RESEARCH · RESEARCH INTEGRITY

Responsible Use of AI Tools in Research Workflows

How generative AI and machine learning tools can be used responsibly at each stage of research, the main risks such as fabricated citations, hallucinated content, data leakage and synthetic respondents, and practical rules for disclosure, verification and accountability.

ModQu research article · 1,048 words · 11 references (APA 7th edition)

Artificial intelligence tools, especially large language models, are now part of many research workflows. They can summarise literature, draft text, suggest code, classify open-ended answers and translate questionnaires. Used well, they save time. Used carelessly, they introduce errors that are hard to see, because the output is fluent and confident whether or not it is correct. This article sets out where AI tools can help, where they create risks and what responsible use looks like.

Where AI tools can help

Experimental evidence suggests real productivity gains for some tasks. Noy and Zhang (2023) gave professionals writing tasks and found that those with access to a generative AI tool completed them faster and produced work rated as higher quality on average. In research, similar gains are possible in drafting, editing, coding routine analysis steps, and organising large volumes of text.

The table below summarises common uses across the research workflow, with their main risks and safeguards.

AI tools across the research workflow

A helpful use, the main risk and a safeguard at each stage

Stage	Helpful use	Risk / safeguard
Literature	Summaries, search terms	Invented references / check each source
Design	Draft and test question wording	Simulated answers are not data / pilot with people
Data collection	Translation, transcription	Confidentiality / approved tools only
Analysis	Code suggestions, text coding	Leakage, silent errors / test and validate
Writing	Editing and drafting	Unsupported claims / author verifies all

Drawing on Walters & Wilder (2023), Kapoor & Narayanan (2023), Bisbee et al. (2024) ModQu.com

Figure 1. Uses, risks and safeguards for AI tools at each stage of research. Table by ModQu, drawing on Walters and Wilder (2023), Kapoor and Narayanan (2023) and Bisbee et al. (2024).

The main risks

Fabricated and inaccurate content

Language models generate text that is plausible, not text that is verified. Ji et al. (2023) surveyed research on **hallucination**, the generation of content that is fluent but unsupported or false, and found it to be a pervasive problem across tasks. In research the most visible example is fabricated citations. Walters and Wilder (2023) asked language models to produce literature reviews with references and found that a large share of the citations did not exist, and that many real citations contained errors. A reference produced by an AI tool must therefore be checked against the source itself before it is used.

False confidence

Messeri and Crockett (2024) warned that AI tools can create **illusions of understanding** in research: researchers may believe they understand more than they do because a tool produced an explanation, may explore a narrower range of questions because tools favour common approaches, and may mistake a tool's output for an objective view. Bender et al. (2021) argued more generally that language models reproduce patterns in their training data, including its biases, without any grounding in meaning.

Errors in machine learning analysis

When machine learning is used for analysis, subtle mistakes can make results look far better than they are. Kapoor and Narayanan (2023) documented **data leakage**, in which information from the test data seeps into model training, across hundreds of published studies in many fields, and showed that it led to inflated claims of predictive accuracy.

Synthetic respondents

Some researchers have explored using language models to simulate survey respondents. Argyle et al. (2023) found that, when prompted with demographic backgrounds, a model could reproduce some patterns in real survey responses. Bisbee et al. (2024), however, found that synthetic responses were less varied than real ones, sensitive to small changes in prompts and unstable over time, and concluded that they were not a reliable replacement for real data. Simulated answers may be useful for testing questionnaire wording, but they are not evidence about what real people think.

Confidentiality

Entering unpublished data, interview transcripts or client material into an external AI service may breach confidentiality agreements, consent terms or data protection law, depending on how the service stores and uses inputs.

Bias and uneven quality

AI tools perform unevenly across languages, groups and topics. A tool that summarises English-language research well may miss or misrepresent work published in other languages. Automated coding of open-ended answers may classify some groups' responses less accurately than others, for example when people use dialect, slang or a second language. Because these errors are systematic rather than random, they can bias results rather than simply adding noise. Checking a sample of automated classifications against human coding, and comparing accuracy across groups, is a basic safeguard.

Loss of skills and judgement

There is also a longer-term risk. Researchers who rely on tools to summarise papers, write code or draft arguments may have fewer opportunities to develop the skills those tasks require. The ability to judge whether an output is right depends on understanding how it should have been produced. Training in research methods therefore remains essential, even as tools take over routine tasks.

Responsible use

Publishers, funders and research organisations have converged on several principles.

Humans remain accountable. AI tools cannot be authors, because they cannot take responsibility for the work. The Committee on Publication Ethics (2023) and leading journals have stated this explicitly (Thorp, 2023). The researcher is responsible for every claim, number and reference, whatever tool helped produce it.

Disclose use. Researchers should state which tools were used, for what purpose and how outputs were checked. The European Commission (2024) guidelines on generative AI in research recommend transparency about substantial use.

Verify everything. Check every citation against the original source, every fact against a reliable reference, and every piece of code by testing it. Treat AI output as a draft from an unreliable assistant, not as a finished product.

Protect data. Use only tools whose data handling meets the study's consent, contractual and legal obligations. Do not upload personal or confidential data to services that may retain it.

Keep a record. Note prompts, tool versions and dates where outputs feed into results, so that the work can be reviewed and, as far as possible, repeated.

Do not substitute for real evidence. Use AI to help collect, organise and analyse evidence, not to generate it.

Applying the rules in commercial research

In market and business research, the same principles apply with particular force, because clients rely on findings to make decisions. A research team should be able to tell a client which parts of a report were produced or assisted by AI tools, how they were checked, and that no client or respondent data were exposed in the process.

Conclusion

AI tools can make research faster and, in some tasks, better. Their central weakness is that they produce confident output without regard to truth. Responsible use keeps humans accountable, discloses how tools were used, verifies every output and protects participants' data. Under those conditions, AI can support research without undermining the trust on which it depends.

References

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, *31*(3), 337–351. https://doi.org/10.1017/pan.2023.2

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Bisbee, J., Clinton, J. D., Dorff, C., Kenkel, B., & Larson, J. M. (2024). Synthetic replacements for human survey data? The perils of large language models. *Political Analysis*, *32*(4), 401–416. https://doi.org/10.1017/pan.2024.5

Committee on Publication Ethics. (2023). *Authorship and AI tools* [Position statement]. https://publicationethics.org/cope-position-statements/ai-author

European Commission. (2024). *Living guidelines on the responsible use of generative AI in research*. Publications Office of the European Union.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), Article 248. https://doi.org/10.1145/3571730

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns, 4*(9), Article 100804. https://doi.org/10.1016/j.patter.2023.100804

Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, *627*(8002), 49–58. https://doi.org/10.1038/s41586-024-07146-0

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science, 379*(6630), 313. https://doi.org/10.1126/science.adg7879

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