UKRI Medical Research Council

Interim guidance for applicants and grant holders developing software and Artificial Intelligence in biomedical research and innovation

UK research funders have issued a joint statement on Use of AI tools in funding applications and UKRI has a policy on Use of generative AI in application preparation and assessment.

This interim guidance is for MRC applicants and grant holders using data about people to develop software and AI during their grants. This includes those developing models that may have applications as research tools, or in the healthcare sector. This guidance will be superseded by UKRI guidelines, which will be produced in due course as initiatives such as Enabling a responsible AI ecosystem and Responsible AI UK deliver recommendations and resources.

We use the term AI in line with the UKRI publication Transforming our world with AI: Artificial intelligence, commonly known as AI, describes a suite of technologies and tools that aim to reproduce or surpass abilities (in computational systems) that would require 'intelligence' if humans were to perform them. This could include the ability to learn and adapt; to sense, understand and interact; to reason and plan; to act autonomously; and even to create.

We outline the principles that as a biomedical research community we should strive towards, and signpost further resources that will help meet these. The main purpose is to raise awareness of the issues, and to help manage the risks associated with working with individual level data in AI and software development. As a research funder we focus on development of models and not their deployment.

It is an expectation that researchers in receipt of MRC funding work to the highest ethical standards and meet all necessary regulatory and policy requirements. It is the responsibility of host research organisations to ensure this, as outlined in our Terms and Conditions of funding.

In software and AI development requirements can be unexpected, complex and sometimes difficult to meet. We outline general principles relevant to all software and AI algorithms developed with data about people. Additional requirements come into play if the software or AI algorithm has a 'medical purpose', here the Medical Devices Regulations, including In Vitro Diagnostics, may apply.

1. General principles for development of software and AI when using data about people

1.1 Trust in AI in health research

Building public trust in the development and use of AI is very important. Keeping up to date with legal requirements and standards, which will evolve as the field develops, helps with this. Of fundamental importance is good transparency with the public.

Transparency includes helping people understand the following:

- why AI is a good solution to a specific problem;
- where the data used in development is from and how it is used;

- the risks of identification;
- how the AI works and makes decisions, whether human input remains an element of decision-making;
- how bias is managed and fairness ensured (bias both in the data used to derive the model and in the data used by the model when performing its task);
- how other risks, responsibilities and accountabilities for new tools or systems are managed, and by whom.

There's been much public dialogue on AI, and understanding public concerns can help with development of good transparency information and planning engagement activities that help build trust in AI.

Some resources to help include:

- Key considerations for the use of artificial intelligence in healthcare and clinical research
- Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness
- MHRA principles for transparency in Machine Learning AI, essential for AI as a medical device, and also more generally relevant for transparency in AI development
- Joint guidance from Information Commissioner's Office and Alan Turing Institute on Explaining decisions made with AI
- Understanding how the public feel decisions should be made about access to their personal health data for AI research | Ipsos Ipsos, Sciencewise and NHS AI Lab Public dialogue on data stewardship, November 2022
- International survey of public opinion on AI safety – GOV.UK (www.gov.uk) Centre for Data Ethics and others, October 2023
- Public attitudes to data and AI: Tracker survey (Wave 3) – GOV.UK (www.gov.uk) Centre for Data Ethics and others UK survey of 4000 people, December 2023
- What do the public think about AI? | Ada Lovelace Institute rapid evidence review of public attitudes. Published October 2023.

1.2 Protecting privacy and confidentiality

When using data about people in the development and training of software and AI models, ensuring the protection of privacy and confidentiality is of utmost importance, including when publishing or sharing models or data. In general we recommend using the 5 safes approach and working with your local information governance and data security teams.

It is important to consider the risk of deep learning models fine-tuned on specific health data sets being vulnerable to attacks aimed at identifying individuals, and the security issues that may arise when such models are made available for use.

The risks may be greater in AI based on very detailed models where sequences of data are embedded in the model.

A good summary of cybersecurity issues with large language models is Security and Privacy Challenges of Large Language Models: A Survey.

Also see the GRAIMATTER Green Paper: Recommendations for disclosure control of trained machine learning models from trusted research environments (a 2022 Data and Analytics Research Environments UK programme output).

If you need access to health data for development, data access processes vary and help can be found from MRC Health Data Access Toolkit; Health Data Research UK (HDRUK) Innovation Gateway and Research Hubs.

Support with the legal framework when working with individual level data is available from MRC Regulatory Support Centre and the AI and Digital Regulations Service comprehensive data compliance checklist. You may be working with Personal Data under GDPR and there may be common law of confidentiality requirements that you need to meet.

The processing of Personal Data, as defined in UK GDPR, is regulated by ICO. Their webpage, Our work on Artificial Intelligence, includes guidance on AI and data protection, and how to use AI and personal data appropriately and lawfully.

1.3 Tools need to be fit for purpose, and the benefits available to all in society

All software and AI-based tools need to be fit for purpose and must be tested and validated to ensure that they perform as expected in a reproducible way. This is no different to development of any research tool or technique. As such, our existing good research practice policies, where relevant, apply to the development of AI and software. Algorithms that change over time must be regularly monitored to ensure that they continue to perform as expected.

UK Reproducibility Network AI case studies may help here.

Guidelines and standards for healthcare trials and clinical decision-making involving AI include:

- Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension
- Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension
- TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods
- Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence

In line with our policy on Embedding diversity in research design, software and AI tools must perform well for a diverse target population (by population we mean the breadth of people, patients, samples, cells, etc, relevant to your research or intended application of the tool).

To reduce bias in outputs, it is very important to develop and train models using broad, representative data that reflects the target population. This is particularly important in algorithms that change over time where bias can be amplified. Depending on the application, this may include involving data from people with relevant lived experience.

HDRUK has information about diversity and inclusion in health data research. NHS England have guidance to improve data quality of protected characteristics and others.

Bias is also from human and systemic sources not only data as outlined in a report from the US National Institute of Standards and Technology. This comments that the course of bias is not only introduced by the data to train an AI but societal factors also influence how technology is developed.

The Gates Foundation focus on equity and access in their principles for AI.

NHS England AI programmes promote healthcare equity.

A call for operationalising fairness in machine learning in healthcare includes practical recommendations.

1.4 Sharing data and models

All research must meet requirements of the MRC Data Sharing Policy, including sharing of code. As datasets grow and change over time, the use of Data Object Identifiers and good metadata are very important.

Sharing of the model developed in the course of the grant may be subject to proprietary considerations. Therefore, a period of exclusivity and/or controlled sharing through legal agreements may be required. You should consult your legal department on how to do this in order to manage any Intellectual Property.

When publishing the research or sharing models, developers should ensure that details are included on the intended application of the model, and the type of data used in its development. This is so that other researchers and innovators can understand any limitations and apply these to their own ideas and purposes. Given the potential for adversarial attacks on fine-tuned models, those responsible for the privacy of the original training data should consider what additional restrictions should be placed on model sharing, for example holding the model code in a secure environment and implementing cybersecurity protection to model access.

1.5 Managing the potential for research misuse

There are concerns of potential misuse around the development of AI including cybersecurity threats and terrorism, and doing a risk benefit analysis of your planned work is considered best practice.

This short article illustrates how using AI to read electrocardiograms can have potential misuses: What’s lurking in your electrocardiogram?

AI projects are likely to be subject to greater scrutiny, as outlined in UKRI Trusted research and innovation and the joint (MRC, BBSRC, Wellcome) statement on managing risks of research misuse. The latter will be updated in due course to reflect technological advances in recent years.

2. Development of software and AI with a 'medical purpose'

If the software or AI has a medical purpose such as the prevention, treatment, diagnosis of disease, it will be classed as a Medical Device or In Vitro Diagnostic (IVD) and appropriate regulations will apply. Requirements may include clinical studies, termed 'clinical investigations' (for devices) or 'performance evaluations' (for IVDs).

Medicines and Healthcare products Regulatory Agency (MHRA) software and AI as a medical device web pages will help you ascertain if you are working in this regulated space. Their standalone software guidance is very helpful, but developers should be aware that this regulatory space is in development with international partners in the US and EU.

MHRA has produced guidelines for communicating clear and relevant information about machine learning-enabled medical devices.

For device clinical investigations and IVD performance evaluations, applications for the necessary approvals are through the Integrated Research Application System, IRAS.

The NHS has support services such as NHS Innovation office, and AI and Digital Regulations Service (AIDRS) to help developers. The AIDRS has a section for developers which will guide you through the whole product life cycle: from coming up with an idea, developing it and placing it on the UK market, followed by guidance on keeping your technology up to date.