USING ARTIFICIAL INTELLIGENCE IN EVALUATION:

A Rapid Evidence review and Proposed Guidelines

Dr Ully-Yunita Nafizah, Dr Hamisu Salihu and Dr Halima Jibril

Warwick University

November 2025

1. Executive summary

This report presents a rapid review of the evidence base on the use of Artificial Intelligence (AI) tools in policy evaluation. It examines current and potential applications of AI globally and across different types of funding organisations and sectors. The report aims to provide UKRI with evidence-based insights to inform appropriate guidelines for the responsible and effective integration of AI into evaluation processes. The report draws on academic and grey literature and develops a funding evaluation cycle as a framework for understanding AI use in i) choosing and prioritising intervention areas; ii) assessing project proposals; iii) programme monitoring and process evaluations; and iv) impact evaluation and value for money assessments.

Use cases and reported benefits of AI in evaluation

The review found notable examples of AI across different phases of the funding evaluation cycle (Section 3). Here we highlight five areas where evidence on AI effectiveness is relatively stronger, as well as areas where its use may be less effective or too risky.

1. International evidence shows that Natural Language Processing tools can effectively support horizon scanning and strategic agenda setting, which rely on analysing large volumes of real-time data. This could help UKRI identify emerging funding priorities.
2. LLMs, ML, and Generative AI can increase efficiency in the administrative stages of proposal assessment, such as pre-screening and classifying applications, thereby reducing administrative burden. However, AI should not be used beyond these stages for peer review or final funding decisions.

3. LLMs can provide efficient and reliable summaries of large documents, supporting evaluators in preparing reports. They are much less reliable for evidence synthesis, which continues to require significant human input.
4. ML tools are effective for real-time monitoring and data collection, though this may be less applicable to UKRI programmes where recipient organisations, such as universities or businesses, operate with a high degree of autonomy in grant use and real time monitoring may be infeasible.
5. LLMs, ML and GenAI tools have performed well in quantitative and qualitative data analysis, including code standardisation and replication, provided that strong data management, security, and governance measures are in place.

Given the limited evidence base, there is no clear consensus supporting a definitive shift to AI tools for specific evaluation functions. Most use cases remain experimental, with effectiveness and safety of AI tools dependent on context and adherence to best practices (outlined in Section 5). Human judgement remains essential in all decision-making functions.

Risks and challenges of using AI in evaluation

Using AI tools in policy evaluation carries significant risks and challenges. We identified structural challenges which include potential biases, gaps in ethical and legal frameworks, difficulties ensuring data privacy and security, and issues with transparency and accountability. At the operational level, challenges include underperformance of some AI tools in maintaining scientific rigour, validity, and reliability, as well as the tendency of Generative AI to produce stilted outputs with low artistic value. There is also a risk that patterns of tool-user interaction may lead to overreliance on AI and erosion of evaluator skills.

How can UKRI create the conditions for safe, transparent and responsible use of AI in evaluation?

Because the existing evidence on the effectiveness of AI in policy evaluation contexts is sparse, and comes predominantly from contexts outside R&I funding, this report recommends a measured and evidence-based approach to any integration of AI into UKRI evaluation activities. UKRI should consider:

1. Defining a clear but adaptable framework outlining appropriate use cases for AI within UKRI evaluations.
2. Adopting an experimental approach through carefully designed policy experiments and pilot schemes that test specific applications, capture any challenges, and generate lessons for wider adoption. This should preferably be done building on a stronger evidence base, for example after Phase 2 of this project when evaluator interviews will provide richer, context-specific insights to inform any UKRI experimentation.
3. Ensuring data security and developing, where feasible, UKRI-specific internal AI systems and platforms.
4. Providing clear accountability structures for all AI-supported outputs in cases of unintended errors or harms arising from AI use.
5. Providing a framework that ensures strong human oversight and ethical safeguards, ensuring that AI tools complement, rather than replace, human judgement
6. Providing clear disclosure norms that incentivise transparency and accurate reporting of AI use (e.g., emphasising that disclosure enhances trust)
7. Maintaining open communication with evaluators on AI use throughout the evaluation process and creating opportunities for learnings and feedback loops
8. Investing in AI literacy and skills training for evaluators and periodically updating standards to reflect emerging best practices.

In addition to these governance and oversight responsibilities for UKRI, the review synthesises best practices and risk mitigation measures for UKRI evaluators including i) understanding of AI tools and assessing utilisation readiness ii) ensuring data privacy, confidentiality and security iii) ensuring transparency and compliance with ethical and legal frameworks iv) ensuring scientific rigour and reliability v) reflection, learning and capacity building, as well as guidelines related to the use of specific AI tools. These are detailed as Evaluator Checklists in Section 6.

Gaps in the evidence base

The overall evidence base on the use of AI in evaluation is sparse and emerging, but there is even less evidence in the specific context of evaluating Research and Innovation policies and programmes in the UK and internationally. There is also limited discussion in the literature about the extent of transparency and disclosure of AI use in evaluation; most of

the reviewed studies set out to explicitly incorporate and test the use of AI. We still need to understand more about the extent, benefits and challenges of 'everyday' adoption of AI tools in evaluation contexts. We also found little evidence on the use of AI to design evaluations.

Next steps

To enhance our understanding of how AI tools are being used in evaluation, there is scope for conducting semi-structured interviews with staff in organisations actively trialling AI tools, consultancy firms experienced in their application, and evaluators within UKRI's own portfolio. These interviews could explore which tools are being used and for what evaluation tasks, the benefits gained compared to manual methods, the challenges and risks encountered, and how these are managed in relation to ethical and legal standards, including any policy and governance aspects within institutions. Interviews could also explore gaps in evaluators' skills and awareness, and examine attitudes around disclosing the use of AI tools, perceptions of how AI-produced outputs are received, and why evaluators may avoid disclosure.

Please tell us what you thought of this report?

Now you have read our report we would love to know if our research has provided you with new insights, improved your processes, or inspired innovative solutions.

Please let us know how our research is making a difference by completing our short feedback form via this link

Thank you

The Innovation & Research Caucus

Authors

- Dr Ully-Nafizah – Warwick University
- Dr Hamisu Salihu – Warwick University
- Dr Halima Jibril – Warwick University

Acknowledgements

This work was supported by Economic and Social Research Council (ESRC) grant ES/X010759/1 to the Innovation and Research Caucus (IRC) and was commissioned by UK Research & Innovation (UKRI). The interpretations and opinions within this report are those of the authors and may not reflect the policy positions of UKRI.

About the Innovation and Research Caucus

The IRC supports the use of robust evidence and insights in UKRI's strategies and investments, as well as undertaking a co-produced programme of research. Our members are leading academics from across the social sciences, other disciplines and sectors, who are engaged in different aspects of innovation and research system. We connect academic experts, UKRI, IUK and the ESRC, by providing research insights to inform policy and practice. Professor Tim Vorley and Professor Stephen Roper are Co-Directors. The IRC is funded by UKRI via the ESRC and IUK, grant number ES/X010759/1. The support of the funders is acknowledged. The views expressed in this piece are those of the authors and do not necessarily represent those of the funders.

Find out more

Contact: info@ircaucus.ac.uk

Website: https://ircaucus.ac.uk/