

## AI and Research Integrity (v.1)

### Definitions

Artificial intelligence is the field of computer science that studies ways that mimic how beings (humans or any other group) can deal with complex information with the aim of processing problems and identify their possible solutions.

Generative AI is specifically the use of a group of techniques called machine learning to produce combinations of outputs to specific inputs which are always based on historical data. The data and outputs can be text, computer code, images, voice, music and video.

Open AI systems in this document refers to those where data inputted by users is retained by the system and is reused, e.g. to contribute to its learning. This is contrasted with private AI systems, where the processing power of the system can be used, but the data inputted by users isn't retained or reused by the system. The latter tends to be a paid-for service, accessible through formal licencing arrangements.

### Challenge

AI presents both opportunities and threats to many aspects of life, including research integrity. This policy note seeks to distinguish what are legitimate uses of it in research, versus what applications of it are potential research misconduct or fundamentally unethical endeavours.

### Scope

This policy applies to all Sheffield Hallam University staff and students (doctoral, masters and undergraduate). This policy note deals only with principles. It does not refer to specific systems or detection solutions, as these are continuing to evolve at a rapid speed.

### Principles

As with all research integrity, the fundamental principles of good research practices remain the same: transparency, informed consent, confidentiality of personal data, effective sharing, accurate authorship/acknowledgement and fair intellectual property.

### Specific Issues

#### 1) Ethical Considerations When Using AI Tools

The use of AI in research can be appropriate and powerful. But at times the use of AI can also be harmful – e.g. when there are unintended consequences on livelihoods, or discrimination by bias systems.

The foundational principle when using AI therefore is that researchers must use AI responsibly. This means that researchers should be aware of the limitations of using AI in their project and the potential consequences this may have, and put any necessary mitigations in place.

Consequently, before using AI, all researchers should consider whether their use of AI upholds the ethical principles of research integrity. To this end, the purpose and potential unintended consequences of any research involving AI must be outlined in the ethics application for the project.

One of the limitations researchers must consider is potential bias because of different representation in training data. Within ethics applications, researchers should therefore actively consider the quality of the input data of AI: what, how and for what purpose is the basis of its programming; what are its biases and discriminations? Efforts should be made to assure that, where it is used, its input data tries to represent the best of human values, not necessarily the sum of them (e.g. where learning is based on the content of the internet).

The choice of AI tool should also be considered in ethics applications. It is after all the combination of input data and algorithm which may result in biased outputs.

For the sake of rigour, transparency, and reproducibility/replicability, especially when AI tools are used to process data, researchers should choose a tool that allows the researcher to understand and clearly explain the decisions or predictions made by the artificial intelligence, so that these decisions and predictions are more understandable and transparent. If this is not possible, artificial intelligence is basically a 'black box', where the relationship between input and output is opaque. This endangers the integrity of the research undertaken.

## **2) Sharing Information with AI Tools**

Researchers should be aware that sharing information with AI systems generally means the data is retained and reused by those systems, and will potentially become widely accessible. This presents opportunities in terms of open research. However it also presents potentially serious concerns regarding compliance with UK data protection laws, intellectual property and commercial sensitivities regarding the data, and adhering to the specific terms of participants' consent under which the data was collected.

### **2.1 Live Data**

Where data derived from any form of human participation (including tissue), live datasets must not be uploaded to open third party AI systems. This would likely be a breach of data protection laws and would constitute serious research misconduct.

Simply removing identifiers from the data (pseudonymising) will likely still contain information that may be identifiable, e.g. quotations and opinions, that the participants will not normally have consented to the wider sharing of. Keys will also still exist to be able to reverse the anonymising. Pseudonymising therefore is not an adequate assurance of data protection compliance to a level sufficient to permit the uploading of live human data to open AI.

Even with non-human data (e.g. from subject areas such as engineering, physical sciences or humanities), researchers should consider the consequences of uploading this – that it could potentially be accessed and used by other actors before they have established any form of intellectual property claim over it (as would normally happen through post-analysis publication of results), or indeed assured themselves of the quality of the data (thereby potentially initiating the propagation of flawed data).

'Open third party' systems in particular is the key restriction – alternative permitted solutions include obtaining licences to use the analytical capabilities of external systems but where the inputted

research data does not consequently feed back into it, and the licence in all other ways guarantees GDPR compliance (so not open); or researchers developing their own systems (so not third party).

Resources are in development to direct staff and students to private/licenced AI systems that are GDPR compliant and may be used for the analysis of live data.

## 2.2 Archive Data

The University is committed to the principle of [open research](#). The sharing of processed/archive research data through the institutional repository ([SHURDA](#)) is encouraged, where this meets relevant standards regarding robust anonymisation, participants' consent and appropriateness. AI systems will mine, learn from and reuse these data sources, and this should be viewed as a public good.

Archive data will also be anonymised, rather than pseudonymised (i.e. all keys enabling the reversing of the anonymisation will have been destroyed), and so it therefore has a different status under GDPR laws, specifically that it ceases to be categorised as personal data.

## 2.3 Publications and Other Outputs

All research outputs should be made freely available under appropriate reuse licences, as per the University's [open access policy](#). AI systems will mine, learn from and reuse these sources, which again is a public good.

Unlike the established courtesies of human use of third-party materials, where AI trains on and reuses research outputs, it will generally not acknowledge or credit its sources. Researchers may be able to pursue intellectual property infringement of their work in any cases where it is reused for commercial purposes e.g. in paid-for educational materials or other media.

### **3) Using Information Produced by AI Tools**

Researchers should never acknowledge AI as a co-author. According to the Committee on Publication Ethics (COPE), "AI tools cannot meet the requirements for authorship as they cannot take responsibility for the submitted work. As non-legal entities, they cannot assert the presence or absence of conflicts of interest nor manage copyright and license agreements."

Instead, authors must acknowledge the use of AI in other ways. In all cases, the use of AI must clearly be described in a way that makes the research transparent and verifiable (where appropriate in a methods or similar section), and it must always be properly referenced.

The unacknowledged use of AI in research constitutes misconduct. Examples of this are when generative AI is used to author research or create imitation datasets (distinct from legitimate modelling). Investigations into research misconduct will, however, always take care to distinguish generative AI against proofreading software used by researchers who are not native speakers of English.

Where available, researchers should also follow the editorial guidance from their publisher with regards to the use of artificial intelligence.

Researchers are warned that AI has major limitations in terms of summarising materials, particularly that while it will predict what is in third party work that it references, it won't check it (which may produce hallucinations). And at times AI will create references to work that does not exist (fake

references). Therefore, to ensure rigour of academic work, use in this capacity is discouraged and its output should, at least, be thoroughly checked by a human. In most cases, it will be more efficient not to use AI for this purpose.

#### **4) Researcher Development**

AI can perform repetitive tasks with data and/or find patterns in large amounts of data without the need for constant human guidance. AI can therefore imitate some research activities and replicate them to a good level, which can create efficiencies and exciting possibilities. However care should be taken to ensure that AI does not replace the 'growth zone', where earlier-career researchers learn to be excellent by undertaking the more mundane and learning the fundamentals. AI should not therefore, for reasons of cost, be used to replace training roles within research ecosystems; rather it should be used responsibly as a tool to complement and enrich human research activity.

University Research Ethics Committee, November 2023