Background

The UK Committee on Research Integrity notes that generative AI is the next stress test of the research sector for integrity, but that the system is familiar in adapting and responding to emerging and disruptive technologies. We also recognise that AI could be a boon to improve research integrity through error checking and analysis of larger and more varied data sets. The committee proposes that the five principles of the concordat to support research integrity (honesty, transparency, rigour, accountability, and care & respect) form an existing framework, familiar to the community that can be used to centre research integrity in response to the rapidly growing use of AI in the sector.

The following was developed by the committee's AI working group and refined further with great thanks to contributors across the UK research sector and internationally. The following references to "AI" refer to generative artificial intelligence tools that are publicly available unless otherwise specified.

How might Generative AI affect Research Integrity?

AI has the potential to deliver huge benefits for research, but also challenge some of our existing practices and methods, and our understanding of integrity.

Policies related to AI are emerging in the UK and internationally, with guiding principles on development, use and safeguards. These policies are not always consistent or easy to navigate, and there are gaps with respect to their applicability for research in different disciplines. Rarely is the focus on research integrity, although there is increasing focus on the importance of non-technical questions including governance and public understanding.

The UK Committee on Research Integrity's interest is in helping the sector to understand how research integrity - the principles of research practice which, together, promote and enhance trustworthiness - is being shaped by AI. And, importantly, how the sector is responding.

For the committee, questions include: what are the risks and opportunities that AI brings for integrity, how do these vary across research settings and disciplines, what are the appropriate mitigations, what do we see as the gaps, and what should the sector be developing to support a thriving research system, and trustworthy research? Researchers are responsible for the robustness of their research - what help do they need to navigate this rapidly developing environment?

In our discussions to date, and reading of the rapidly expanding literature, we have identified the following themes and observations for wider discussion. We wish to stimulate conversation across the sector, in particular: *Are these the right issues and if not, what is missing?*

1. Sector: Governance
It is important for cross-disciplinary coordination and a decision on convergence or divergence of UK/EU/international standards and alignment of professional standards and codes of conduct (including identification of misconduct).

Clarity is needed on who sets standards; what good practice looks like; the policy gaps, who should fill them and whether they should they be discipline specific.

Recognising that this is a rapidly evolving environment, where should responsibility for addressing these issues lie?

2. People: Roles and responsibilities
Across the sector there are many actors developing policies and guidance, such as: funders; HEIs; professional bodies; and publishers. There are also gaps, for example STEM subjects are better reflected in the literature than the arts and humanities. Guidance for research is less prominent than are policies for teaching and student assessment. And there is more emphasis on outputs (e.g., use of AI in papers submitted for publication) than on inputs (e.g., assessing quality of data used in models, or the completeness of AI-generated literature reviews. See also six below). There are also inconsistencies in the guidance produced.

What can be done to better foster better collaboration across communities to improve communication and share best practice?

3. Skills and training
AI is starting to be seen as part of the research 'toolkit', however, understanding is needed on the level of AI literacy amongst researchers, institutions, funders, and users. We need clarity on what questions should be asked to design or interpret research which includes AI generated content. There is a need to evaluate what training is available and does it, for example, reflect developments in AI and integrity in early career training.

What new skills are required, by whom, and how should training be provided?

4. Public understanding and expectations re trust and trustworthiness

The term AI is widely referred to, but it is not clear what people understand by the term, and how people's trust in the outputs from research are affected by this understanding. Researchers too need to be able to clearly explain their sources, methods, and limitations of their research - both for the research record and other researchers evaluating findings or planning their own studies, and for users more broadly.

What new challenges does AI bring to explaining research?

5. Attribution and ownership

Concerns about attribution, ownership and IP emerge when writing or evaluating research proposals. Existing intellectual property laws and statements of use may be better utilised to address these kinds of challenges. In addition, various policies are emerging from academic publishers and funders, but they are not always consistent.

What should consistent good practice look like?

6. Reliability and quality of data inputs and models

Researchers want to understand metadata and data provenance assurances and their role in reproducibility and replicability. Although there have been early discussions about reliability of data outputs, prompts and training sets, there may be benefit in bringing together what different disciplines/institutions/professional bodies are doing to help researchers assure the quality of model inputs. Bias (arising from unrepresentativeness of training or other data, or in the asking of research questions) is also widely documented. However further work may be necessary to develop or share tools/methods that can help researchers assess, and correct for, bias.

How is rigour embedded throughout the research life cycle?

7. Research on research integrity

Much research is being conducted into the development of AI tools, applications, and methods for safeguarding privacy. Little research has been carried out specifically on the implications for integrity and trustworthiness in different disciplines across the research life cycle or emergent behaviours arising from use of new AI tools. We welcome the DSIT-UKRI metascience research funding call with a focus on "research integrity, including issues of reproducibility and the impact of interventions such as preregistration, open data, and reproducible code."

What does the UK need to prioritise for further research?

Conclusion

A unifying theme across all these points is the importance of transparency and the challenges of implementing transparency across the research lifecycle. The sector would benefit from pulling together all the outputs of activities taking place on AI across the UK.

October 2024
Professor Ian Gilmore, FMedSci
Dame Jil Matheson, DCB FaCSS
Professor Jeremy Watson, CBE FREng