GRAIL: Developing responsible practices for AI and machine learning in research funding and evaluation with a community of learning

Denis Newman-Griffis, PhD[1,2*]

[1]*University of Sheffield, Sheffield, United Kingdom*
[2]*Research on Research Institute, London, United Kingdom*
d.r.newman-griffis@sheffield.ac.uk

1 The GRAIL project

Research funding agencies are tasked with bridging the gap between the dual, rapidly-evolving areas of public policy and scientific advancement. As the growth of the scientific community continues to outpace uplifts in funding allocations, and increasingly data-driven policy creates demands for stronger evidence of efficacy and impact, there is a clear need for better tools to enable funders to ensure their decision-making stays well aligned with current science and policy needs.

The maturation of artificial intelligence (AI) and machine learning (ML) technologies is opening new avenues for funders to learn from the rich data sources and internal expertise they have curated over decades, and to develop new data-driven practices to support responsiveness to rapid scientific development as well as changing policy environments. However, there is a lack of shared experience and best practice in using AI and ML in the work of research funding and evaluation, and it is often unclear how developments in AI Safety and Responsible AI discourses translate into practical insights for complex organisations like research funders.

The Research on Research Institute's GRAIL project (Getting Responsible about AI and Machine Learning in Research Funding and Evaluation) is an ongoing effort drawing on a community of learning among research funding organisations to develop specific insights, pathways, and critical questions to guide responsible use of AI and ML in the research funding ecosystem.

The GRAIL project brings together **nine government research funders** (Australian Research Council, Austrian Science Fund, Dutch Research Council, German Research Foundation, Research Council of Norway, Social Sciences and Humanities Research Council of Canada, Swedish Research Council, Swiss National Science Foundation, UKRI) and **four philanthropic funders** ("la Caixa" Foundation, Novo Nordisk Foundation, Volkswagen Foundation, Wellcome Trust), each of whom is at different points in the process of exploring, adopting, deploying, and evaluating AI and ML approaches in their work. The discussions and investigations in the project are producing new insights into how research funders and other public bodies can effectively navigate the sociotechnical systems and processes required to bring AI and ML technologies to bear effectively in their work whilst maintaining the highest standard of ethics and social responsibility.

2 GRAIL workshop series

The core structure of the GRAIL project is a series of virtual, co-productive workshops held with staff from the participating research funders. Each workshop is hosted by one of the project partners and is organised around a specific topic regarding an area of AI/ML application in research funding and/or a particular challenge in effectively and ethically managing AI/ML use.

Workshops are held as closed sessions with limited external data sharing and a strong focus on protected conversation with the freedom to discuss challenging

topics and experiences. The host organisation for each workshop may invite external presenters and additional guests as relevant, with all attendees agreeing to abide by a co-produced set of ground rules. Workshop discussions are noted by the project team, with anonymised versions of notes produced for sharing to attendees after the workshop. Attendees are also invited to complete an anonymous feedback survey reflecting on the presentations/activities and discussion topics in each workshop and highlighting specific learning to carry forward.

Workshop notes are reviewed by the project Steering Group, consisting of nominated representatives from participating organisations, and analysed by the Steering Group together with the project team. Emergent themes and recurring topics are identified by group discussion and used to update evolving documentation of key topics and challenges in navigating implementation and management of AI/ML in research funding and evaluation. As the workshop series continues, discussions are increasingly reflecting specifically on this developing documentation and the insights and recommendations produced.

3   Key findings

The current workshop series has developed to respond to and expand upon the insights generated from a three-session set of workshops in 2021, summarised in a recent RoRI report (Holm et al., 2022). Five GRAIL workshops have been held as of May 2024 (Table 1).

A further programme of workshops has been scheduled through Spring 2025, addressing topics including (a) the role of funders in producing generative AI guidance in science; (b) evaluating efficacy and resilience in applications of AI/ML in research funding contexts; (c) building and managing AI competencies in funding organisations; and (d) documentation, reporting, and scientific transparency in use of AI/ML.

Key themes from discussions to date include:

Theme 1: Disconnects between AI performance evaluation and organisational impact. It is not clear when an AI model that has been developed can be considered "good enough" to use, or how to effectively measure the balance of risk between AI adoption and continuing with current processes. The question of measuring reliability of AI systems, and developing organisational and sector understandings of what reliability should mean in the context of research funding and evaluation, is also a significant challenge that must be explored.

Month	Workshop topic
June 2023	Generative AI in the research funding ecosystem
November 2023	AI/ML in national research assessment exercises
January 2024	Developing meaningful AI/ML guidance for research funders
February 2024	Natural language processing applications in research funding and evaluation
April 2024	Policy and responsible use of AI/ML

Table 1. GRAIL workshops held as of May 2024.

Theme 2: AI as one of many tools. Use of AI is often discussed as a direct improvement of (or replacement for) existing decision-making processes, however workshop attendees have highlighted getting more value from AI as a tool for process insight than efficiency improvement in many cases.

Theme 3: Managing across competencies. AI use is not a technical problem alone, but must combine experiences and perspectives from technical, strategic, and operational components of funding organisations. New work is needed to develop best practice for interprofessional team management and communicating across competencies within organisations to achieve effective AI use.

The ongoing community of learning in the GRAIL project is well positioned to build on these initial themes to develop best practice and greater insight for using AI and machine learning in research policy ecosystems.

Acknowledgements

References

Holm, J., Waltman, L., Newman-Griffis, D. & Wilsdon, J. (2022). Good practice in the use of machine learning & AI by research funding organisations: insights from a workshop series. RoRI Working Paper No. 10. December 2022. DOI: 10.6084/m9.figshare.21710015