

The future of reviews writing in the AI era

Zhiling Zheng

 Check for updates

Agentic workflows powered by large language models are beginning to assist chemists in literature search, summarization, and outline drafting. Though they remain unable to replace expert insight, these systems promise to reshape how reviews are prepared – shifting the human role from exhaustive curator to creative synthesizer, empowered by intelligent, always-on review-copilots.

Large language models (LLMs) are getting better at more than just answering questions¹. They are starting to help chemists find papers^{2,3}, summarize literature^{4,5}, and even draft outlines of review articles^{6,7}. Indeed, it is tempting to imagine a future where the hard part of writing a review – the labour-intensive sifting, sorting, and summarizing – is offloaded to machines.

But if review articles were just summaries, we would not value them as much as we do. A good review is an interpretation, not just a report. It finds patterns, raises new questions, teaches the reader something new, and suggests where a field might go next, and it is these qualities that make the best reviews last for years, even decades, before becoming outdated. Clearly, today's LLMs can map the territory, but they may not yet be able to chart new paths forward. What they can do, though, is reshape how we get there.

Augmentation rather than replacement

Right now, writing a review is a mostly manual, linear process: search the literature, read papers, take notes, build an outline, draft sections, revise, submit. It is slow and labour-intensive. Worse, we are entering an era when many times more new papers are published every day than two decades ago, and even the best human review teams struggle to be exhaustive or timely. This is where agentic systems or workflows with artificial intelligence (AI) models taking action autonomously based on goals could change the game. In a recent *Thesis* article in *Nature Chemistry*, Bruce Gibb envisioned a hypothetical AI assistant, 'Eric', that could free researchers from paperwork and mundane tasks while also serving as intellectual muses⁸. Likewise, the most valuable AI-augmented reviews will not only digest the known literature at a scale surpassing what humans can do, but also highlight unexpected connections across disciplines, bring attention to contradictions that might otherwise be overlooked, and inspire human authors with new ideas.

The step forward is not simply replacing human effort with automation but reimagining how researchers approach their work. Instead of reading every paper manually, a scientist might guide an AI to scout the field: identify key papers, group them into topics, extract trends, and highlight contradictions. Instead of drafting an article from scratch, one could receive an auto-generated outline, key

papers selected for human approval, and a first-pass narrative ready to be sculpted (Fig. 1). Going beyond just natural language processing (NLP), today's frontier agentic systems like Deep Research powered by OpenAI advanced reasoning LLMs and Internet search modules are already exploring this territory, but they act more as tireless librarians than as philosophers. As yet, they are unable to produce human-level, camera-ready reviews, but they hint at what might come. Well-crafted reviews do more than compile data⁹, but human creativity and productivity can be enhanced when the prior preparation of relevant papers becomes more comprehensive and faster.

Degrees of autonomy

To understand where we are heading, it helps to think in stages (Fig. 1). At level 0, humans do everything. At level 1, AI suggests papers based on keywords. At level 2, it starts summarizing sections of papers. At level 3, it drafts large blocks of text under close human supervision. Level 4 would mean an AI writing an entire review with only strategic human steering. Level 5 would mean fully autonomous review writing along with new insights and fresh hypotheses, and remains far beyond today's frontier. Most systems today operate somewhere between levels 1 and 3. They can find papers, summarize key ideas, suggest outlines; they might even write a first draft, but they still need heavy human prompting, steering, and checking.

Will highly intelligent AI reviewers at levels 4 and 5 eventually eclipse human creativity? This seems unlikely. For labour-intensive comprehensive surveys, automation will increasingly handle the heavy lifting. But for truly insightful perspective pieces, the future points toward collaboration rather than replacement. Human intuition, which is still irreplaceable and unique at this moment, will set the aesthetic and conceptual direction. If anything, the rise of agentic AI systems will make good human reviewers even more important. Those less appealing surveys that simply list papers without insight will increasingly be deprioritized as it becomes easier to do so with AI. But the best reviews – the ones that organize chaos into clarity – will become even more valuable. Researchers who know how to work with AI, using it, for example, to gather, filter, and organize information, while adding their own insight and creativity, will be able to write better, faster, and deeper reviews than ever before.

Current agentic AI systems face several significant limitations in producing high-quality reviews. While the 'hallucination problem', where models seem to make up facts, functions, or citations, can be alleviated through various strategies, challenges exist for them to cross the level 3 to 4 chasm from just survey reports to valuable reviews and perspectives. Even with web search capabilities or retrieval-augmented generation (RAG) using local downloaded literature databases, they often struggle to access the latest non-open-access articles updated daily, posters and oral presentations from conferences, or complex raw data in repositories. They may also have difficulty distinguishing between groundbreaking papers and less significant publications, sometimes overgeneralizing claims by extending the conclusions of the source material to a broader context than is supported, being misled by papers with incorrect data and false claims,

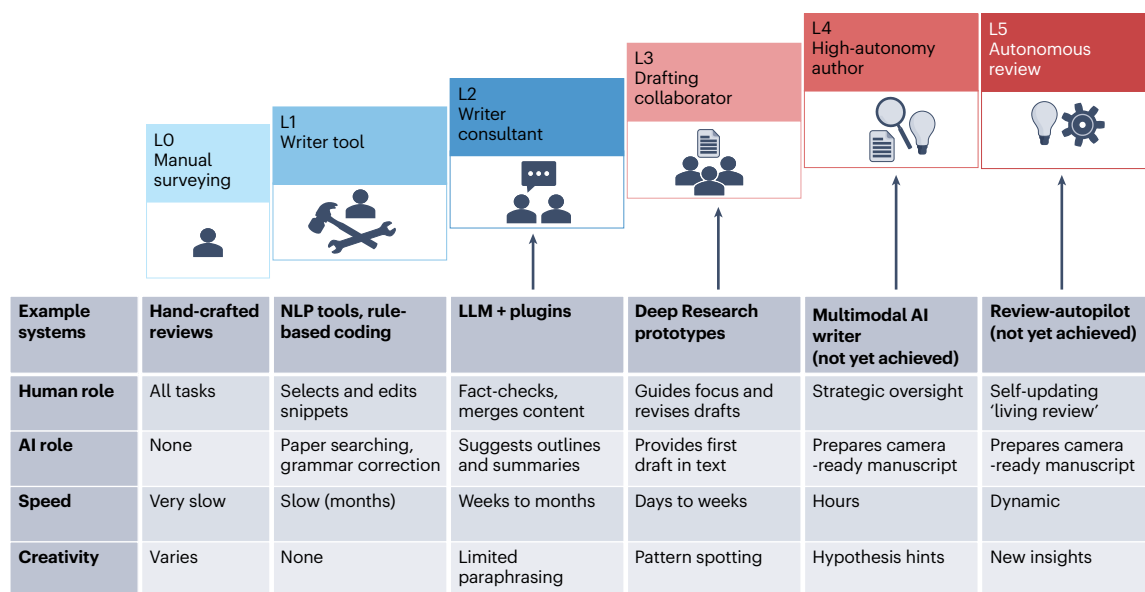


Fig. 1 | Ladder of autonomy in AI-assisted review writing. Current systems operate mainly at levels 1–3 (L0–L3), supporting literature search, clustering, summarization, and first-pass drafting. Levels 4 (L4) and 5 (L5), which involve

writing human-expert-level manuscripts and, more importantly, offering original hypotheses and insights, remain unachieved.

or occasionally misinterpreting or omitting details from figures or complex illustrations and conflating similar compound names.

Outlook

Realizing this potential will require careful stewardship as agentic AI systems grow in sophistication. It is essential that these systems be imbued with strong norms around transparency, reproducibility, and alignment with human values. Achieving true ‘review-autopilots’ that dynamically update themselves and autonomously combine information into novel insights is a daunting challenge that will require major innovations in AI architectures, knowledge management, and computational infrastructure. But even in the near term, with more and more researchers begin to embrace them at least as a tool in daily research¹⁰, AI can meaningfully enhance the review-writing process by serving as a tireless research assistant, able to rapidly map the contours of a field and flag contradictory or anomalous results for human consideration.

We are on the cusp of a new era in scholarly communication. Still, the direction is clear. As AI gets better at gathering and organizing information, researchers will spend less time on clerical work and more time on creative thinking. Writing a review could become less about wrestling with information overload, and more about shaping insights. In other words, this transformation will not make human expertise obsolete. Rather, it will recast the role of the human reviewer from one of laborious curation to creative synthesis. In the end, the most important reviews of tomorrow will still carry the fingerprint of human curiosity, intuition, and vision. They will just be written with more help – and perhaps, a lot more speed and new ideas.

Zhiling Zheng 

Department of Chemistry, Washington University, St. Louis, MO, USA.

✉ e-mail: z.z@wustl.edu

Published online: 01 July 2025

References

1. Caldas Ramos, M., Collison, C. J. & White, A. D. A review of large language models and autonomous agents in chemistry. *Chem. Sci.* **16**, 2514–2572 (2025).
2. Boiko, D. A., MacKnight, R., Kline, B. & Gomes, G. Autonomous chemical research with large language models. *Nature* **624**, 570–578 (2023).
3. Lala, J. et al. PaperQA: Retrieval-augmented generative agent for scientific research. Preprint at <https://arxiv.org/abs/2312.07559> (2023).
4. Bran, A. M. et al. Augmenting large language models with chemistry tools. *Nat. Mach. Intell.* **6**, 525–535 (2024).
5. Zheng, Z. et al. Large language models for reticular chemistry. *Nat. Rev. Mater.* **10**, 369–381 (2025).
6. Huang, J. & Tan, M. The role of ChatGPT in scientific communication: writing better scientific review articles. *Am. J. Cancer Res.* **13**, 1148–1154 (2023).
7. Shao, Y. et al. Assisting in writing Wikipedia-like articles from scratch with large language models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1: Long Papers, 6252–6278 (Association for Computational Linguistics, 2024).
8. Gibb, B. C. Eric the AI research assistant for chemists. *Nat. Chem.* **17**, 157–158 (2025).
9. How to write an excellent Review article. *Nat. Rev. Bioeng.* **2**, 907–907 (2024).
10. Mishra, T. et al. Use of large language models as artificial intelligence tools in academic research and publishing among global clinical researchers. *Sci. Rep.* **14**, 31672 (2024).

Acknowledgements

Z.Z. extends special gratitude to J. Weng from OpenAI and K. Dong from Stanford University for valuable discussions.

Competing interests

The author declares no competing interests.