perspectives on Artificial Intelligence in Medical Publishing: A Survey of Medical Journal Editors. *J Cardiovasc Pharmacol*. 2025;86:374-383. doi:10.1097/FJC.0000000000001738

2. Mitchell TM. *Machine Learning*. McGraw-Hill; 1997.

3. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer; 2009.

4. Orru G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews*. 2012;36:1140-1152. doi:10.1016/j.neubiorev.2012.01.004

5. Sadeghi N et al. Brain phenotyping in Moebius syndrome and other congenital facial weakness disorders by diffusion MRI morphometry. *Brain communications*. 2020;2:fcaa014. doi:10.1093/braincomms/fcaa014. PMID:32328577

6. Vaswani A et al. Attention is all you need. *Advances in neural information processing systems*. 2017;30.

7. Brown T et al. Language models are few-shot learners. *Advances in neural information processing systems*. 2020;33:1877-1901.

8. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. *nature*. 2021;596:583-589. doi:10.1038/s41586-021-03819-2. PMID:34265844

9. Ren F et al. AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. *Chemical science*. 2023;14:1443-1452. doi:10.1039/D2SC05709C. PMID:36794205

10. Ding AW, Li S. Generative AI lacks the human creativity to achieve scientific discovery from scratch. *Sci Rep*. 2025;15:9587. doi:10.1038/s41598-025-93794-9

11. Marek S et al. Reproducible brain-wide association studies require thousands of individuals. *Nature*. 2022;603:654-660. doi:10.1038/s41586-022-04492-9. PMID:35296861

12. Norman LJ et al. Cross-sectional mega-analysis of resting-state alterations associated with autism and attention-deficit/hyperactivity disorder in children and adolescents. *Nature Mental Health*. Published online 2025:1-15. doi:10.1038/s44220-025-00431-5. PMID:41446723

13. Koike S, Tanaka SC, Hayashi T. Beyond case-control study in neuroimaging for psychiatric disorders: Harmonizing and utilizing the brain images from multiple sites. *Neuroscience & Biobehavioral Reviews*. Published online 2025:106063. doi:10.1016/j.neubiorev.2025.106063

14. Sudre G et al. Estimating the heritability of developmental change in neural connectivity, and its association with changing symptoms of attention-deficit/hyperactivity disorder. *Biological psychiatry*. 2021;89:443-450. doi:10.1016/j.biopsych.2020.06.007. PMID:32800380

15. Huynh N, Lin B. Large Language Models for Code Generation: A Comprehensive Survey of Challenges, Techniques, Evaluation, and Applications. *arXiv preprint arXiv:2503.01245*. Published online 2025.

16. Liu J, Xia CS, Wang Y, Zhang L. Is your code generated by chatgpt really correct? rigorous evaluation of large language models for code generation. *Advances in Neural Information Processing Systems*. 2023;36:21558-21572. doi:10.52202/075280-0943

17. Agarwal V, Pei Y, Alamir S, Liu X. Codemirage: Hallucinations in code generated by large language models. *arXiv preprint arXiv:2408.08333*. Published online 2024.

18. Lee Y et al. Hallucination by Code Generation LLMs: Taxonomy, Benchmarks, Mitigation, and Challenges. *arXiv preprint arXiv:2504.20799*. Published online 2025.

19. Rempe M, Heine L, Seibold C, Hörst F, Kleesiek J. De-identification of medical imaging data: a comprehensive tool for ensuring patient privacy. *European radiology*. Published online 2025:1-10. doi:10.1007/s00330-025-11695-x. PMID:40481871

20. Tian S et al. Opportunities and challenges for ChatGPT and large language models in biomedicine and health. *Briefings in Bioinformatics*. 2023;25. doi:10.1093/bib/bbad493. PMID:38168838

21. Kwon D. Scientists Split on Ethics of AI Use. *Nature*. 2025;641:574-578. doi:10.1038/d41586-025-01463-8

22. Taylor R et al. Galactica: A large language model for science. *arXiv preprint arXiv:2211.09085*. Published online 2022.

23. Giglio AD, Costa M. The use of artificial intelligence to improve the scientific writing of non-native english speakers. *Rev Assoc Med Bras (1992)*. 2023;69:e20230560. doi:10.1590/1806-9282.20230560

24. Alyasiri OM, Salman AM, Akhtom D, Salisu S. ChatGPT revisited: Using ChatGPT-4 for finding references and editing language in medical scientific articles. *J Stomatol Oral Maxillofac Surg*. 2024;125:101842. doi:10.1016/j.jormas.2024.101842

25. Gibney E. What are the best AI tools for research? Nature's guide. *Nature*. Published online 2025. https://www.nature.com/articles/d41586-025-00437-0

26. Lingard L. Writing with ChatGPT: An Illustration of its Capacity, Limitations & Implications for Academic Writers. *Perspect Med Educ*. 2023;12:261-270. doi:10.5334/pme.1072

27. Kousha K, Thelwall M. Artificial Intelligence to support publishing and peer review: A summary and review. *Learned Publishing*. 2023;37:4-12. doi:10.1002/leap.1570

28. Jacques T, Sleiman R, Diaz MI, Dartus J. Artificial intelligence: Emergence and possible fraudulent use in medical publishing. *Orthop Traumatol Surg Res*. 2023;109:103709. doi:10.1016/j.otsr.2023.103709

29. COPE Council. COPE position - Authorship and AI - English.

30. Thelwall M, Yaghi A. Evaluating the predictive capacity of ChatGPT for academic peer review outcomes across multiple platforms. *Scientometrics*. 2025;130. doi:10.1007/s11192-025-05287-1

31. Farber S. Enhancing Academic Decision-Making: A Pilot Study of AI-Supported Journal Selection in Higher Education. *Innovative Higher Education*. 2025;50:1813-1831. doi:10.1007/s10755-025-09791-3

32. Li Z et al. Use of artificial intelligence in peer review among top 100 medical journals. *JAMA Netw Open*. 2024;7:e2448609. doi:10.1001/jamanetworkopen.2024.48609. PMID:39625725

33. Frontiers Media. *Unlocking AI's Untapped Potential: Responsible Innovation in Research and Publishing.*; 2025.

34. Naddaf M. More than half of researchers now use AI for peer review — often against guidance. *Nature*. 2026;649. doi:10.1038/d41586-025-04066-5

35. Perlis RH et al. Artificial intelligence in peer review. *JAMA*. 2025;334. doi:10.1001/jama.2025.15827

36. Ye R et al. Are we there yet? revealing the risks of utilizing large language models in scholarly peer review. *arXiv preprint arXiv:2412.01708*. Published online 2024.

37. Zhu L et al. Evaluating the potential risks of employing large language models in peer review. *Clinical and Translational Discovery*. 2025;5:e70067. doi:10.1002/ctd2.70067

38. Lin Z. Hidden prompts in manuscripts exploit ai-assisted peer review. *arXiv preprint arXiv:2507.06185*. Published online 2025. doi:10.31234/osf.io/nea5u_v2

39. Gottweis J et al. Towards an AI co-scientist. *arXiv preprint arXiv:2502.18864*. Published online 2025.

40. Huang K et al. Biomni: A general-purpose biomedical ai agent. *biorxiv*. Published online 2025. doi:10.1101/2025.05.30.656746. PMID:40501924

41. Karakuzu A et al. NeuroLibre: A preprint server for full-fledged reproducible neuroscience. Published online 2022. doi:10.31219/osf.io/h89js

42. Karakuzu A. Toward a woven literature: Open-source infrastructure for networked scientific publishing. *Aperture Neuro*. 2025;5. doi:10.52294/001c.146072

43. Villaescusa-Navarro F et al. The Denario project: Deep knowledge AI agents for scientific discovery. *arXiv preprint arXiv:2510.26887*. Published online 2025.

44. Lu C et al. The ai scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint arXiv:2408.06292*. Published online 2024.