
Monograph:
Uttley, L. (2026) Review and Analysis of UK Institutional Guidance for Artificial Intelligence Use in Research: A Report on Strategy, Policy and Guidance Across Higher Education Institutions and Research Organisations. Report. UK Committee on Research Integrity

---

Review and Analysis of UK Institutional Guidance for Artificial Intelligence Use in Research

A Report on Strategy, Policy and Guidance Across Higher Education Institutions and Research Organisations

Prepared for:

UK Committee on Research Integrity (UKCORI)

Authored by:

Dr Lesley Uttley

April 2026

INTRODUCTION

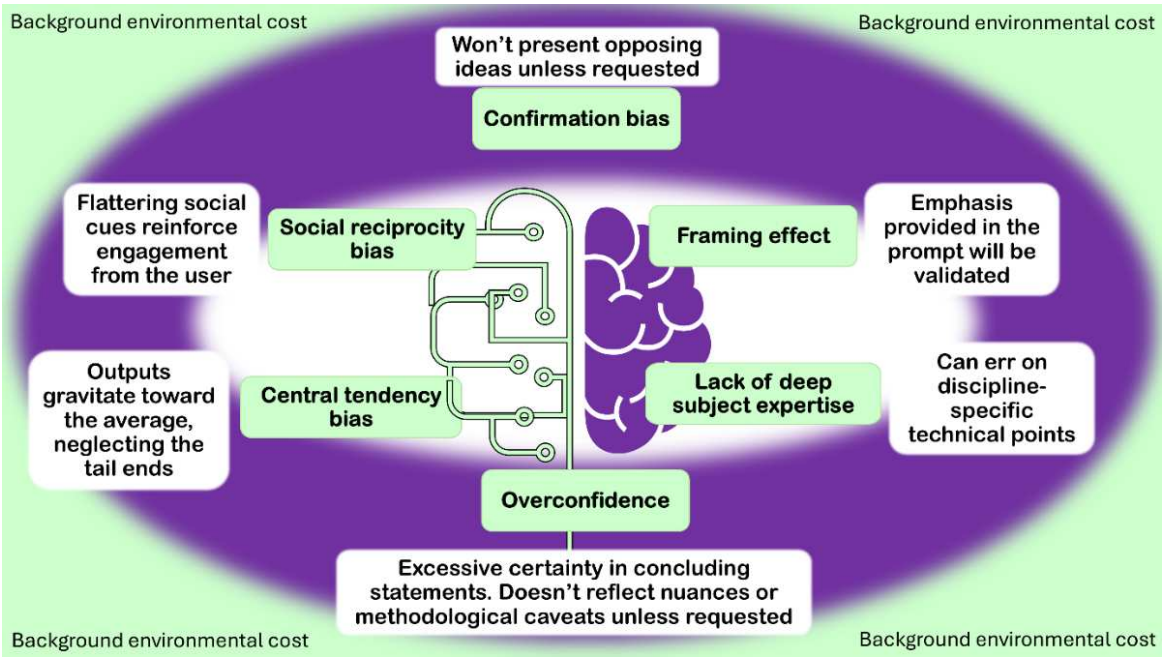
Research institutions across the UK are developing policies to accommodate the emergence and widespread availability of artificial intelligence (AI) to researchers. This report provides an analysis of publicly accessible AI-related policies, strategies, and guidance across a sample of **44 higher education institutions** and **23 research-focused organisations**. The analysis examines provision without attributing content to individual institutions and recognises that the absence of published guidance does not imply that internal policy does not exist. The aim is to identify patterns of provision, areas of commonality, and points of fragmentation or overlap in how organisations are supporting researchers' use of AI.

AI systems have the potential to enhance productivity, accelerate discovery, improve accuracy, manage large volumes of information and foster research innovation across disciplines. AI models (including but not limited to generative AI (GenAI) models), continue to evolve rapidly, yet they retain structural and methodological limitations that have implications for research integrity. Researchers integrating AI into research therefore require a clear understanding of these constraints. The report begins by outlining key areas in which current AI technologies present challenges and opportunities for research integrity, aligned with the core principles of the UK Concordat to Support Research Integrity. (1)

Implications of AI for Research Integrity Principles

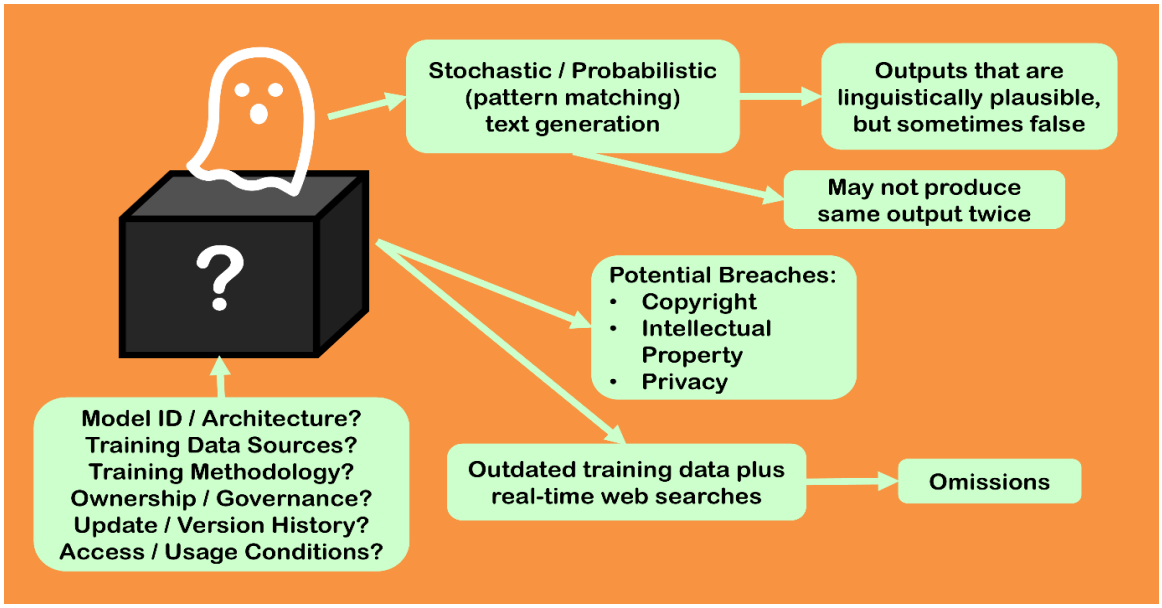
AI technologies can produce fluent outputs but can operate in ways that are not always transparent or reproducible. (2,3) These limitations are common across many forms of AI and may persist even as specific models improve. (4,5) They may omit relevant information or generate responses that appear authoritative but contain technical inaccuracies. Many models are conversationally optimised, meaning they tend to produce agreeable or confirmatory replies to reinforce the framing provided by users, rather than challenge assumptions or present alternative perspectives, operating as an 'echo chamber' for users as represented in Figure 1. (6) General-purpose AI models may lack specialist knowledge in fields requiring fine-grained expertise, therefore AI-generated outputs may require interpretation, verification, and context-setting by subject experts, and they do not yet provide a dependable mechanism for autonomous decision-making. (7) These factors combined highlight a need for human oversight and critical thinking when incorporating AI into research.

Figure 1. IMPLICATIONS FOR OBJECTIVE ACCOUNTABILITY IN RESEARCH



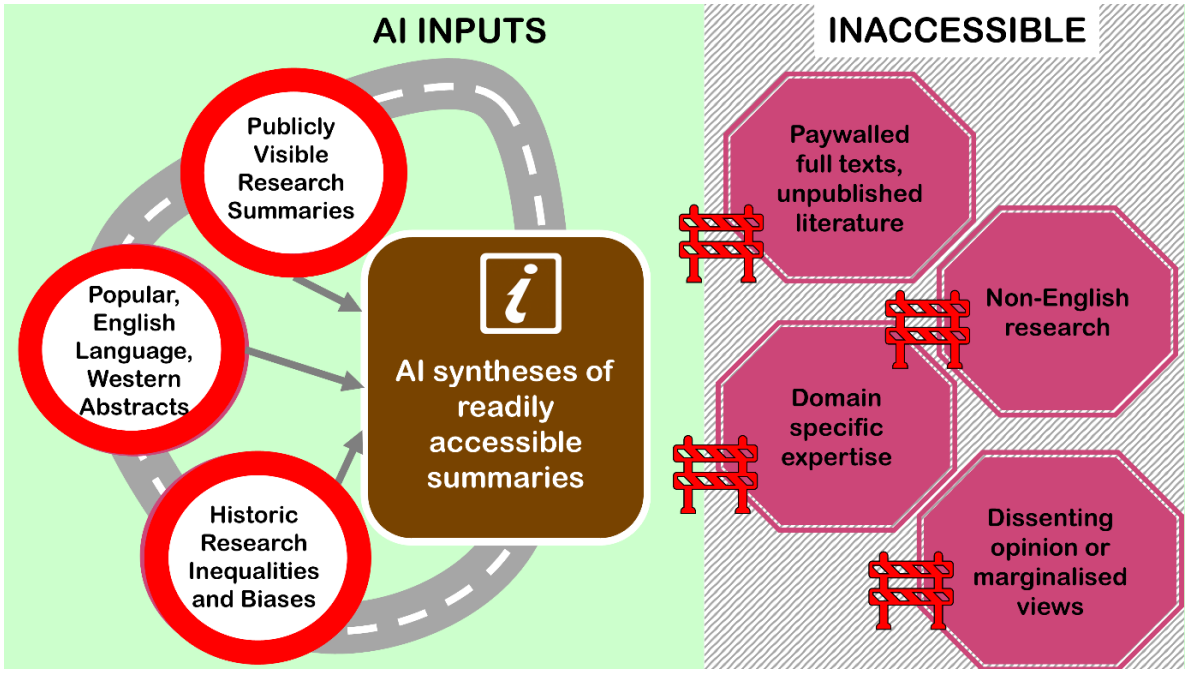
AI tools can support research processes, but their outputs depend heavily on patterns within training data, model architectures, and system design choices that may not be fully disclosed. Many models rely on fixed training cut-offs, meaning their historical knowledge may not reflect the most recently documented developments, as shown in Figure 2. (10) Training processes, including data curation, labelling, and model optimisation, can introduce or reinforce biases. (11) Tools that supplement model outputs with real-time search may reduce information gaps, but the underlying model architecture remains shaped by earlier data. (12) System updates, version changes, and stochastic generation introduce variability in outputs, limiting reproducibility. Such differences across versions or models can result in divergent outputs to identical queries. AI systems can produce plausible sounding but factually incorrect information, perpetuate biases, and breach confidentiality. (7,13,14) These factors collectively affect traceability, comparison across studies, and long-term reproducibility of research.

Figure 2. IMPLICATIONS FOR RANSPARENCY IN RESEARCH



AI outputs reflect the scope, structure, and underlying biases of training data. (6) Publicly available sources, English-language material, and highly indexed content are often overrepresented, while grey literature, minority-language research, and paywalled sources are less visible. This can lead to an unbalanced representation of knowledge as represented in Figure 3. Methods for information retrieval and synthesis differ fundamentally from established research methods such as systematic reviews, where search strategies and inclusion criteria are explicit and reproducible. AI systems can also produce fabricated citations or incorrect details, and they may not distinguish between reliable and unreliable sources unless externally constrained.(12,13)

Figure 3. IMPLICATIONS FOR RIGOUR IN RESEARCH



Access to advanced AI tools often depends on subscription models, infrastructure, or institutional capacity, creating disparities between researchers and across regions.(9) Differential access to digital skills, data literacy, and training opportunities affects who can benefit most from AI-supported research practices. Some communities, such as those with limited digital access or those underrepresented in training datasets, may be less visible in AI-generated outputs, with implications for the inclusivity and fairness of research. (15) As AI use expands, new forms of hidden labour may emerge in the form of verification, oversight, and quality assurance.

Review Objective

As AI models evolve and are increasingly integrated into the research life cycle, their ongoing limitations described above highlight the importance of regularly reviewing guidance for researchers to ensure AI use aligns with the principles of research integrity. This report reviews a sample of current policy and guidance for using AI in research across the UK (for full methodology see the *Supplementary Methods* appendix*)*. It synthesises key principles and best practices from information available from UK universities and research bodies to support researchers in leveraging AI tools responsibly while upholding the established standards of research integrity.

FINDINGS

1. Current Research Landscape

Across the 44 universities sampled, guidance on AI use in research varies in scope and relevance to research.

- For **27 institutions (61%)** available guidance is **focused on education**/student assessment contexts.
- For **12 institutions (27%)** provision of **research-specific policy** or guidance, discrete from education, demonstrates growing recognition of distinct research considerations such as addressing research integrity, separate from broader misconduct frameworks.
- **Six institutions (14%)** provide **detailed guidance to support researcher** competency development beyond permission frameworks.
- **One institution (2%)** provides **comprehensive** integration of a position statement, strategy, policy, and detailed guidance for research innovation.

Of the 23 national bodies sampled (which included government, research funders and councils, academies, learned societies, and independent policy organisations) **18 (78%) had** a **strategy, policy or think piece on AI in research**. Topics of focus included grant applicants and peer review, responsible use, accelerating UK science, sustainability and governance of AI in research.[†1]

Strategic Positions and Policy Divergence

The review showed that institutions occupy a spectrum between **proactive** and **reactive**. National organisations tended to regard AI as an opportunity to enhance research capacity whilst addressing systemic and global challenges such as competitiveness, infrastructure, and ethical governance. Institutional policies tended to emphasise operational risks, including academic standards and data protection, focusing primarily on misconduct and compliance.

The strategic disconnect produces a **patchwork of vision and practice** across universities. Where policies existed, they occasionally overlapped with academic integrity or misconduct guidance and were inconsistently applied across departments. Access restrictions and staff-only portals occasionally limit public visibility of policies that could help to build public confidence and reputation.

As observed, some policies required detailed disclosure of AI use (tools, prompts, outputs), while others accepted brief acknowledgements. The threshold for declaring “substantial use”

[†1] Analysis is based on information available through open internet searches and does not include content requiring login credentials. Available documents or web content were then taken forward as the sample for analysis in this report. Where the report details gaps or commonality it is in relation to the sample taken.

also varied, with some limiting it to interpretive or analytical applications, and others requiring disclosure of any AI-assisted content beyond basic proofreading.

2. Common Principles

Across the sample, there is broad consensus across three principles:

1. Accountability: Human researchers retain ultimate responsibility for research integrity, accuracy, and ethical conduct. AI is positioned as a tool requiring critical oversight, with users accountable for any errors, biases, or breaches of acceptable standards in AI use. There was agreement that AI systems cannot be authors, and their use in confidential peer review is prohibited.

2. Transparency: The guidance reviewed required researchers to openly declare substantive AI use in their work, and adhere to open science practices, enabling impartial evaluation of research outputs.

3. Maintaining Standards of Integrity: The analysis found that guidance highlights risks associated with undisclosed AI use, inappropriate handling of research data, and non-compliance with data protection, intellectual property, or confidentiality requirements. These considerations were framed not as questions of intent, but as matters of adherence to existing research integrity standards and governance obligations. AI use must comply with established ethical, legal and regulatory research standards.

3. Risk Awareness: Shared Understanding

This analysis found that policies across HE institutions and national bodies demonstrated consistent awareness of key risks:

- **Data Security and Confidentiality**: Input of sensitive, personal, or unpublished data into public AI tools can breach GDPR and violate confidentiality agreements.
- **Intellectual Property and Copyright**: Upload of original work may inadvertently transfer IP rights. AI-generated content may infringe existing copyrights.
- **Accuracy and Bias**: AI systems can fabricate plausible but non-existent information and references. Models may inherit and amplify societal biases from training data.
- **Reproducibility**: The probabilistic nature of AI means identical prompts can produce different outputs, challenging reproducibility principles.

The review found that there was broad **scepticism about AI-detection software**, which is viewed as unreliable and potentially discriminatory whereby such systems may disproportionately flag writing from non-native English speakers or individuals whose writing styles diverge from training data. Consequently, HE institutions favoured education, fostering a culture of integrity, and assessment design to deter misuse.

4. Disciplinary Diversity and Responsible Adaptation

Across the guidance reviewed, examples more frequently framed AI use within STEM-oriented research contexts such as data-intensive, computational, and laboratory-based

research. This suggests a need to ensure that guidance also reflects the distinct requirements of practice-based, creative, and other non-STEM research approaches.

Targeted analysis of discipline-specific issues indicated potentially unique challenges to preserving diverse approaches for research incorporating AI:

Arts, design, and humanities disciplines emphasised creativity, authorship, and originality, highlighting risks to intellectual property and cultural diversity. Creative disciplines face unique considerations including ideation, rendering, and computational design applications. Distinct concerns included IP threats, authorship questions, creative field job displacement, and aesthetic homogenisation risks.

Law, politics, and social sciences stressed ethical and societal implications, from fairness and democracy to data confidentiality. The guidance referred to AI as both a tool and object of study, with a focus on legal implications (IP, criminal law), political impacts (armed conflict, democratic processes), requiring social science disciplinary expertise for societal consequence analysis.

STEM and computer science fields focused on reproducibility, explainability, and the technical integration of AI as a research tool.

Engineering and environmental sciences highlighted the resource costs, sustainability and ecological impact of large-scale AI systems such as critical material consumption from data centres and hardware demands.

Social sciences highlighted specific ethical concerns in qualitative research regarding AI use for transcription and analysis. It was found that identified risks included confidentiality breaches and the potential for misinterpretation and oversimplification of cultural and contextual nuance when public AI tools are used. Opportunities exist to develop guidance in qualitative research, archival work, textual analysis, and other humanities and social science methods.

5. Care and Respect

Issues relating to the inclusivity of the research community are addressed only to a limited extent in the HE institutional guidance reviewed. In particular, the following areas are rarely covered, or are absent from current guidance:

- Potential impacts on **career pathway evolution,** entry-level research positions and early-career opportunities as AI capabilities expand.
- Ensuring equitable access to skills and training for AI literacy development **across career stages** and roles including technical staff and professional services, with attention to potential amplification of existing educational or structural inequalities without inclusive practices.
- Consideration of how AI tools may differentially impact the **diversity of research populations** with different backgrounds, languages, working patterns, or accessibility needs and the importance of preserving independent voices, methods and perspectives in research.

- Consideration of **ethical sourcing** and equity, including concerns relating to labour conditions, data colonialism and the broader impacts of AI development on the Global South.

6. Governance, Roles, and Responsibilities

Effective governance requires dynamic, non-prescriptive frameworks built on high-level principles and regular review. The most detailed examples for higher-education institutions from the sample combine:

- Clear guidance on **permitted and prohibited** uses
- **Transparent** disclosure requirements
- **Training and technical support** for safe, innovative research

Across the guidance reviewed, roles and responsibilities relating to AI use in research were framed as extending across the research ecosystem. The guidance commonly attributed the following responsibilities to different stakeholder groups:

- Researchers were described as retaining **accountability** for AI-assisted outputs, including responsibility for critically evaluating results, protecting research data, and disclosing substantive AI use.
- Academic and professional staff were positioned as supporting **ethical learning environments**, including through assessment design and the development of AI literacy.
- Institutional **leadership was characterised as having responsibility for setting strategic direction**, aligning policy and practice across departments, and promoting transparency in governance.
- Developers and private sector partners were referenced in relation to the design of **trustworthy systems** and explainable AI systems.
- Publishers and reviewers are consistently associated with obligations to **protect confidentiality** and to avoid the use of AI tools when handling unpublished material.

Several institutions are developing secure AI infrastructure that protects data while enabling innovation:

- Some institutions were providing **centrally managed, enterprise-level AI tools** where data is not retained for external model training, offering secure alternatives to public versions.
- Some institutions had developed **bespoke proprietary secure platforms** enabling researchers to work with confidential information without external exposure risk.
- **Approved AI tool catalogues** linked to institutional data classification schedules provide clear guidance on appropriate AI use for different data sensitivity levels.
- Recognition **of AI detection tools' unreliability** led many institutions to explicitly prohibit their use, favouring human-led review and discussion approaches to suspected breaches.

7. Skills and Training

AI can be a supportive tool that can enhance ideation, analysis, coding, writing, and dissemination, particularly benefiting non-native English speakers and interdisciplinary collaboration. Moreover, investment in AI literacy and skills development is increasingly framed as a mechanism for strengthening research capability and supporting equitable access to emerging digital competencies across the research workforce.

Across the guidance reviewed, a broadly consistent set of competencies was identified as relevant to developing responsible AI practice in research. These were typically framed in relation to researchers' ability to critically engage with AI tools, rather than technical mastery alone. The core competencies most referenced included:

- An understanding of **how AI tools work** and their limitations and potential sources of bias
- Emphasis on the continued importance of human judgement, including **original thinking, critical analysis,** creativity, problem-solving; alongside AI-supported methods
- Awareness of ethical considerations in interacting with AI tools, including **prompt design and task framing**
- Proficiency in evaluating the quality, accuracy and appropriateness of **AI-generated outputs**
- **Knowledge of data protection**, intellectual property, and responsible authorship when using AI in research
- Recognition of wider **environmental and societal** considerations associated with AI use in research contexts.

The extent to which these competencies are articulated varied across sources, with some guidance addressing them implicitly rather than through explicit skills frameworks. Training spanning students, researchers, and professional staff is identified in the guidance reviewed as being key to promoting sustainable good practice. Some institutions were embedding AI literacy into formal curricula, workshops through libraries/skills centres, and professional development schemes, supported by national skills programmes and doctoral training initiatives.

8. Public Trust and Engagement

Across several national-level documents reviewed, the level of public confidence in AI was frequently characterised as low, with concerns commonly linked to bias, job displacement, misinformation, and data privacy. Guidance most often associated public confidence with the following features:

- Promoting standards expected for research trustworthiness including transparency and explainability, and the importance of **informed consent** when data from public contributors are participants in research.
- Communicating clearly how AI is used, alongside expectations for data stewardship and the documentation of AI-assisted methods within **ethics and governance** processes.

- Emphasis on engaging meaningfully with communities; **co-designing** technologies with public contributors and deliberating broader societal implications of AI-enabled research.
- Recognition of the role of AI literacy beyond specialist developer communities to ensure that researchers and public citizens understand **capabilities, limitations, and responsibilities** of AI use.

9. Attribution, Ownership, and Legal Considerations

There is strong consensus across guidance and wider policies reviewed that AI systems **cannot hold authorship** or inventorship. Legal precedent reinforces that accountability lies solely with the human user. (16)

Sources within the guidance reviewed describe expectations for researchers to **disclose all substantial AI use** in publications, proposals, and outputs, detailing the tools and methods employed. However, no current consensus on formal AI citation emerges from the guidance. Some institutions suggested treating AI use as "personal correspondence" (non-retrievable output). Others explicitly discouraged citing AI as authoritative source (comparing it to search engines), arguing use should only appear in acknowledgements.

Guidance materials highlighted that intellectual property and copyright remained as areas of uncertainty and that **key risks arise from inputting confidential or copyrighted** material into public tools, as well as from AI outputs that may replicate protected content.

Text and Data Mining (TDM) exception supports using others' works in building/training models for research but doesn't grant republishing rights. Researchers must still cite and attribute original works and TDM is not licensed to pass off mined material as newly generated content. AI providers may claim output ownership and outputs risk inadvertently plagiarising copyrighted training data works. The guidance reviewed highlights the use of **secure, enterprise-level platforms** or **institutionally approved AI tools** to protecting research data and confidential information.

10. Reliability and Quality of AI Models: Need for Accountable Objectivity

The analysis found that AI outputs are noted across several sources to be probabilistic, requiring human oversight to check for accuracy. Emphasis was placed on the importance of researchers applying **rigorous critical evaluation** to AI-assisted outputs, including the use of verification techniques, and the maintenance of audit trails to support transparency and reproducibility.
Key challenges noted variably across the guidance reviewed include:

- Bias and data quality: inherited from incomplete, unverified, copyrighted content, or unrepresentative training data.
- Model collapse: AI models becoming less accurate, biased, misleading or homogeneous over successive generations trained on synthetic data, potentially contaminating internet evidence sources.
- Data poisoning: intentional manipulation of training data.

- Environmental and societal impact that unmitigated AI use can have through energy and water consumption of AI infrastructure and ethical concerns regarding labour practices in data-labelling supply chains.
- Explainability of the opaque “black box” design: limiting reproducibility and accountability. Unpacking how AI systems operate can demonstrate research sector awareness of limitations.
- Superficiality of AI-generated content: lacking depth, originality, or critical perspective.

Guidance highlighted the critical role of users in developing "prompt engineering" skills, including the ability to provide clear, specific and context-rich prompts which generate more relevant outputs. Incorporating counter-prompts for ethical considerations and advanced search techniques (Boolean logic, controlled vocabularies, specialist databases) can alleviate some known limitations. In addition, researchers were advised in the guidance reviewed to never input personal, sensitive, confidential or copyrighted material into public AI tools.

Summary of Findings Across UKCORI Themes; Risks and Opportunities

A summary of the analysis from HE institutions and national bodies across seven UK Committee on Research Integrity key themes, aligned with relevant principles of the Concordat for research integrity is presented in Table 1. Table 2 outlines potential risks, consequences and wider impacts posed by AI whilst Table 3 highlights possible opportunities, productivity gains and broader benefits of AI to research.

Table 1. UKCORI Themes Across UK Research Integrity Concordat Principles

	ACCOUNTABABILITY	**TRANSPARENCY, HONESTY**	**CARE & RESPECT**	**RIGOUR**
Sector: Governance	Ethical use of AI Legal use of AI Russell Group Principles TUC 2021	Align principles across relevant policies: student facing / staff facing Balance innovation with risk	Bias audits for outputs (Language, Geographic, Equality Act Demographic, Disciplinary Epistemic biases)	Unity across acts: National Security and Investment Act (2021); EU AI Act UK AI Regulation Bill GDPR/Data Protection Act
People: Roles & responsibilities	Informed consent for human participants Environmental impacts Ethical oversight	Institutional consistency (faculty/departmental) on acknowledgement of AI use	Manifesto on AI (employment) Prioritise human creativity, curiosity, judgement	Human oversight & due diligence when using AI tools Consult funder/publisher guidelines Learn from breaches & security incidents
Skills & training	Responsible proportionality of AI use Ethical & self-critical prompting	Mandatory training Aligning: information governance, protecting information, ethics	Continuing professional development & digital upskilling Differentiated learning pathways	AI literacy AI Critical Thinking Horizon scanning rapidly changing digital landscape
Public understanding	Trust, legitimacy and acceptance	Transparency about standards and expectations for AI's role in research	Civic education & societal dialogue with the public (not just to the public)	Diversify visualisation and dissemination strategies of research results Patient/participant consent issues
Attribution & ownership	IP Rights Proprietary ownership of data/models Potential conflicts of interest	No AI authors Data provenance & citation Full declaration including software version	Recognition for currently "hidden" labour (data curation and validation, prompt engineering)	Copyright, Designs and Patents Act (1988) Text and Data Mining (TDM) exception
Reliability & quality of inputs/models	Document potential risks and harms	FAIR data principles Monitoring & reporting (AI tool usage)	Enhanced Search Techniques Considered, nuanced user prompts	Audit trail of raw AI output Monitoring AI data quality, data contamination /model collapse
Research on research integrity	Longitudinal monitoring incentives, productivity pressure Emerging questionable research practices	Transparent reporting of methodology in research (and meta-research) of AI	Epistemically inclusive design principles Interdisciplinary collaborations Safeguard linguistic & cultural diversity	AI Meta-Research Integrity Policy AI output provenance

DISCUSSION

This analysis captures a snapshot of 67 UK organisations' guidance provision as of September 2025, in a dynamic, rapidly developing landscape. Early indications from this work highlight a potential disconnect between the aspirational, innovation-focused national landscape and the more variable institutional approaches to operationalising AI use in research.

Table 2: Summary of Convergence Between HE Institutions and National Bodies Sampled

	Research Organisations (ROs)	**HE Institutions**	**Convergence / Difference**
Current focus of AI guidance	Mostly frames AI in relation to research strategy, capability building, and long-term innovation.	Primarily focused on student assessment and education-related frameworks (61%), with a smaller proportion (27%) developing research-specific guidance.	Difference: ROs tend to adopt a research-first perspective, while many institutions remain education-focused.
Overarching principles	Consistent emphasis on accountability, transparency, and integrity as foundational principles for AI use in research.	Alignment of core principles across institutional guidance, particularly in relation to responsibility and the need for honest disclosure.	Convergence: Universal agreement on core principles of integrity providing a shared foundation.
Approach to innovation and opportunity	AI is frequently positioned as both an opportunity and a responsibility, requiring enabling frameworks to support responsible research innovation.	Increasing recognition of AI's potential beyond risk management, particularly among institutions developing research-specific guidance.	Partial convergence: Shared recognition exists, but institutional vision varies.
Infrastructure and governance models	Greater emphasis on coordinated approaches, including infrastructure, governance, and skills development at scale.	Some emerging examples of integrated approaches, including secure platforms, governance mechanisms, and cultural initiatives, though provision is uneven.	Difference: Multiple models exist, but inconsistent adoption across institutions.
Skills, training, and culture	Skills development and research capability are commonly viewed as strategic enablers of responsible AI use.	Growing attention to AI literacy and competency development, particularly where institutions move beyond compliance-focused frameworks.	Emerging convergence: Increasing alignment as institutions expand beyond risk-based approaches.

Future Directions for Research on Research Integrity

Future research on research integrity can monitor the impacts and broader integration of AI into research across the Concordat to support research integrity principles. Potential research on research suggestions below are posited from challenges discussed across the guidance as well as suggestions from interpretation of areas less represented across the sources.

1: Accountability

How AI potentially influences researcher behaviour, incentives, and productivity pressure.

- AI literacy effectiveness and potential detrimental impacts on researchers' cognitive and analytical skills.
- Effectiveness of different training interventions on researcher behaviour and confidence in critical and ethical AI use.
- Extent researchers actively verify AI-generated content; prevalence of "hallucinated" citations in published works; researcher fact-checking methods.
- Longitudinal monitoring of evolving authorship and contribution norms across career stages and disciplines.

2: Transparency and Honesty

While consensus exists within guidance reviewed on honest disclosure of AI use, less agreement exists on what constitutes declarable use and appropriate disclosure format/detail level.

- How "substantive AI use" is interpreted across fields; factors influencing researchers' disclosure decisions.
- Surveying existing practices and proposing harmonized standards for citing non-retrievable AI outputs ensuring clarity and consistency.
- How the "black box" nature of AI models impacts research reproducibility, creating new standards for documenting and sharing AI-driven research enabling replication.
- How researchers can accurately and meaningfully describe AI's role in their work, developing standardized taxonomy for acknowledging different AI contribution levels (from minor editing to substantive analysis).

3: Rigour

Principles to ensure rigour, quality and respect (for data, participants, IP) emphasised in the guidance reviewed may be challenged by AI's technical limitations and misuse potential. Rapid policy development by universities, funders, and publishers provides areas for comparative and evaluative research.

- Peer reviewer community awareness of AI use prohibition in peer review; meta-research evidence of usage; qualitative research on perceived temptations to use AI for efficiency; enforcement mechanism effectiveness.
- Monitoring safety and efficacy of rapidly emerging AI software through wider collaborations.
- Prevalence of risky behaviours (e.g., uploading sensitive data); effectiveness of institutional guidance and secure in-house AI platforms in mitigating data security vulnerabilities.

4: Care and Respect for Inclusivity

Research identifying, quantifying, and mitigating the threats AI poses to research discussed across the sources reviewed such as output reliability and generalisability to real-world populations and scientific workforce.

- Gaining multiple perspectives from ECRs, non-STEM disciplines, marginalised end-users, AI developers, data scientists and other stakeholders.
- Meta-research on risks of AI harming creative or traditional methods and opportunities for diversifying excellent research outputs.
- How biases inherited from AI training data can be identified and mitigated in research outputs; developing and testing frameworks for "bias auditing" in AI-assisted research.
- Integrating broader ethical considerations to monitor how AI might help or hinder fostering inclusion in research workforce and cultivating trust from wider society.

Supporting UK Research Excellence

The guidance landscape analysed reflects a sector positioned to support UK ambitions for global research leadership in the AI era. By building on strong foundational consensus, learning from emerging best practice, expanding frameworks to encompass care and respect dimensions fully, and fostering sector-wide collaboration, the UK research community can continue developing governance approaches that enable innovation while upholding the highest standards of research integrity.

Where institutions have developed more comprehensive AI guidance, this is reflected in clearer alignment between strategy, policy, and operational guidance, often supported by visible governance structures, training provision, and approved toolsets. Where these

elements are made publicly available, they offer practical reference points that other institutions can adapt supporting collective capability while respecting institutional diversity. As global technology, standards and practice continue evolving, commitment to regular review, collaborative knowledge sharing, and flexible, enabling frameworks positions the sector well for continued development.

REFERENCES

1. The Concordat to Support Research Integrity – UKCORI [Internet]. [cited 2025 Oct 13]. Available from: https://ukcori.org/research-integrity-concordat/

2. Arar KH, Özen H, Polat G, Turan S. Artificial intelligence, generative artificial intelligence and research integrity: a hybrid systemic review. Smart Learn Environ. 2025 July 22;12(1):44.

3. Nicholas D, Herman E, Clark D, Abrizah A, Revez J, Rodríguez-Bravo B, et al. Integrity and Misconduct, Where Does Artificial Intelligence Lead? Learned Publishing. 2025;38(3):e2013.

4. Doskaliuk B, Zimba O, Yessirkepov M, Klishch I, Yatsyshyn R. Artificial Intelligence in Peer Review: Enhancing Efficiency While Preserving Integrity. J Korean Med Sci [Internet]. 2025 Feb 10 [cited 2025 Sept 24];40(7). Available from: https://synapse.koreamed.org/articles/1516090023

5. Lund B, Lamba M, Oh SH. The Impact of AI on Academic Research and Publishing [Internet]. arXiv; 2024 [cited 2025 Sept 24]. Available from: http://arxiv.org/abs/2406.06009

6. Lopez-Lopez E, Abels CM, Holford D, Herzog SM, Lewandowsky S. Generative artificial intelligence–mediated confirmation bias in health information seeking. Annals of the New York Academy of Sciences. 2025;1550(1):23–36.

7. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. npj Digit Med. 2023 July 6;6(1):120.

8. Kansal N, Kansal S. The Energy Burden of AI Health and Equity Risks for Resource Poor Nations. IJSAT - International Journal on Science and Technology [Internet]. 2025 June 22 [cited 2025 Oct 15];16(2). Available from: https://www.ijsat.org/research-paper.php?id=6265

9. Regilme SSF. Artificial Intelligence Colonialism: Environmental Damage, Labor Exploitation, and Human Rights Crises in the Global South. SAIS Review of International Affairs. 2024;44(2):75–92.

10. Bortnyk K, Yaroshchuk B, Bahniuk N, Pekh P. Overcoming challenges in artificial intelligence training: data limitations, computational costs and model robustness. COMPUTER-INTEGRATED TECHNOLOGIES: EDUCATION, SCIENCE, PRODUCTION. 2023 Dec 16;(53):37–43.

11. Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digital Health. 2023 June 22;2(6):e0000278.

12. Gallifant J, Afshar M, Ameen S, Aphinyanaphongs Y, Chen S, Cacciamani G, et al. The TRIPOD-LLM reporting guideline for studies using large language models. Nat Med. 2025 Jan;31(1):60–9.

13. Zack T, Lehman E, Suzgun M, Rodriguez JA, Celi LA, Gichoya J, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. The Lancet Digital Health. 2024 Jan 1;6(1):e12–22.

14. Omiye JA, Gui H, Rezaei SJ, Zou J, Daneshjou R. Large Language Models in Medicine: The Potentials and Pitfalls : A Narrative Review. Ann Intern Med. 2024 Feb;177(2):210–20.

15. Lewis JE, Abdilla A, Arista N, Baker K, Benesiinaabandan S, Brown M, et al. Indigenous Protocol and Artificial Intelligence Position Paper [Internet]. Honolulu, HI: Indigenous Protocol and Artificial Intelligence Working Group and the Canadian Institute for Advanced Research; 2020 [cited 2025 Oct 15]. Available from: https://spectrum.library.concordia.ca/id/eprint/986506/

16. Thaler (Appellant) v Comptroller-General of Patents, Designs and Trademarks (Respondent) - UK Supreme Court [Internet]. [cited 2025 Oct 2]. Available from: https://www.supremecourt.uk/cases/uksc-2021-0201

Funding

This document was commissioned by the UK Committee on Research.

Licensing

Suggested citation: Uttley, L. (2025). Review and Analysis of UK Institutional Guidance for Artificial Intelligence in Research. Report commissioned by the UK Committee on Research Integrity and UKRI Research Integrity Working Subgroup. October 2025.

Abbreviations

AI: Artificial Intelligence

GenAI: Generative Artificial Intelligence

HE: Higher Education

IP : Intellectual Property

STEM: Science, Technology, Engineering, and Mathematics