Part A: Ethical Use of AI in Research

Note: This guidance contains suggested principles and ethical considerations for the research community when using AI in research, rather than formal policy. It is intended to be updated in line with emerging digital policy, aligned with AI Steering Group developments and to be informed by a project commissioned by the UK Committee on Research Integrity (UKCORI).

The guidance should be read in conjunction with Part B: Use of Generative AI in Research Degree Programmes & Assessments.

Introduction

This guidance aims to help UoW Researchers to engage with, and benefit from, the opportunities of Generative AI tools for Research, whilst protecting against potential ethical, legal, security and integrity issues. Researchers must have due regard to matters of research integrity as detailed in our Code of Good Research Practice.

Artificial Intelligence (AI)

The umbrella term for computer systems designed to simulate human intelligence, reasoning, and problem-solving, including machine learning, computer vision, and natural language processing.

Generative AI (Gen AI)

Generative (AI), a subset of AI often referred to as Gen AI, is an overarching term for web-based tools that use algorithms, data, and statistical models to draw reasonable inferences to generate content of their own. These tools rely on large amounts of training data to generate content using natural language that gives the appearance of a human response. This content can include text, images, video, or code. It operates reactively to user prompts, providing a "one-and-done" output.

It is important to remember that however real the response, it is the outcome of probability maths and programming. There is no wider understanding of the world or the context, and this leads to many of the errors Gen AI produces.

Agentic AI

Systems that use Gen AI as a foundation which are designed to operate autonomously, making decisions, planning, and using tools to achieve complex, long-term goals with minimal human supervision. It is proactive, capable of multi-step reasoning, focusing on accomplishing specific tasks or goals (e.g., scheduling, managing workflows) acting on behalf of the user.

Typical Use of Gen AI in Research

Gen AI can improve efficiency and productivity in research. Typical uses include:

- facilitating idea generation and research design
- improving content and structuring
- supporting literature review and synthesis
- enhancing data management and analysis
- supporting editing, review, and publishing
- assisting in communication, outreach, and ethical compliance.

Microsoft Copilot and other tools

If researchers decide to use Gen AI to assist with their research, the University **strongly recommends using Copilot for Microsoft 365** which is the University's institutionally licensed GenAI platform. This platform is free to use for all staff and students and comes with commercial data protection. It is very important to be careful about the information you provide to generative AI tools; using Copilot minimises data protection and privacy risks.

Log in with your University login details to ensure you are working within the Enterprise version, which means the information you provide to Copilot is not stored or used except to answer your prompts.

The University currently **prohibits** researchers from inputting any Personal (including Special Category Data), Confidential, Third Party, or UoW Business Critical Data / information / material into a Gen AI tool, without prior approval from the Data protection Team (dataprotection@wlv.ac.uk) to ensure compliance with Data Protection legislation and to confirm the security of the tool. This is likely to include almost all research data.

There are risks associated with using Co-Pilot Enterprise that researchers should be aware of:

- **Data Oversharing and Exposure:** Copilot accesses data via Microsoft Graph, meaning if research files, datasets, or notes have broad permissions (e.g., "shared with everyone in the company"), Copilot will surface that data to any user who asks, even if they aren't part of the research team.
- **Data Misclassification and Leakage:** Inconsistent sensitivity labels can result in restricted research data being improperly processed or exposed.
- **Intellectual Property and Data Retention:** While Microsoft offers enterprise data protection, risks remain around how long data is cached or if confidential data is utilised in prompts that are logged.
- **Lack of Transparency:** It may be difficult to trace exactly what research data Copilot accessed or used to generate a specific summary, complicating audits

Ethical Use of AI in Research

It is crucial to exercise appropriate caution and adopt responsible practices when utilising Gen AI tools. Ethical AI use in research requires human accountability, transparency, and data privacy, ensuring tools augment rather than replace human judgment. Researchers should ask themselves:

- Why are you using this tool? What are the benefits, and do you know how to use it properly?
- What are the risks of using this tool and how can you mitigate these?
- Do you have access to the information and resources you need to answer these questions honestly and accurately?

Gen AI can help us achieve great things, but it can come at an ethical, social, environmental and human cost.

Principles for Ethical AI Use

1) Data Use
2) Transparency
3) Quality
4) Responsibility

1) Data Use

Privacy, confidentiality, security and ownership.

Sensitive or confidential information should not be entered into public AI tools (i.e. AI tools designed for widespread, easy access by the general public, often available through websites, apps, or APIs) as this can be exposed or used for AI training.

Critical questions:

Where is data uploaded to this tool stored? Who has access to this?

Does the tool support privacy and adhere to GDPR?

Who owns the data generated by this tool? How will they use this?

Does this tool use 'data scraping'? Has it been trained on copyrighted material?

Copyright, ownership and intellectual property

Gen AI output imitates or summarises existing content, mostly without the permission of the original content owners. The output's appearance of creativity and originality generates challenges. It is important to keep in mind when using Gen AI tools that there are issues of copyright, ownership, intellectual property and lack of legislation in this area.

You should not copy and paste any copyrighted text, or other sensitive or personal data, into an AI tool for it to use. The AI tool could incorporate this data into its training dataset, which could then be used illegally or unethically.

Due to the large amounts of data being processed by Gen AI tools, it is easy for researchers to inadvertently use or reveal sensitive information hidden among anonymised data in the tool. Researchers should be careful to only input the amount of anonymised data they need for their Research purposes to reduce the likelihood of a 'linkage' between datasets in the Gen AI tool, that could enable re-identification of anonymous data records.

Uploading confidential information into AI tools for any work that may be patentable carries significant legal and commercial risks. Many AI platforms process inputs on external servers, and unless strict safeguards and enterprise agreements are in place, the information provided may be stored, reused for model training, or potentially accessed by third parties. This can jeopardize the novelty requirement for patentability, as unintended disclosure, even without formal publication, may be interpreted as prior art. Members of staff should therefore exercise caution when using AI, to insure only information that would not count as a disclosure is shared.

Data that should <u>never</u> be input into any AI software (including Co-Pilot)

- passwords and usernames
- personally identifiable information or other sensitive or confidential material
- personally identifiable or confidential information that has been processed (including but not limited to summarisation or pseudonymisation) unless specific consent for use with AI tools has been gained from the research participant
- any data that is not fully consistent with University's policies on Data Protection, Research Data Management Policy, GDPR, Academic Authorship Policy and Ethics Policy & Guidance
- any data that is protected by Copyright, unless explicit permission for its use with AI tools has been obtained
- any data which might result in reputational damage to UoW
- any data from third parties where the individual has not explicitly consented for their data to be used with AI, with the exception of data that is clearly already in the public domain
- any data from third parties where the explicit use of the data with AI has not been authorised by a University Ethics Subject Panel, irrespective of whether the data is in the public domain.

2) Transparency

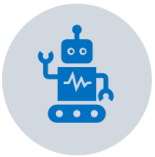
Reporting, explainability, awareness, reproducibility

- Researchers should declare the use of AI tools**.**
- AI cannot be listed as an author or co-author, as it cannot take responsibility for the research.

Critical questions:

How and when will you report your use of AI tools?

How will authorship be accredited if you have used generative AI?



Can you explain how the AI tool arrives at a decision?

Who is it important to be transparent to? How will you support their understanding or awareness of your use of AI?

Acknowledging use of Generative AI

Researchers are required to truthfully acknowledge what elements of their work is wholly or partially generated, modified or proof-read using an AI tool. This should be declared clearly in the document in which it occurs, whether the document is for internal or external use.

Researchers must also detail any use of Generative AI in collecting, analysing or otherwise processing Research Data in their Data Management Plan. Researchers should explain the reasons for using an AI tool, including an evaluation of the risks associated with using that tool.

Gen AI is evolving rapidly and there is not yet consensus on how to acknowledge and reference it. The minimum requirement to include is:

- Name and version - Copilot
- Publisher - Microsoft
- URL - https://m365.cloud.microsoft/chat
- Brief description (single sentence) of the context in which you used the tool.

Further guidance on how you should cite and reference Gen AI can be found on the Cite Them Right **website**. The Harvard section has the most detailed guidance on ways of citing Gen AI, and therefore it is recommended that Harvard guidance is applied where no guidance exists.

If you have used a Gen AI tool to produce part of your work, but have not acknowledged it, this could be investigated under the Procedure for the Investigation of Allegations of Misconduct in Research

Reproducibility
Generative AI poses challenges to traditional scientific reproducibility due to the "black box" nature of models, random components in inference, and lack of access to training data.

Using GenAI while maintaining reproducibility requires documenting the specific AI model, version, prompts, and datasets used. Outputs should be audited for hallucinations.

Recording prompt texts
Prompt texts determine how the Gen AI tool creates its response. Different prompts produce different results and therefore prompt text and outputs that the Gen AI tool has produced should be saved. If the prompt texts are recorded and shared, then others can understand more about the process of generating the response. This will aid transparency, integrity and open research practices. You should be aware that you will receive different outputs when using the same prompts.

Research publications & funding applications
Authors are accountable for the accuracy, integrity and originality of their research outputs, including any use of Gen AI. Research outputs must be the authors' own work, not presenting others' work or output from Gen AI tools without appropriate citation and referencing.

Funders also advise caution in relation to the use of Gen AI tools in developing funding applications including with collaborators. Individual journals, publishers and funding bodies may have specific requirements which must be followed.

Using GenAI in the peer review process.
The peer review process is a cornerstone of the research cycle. It is underpinned by high standards of rigour and integrity. Many funders and publishers prohibit the use of GenAI in the assessment of funding applications or research publications. The nature of peer review at any stage involves high ethical standards and confidentiality to safeguard academic discourse and intellectual property. Funding applications and their assessment remain confidential throughout the assessment process and therefore this data must not be inputted into any GenAI tool.

Many journals also take this stance, restricting or prohibiting the use of GenAI in peer review in order to: stop breaches of confidentiality, loss of rigour and specificity, fraudulent misrepresentation of outputs and peer review manipulation. Ultimately, it is human input that provides rigorous assessment, and both allow for evaluation and innovation.

3) Quality

Bias, accuracy, oversight.

- Researchers must actively identify and control for biases in AI algorithms, which can amplify existing prejudices in training data.
- AI-generated content must be rigorously checked for accuracy.

Critical questions

How will you know if the outputs of the AI tool are accurate?

Is there a risk the tool will reproduce harmful biases?

How will you maintain oversight and accountability over the use of AI?

Can you ensure that the use of this tool complements human capabilities, rather than diminishing them?

Lack of trust and authenticity
Gen AI can generate information that appears factual but is often inaccurate. This is often called AI hallucinations. We must remember that:

- although Gen AI models appear to understand the content that they use and generate, they do not understand it
- the data that Gen AI models use for training have lots of inaccuracies and biases in them
- Gen AI can also easily create fake news, misinformation and 'deep fakes'.

Most AI models are created to provide a likely output based on prompts and training. Their outputs are designed to appear convincing even when there is no factual basis for the output. Consequently, 'facts' provided by these tools may appear to be trustworthy, but that appearance is false. Both input data and prompts can lead to bias in the output. As a result, ALL outputs from all AI tools must be independently verified for truthfulness.

Data accuracy
Researchers must take all reasonable steps to make sure that any personal data that is entered into a Gen AI tool is not "incorrect or misleading as to any matter of fact".

Biases
Gen AI tools can inadvertently perpetuate or amplify societal biases due to biased training data or algorithmic design. To minimise discrimination and ensure fairness, it is crucial for Researchers to mitigate against the Gen AI tool creating biases or discriminatory outcomes at the point of application by, for example, carefully checking the quality of the data for any biases, or their coding if they are building / developing a Gen AI tool.

Researchers using large language models (LLMs) face a significant risk of inadvertently reinforcing, amplifying, or recycling societal prejudices, stereotypes, and biases present in training data. Because LLMs are trained on vast, uncurated internet datasets, they inevitably inherit and can perpetuate biases related to gender, race, culture, and age.

4) Responsibility

Harm, accountability, fairness, integrity, environmental and societal impact.

- Researchers are fully responsible for the integrity of their work, requiring them to check for AI-generated hallucinations, fabricated citations, or plagiarised content.
- AI should not be used to fabricate or falsify data.
- Research involving AI should include a risk assessment for potential harms. We cannot realistically remove every risk when using a Generative AI tool, but researchers need to assess the potential risks and implement measures to sufficiently mitigate against them.

Critical questions:

Have you identified any risks of harm that might arise from using this tool?

Does your use of AI exclude anyone, or worsen inequalities?

What are the potential negative environmental or societal impacts of this use of AI? How can you address these?

Environment & Sustainability
Researchers should always consider the sustainability of their research in line with our Environmental, Social & Governance Strategy 2035. Globally there are both positive and negative impacts of AI on sustainability development goals. These are clearly articulated in the Nature article 'The Role of Artificial Intelligence in Achieving the Sustainable Development Goals'.

Ethical, Social and Human costs
After training, the Gen AI model is often checked and refined in a process known as Reinforcement Learning from Human Feedback (RLHF). In RLHF, human beings review the Gen AI responses and validate them. This ensures that the Gen AI responses are appropriate, accurate and align with the intended purpose. There have been issues of exploitation of workers which had a massive negative impact on many of those who were involved, including experiencing trauma.

Gen AI also tends to output standard answers that replicate the values and biases of the creators of the data used to train the models. This may constrain the development of opinions and further marginalise already marginalised voices.

Further Reading

Internal: Canvas staff help & guidance 'UoW Introduction to Generative AI guidance '

External: For further reading on ethical guidelines, you can consult resources from:

- UNESCO Ethics of Artificial Intelligence
- UK Government Understanding artificial intelligence ethics and safety
- UKRIO Embracing AI with Integrity
- OECD Principles for Trustworthy AI
- The Russell Group Principles on the use of generative AI in education, 2023.
- COPE position statement on the role of GenAI in publications, 2023.