ADAPT
Engaging Content
Engaging People

Taighde Éireann
Research Ireland

Responsible AI in Research Assessment: Aligning Research Practices with FRIA under the AI Act

HOST INSTITUTIONS

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

DCU
Ollscoil Chathair
Bhaile Átha Cliath
Dublin City University

PARTNER INSTITUTIONS

MTU
Maynooth University
DUBLIN
TUS
University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath

Responsible AI in Research Assessment: Aligning Research Practices with FRIA under the AI Act

Advances in AI, particularly the introduction of generative large language models, are a major disruptor of academic research posing novel challenges for responsible research assessment. Researchers must now carefully consider the risks of using AI in the conduct, sharing, and assessment of research upon values of reliability, honesty, respect, and accountability. In Europe, such AI risk assessments may need to address new regulatory requirements to undertake a Fundamental Rights Impact Assessment (FRIA) under the AI Act. It also requires rapid learning and exchange of best practice in conducting and attesting to AI risk management in research and its integration into responsible research conduct, access to, and assessment of policies developed by institutes and their funders.

Background

These issues are being addressed by several international organisations representing policy and developing practice for universities, including the League of European Research Universities (LERU) and the European Association of Research Managers and Administrators (EARMA)[1]. Of particular relevance is the policy work of the Coalition for Advancing Research Assessment (CoARA)[2], in particular those outputs developed in its working groups on Ethics and Research Integrity Policy for Responsible Research Assessment in Data and Artificial Intelligence (ERIP)[3] and Open Infrastructures for Responsible Research Assessment (OI4RRA)[4]. Here, we outline how this policy work on research integrity and open science may be aligned with the regulatory requirements for FRIA in general. We then make recommendations on the need for a machine+human readable vocabulary and self-assessment tool for responsible AI research that can support local compliance officers, peer-review leaders and research ethics review board members as FAIR data resource.

Researchers and research institutions need to engage with policy development on responsible research practice for situations where AI is used in the conduct, assessment, communication and sharing of research. There is an urgent need to

[1] https://earma.org/ResearchManagementandAITools/
[2] https://coara.eu/
[3] https://coara.eu/working-groups/working-groups/wg-erip/
[4] https://coara.eu/working-groups/working-groups/wg-towards-open-infrastructures-for-responsible-research-assessment-oi4rra/

accelerate policy[5] [6] and best practice development[7] [8] in addressing challenges to the responsible use of AI in academic writing and editing,[9] data analysis, and literature reviews.[10] Such policy development should support the efficient and transparent risk assessment of potential research harms in the use of AI, including those arising from amplification of inaccuracies and embedded biases[11] and reduction of reproducibility and transparency of results.[12] It should develop researcher training and awareness;[13] [14] mitigate the misuse and harmful application of results,[15] and promote copyright and data protection.

Support for research on advancing researcher practice in the responsible use of AI is needed to provide critical evidence and information sharing between Research Performing Organisations (RPO), Research Funding Organisations (RFOs) and regulatory oversight bodies for the AI Act, e.g. national competent authorities and the AI Office. Revised practice must ensure that research risk assessment for AI and other digital research outputs is mapped into both an ethics and fundamental rights impact analysis. The information produced from such AI risk management should be made available as a FAIR resource to accompany existing open research practices.

Recommendations

Research is required into the interactions between stakeholders involved in the production, peer-review, publication and citation of research, as well as on how their behaviour may change as a result of AI adoption and how this impacts research assessment. Within the framework of the ERA Report on Research Assessment, this could be achieved through an analysis of relationships between stakeholders with differing objectives using the stakeholder theory.[16] This would model publicly funded research teams as the centre of the network of research production, peer-review, publication, and citation stakeholders that form a mutually beneficial scheme of co-operation. The aim is to understand

[5] https://doi.org/10.1016/j.jrt.2023.100060
[6] https://doi.org/10.1016/j.ajmo.2023.100036
[7] https://doi.org/10.1016/j.heliyon.2023.e19688
[8] https://doi.org/10.1007/s42438-023-00440-6
[9] https://doi.org/10.1016/j.ijinfomgt.2023.102642
[10] https://doi.org/10.53880/2744-2373.2023.4.37
[11] https://doi.org/10.48550/ARXIV.2305.07605
[12] https://doi.org/10.1002/asi.24750
[13] https://doi.org/10.1007/s40593-021-00239-1
[14] https://doi.org/10.1080/14703297.2023.2271445
[15] https://doi.org/10.1145/3531146.3533780
[16] Phillips, R., 2003. Stakeholder Theory and Organizational Ethics. 1st ed. San Francisco: Berrett-Koehler Publishers Inc

self-assessment of AI risk in a space that requires sacrifice and contribution from stakeholders, while maintaining obligations of fairness and reciprocity to sustain the network. Stakeholders considered should include academic/nonacademic collaborators; RFOs; peer-reviewers; publishers; peer-review committees; research AI system developers; research output aggregators; regulators; researcher licensers/users and the general public. Explicitly modelling the assessment of stakeholders' outputs in the digital space and addressing open science objectives would enable such a stakeholder analysis adaptable to the contexts of different RPOs and RFOs.

Further, the grounding policy development for research AI risk assessment in the fundamental rights impact assessment requirements of the AI Act means the approach will address concerns of RPO operating as public bodies to ensure compliance with the AI Act and to facilitate compliance of other organisations seeking to use research outputs within the jurisdiction of the Act. Though AI used *solely* for research purposes is exempt from the Act, RPO will naturally aim to be at the forefront of developing research assessment practices for the responsible use of AI in general. They should also be prepared to shoulder the ethical responsibility of ensuring that development of digital research outputs including datasets, software and AI/statistical models and published scientific results, offer minimum risk of harm when applied by other scientists and academics, whether licensed to them under commercial or open access license terms.

The sustainability and impact of any AI risk self-assessment policy will be enhanced if tools to implement it are based on open, FAIR research information principles.

Below we break down this approach to developing policy and open tools for AI risk assessment for research into a set of recommendations. These aim to encourage investment in and coordination of efforts to ensure responsible integration of AI alongside human research expertise. Specifically, these recommendations aim to ensure that AI enhances rather than replaces human capabilities for the benefit of the academic community and supports our understanding of how the networks of collaborating stakeholders that produce, peer-review, publish, and cite publicly funded research can evolve to maintain quality and integrity while beneficially employing AI systems. The recommendations aim to promote an open, legally compliant and FAIR-compatible framework for transparently evaluating and attesting to the responsible use of AI in research. This will require support for interdisciplinary research and development, combining expertise in protections for fundamental rights in digital regulation and open FAIR Data Principles associated with AI use in knowledge production.

Recommendation 1: Undertake Requirements Analysis for Risk Management for AI Use in Research

Support should be provided to advance research in understanding the requirements for maintaining research integrity and protections of fundamental rights as AI is adopted by the research community, e.g. through research tenders or Horizon Europe RIA topics. Analysis of requirements for AI risk assessment for responsible research should leverage existing policy development such as CoARA policy development, checklists and tools in the ERIP WG[17] and OI4RRA WG.[18] [19] Placing such requirement in the European context by aligning these with the AI Act and specifically its requirements for FRIA can also avail of emerging research in this area.[20] [21] RPO requirements should be co-created with university compliance officers, experienced peer-review leaders, academic practice professionals, and research ethics review board members. This can also build on examples of existing policy for AI adoption from individual RPOs[22] [23] and their expertise in leading national debates on AI risks for universities[24] and impact on research management.[25] The requirements should also be informed by the ERA Forum's living guidelines on genAI use in research[26] and the UNESCO guidance on GenAI use in education and research.[27] The analysis of requirements in the digital space will include the stakeholder theory to assess how researchers should structure and communicate their AI risk and FRIA to other stakeholders (depicted below) and the impact it may have on their behaviour (including their own use of AI) in line with the broader objectives of achieving a strong and responsible research ecosystem based on reformed research assessment practices addressing AI-driven disruptions.

The scope of such an analysis should include the use of AI for literature review, research project ideation, hypothesis generation, data analysis, experimental

---

[17] https://coara.eu/working-groups/working-groups/wg-erip/
[18] https://zenodo.org/records/14844582
[19] https://infrafinder.investinopen.org/solutions
[20] https://doi.org/10.1016/j.clsr.2025.106113
[21] https://www.medialaws.eu/assessing-the-impact-of-artificial-intelligence-systems-on-fundamental-rights/
[22] https://www.tcd.ie/academicpractice/assets/pdf/college-statement-on-genai.pdf
[23] https://research.kuleuven.be/en/integrity-ethics/integrity/practices/genai/genAI
[24] https://www.postgrad.ie/blog/ai-embracing-the-opportunities-and-threats-of-ai-within-universities/
[25] https://www.adaptcentre.ie/news-and-events/generative-ai-in-research-management/
[26] https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf
[27] https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

design and execution, research subject interaction, reporting of results in publications, and licensing of AI-based results. Particular attention is given to the impact on specific fundamental rights, including privacy and data protection; non-discrimination; freedom of expression, association and academic freedom; health and safety; individual dignity and autonomy; intellectual property, and access to healthy environments, justice, and democracy.

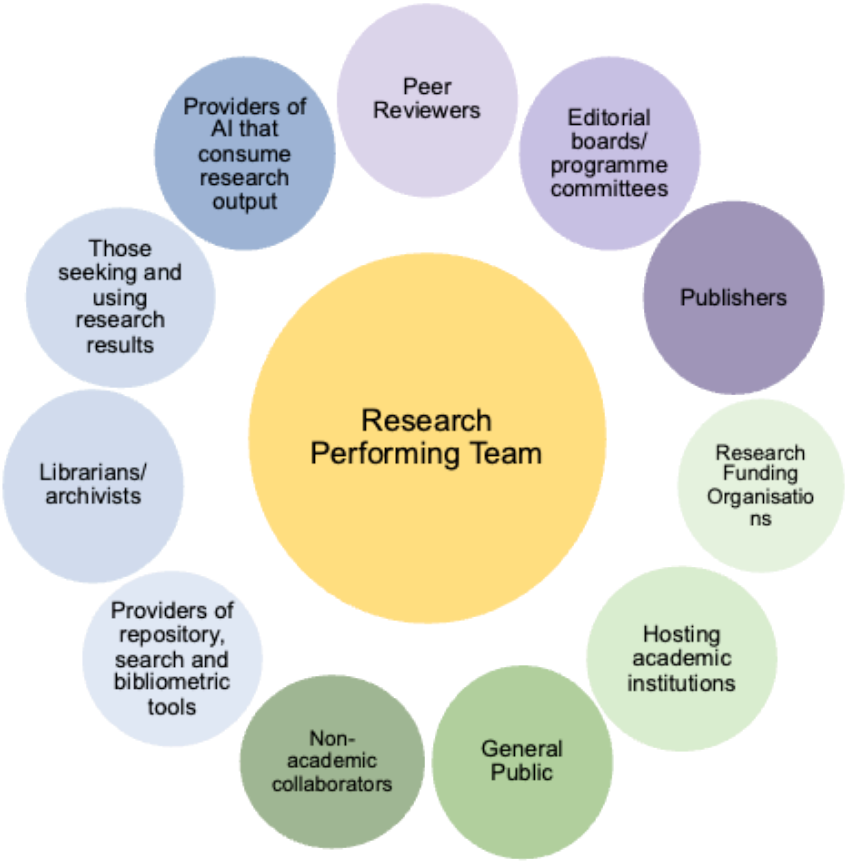

Figure: Stakeholders considered in requirement for AI risk assessment by researchers practicing responsible research

Recommendation 2: Support Development of Open AI Risk and FRIA Vocabulary and Self-assessment Tool

In line with existing open science policies, any analysis of requirements based on stakeholder interaction should be used to drive development of an open semantic vocabulary for capturing and signalling AI usage between stakeholders. This could be further supported through development of a prototype form template that will extend and implement the ERIP self-assessment checklist for RPO, i.e. one that can be used by research teams to self assess the potential risks of AI use in their research.

The development of an open vocabulary for AI risk management and FRIA for responsible research will be informed by international standards;[28] international best practices;[29] [30] literature on FRIA for the AI Act;[31] FRIA experiences in other public sector domains;[32] and existing open semantic models for addressing regulatory risk and compliance for both GDPR[33] and the AI Act.[34]

Such as open semantic vocabulary would be best formalised using the semantic web open knowledge graph Resource Description Framework (RDF) standardised by the World Wide Web Consortium.[35] This benefits from an ecosystem of standards and tools for querying, cataloguing, and sharing FAIR information,[36] including scholarly knowledge and research data[37] in paper repositories such as Semantic Scholar.[38] The well-established LOT ontology engineering framework[39] could be followed in developing such an open semantic model to ensure an extensible, maintainable model that can be reused and extended by others, including open source and commercial research support tool developers.

Supporting the development of a self-assessment tool based on these models will prompt researchers to consider the risks and fundamental rights impacts of AI use in a research project. For example, an online tool could implement and broaden the CoARA-ERIP self-assessment checklist with that of AI fundamental rights risks and their mitigations. Ideally, such tools will generate a human and machine readable attestation in RDF to accompany research outputs. Such attestations offer an evidence base for ongoing research assessment policy development by RPO, RFO and international research policy groups such as CoARA, LERU and EARMA. Such tools may also support evidence collection to assess risks from AI use to the fundamental right of academic freedom[40] and the level of its protection under the AI Act.

Recommendation 3: Support for Trialling of self-assessment tool and revision of vocabulary

---

[28] https://www.iso.org/standard/77304.html
[29] https://airisk.mit.edu/
[30] https://oecd.ai/en/site/risk-accountability
[31] https://doi.org/10.1016/j.clsr.2024.106020
[32] https://www.government.nl/documents/reports/2022/03/31/impact-assessment-fundamental-rights-and-algorithms
[33] https://arxiv.org/abs/2404.13426
[34] https://doi.org/10.3233/SSW220008
[35] https://www.w3.org/2001/sw/
[36] https://cacm.acm.org/research/a-review-of-the-semantic-web-field/
[37] https://doi.org/10.52825/cordi.v1i.272
[38] https://arxiv.org/abs/2301.10140
[39] https://lot.linkeddata.es/
[40] https://www.leru.org/publications/academic-freedom-as-a-fundamental-right

While support for the development of open semantic models and resulting risk assessment tools is an important step, further support, e.g. through EC tenders or Horizon innovation actions, will be required to assess their usability and efficacy and subject designs to co-created revision. For example, trials could be conducted with pools of RPO researchers who would be asked to use tool offerings before and after conducting research activities and a co-creation expert group could be assembled to assess the output for usefulness and clarity in assessing risks and the model revised accordingly. The pattern of such assessment trials and co-creation of revisions is necessary also to allow an open approach to risk assessment of AI in research to respond to the rapid development in the technology and to ongoing regulatory leaning and legal determinations in the protections of fundamental rights under the AI Act.[41] Such trials may benefit from provisions to enable real-world trials of AI systems under the AI Act (Article 60) and thus encourage AI-based research tool vendors (especially European SMEs as per Article 62) to engage with such AI risk assessment management trials.

The AI usage scenarios and requirements for research assessment using open science, the open attestation vocabulary and resulting demonstrations as well as evaluation trials should be widely disseminated to inform emerging policy discussions in RPO, RFO and international research fora. The resulting semantic model, self-assessment tool, and its usage and trialling protocols should be made available under a CC-BY-4.0 license in standardised formats including RDF in order to align with Open Science guidelines and FAIR Data Principles.

Acknowledgements:

This submission was developed with contributions from Dave Lewis (delewis@tcd.ie), Marta Lasek-Markey (mlasek@tcd.ie), Delaram Golpayegani (golpayes@tcd.ie) and Beyza Yaman (beyza.yaman@adaptcentre.ie) of the ADAPT Centre at Trinity College Dublin. We are especially grateful for the insights offered through participation in the CoARA ERIP working group and especially wish to thank its chair Francis P. Crawley for his advice and comments. This work was developed with support from European Commission's Horizon Europe Research and Innovation Programme under grant agreement No. 101177579 (FORSEE) and No. No 101169409 (HARNESS) and from the ADAPT Centre for Digital Media Technology, which is funded by Research Ireland and is co-funded under the European Regional Development Fund through Grant #13/RC/2106_P2.

[41] https://arxiv.org/abs/2503.05787