THE GENERATION OF KNOWLEDGE · TRUST, QUALITY AND THE FUTURE

Artificial Intelligence and the Future of Knowledge Production

How artificial intelligence is changing the production of knowledge, from scientific discovery to survey research, the risks it brings, including fabricated content and illusions of understanding, and the human checks that remain essential.

ModQu research article · 726 words · 7 references (APA 7th edition)

Artificial intelligence (AI) is changing how knowledge is produced. Systems that find patterns in vast datasets, summarise literatures and write fluent text are now used at every stage of research. This article considers what AI adds, what risks it brings and how producers of knowledge can use it responsibly.

What AI can do

The most striking achievements come from systems trained for specific scientific problems. Jumper et al. (2021) described AlphaFold, a system that predicts the three-dimensional structure of proteins from their sequence of amino acids with accuracy often comparable to laboratory methods. A problem that had resisted decades of work became, for many proteins, a matter of computation. Its predictions for hundreds of thousands of proteins were made freely available, and researchers now use them as a starting point for laboratory work.

More general systems, trained on very large amounts of text, are known as **foundation models**, because they can be adapted to many tasks (Bommasani et al., 2021). Researchers use them to search and summarise literature, write code, code open-ended survey responses and draft reports. Argyle et al. (2023) even found that a language model, prompted with the characteristics of real survey respondents, produced answers resembling those of the corresponding groups of people. Such "silicon samples" might help in designing and testing questionnaires, but they reproduce patterns from past data and cannot replace asking real people about new situations.

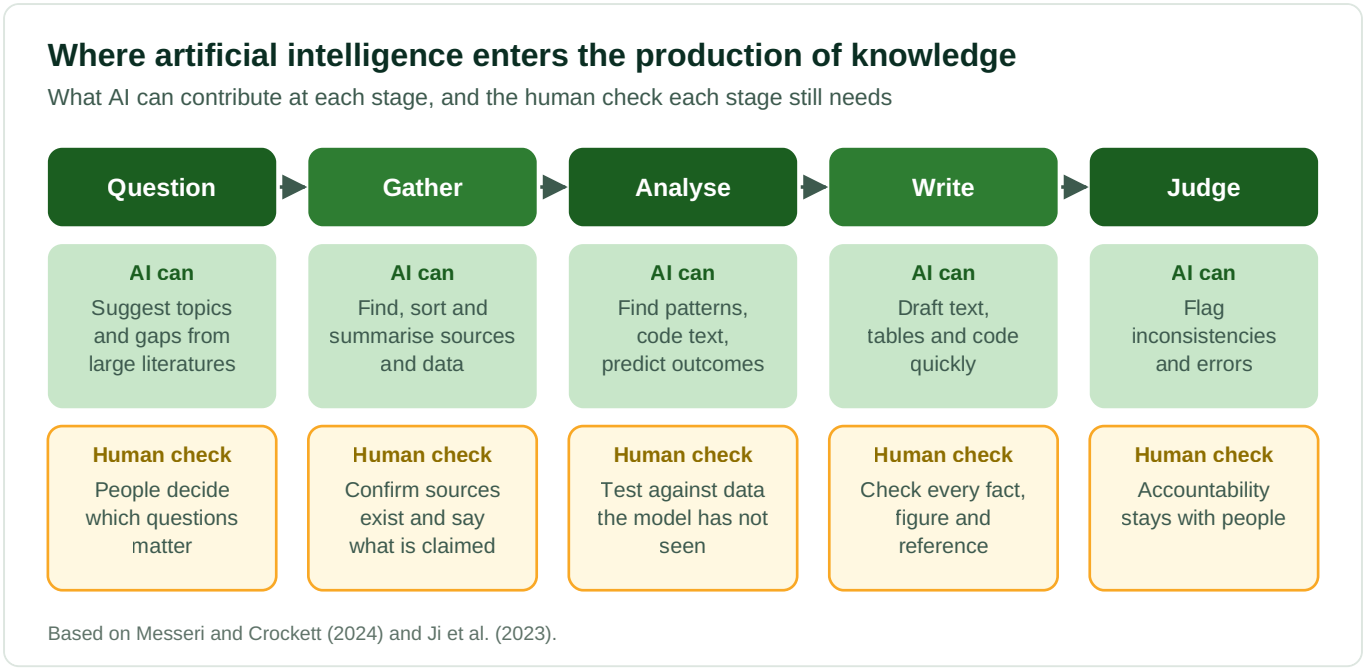


Figure 1. Where AI enters the production of knowledge and the human checks each stage needs. Based on Messeri and Crockett (2024) and Ji et al. (2023); diagram by ModQu.

The risks

Language models generate text that is plausible, not text that is checked. Ji et al. (2023) reviewed research on **hallucination**, the generation of content that is fluent but false or unsupported by the source material. Fabricated references, invented statistics and confident misstatements are common unless outputs are checked. Bender et al. (2021) warned that such systems reproduce the patterns, including the biases, of their training data without any understanding of meaning.

A subtler risk concerns researchers themselves. Messeri and Crockett (2024) argued that AI tools can create **illusions of understanding**: researchers may believe they understand more than they do, because a tool has produced an answer they cannot fully explain. If many researchers rely on the same tools, the range of questions and methods may also narrow.

A further risk concerns the future supply of knowledge. Shumailov et al. (2024) showed that models trained repeatedly on text generated by earlier models gradually lose information about rare events and less common patterns, a process they called model collapse. Original human-generated data, including well-designed surveys, may therefore become more rather than less valuable.

AI in survey and market research

In survey and market research, AI is already used to translate questionnaires, check them for ambiguous wording, classify thousands of open-ended answers, detect careless or fraudulent responses, and summarise findings for clients. Each use saves time, and each needs checking. An automatic classification of open answers should be compared with a

sample coded by people. A translation should be reviewed by a fluent speaker who knows the subject. A summary should be read against the tables it claims to describe. The aim is to let AI take on repetitive work while people keep control of the judgements that determine what the findings mean.

Using AI responsibly

Several principles follow. Every fact, figure and reference produced with AI should be checked against original sources. Methods should disclose where and how AI was used. Models should be tested against data they have not seen. And responsibility for conclusions should remain with people who understand them and can explain how they were reached. Organisations should also keep records of which tools and versions were used, since a model that changes from one month to the next can give different answers to the same request, which makes results hard to reproduce. Sensitive data, such as survey responses that could identify people, should not be sent to external services without the consent and safeguards that apply to any other handling of personal information.

Conclusion

AI can speed up many stages of knowledge production and make some discoveries possible for the first time. But it produces plausible output, not verified knowledge. Its value depends on human questions, real data and careful checking, which remain the foundations on which reliable knowledge is built.

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

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2021). *On the opportunities and risks of foundation models* (arXiv:2108.07258). arXiv. https://doi.org/10.48550/arXiv.2108.07258

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

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

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y