UKRIO

RESEARCH INTEGRITY OFFICE

Promoting integrity and high ethical standards in research
Providing confidential, independent and expert support

Embracing AI with integrity

A practical guide for researchers

Version No.: 01

Publication Date: 27/06/2025

DOI: https://doi.org/10.37672/UKRIO.2025.06.embracingAIwithintegrity

Table of Contents

UKRIO
RESEARCH INTEGRITY OFFICE

Research integrity and Artificial Intelligence (AI)

Research integrity encompasses the principles and practices that underpin responsible research and help foster trust and confidence in the research process [1]. These principles form the foundation of good research practice and must be upheld regardless of the tools or technologies employed. When research integrity is compromised, public trust in research findings is eroded, and the research record may be distorted, negatively impacting future research, policy development, legal frameworks, clinical practice, and societal norms.

Artificial Intelligence (AI), as described by the Information Commissioner's Office (ICO), is *'an umbrella term for a range of algorithm-based technologies that solve complex tasks by carrying out functions that previously required human thinking'* [2]. The use of AI in research is not a new phenomenon. Many disciplines have long integrated AI methods, ranging from the application of machine learning for large-scale data analysis to the use of generative adversarial networks in the creation of digital art.

As with any emerging technology, the integration of AI into research will require time to establish accepted practices, applications, and norms. The rapid pace of AI development and its expanding role in research make it challenging to define consistent standards [3]. However, as national and international guidelines continue to evolve, the research community is steadily working to define, adopt, and refine frameworks for the responsible use of AI [4].

The UK Department for Science, Innovation, and Technology policy paper, *A pro-innovation approach to AI regulation* (2023), outlines five guiding principles for the responsible use of AI: Safety, security and robustness, Appropriate transparency and explainability, Fairness, Accountability and governance, and Contestability and redress [5]. These principles provide a useful framework for considering how to use AI responsibly in your work and identifying potential risks from the outset.

Scope of this guidance

This guidance does not address deliberate breaches of research integrity that AI might enable. Instead, it aims to raise researchers' awareness of the potential risks AI poses to research integrity. It highlights key areas where researchers can proactively take steps to mitigate these risks and thoughtfully navigate ethical dilemmas.

The focus is on the current use of AI in research. By referring to AI in general terms, the guidance remains broadly applicable rather than limited to specific types, such as generative AI. At the time of writing, researchers typically employ AI in two main ways:

- As an *assistant* – to support tasks such as literature mapping, improving writing style, or correcting grammar
- As a *research tool* – to generate, interpret, analyse, or process data, and to predict outcomes [6]

Other stakeholders are also integrating AI. For example, publishers use AI to support editorial oversight by screening article submissions, detecting manipulated images, and identifying papers from fraudulent sources [7].

Where to begin when using AI in your research

With any emerging technology used in research practice, it is important to consider the following three questions:

1. What are the tangible benefits?
   Will it save time, streamline your processes, or improve how you visualise or communicate your ideas?
2. What is the impact?
   It is an appropriate, useful, and necessary tool for this particular task or phase of research? Might its use raise ethical, legal, or integrity-related concerns? Can you adequately document and disclose use?
3. Is it the only way to achieve the desired outcome?
   Could the same goal be reached through alternative methods that might better support transparency, critical thinking, or originality?

Types of Artificial Intelligence

This section outlines commonly used types of AI and gives examples of their applications across different research disciplines:

Narrow AI

Narrow AI is designed to perform specific tasks or a closely related set of tasks. Real-world examples include virtual assistants like Amazon's Alexa, BMW's in-car iDrive system, and Netflix's streaming recommendations. In research settings, narrow AI is used to analyse large datasets and images, enhancing efficiency and supporting informed decision-making. In medical research, it is often used to analyse images and classify cell phenotypes. In the digital humanities, it aids textual analysis to identify patterns or themes. In archival research, it can be used to create interactive digital archives, maps, and photographs.

Machine learning

Machine learning is a broad term that refers to AI that learns from data without being explicitly programmed. This learning can be unsupervised or human-instructed/'human-in-the-loop' (reinforcement learning). These algorithms are trained on specific data sets. Deep learning is a more advanced subset of machine learning that interconnects layers of artificial neurons. One example is convolutional neural networks (CNNs), which use three-dimensional data for image classification and object recognition tasks.

Natural language processing (NLP)

NLP enables machines to understand, generate, and manipulate human language.

Computer vision (CV) and visual AI

Computer vision (CV) and visual AI apply deep learning to 'see' and comprehend the visual world, analysing images and videos like humans. CV algorithms analyse

images and videos for tasks like object detection, face recognition, and autonomous driving.

Generative AI

Generative AI, which has gained significant public attention and use in recent years, is AI that can create new text, images, audio, video, and code. A widely known example is ChatGPT, a unimodal, large language model (LLM) AI designed to receive text cues or prompts (inputs) and generate outputs using NLP.

Multimodal AI

Multimodal AI uses multiple AI systems to process different types of data simultaneously. An example is ChatGPT 4.0, which can process text, audio and image inputs. There are many applications of multimodal AI in research, including inorganic materials design [8], monitoring and diagnosing disease [9], text correction for historical documents [10], and nature and climate finance [11,12].

Explainable AI (XAI)

XAI is an area of research and practice that refers to processes and methods attempting to improve comprehension and trust in the outcomes of AI systems. XAI aims to gain insight into AI decision-making processes to enhance their interpretability, ensure compliance with laws and regulations, and mitigate biases and errors associated with the 'black-box' nature of many AI models [13]. A published manifesto outlines 28 open challenges and interdisciplinary research directions for XAI, offering a potential roadmap for its development, supported by real-world examples [14].

Key challenges to research integrity

When considered through the lens of research integrity principles, many of the challenges associated with the use of AI in research can be categorised into five key themes:

1. Breaching laws, regulations, and conditions
2. Ethical considerations
3. Protecting the research record
4. Research dissemination
5. Creativity and critical thinking

These themes provide a practical framework for identifying and addressing potential integrity risks associated with AI. Each will be explored in the following sections, along with suggested advice and guidance to support understanding and application. The suggestions provided are not exhaustive checklists. Rather, they are intended as prompts to help you take an informed, considered, responsible, and ethical approach to integrating AI into your research.

Breaching laws, regulations, and conditions

This section aims to help researchers understand and navigate the legal and regulatory risks associated with using AI in research, including potential breaches of privacy, data protection and security, intellectual property rights, copyright and licensing agreements, and the terms and conditions set by funders or employers.

Funder and employer requirements

When conducting research, researchers must be aware of the obligations tied to their funding and employment. Failing to meet the conditions set by funders or breaching their employer's research codes of practice can have significant consequences.

It is common for research organisations to have established policies, guidelines, and processes for the use of AI across both research and non-research contexts, such as teaching and learning. These may include specific guidance on particular types of AI to help ensure compliance with legal and ethical standards.

Some organisations offer access to approved AI tools for broad use, such as a version of Microsoft Co-Pilot that operates within a closed or secure environment. Others have established formal processes to evaluate the suitability of AI tools for specific research questions or projects, ensuring alignment with organisational values, research ethics and disciplinary standards. For examples of organisational policies and guidance, see references [15–18].

> **Advice**: Check your organisation's policies, procedures, tools, and resources relating to the use of AI.

All research employers and funders require compliance with laws and regulations concerning privacy, data protection and security, intellectual property, copyright, and license agreements.

> **Advice**: Reflect carefully on whether your use of AI could put you at risk of breaching any of your funder's conditions. If you are unsure, seek expert advice from within your organisation.

Laws and regulations

When using AI in research, it is essential to carefully consider issues related to both personal data and intellectual property. A key concern is the risk of data breaches, which can lead to privacy violations, identity theft, reputational damage, financial loss, and even health risks.

When using AI in your research, ask the following questions:

- Who owns the information you are inputting into the AI?
- Who owns the information once it is in the AI?
- Who owns the outputs from the AI?

These questions are not always straightforward to answer and are the subject of active debate and legal challenge.

Intellectual property

According to UK law, you own intellectual property if you created it (and it meets the criteria for copyright, patents, or design rights. It can also be acquired by purchasing rights from the original creator or a previous owner. Brands, such as well-known product names, can also be protected as trade marks. It's important to note that your employer and/or research funder may also claim intellectual property rights arising from your research.

When it comes to research involving AI, AI providers may claim ownership over both the inputs and resulting AI outputs, considering them part of their intellectual property. Commercial and academic AI tools may record your inputs, including data sets, code, algorithms, videos, images, or text prompts. These inputs may be used to train AI algorithms, improve the performance of AI systems, or even be sold to third parties for purposes such as data mining. There is also a risk that unethical actors, such as paper mills, might exploit unintentional leaks of intellectual property.

> **Advice**: To protect your research and comply with legal and ethical standards, never upload copyrighted, personal, or sensitive data into an AI tool without appropriate permissions or consent. This includes unpublished works that you are peer reviewing (e.g. articles or grant proposals), as uploading these materials would likely breach confidentiality agreements and release the works into the public domain.

Personal data

A key concern when using AI in research is ensuring the protection of personal data. Many AI tools require access to vast amounts of data, some of which may be sensitive or personal. Personal data includes names, dates of birth, and addresses to photographs, health and dental records, and other information that could identify an individual. In the UK, the use and protection of such data is governed by the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018.

As AI continues to evolve – particularly through multimodal approaches and cross-system data integration – there is growing potential for both personal and non-

personal information to be reused or cross-referenced. This increases the risk of individuals being reidentified, for example, by linking anonymised data with other datasets.

Under existing legislation, individuals have specific rights concerning their personal data, subject to certain exceptions. These include the right to:

- Be informed about how their data is being used
- Access their personal data
- Have incorrect data updated
- Have data erased (see *Right to erasure, Article 17*)
- Stop or restrict the processing of their data (see *Right to restriction of processing, Article 18*)
- Data portability (allowing individuals to get and reuse their data for different services)
- Object to how their data is processed in certain circumstances

While the UK GDPR includes specific exemptions for research purposes [19], complying with these rights when using personal data in AI systems can be challenging to navigate.

It is necessary to consider how special category data (i.e., sensitive data) is kept safe and secure, both during a research study and after it concludes. When erasing inputs put into AI, it can be difficult to ensure complete deletion due to how the model may store, process, or integrate the inputs into its learning. Similarly, stopping or restricting processing can be challenging, as AI systems often handle data in ways that are neither transparent nor easy to control, making it hard to restrict processing, particularly when the system is continuously learning or evolving based on new data inputs.

Advice: When using an AI tool for the first time, begin by reviewing the information provided by the tool's developer. Document your findings, including their terms and conditions and privacy policy. Assess whether and how you can:

- Opt out of having your inputs used for training the AI
- Delete inputs permanently if needed
- See clear information about data security

Navigating this landscape can be challenging. Before proceeding, we recommend seeking expert advice – such as from IP specialists, or your organisation's research governance, information security, or compliance officers – to ensure you meet legal, institutional, and funder requirements. If your organisation lacks in-house expertise, seek advice from the Information Commissioner's Office (ICO), which has published guidance on data protection and the use of AI.

Ethical considerations

This section aims to highlight and prompt discussions about the ethical concerns that may arise from using AI in research, including potential harms to individuals, society, and the environment.

At present, there are many ethical concerns associated with AI use in research. A key concern is bias – particularly the uncertainty around how AI tools address, or may reinforce, existing biases and discrimination. This includes concerns about the lack of diversity in the teams and processes that develop these tools.

Additional ethical concerns include:

- Challenges around ensuring appropriate informed consent
- The risk of inadvertent linkage that could lead to future disclosure of sensitive information
- The potential for incorrect diagnoses in healthcare settings and impacts on mental health
- Inadequate management of conflicts of interest
- Broader societal and environmental consequences
- Unclear lines of accountability

Assessing ethical implications

If you are considering using AI in your research, it is essential to critically reflect on the ethical implications early in the planning phase – before submitting your study for formal review by a Research Ethics Committee (REC). An effective approach is to review relevant literature and seek advice from colleagues or peers, particularly those with expertise in your specific field or discipline.

Resources such as the *Transparent Reporting of Ethics for Generative AI: the TREGAI Checklist* (see Figure 1) can support ethical planning and reflection. Additionally, if your research involves human participants, consult UKRIO's *Research Checklist of Ethics Applications for Research with Human Beings* for further practical guidance.

Definitions of 10 ethical principles for generative artificial intelligence (GenAI) in a health-care context*

Accountability

- The explicit clarifications of to whom and to what extent responsibility or legal liability, or both fall
- The mandated and moral duty to establish regulatory mechanisms to prevent potential adverse effects on patients from the use of GenAI

Autonomy

- The preservation and fostering of patients' dignity, rights for self-determination, and capacity to make informed decisions
- Provision of understandable information to enable patients to use according to their values

Beneficence

- The putative benefits that AI tools offer and the limits of these benefits

Equity

- The use of GenAI to promote equity according to some notion of fairness (equality of opportunity, outcomes, etc) in health or health resources across diverse groups of patient populations, and to actively prevent or remedy systemic, unfavourable outcomes in specific patient populations
- The equitable access to AI or GenAI technology

Integrity (in medical education and quality of clinical research)

- The commitment to intellectual honesty and personal responsibility to abide by responsible research conduct, including data integrity, to establish accountability and prevent harm
- The rightful acknowledgment of contributions to and ownership of intellectual work when GenAI is used in clinical research

Non-maleficence

- The prevention of harm and potential risks to patients associated with GenAI use in health care, such as incorrect or misleading outputs (hallucinations)

Privacy

- The protection of patients' information from illegitimate access, and of confidentiality of personal sensitive information

Security

- The protection of health data integrity and safety, through careful assessments of vulnerabilities in data systems and the prevention of data breaches, cyberattacks, or other threats

Transparency

- The full disclosure and thorough documentation of information regarding GenAI development, including its dataset and evaluation of performance
- The ability to access and understand the processes underlying models' outputs, especially pertaining to black-box models, in so far as this is possible

Trust

- The confidence of users in GenAI and its developers, and expectations that the model is competent in performing its prespecified tasks and behaves in ways that serves patients and the medical community
- Evidence of performance and its limitations
- The willingness to accept and integrate GenAI tools to assist delivery of care or research
- Trustworthy GenAI processes and exhibition of a range of ethically reliable properties, including performance robustness, fairness, and security, among others

*Adapted from panel 1 of Ning et al (2024) to include the principle of beneficence from the TREGAI checklist.

Figure 1: Definitions of 10 ethical principles for generative AI in a health-care context. Adapted from Panel 1 in Ning *et al* (2024) [20], with the addition of the principle of *beneficence* – denoting the moral duty to promote the well-being of others – from the TREGAI checklist.

Advice: Identify and document potential ethical risks and harms, considering the potential impact on everyone involved, including participants and researchers. For example, researchers could be exposed to sensitive or disturbing information generated by the AI, such as the re-identification of personal data or the unintended disclosure of a predicted diagnosis that a participant has explicitly chosen not to know. This situation can be particularly distressing for researchers, who may become aware of a participant's diagnosis but be unable to disclose it to them.

Seeking an ethical opinion

Before any research is undertaken, its potential ethical implications must be carefully examined. Within research organisations, this is typically addressed through governance and ethical review processes. These include evaluating risks related to information security, intellectual property, and data protection (as noted earlier), all of which should be considered before a formal ethical review by an REC. Considering such risks can help identify practical concerns, allowing RECs to focus on the ethical issues associated with these concerns – for example, inadequate data security. An REC will critically assess the ethical dimensions of your proposed research, offer feedback, and may raise risks or concerns that you had not previously considered.

> **Advice**: Review your organisation's policies and processes to understand the steps required to gain an ethical opinion on your research, including who to contact and when. Rather than viewing this as a burdensome task, consider it an opportunity to enhance your research and ensure it meets the highest ethical standards.

Given the rapidly evolving nature of AI, an ethical approval granted at the outset of a study may not remain valid throughout its duration – especially for long-term projects spanning multiple years. We strongly encourage ongoing ethical reflection and dialogue in this fast-moving landscape. Researchers should not assume that an initial favourable ethical opinion will remain sufficient for the entirety of their project. As an example, the University of East Anglia grants a favourable ethical opinion for a research project using AI for one year rather than five [15].

> **Advice**: Researchers should regularly check for changes in the terms and conditions of AI tools to ensure ongoing compliance with their data management plans and effective risk management. If any changes affect informed consent or other ethical considerations, they should seek support from governance teams and their REC.

Some research proposals may be deemed unacceptable by organisations or sponsors if the use of AI poses irreversible harm to individuals, society, or the environment [21]. Just as some sponsors refuse funding from the tobacco industry, they may also decline to support research using AI that conflicts with their values, ethical standards, or organisational principles.

Mitigating ethical concerns

To support critical reflection on the ethical risks of using AI in research, the table below outlines key areas where risks to individuals, society, and the environment may arise, alongside mitigation strategies to address them.

Impact on individuals	Mitigation strategies
Bias: AI systems can inherit and amplify human biases, producing incorrect or misleading outcomes, especially if the data they're trained on is not diverse or representative [22]. For example, in healthcare, biased AI tools used in decision-making may fail to accurately diagnose or treat patients from underrepresented groups, putting their health and well-being at risk.	Identify and reduce sources of bias by critically examining the AI algorithms, data sets, and prompts you use. This process enhances transparency about the limitations of these tools and allows you to make necessary adjustments – such as refining prompt language – to mitigate biased outcomes [23,24]. Additionally, disclose and openly discuss any identified biases [25].
Lack of transparency: It may not be clear who has access to private or sensitive information or how AI companies are accessing and using that data.	Adopt and implement Explainable AI (XAI) to make the decision-making processes of the AI system more transparent and understandable [13,14]. Stay informed about emerging best practices for the responsible handover and oversight of AI decision-making systems [26].
Loss of control over personal information: Once data is entered into an AI system, it can be difficult to remove, especially as the cross-linking capabilities of AI systems evolve.	Be aware of the limitations surrounding data withdrawal, as in some cases, complete removal from an AI system may not be feasible. Ensure participants receive full disclosure and clear documentation about how AI will be used in your research, including: • The risks and limitations of the AI system (e.g., limited accuracy or benefits) • Evidence of the system's performance and any associated uncertainties • The extent of human oversight compared to autonomous decision-making by the AI • Potential risks of future data breaches or unintended data linkage

Impact on society	Mitigation strategies
Influencing public policy and institutions: AI may shape policy and decision-making in areas such as economics, education, national security, law enforcement, and employment.	Aim to maintain a 'human-in-the-loop' approach by developing robust processes to ensure human oversight remains central to AI decision-making in your research. This can be achieved by establishing clear protocols for when and how humans intervene, ensuring AI does not replace but rather supports human judgment during the research process.
Dual-use and misuse: The use of AI in research carries the risk that research outputs may later be used for harmful purposes. For example, AI-generated outputs could be repurposed to spread false or misleading information or to support malicious applications (e.g., the misuse of synthetic data). A notable case is the misuse of AI-powered drug discovery tools to design toxic or harmful chemical compounds, illustrating the dual-use potential of such technologies [27].	Be aware of the *National Security and Investment Act (2021)*, which mandates the UK government to review and scrutinise investments involving sensitive sectors, such as AI, to protect national security. Seek guidance from relevant teams within your organisation to ensure compliance and assess potential risks.
Social bias and inequality: Biases within AI systems can amplify social disparities and exacerbate discrimination against minority groups. **Impacts on individual societies:** Some workforces may be subject to unethical practices, where they are poorly paid to label content or identify toxic content to help train AI models.	Involve diverse communities and stakeholders throughout the research process through co-production methodology to support equitable outcomes and avoid favouring specific populations [22]. In some cases where AI is purchased as a product (e.g., software), ethical procurement practices should be followed. This includes conducting proportionate due diligence on the AI's production, provenance, and continued development. Seek advice from your organisation on ethical procurement processes. *See the above section on bias in relation to individuals for additional mitigation strategies.*

Impact on the environment	Mitigation strategies
High energy consumption: Maintaining AI servers consumes vast amounts of power.	Research using AI should also be evaluated for its environment – particularly as human health is directly connected to climate change [28]. Seek guidance from within your institution to support responsible practice.
Electronic waste: The production and disposal of AI infrastructure adds to global e-waste.	
Resource depletion: AI development relies on the consumption of rare earth elements and minerals and often consumes large volumes of water [29].	
Contamination: Use of AI systems – particularly those lacking sufficient human oversight – can lead to the release of hazardous materials with harmful environmental consequences. A specific example is research involving AI-driven weapons and robotic systems. If deployed in real-world settings, these technologies could accelerate the pace and intensity of warfare, resulting in severe human costs and significant environmental damage, such as rapid soil and groundwater contamination from hazardous materials [29].	

Protecting the research record

This section aims to highlight and prompt discussions about the challenges AI poses to maintaining a robust and reliable research record, which is fundamental to research integrity. These challenges can be grouped into three main areas of concern:

1. Proliferation of poor-quality or fabricated research
2. Lack of robustness or explainability
3. Facilitation of fraudulent publishing

Proliferation of poor-quality or fabricated research

AI can incorporate or cite fake, retracted, or low-quality research into its outputs. In some cases, it can fabricate sources entirely – a phenomenon known as 'AI hallucination'.

> **Advice**: Verify your citations carefully to ensure they are not the result of fraudulent publication practices. Red flags indicating a potentially illegitimate article may include strange or oddly worded titles, unrealistically short peer review times, outdated or excessive self-citations, authors with unverifiable identities or affiliations unrelated to the paper's subject, and duplicated images, figures, or data.

Lack of robustness or explainability

Research outputs generated by AI systems are unlikely to be robust or reliable if they are based on incomplete, inappropriate, or poorly documented training sets. This concern is especially pronounced with AI systems that lack transparency – 'black box' models.

> **Advice**: Assess whether a research output clearly documents its use of AI, including a description of the AI system's limitations and the measures taken to ensure sufficient human oversight throughout its application.

Facilitation of fraudulent publishing

AI can be misused by malicious actors to enable fraudulent publishing practices. For example, 'paper mills' may leverage AI to generate fabricated manuscripts that undermine the integrity of the research record. Researchers should exercise caution when using AI to automate tasks such as literature reviews or evidence mapping. Rigorous human oversight is essential to verify the validity and reliability of AI-suggested materials, including checking for retracted, fabricated, or otherwise unreliable materials.

Advice: Use the following resources to help ensure you are maintaining sufficient human oversight when using AI to automate tasks in your research:

- Retraction Watch Database – A searchable database that tracks retracted articles and helps identify problematic research
- PubPeer – A platform for post-publication peer review where researchers can discuss and critique published work
- RedacTek – A tool that flags retracted papers, articles with high author self-citation rates, and publications with comments of concern on PubPeer
- Google Lens – A tool that searches for images across the internet and can be used to assess whether images or figures exist elsewhere

Research dissemination

This section aims to raise awareness of research integrity concerns related to the use of AI to support the dissemination of research. It focuses on AI applications during the writing, visualisation, and dissemination stages, excluding data collection. It addresses key issues related to publication ethics, including authorship and peer review, as well as funding applications and their evaluation. A brief subsection also outlines steps researchers can take to transparently report AI use, particularly in writing systematic reviews.

It is important to note that the concerns and benefits discussed here apply equally to authors across all disciplines, particularly as AI use becomes increasingly widespread throughout the research community.

Accountability and oversight as an author

AI tools, especially LLMs, are increasingly capable of producing human-like prose. However, using these tools to generate original text carries risks, as they are not designed to create accurate or internally consistent content and may produce 'hallucinations'. When determining the acceptable level of AI assistance in writing, researchers must carefully balance the benefits against the potential risks.

It is important to understand that authors remain fully responsible for everything published under their name, regardless of whether it was generated by AI. Intentionally presenting AI-written work as one's own without proper disclosure constitutes a serious breach of good research practice and is likely to be considered plagiarism or misrepresentation.

There are numerous scenarios in which improper use of an AI tool could lead to allegations of research misconduct. Such cases may fall within commonly accepted definitions of research misconduct, including those outlined in the UK's *The Concordat to Support Research Integrity* [30]. For further information on research misconduct, see UKRIO's short guide.

Advice: AI tools do not fulfil the criteria for authorship or co-authorship of research outputs, as they are not legal entities that can sign publishing agreements, and they cannot take responsibility for the content or integrity of the work. Therefore, it is unacceptable to list any AI or LLM, such as ChatGPT, as an author or co-author.

For further guidance on proper accreditation when using AI tools, refer to COPE's statement on authorship and AI tools [31].

When using AI tools in writing, it is preferable to limit their use to improve spelling, grammar, or readability without altering meaning or content. In some instances, they can improve clarity if the author is writing in a second language, has challenges with writing due to neurodiversity (e.g., dyslexia), or needs assistance to adjust text into a certain style (e.g., writing for a lay audience).

There are also several powerful and useful AI tools available for visualisation, image, and video creation. However, it is important to stress that authors must check and verify all content. Authors can use the Currency, Relevance, Authority, Accuracy, and Purpose (CRAAP) test to help evaluate the reliability of such materials [32].

Existing guidance and reports [3,4] along with the EU AI Act [21], emphasise the importance of human oversight and agency when using AI tools for writing assistance. AI should never replace a researcher's ability to generate insights, make interpretations, and draw conclusions.

When writing systematic reviews, AI tools offer the potential to save time, and enhance literature searches, helping to streamline workflow. However, this advantage may currently be offset by the time required to select an appropriate tool and verify the AI-selected literature. As of the time of writing, there is no widely accepted, standardised approach for conducting or transparently reporting AI-supported systematic reviews [33], and discussions are ongoing about their appropriate use [34]. Researchers should be aware that AI tools may draw on content from preprints that have not undergone peer review.

Declaring use for transparency

For transparency, any use of AI tools in writing (including code generation), editing, and the creation or visualisation of images and videos should be clearly declared. This disclosure can be made in the Methods or Acknowledgements sections, or in a separate section as specified by the guidelines or policies of the publisher, funder, employer, or ethical review body. It is good practice to declare the use of generative AI or AI-assisted technologies when developing initial research, such as ideas and theories.

Some journals, publishers, funders and employers have established strict guidelines or policies regarding AI use in academic writing and visualisation (e.g., research funders joint statement) [35,36]. Since this is an evolving area, it is important to stay updated on any changes and verify requirements frequently, especially prior to submission.

Advice: Refer to guidance from publishers or other bodies on how to disclose AI use in research outputs, particularly those tailored to your field or discipline.

Useful examples include:

- The JAMA Network's *Reporting Use of AI in Research and Scholarly Publication*
- Nature's AI Policy
- Taylor and Francis' AI Policy

Researchers should stay informed of evolving best practices and seek clarification from the providers of AI tools when needed. As with all research, detailed record-keeping is essential. Employing version control and tracking changes in writing files can enhance transparency.

Advice: Seek up-to-date guidance on best practice for tracking and recording AI use. When developing your data management plan, include strategies for transparently documenting how AI is used in your research – such as cataloguing AI-generated prompts and responses.

Two useful resources for this are Edith Cowan University's Prompt anatomy and AI Data Management template.

Reporting methodology

Like any equipment or methodology, the use of AI must be accurately reported in research outputs to support transparency, replicability, and open research practices. However, implementing this principle is often challenging due to the diversity and rapid evolution of AI tools, and the limited transparency of many commercial systems (e.g., black-box algorithms), which can make full disclosure difficult. Despite these challenges, best practices for reporting the use and development of AI in research are actively evolving, as accurate reporting is critical for maintaining trust, ensuring reproducibility, and upholding research integrity. A recent Royal Society report highlights the persistent barriers to reproducibility in this area [3].

Accurate reporting of data collection processes, methodology, and results is especially critical when using AI (e.g., predictive AI) in research that directly impacts clinical care pathways [37]. The JAMA Network has published helpful guidance on reporting AI use in research and scholarly publications, including examples of AI-related reporting guidelines specific to clinical research [38].

Advice: Acknowledgement should specify what type of AI was used, how it was used, and what role it played in shaping the final output. Transparency builds trust and helps protect the credibility of your work.

Publishers' and peer reviewers' use of AI

Publishers are increasingly using AI to support various editorial tasks, including proofreading, editing, formatting, technical screening, reviewer assignment (e.g., Springer Nature's Reviewer Finder), and even aspects of editorial decision-making, such as assessing journal fit or determining acceptance/rejection outcomes [31,39]. The ethical risks previously discussed – especially those relating to bias and privacy – are equally relevant to editors and reviewers using AI. However, publicly available guidance on the responsible and transparent use of AI in editorial processes remains limited. This landscape is likely to evolve as publishers continue to develop and refine their policies [40]. Some, such as SAGE Publishing, have already introduced clear boundaries: for example, advising editors not to use AI tools to triage manuscripts, summarise reviews, or draft decision letters.

When acting as a peer reviewer – whether for scholarly content or funding applications – you are entrusted with privileged access to research that has not yet entered the public domain. If you are considering using AI tools to support your review, it is essential to reflect on the implications. Key questions to consider include:

- Who owns the materials you are inputting into the AI?
- Could using AI breach the confidentiality agreement you accepted as a peer reviewer?
- Might you unintentionally release research into the public domain, potentially compromising the authors' ability to publish or compromising their intellectual property?

Advice: In the absence of comprehensive policies to refer to, authors and peer reviewers are encouraged to seek clarity from editors and publishers about the use of AI in editorial processes and peer review. This includes understanding which tools are used, how they are applied, and what constitutes acceptable use.

Creativity and critical thinking

This section explores how the use of AI in research may affect creativity and critical thinking. It also emphasises the importance of AI literacy and training, and it provides resources to support responsible and informed use.

Impact on creativity and critical thinking

Creativity and critical thinking are fundamental to the research process. Researchers develop these skills as they advance their careers, enabling them to formulate original questions, evaluate evidence with discernment, and develop innovative solutions to complex problems. While AI can support these skills, it also has the potential to undermine them and their development [41].

As Carobene *et al*. (2024) observe, there is a need to 'maintain a balance between AI's capabilities and fostering independent reasoning and creativity' [42]. Overreliance on AI in research risks discouraging novel or unconventional lines of inquiry, dampening creative insights and intellectual diversity, and reducing critical engagement with findings. There is a concern that if AI is used as a substitute for deep analytical and conceptual work, early-career researchers may not fully develop the foundational skills necessary for rigorous, independent research.

The use of AI to assist in writing funding proposals or ethics review applications is becoming increasingly common. However, undisclosed use may be viewed unfavourably. For example, if a student relies on AI to complete an ethics application without disclosure, it may be perceived as bypassing the educational purpose of the assignment. The use of AI should also be critically assessed in contexts that require a nuanced understanding of human experience, such as clinical case studies [43].

> **Advice**: Consider carefully when and how to use AI. For example, using AI tools during confidential peer review processes – such as for grant or publication reviews – may breach ethical or professional standards.

Assessing training needs

Developing AI literacy is essential to understanding the capabilities, limitations, and responsible use of AI within your discipline. Reflect on your current level of knowledge and identify any areas where training or further support is required – particularly on topics such as bias and transparency [44]. Where appropriate, seek opportunities for more in-depth learning and engage in broader discussions or case study analyses, with colleagues as recommended by Paschke *et al* (2024) [45].

Many research organisations are expanding their training provision to researchers in acquiring valuable hybrid skills for interdisciplinary work involving AI. There are a range of resources, open-access online training courses, and case studies available, including:

- MIT – *MIT FUTURE OF AI*

- University of Glasgow – *Generative AI for Students: Ethics and Academic Integrity* (*Note: access is free, but a certificate of completion requires payment*)
- Princeton University – *Case Studies in AI Ethics*
- Eneri – *AI and Ethics Case Study*
- Ada Lovelace Institute – *Looking Before We Leap: Case Studies in AI Ethics*

Conclusion

A comprehensive and responsible approach to the use of AI in research is essential to safeguard research integrity. AI holds great potential to enhance research processes, but these benefits must be carefully weighed against potential risks to ethical standards, the rigour of the research record, and the wider societal and environmental impacts.

As AI becomes more integrated into research practice and everyday activities, researchers must critically assess its role and influence in their work. Upholding principles of transparency, accountability, and ethical responsibility is vital to maintaining trust – both within the research community and with the public.

This guidance does not offer an exhaustive guide on the use of AI in research. Instead, it highlights the importance of using AI with care, intention, and openness to protect research integrity. Researchers have a shared responsibility for ensuring AI is integrated into research practice in a way that is thoughtful and informed. Doing so will help ensure that AI contributes positively to the research process, upholding the highest standards of integrity and producing work that benefits society.

References

1. *What is Research Integrity? - UK Research Integrity Office*. (n.d.). Retrieved 10 June 2025, from https://ukrio.org/research-integrity/what-is-research-integrity/

2. *Definitions | ICO*. (n.d.). Retrieved 14 June 2025, from https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/definitions/

3. The Royal Society. (2024). Science in the age of AI: How artificial intelligence is changing the nature and method of scientific research. *The Royal Society*. https://royalsociety.org/-/media/policy/projects/science-in-the-age-of-ai/science-in-the-age-of-ai-report.pdf

4. European Commission. (2024). Living guidelines on the responsible use of generative AI in research. *European Commission*. https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf

5. *A pro-innovation approach to AI regulation - GOV.UK*. (n.d.). Retrieved 26 June 2025, from https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#part-2-the-current-regulatory-environment

6. *Research your AI research tools - UK Research Integrity Office*. (n.d.). Retrieved 10 June 2025, from https://ukrio.org/ukrio-resources/research-your-ai-research-tools/

7. Hosseini, M., & Resnik, D. B. (2024). Guidance needed for using artificial intelligence to screen journal submissions for misconduct. *Research Ethics*, *21*(1), 1–8. https://doi.org/10.1177/17470161241254052

8. Zeni, C., Pinsler, R., Zügner, D., Fowler, A., Horton, M., Fu, X., Shysheya, S., Crabbé, J., Sun, L., Smith, J., Nguyen, B., Schulz, H., Lewis, S., Huang, C.-W., Lu, Z., Zhou, Y., Yang, H., Hao, H., Li, J., … Xie, T. (2023). *MatterGen: a generative model for inorganic materials design*. https://arxiv.org/abs/2312.03687v2

9. Gkoumas, D., Wang, B., Tsakalidis, A., Wolters, M., Purver, M., Zubiaga, A., & Liakata, M. (2024). A longitudinal multi-modal dataset for dementia monitoring and diagnosis. *Language Resources and Evaluation*, *58*(3), 883–902. https://doi.org/10.1007/S10579-023-09718-4/TABLES/5

10. *Text Correction for Historical Documents | Sheffield AI Research Engineering*. (n.d.). Retrieved 21 November 2024, from https://shef-aire.github.io/project/historical-doc-text-correction/

11. Tkachenko, N. (2024). Opportunities for synthetic data in nature and climate finance. *Frontiers in Artificial Intelligence*, *6*, 1168749. https://doi.org/10.3389/FRAI.2023.1168749

12. *Spatial Finance 2.0: Combining Large Language Models (LLMs) with remote sensing to discover the most polluting economic assets | Research Communities by Springer Nature*. (n.d.). Retrieved 21 November 2024, from

https://communities.springernature.com/posts/spatial-finance-2-0-combining-large-language-models-llms-with-remote-sensing-to-discover-the-most-polluting-economic-assets

13. Royal Society, T. (n.d.). *Explainable AI: the basics POLICY BRIEFING*.

14. Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Ser, J. Del, Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., Jiang, R., Khosravi, H., Lecue, F., Malgieri, G., Páez, A., Samek, W., Schneider, J., Speith, T., & Stumpf, S. (2024). Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion*, *106*, 102301. https://doi.org/10.1016/J.INFFUS.2024.102301

15. University of East Anglia, Research and Innovation Services, & Postgraduate Research Service. (2024). Generative AI Policy for Research and Innovation. https://www.uea.ac.uk/documents/20142/1553813/uea_generative_ai_policy_for_research_and_innovation_-_20_03_24.pdf/8ddb7e71-9cd4-b689-5e0f-3fa36a155b76?t=1711376037035

16. *AI in Research*. (n.d.). Retrieved 13 June 2025, from https://warwick.ac.uk/services/ris/research-integrity/airesearch/

17. *University of Glasgow - Research - Strategy and policies - Our policies & guidance - AI Guidance for Researchers*. (n.d.). Retrieved 13 June 2025, from https://www.gla.ac.uk/research/strategy/ourpolicies/ai-for-researchers/

18. *Policy and guidance for staff on the use of artificial intelligence (AI) | University of West London*. (n.d.). Retrieved 13 June 2025, from https://www.uwl.ac.uk/about-us/policies-and-regulations/staff-use-artificial-intelligence-AI

19. *GDPR exemptions | Research Support*. (n.d.). Retrieved 13 June 2025, from https://researchsupport.admin.ox.ac.uk/policy/data/exemptions#collapse470196

20. Ning, Y., Teixayavong BSS, S., Shang, Y., Miao, D., W Ting, D. S., Liu, M., Vaughan, R., Liu, N., in Health Services, P., Research Vaughan, S. R., H Ong MPH, M. E., Savulescu, J., Dunn, M., Ning, Y., Teixayavong, S., Shang, Y., Savulescu, J., Nagaraj, V., Miao, D., … Liu, N. (2024). Centre for Quantitative Medicine. Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist. *The Lancet Digital Health*, *6*, e848–e856. https://doi.org/10.1016/S2589-7500(24)00143-2

21. EU Artificial Intelligent Act. (2024). *The EU Artificial Intelligence Act*. https://artificialintelligenceact.eu/

22. Glickman, M., & Sharot, T. (2024). How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour 2024*, 1–15. https://doi.org/10.1038/s41562-024-02077-2

23. *Bias in AI | Chapman University*. (n.d.). Retrieved 16 June 2025, from https://www.chapman.edu/ai/bias-in-ai.aspx

24. *Understanding Prompt Bias and How to Overcome It - Future Skills Academy*. (n.d.). Retrieved 16 June 2025, from https://futureskillsacademy.com/blog/prompt-bias-in-ai/

25. *Living guidelines on the responsible use of generative AI in research | Research and innovation*. (n.d.). Retrieved 21 November 2024, from https://research-and-innovation.ec.europa.eu/document/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en

26. *How to share and implement data analytics and AI tools*. Retrieved 29 October 2024, from https://senseaboutscience.org/wp-content/uploads/2024/09/Responsible_handover_guide.pdf

27. Urbina, F., Lentzos, F., Invernizzi, C., & Ekins, S. (2022). *Dual Use of Artificial Intelligence-powered Drug Discovery HHS Public Access*. *4*(3), 189–191. https://doi.org/10.1038/s42256-022-00465-9

28. Lannelongue, L., Aronson, H. E. G., Bateman, A., Birney, E., Caplan, T., Juckes, M., McEntyre, J., Morris, A. D., Reilly, G., & Inouye, M. (2023). GREENER principles for environmentally sustainable computational science. *Nature Computational Science 2023 3:6*, *3*(6), 514–521. https://doi.org/10.1038/s43588-023-00461-y

29. Programme, U. N. E., & Council, I. S. (2024). *Navigating New Horizons: A global foresight report on planetary health and human wellbeing*. 95. https://wedocs.unep.org/xmlui/handle/20.500.11822/45890

30. *The Concordat to Support Research Integrity*. (n.d.). Retrieved 16 June 2025, from https://www.universitiesuk.ac.uk/what-we-do/policy-and-research/publications/concordat-support-research-integrity

31. COPE. (2021). Artificial intelligence (AI) in decision making. *Committee on Publication Ethics*. https://publicationethics.org/sites/default/files/ai-in-decision-making-discussion-doc.pdf

32. California State University Meriam Library. (2010). Evaluating Information - Applying the CRAAP Test. *California State University Meriam Library*. https://libguides.csuchico.edu/c.php?g=414315&p=2822716

33. van Dijk, S. H. B., Brusse-Keizer, M. G. J., Bucsán, C. C., van der Palen, J., Doggen, C. J. M., & Lenferink, A. (2023). Artificial intelligence in systematic reviews: promising when appropriately used. *BMJ Open*, *13*(7), e072254. https://doi.org/10.1136/bmjopen-2023-072254

34. Fabiano, N., Gupta, A., Bhambra, N., Luu, B., Wong, S., Maaz, M., Fiedorowicz, J. G., Smith, A. L., & Solmi, M. (2024). How to optimize the systematic review process using AI tools. *JCPP Advances*, *4*(2). https://doi.org/10.1002/jcv2.12234

35. Wellcome Trust. (2023, September). *Funders joint statement: use of generative AI tools in funding applications and assessment*. https://wellcome.org/what-we-do/our-work/joint-statement-generative-ai

36. Jocelyn Kaiser. (2023). Funding agencies say no to AI peer review. *Science*, *381*(6655), 261–261. https://www.science.org/content/article/science-funding-agencies-say-no-using-ai-peer-review

37. Kolbinger, F. R., Veldhuizen, G. P., Zhu, J., Truhn, D., & Kather, J. N. (2024). Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. *Communications Medicine*, *4*(1), 71. https://doi.org/10.1038/s43856-024-00492-0

38. Flanagin, A., Pirracchio, R., Khera, R., Berkwits, M., Hswen, Y., & Bibbins-Domingo, K. (2024). Reporting Use of AI in Research and Scholarly Publication—JAMA Network Guidance. *JAMA*, *331*(13), 1096. https://doi.org/10.1001/jama.2024.3471

39. Leung, T. I., de Azevedo Cardoso, T., Mavragani, A., & Eysenbach, G. (2023). Best Practices for Using AI Tools as an Author, Peer Reviewer, or Editor. *Journal of Medical Internet Research*, *25*, e51584. https://doi.org/10.2196/51584

40. Taylor Swaak. (2023). 'We're All Using It': Publishing Decisions Are Increasingly Aided by AI. That's Not Always Obvious. *The Chronicle of Higher Education*. https://www.chronicle.com/article/were-all-using-it-publishing-decisions-are-increasingly-aided-by-ai-thats-not-always-obvious?

41. Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments 2024 11:1*, *11*(1), 1–37. https://doi.org/10.1186/S40561-024-00316-7

42. Carobene, A., Padoan, A., Cabitza, F., Banfi, G., & Plebani, M. (2024). Rising adoption of artificial intelligence in scientific publishing: evaluating the role, risks, and ethical implications in paper drafting and review process. In *Clinical Chemistry and Laboratory Medicine* (Vol. 62, Issue 5, pp. 835–843). Walter de Gruyter GmbH. https://doi.org/10.1515/cclm-2023-1136

43. Phillips, N., & Rees, M. (2023). Artificial intelligence, authorship and publication of case reports. *Case Reports in Women's Health*, *39*, e00553. https://doi.org/10.1016/j.crwh.2023.e00553

44. Paschke, M., Petterson, A., Mihalka, R., & Sudau, M. (2024). *Teaching Collection: Exercises and hands-on examples for ethical use of generative AI*. https://doi.org/10.3929/ETHZ-B-000672206

45. Paschke, M., Mihálka, R., & Sudau, M. (2024). *Cases for Research Integrity: Generative AI*. https://doi.org/10.3929/ETHZ-B-000664648

Authorship and contributions

Authors

Dr Josephine Woodhams – *conceptualisation, investigation, writing*

Contributors and Reviewers

Kathryn Dally – *reviewing*

Stephanie Neave – *reviewing*

James Parry – *reviewing*

Jasper Scott – *copyediting, formatting*

Competing Interests

Authors and contributors are employees of UKRIO.

Funding

This work was funded by UKRIO.

Additional contributions

This guidance was also reviewed by representatives from UKRIO's subscriber organisations who took part in our roundtable on AI use in research, held on 27 February 2025. We are grateful to all participants for their valuable insights and feedback.

We would like to give special thanks to Dr Helen Brownlee (Research Integrity Manager, University of East Anglia), Dr Andrew Porter (Research Integrity and Training Adviser, Cancer Research UK Manchester Institute), and Dr Scott Summers (chair of the UEA Gen-AI Light Touch Review Group, University of East Anglia) for their continued support, providing expert knowledge that has helped shaped UKRIO's priority work stream on AI.

Disclaimer on external resources and third-party content

This guidance document includes references and links (including hyperlinks) to external training resources, tools, and materials intended to support researchers in understanding the evolving use of AI in research. These signposted resources are provided for convenience and are not intended to be comprehensive.

Inclusion of any external link or reference does not imply endorsement or approval by UKRIO of the linked website, its operator, or its content. UKRIO is not responsible for the accuracy, content, or quality of third-party materials, nor for any websites outside of www.ukrio.org.

Researchers are encouraged to use their professional judgment when engaging with external content and to assess the relevance, reliability, and suitability of each resource within their specific context.

UKRIO

RESEARCH INTEGRITY OFFICE

Promoting integrity and high ethical standards in research
Providing confidential, independent and expert support

The UK Research Integrity Office (UKRIO) is an independent charity, offering support to the public, researchers and organisations to further good practice in academic, scientific and medical research. We pursue these aims through a multi-faceted approach:

- Education via our guidance publications on research practice, training activities and comprehensive events programme.
- Sharing best practice within the community by facilitating discussions about key issues, informing national and international initiatives, and working to improve research culture.
- Giving confidential expert guidance in response to requests for assistance.

Established in 2006, UKRIO is the UK's most experienced research integrity organisation and provides independent, expert and confidential support across all disciplines of research, from the arts and humanities to the life sciences. We cover all research sectors: higher education, the NHS, private sector organisations and charities. No other organisation in the UK has comparable expertise in providing such support in the field of research integrity.

UKRIO welcomes enquiries on any issues relating to the conduct of research, whether promoting good research practice, seeking help with a particular research project, responding to allegations of fraud and misconduct, or improving research culture and systems.

UK Research Integrity Office

Impact Hub London Euston, 1 Triton Square, London NW1 3DX
Email: info@ukrio.org Web: www.ukrio.org
Registered Charity No: 1147061 Registered Company No: 7444269

For the full list of UKRIO publications, visit www.ukrio.org