RESEARCH · COMMUNICATING RESEARCH

The Future of Research: Automation, AI and the Researcher's Role

What the evidence shows about how artificial intelligence is changing research work, where it helps, where it fails, and which parts of the researcher's role remain essential.

ModQu research article · 1,249 words · 12 references (APA 7th edition)

Automation has been part of research for decades. Statistical software replaced hand calculation, online panels replaced much face-to-face interviewing, and reference managers replaced card indexes. Generative artificial intelligence, especially large language models (LLMs), extends automation to tasks that once required a trained person: summarising literature, drafting questionnaires, coding open-ended responses and writing reports. This article reviews what is known so far about the effects of these tools on research work and on the role of the researcher.

What AI can already do in research

In some fields, AI has produced results that would have been very hard to reach otherwise. AlphaFold predicted the three-dimensional structures of proteins with accuracy comparable to experimental methods in many cases (Jumper et al., 2021). Wang et al. (2023) review uses of AI across the sciences, including generating hypotheses, designing experiments and analysing large datasets, and note that these uses also bring problems of data quality, interpretability and reproducibility.

In everyday research work, the main change is in writing and analysis tasks. A survey of more than 1,600 researchers by *Nature* found that many expected AI tools to become very important to their field, while also reporting concerns about errors, bias and the spread of misinformation (Van Noorden & Perkel, 2023).

The scale of potential change is large. Eloundou et al. (2024) estimated that around 80% of the U.S. workforce could have at least 10% of their work tasks affected by LLMs, and about 19% could have at least half of their tasks affected. Occupations involving writing, analysis and information processing, which describe much of research work, were among the most exposed.

Evidence on productivity

Several controlled studies have measured the effect of generative AI on professional work similar to research tasks.

- Noy and Zhang (2023) gave 453 college-educated professionals writing tasks typical of their jobs. Those with access to ChatGPT took 40% less time, and the quality of their work, as rated by evaluators, rose by 18%.

- Brynjolfsson et al. (2025) studied customer support agents using an AI assistant. Productivity, measured as issues resolved per hour, rose by about 14% on average and by about 34% for novice and less-skilled workers, with little effect for the most experienced.
- Dell'Acqua et al. (2023) ran a field experiment with 758 consultants at Boston Consulting Group. For tasks within the capabilities of the AI, consultants using it completed 12.2% more tasks, 25.1% faster, with more than 40% higher quality. For a task deliberately chosen to lie outside those capabilities, consultants using AI were 19 percentage points less likely to produce a correct answer than those working without it.

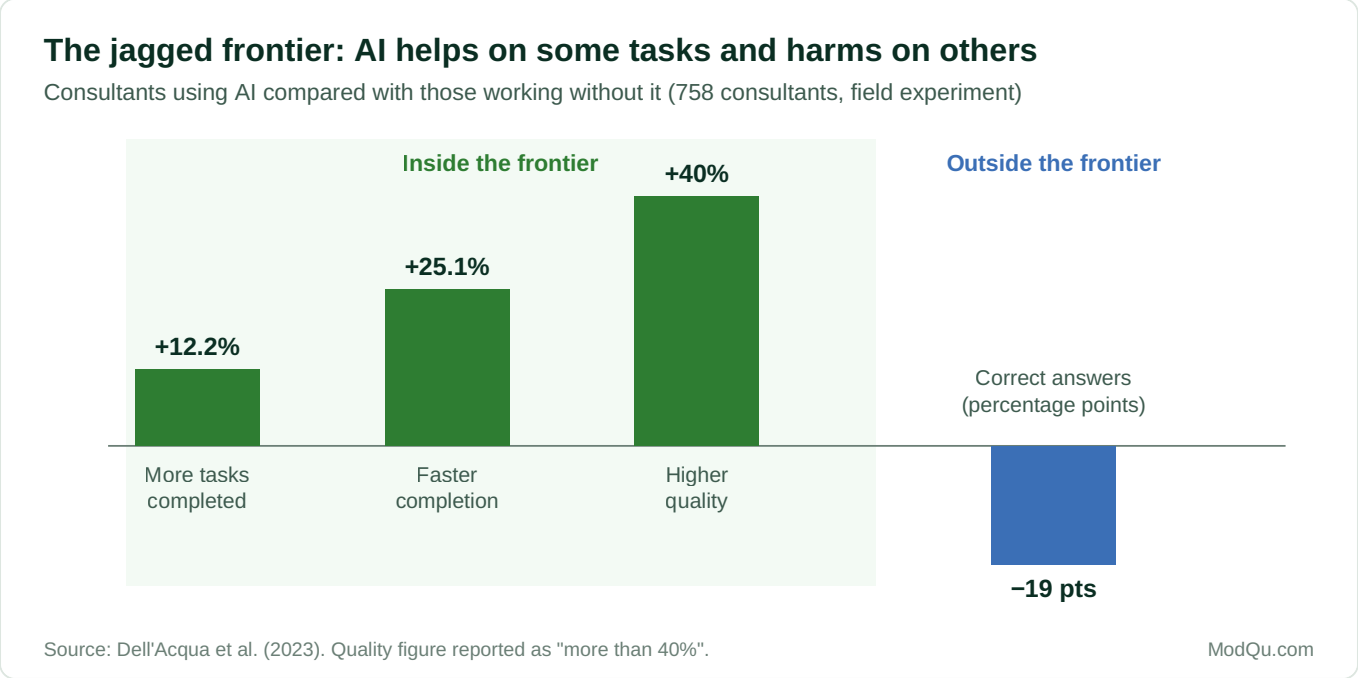


Figure 1. Effects of AI assistance on consultants' work inside and outside the AI's capabilities. Source: Dell'Acqua et al. (2023). Chart by ModQu.

The authors called this pattern a "jagged technological frontier": AI performs well on some tasks and poorly on others of apparently similar difficulty, and it is not always obvious in advance which is which. For researchers, the practical lesson is that gains depend on knowing where the tool is reliable and checking its output where it is not.

Where AI fails

Fabricated sources. LLMs can produce references that look plausible but do not exist. Walters and Wilder (2023) asked ChatGPT to write short literature reviews and checked 636 of the citations it produced. With GPT-3.5, 55% of the cited works were fabricated; with GPT-4, 18% were. Many of the genuine citations also contained errors. Any reference produced by an AI tool needs to be checked against the original source.

Synthetic respondents. Some researchers have proposed using LLMs to simulate survey respondents. Argyle et al. (2023) found that, when prompted with detailed demographic backgrounds, a language model could reproduce some patterns of opinion in particular U.S. population groups. Bisbee et al. (2024) tested the idea more critically. They found

that synthetic responses had much less variation than real survey data, that statistical relationships estimated from them often differed from those in real data, and that results changed when the same prompts were run at different times as the underlying model was updated. Synthetic data may help with testing questionnaires, but it cannot replace real respondents for estimating what people think.

Illusions of understanding. Messeri and Crockett (2024) argue that AI tools can give researchers the impression that they understand more than they do, and that heavy reliance on the same tools may narrow the range of questions and methods used in a field, which they describe as a "monoculture" of knowing. Their concern is that science could produce more while understanding less.

The researcher's changing role

The evidence so far suggests that AI changes the distribution of research work more than it removes the need for researchers. Tasks that are repetitive, well defined and easy to check, such as first drafts, transcription, translation and initial coding, are increasingly automated. Tasks that depend on judgement, context and accountability remain with people.

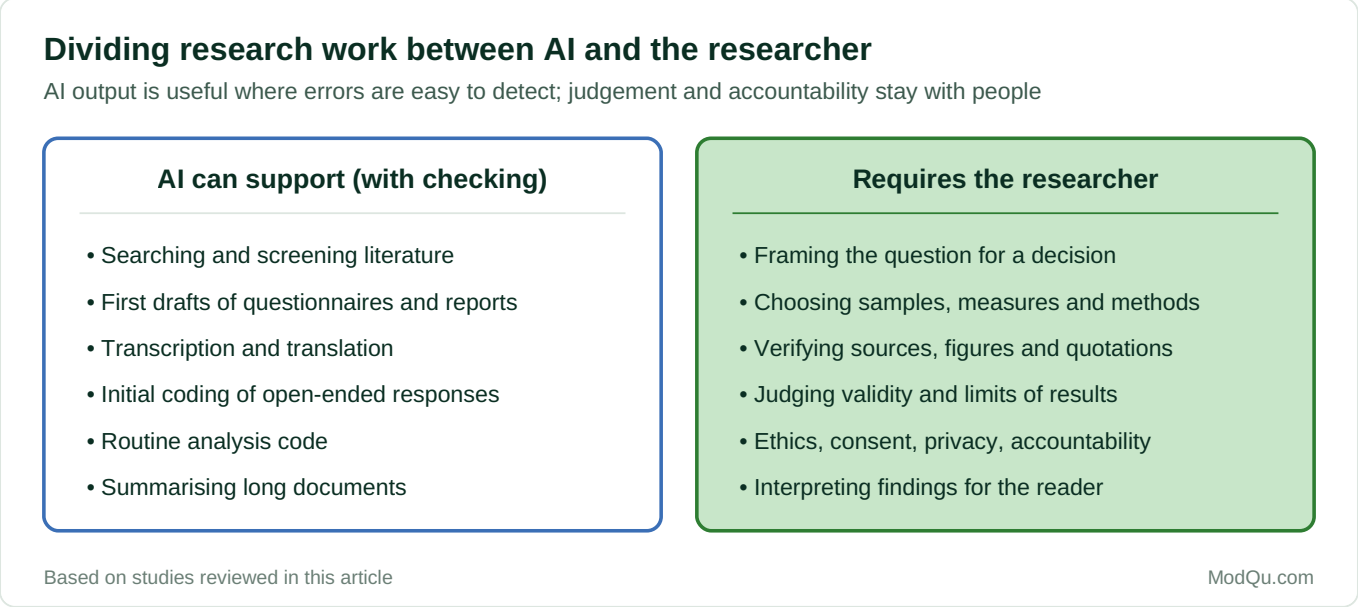


Figure 2. A division of research tasks between AI support and researcher judgement, based on the studies reviewed in this article; diagram by ModQu.

Several parts of the role are likely to become more important:

- **Framing the question.** Deciding what is worth studying and what a decision maker needs to know is not a task that a tool can perform without context.
- **Verification.** Checking sources, figures and claims becomes a larger share of the work when first drafts are produced quickly.

- **Methodological judgement.** Knowing when a sample, measure or model is fit for purpose, and when an AI output falls outside the tool's reliable range.
- **Ethics and accountability.** Researchers remain responsible for consent, privacy and the accuracy of what they publish, whatever tools they use.
- **Interpretation.** Explaining what findings mean for a particular organisation or policy, and with what certainty.

The productivity studies also show that less experienced workers gained most (Brynjolfsson et al., 2025; Dell'Acqua et al., 2023). This may help new researchers produce competent work sooner, but it raises a training question: if routine tasks are automated, junior researchers need other ways to develop the skills that experienced researchers use to judge whether output is correct.

Authorship and disclosure

Publishers have moved quickly to set rules. In early 2023 the editor-in-chief of *Science* announced that text generated by AI tools could not be used in the journal's papers without explicit permission, and that an AI program could not be an author (Thorp, 2023). Other publishers adopted similar positions: AI tools cannot take responsibility for a paper, so they cannot meet the criteria for authorship, and their use should be disclosed in the methods or acknowledgements.

The same principles apply to commercial research. A client who receives a report is entitled to know whether open-ended responses were coded by a person or a model, whether a summary was drafted by AI, and how the output was checked. Disclosure does not weaken a report; it allows the reader to judge how much weight to give it, and it protects the researcher if an error is later found.

Practical guidance

- Use AI for tasks where errors are easy to detect, and check all outputs before use.
- Verify every reference, quotation and statistic against its original source.
- Do not use synthetic respondents as a substitute for real data when estimating opinions or behaviour.
- Record which tools were used and for what, so that work can be reviewed and reproduced.
- Keep sensitive or confidential data out of tools that do not provide adequate protection.
- Invest in the skills that AI does not replace: question framing, method, verification and interpretation.

AI is changing how research is done, and the changes are likely to continue as the tools improve. The evidence to date shows real gains for well-chosen tasks and real risks where output is accepted without checking. The value of a researcher increasingly lies in knowing the difference.

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

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

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics*, *140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Dell'Acqua, F., McFowland, E., III, Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality* (Working Paper No. 24-013). Harvard Business School.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, *384*(6702), 1306–1308. https://doi.org/10.1126/science.adj0998

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

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

Van Noorden, R., & Perkel, J. M. (2023). AI and science: What 1,600 researchers think. *Nature*, *621*(7980), 672–675. https://doi.org/10.1038/d41586-023-02980-0

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

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., Anandkumar, A., Bergen, K., Gomes, C. P., Ho, S., Kohli, P., Lasenby, J., Leskovec, J., Liu, T.-Y., Manrai, A., … Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47–60. https://doi.org/10.1038/s41586-023-06221-2