

Guide for assessing research involving Artificial Intelligence, Machine Learning, and Large Language Model Technology (collectively “AI”)

Artificial intelligence (AI) in research

The rise of Artificial Intelligence (AI) is producing transformational change in almost all sectors of society. The use of AI tools in research is rapidly increasing to the point where they will soon become ubiquitous. Currently, many of the most used tools are Large Language Models (LLM), a machine learning model trained on vast amounts of text. AI tools, including LLMs, are increasingly built into software that is routinely used by researchers, such as search engines (e.g. Google). Many researchers use widely available LLM interfaces, referred to as chatbots (e.g. ChatGPT) to assist in a range of tasks including transcription of audio data or drafting text (e.g. for development of participant information sheets, interview questions, or publications). Others use LLMs for reviewing literature, hypothesis generation, and data analysis. In many of these uses, AI is taking on a role analogous to a research assistant.

NHMRC recognizes that HRECs are seeking guidance on how to approach the ethics review of research involving AI. This guide is aimed at assisting HRECs in identifying the uses of AI that may pose distinctive ethical risks to research participants. In these cases, HRECs may consider requesting additional information from the researcher to assess these additional ethical risks.

For the purposes of this guide, the uses of AI in research can be broadly categorised as ‘routine’ and ‘non-routine’. Routine uses would not usually require researchers to provide additional information to the HREC about specific ethical risks pertaining to AI. It is foreseeable that the content of these categories will rapidly change from year to year – even from month to month – as AI becomes more integrated into systems and processes at all levels of society. Uses that might now be considered ‘non-routine’ may become ‘routine’ in the future. Conversely, uses that would currently be considered ‘routine’ may be found in the future to have negative effects that require mitigation. The changing nature of AI use means that this guide will require revision periodically.

Routine and non-routine use of AI in research

The use of AI tools that are built into widely used software, such as word processing, email programs, and search engines, are considered **routine use of AI**. While there may be ethical concerns about using these tools (such as data security, biased or unethically sourced training data, and environmental impacts), these are best conceptualised as research integrity issues and managed by the institutions where research takes place and which make the AI tools available to researchers. At this stage, application of this guide to assist ethics review would not ordinarily be required for such uses.

Routine use is one end of a spectrum of uses. At the other end is the use of AI as an intervention.

Researchers are using AI in both clinical and non-clinical research to generate synthetic data sets to preserve privacy and/or train AI models. Researchers in fields such as digital health or translational medicine use AI as an intervention to directly improve health; for example, using AI to analyse radiological images. Applications where an AI model is trained as part of the research project or used as an intervention would be considered non-routine use of AI and reference to this guide would ordinarily be required.

Other uses may be considered as falling somewhere between routine and non-routine uses, such as the use of LLMs for hypothesis generation or data analysis. In these cases, researchers could be asked to declare their use of AI followed by a set of questions about that use, while recognising that they may not be expected to answer certain ethically relevant questions such as those related to system/tool validation or which data sets the LLMs were trained on.

Expectations

It is reasonable to expect that researchers disclose information about AI use in research to those reviewing that research to identify ethical issues that might be present and to explore ways to mitigate any risks that may arise from those uses. To facilitate that process, researchers are expected to have considered the uses and identified these issues as part of developing their research proposal and application for ethics review. Consequently, the guidance that follows is addressed both to researchers and reviewers and should be integrated into the broader interchange between researchers and reviewers that characterises the ethics review process.

On a broader level, the transformational nature of the advent of AI in research will require all its users to be mindful of the potentially far-reaching impact of decisions about AI use on our societies. It requires us to reflect on what we are doing with AI, be responsible for what we do with it, and be diligent about maintaining the standards that are established for its use, including relevant institutional policies and procedures.

Introduction to questions

These questions are intended to assist researchers in developing proposals to conduct research involving AI, and reviewers undertaking ethical assessment of the research proposals. The tool only includes questions that would not be covered by ethics review of a research project that does not involve the use of AI.

Research “involving AI” or “involving the use of AI” means the use of AI in any aspect of the research, whether or not AI is the subject of the research. However, caution should be exercised so as not to include in the scope of this tool the design, development, testing or validation of AI technology that would not otherwise be considered human research.

Users of this tool who are interested in international work on the ethics of AI and its use in health research may wish to refer to the work of the World Health Organization and others, current as of 2023.^{1,2,3}

The questions address themes or areas of concern that are relevant to the use of AI in research, as outlined in Box 1. As indicated, the core concern is the risk to research participants, researchers and to society presented by the use of AI at various stages of a research project. Six areas of concern are identified.

Box 1: Themes/areas of concern

- | | |
|---|--------------------------------|
| 1. Overarching concern = Risk (individual + collective) | 5. Technical robustness/safety |
| 2. Human agency/oversight | 6. Data reliability |
| 3. Transparency/explainability | 7. Data governance/security |
| 4. Bias/representativeness | |

These areas of concern can arise at any stage of a research project. Stages of research where AI may be involved are listed in Box 2.

Box 2: Stages of research

- | | |
|---|------------------------------|
| - Discovery (concept generation and design) | - Intervention (if relevant) |
| - Literature review | - Data collection |
| - Planning and development | - Data management |
| - Ethics and/or scientific review | - Data analysis |
| - Recruitment | - Publication |

1 https://www.itu.int/dms_pub/itu-t/opb/fg/T-FG-AI4H-2023-3-PDF-E.pdf

2 Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022;377:e070904. Published 2022 May 18. [doi:10.1136/bmj-2022-070904](https://doi.org/10.1136/bmj-2022-070904)

3 Kwong JCC, Erdman L, Khondker A, et al. The silent trial - the bridge between bench-to-bedside clinical AI applications. Front Digit Health. 2022;4:929508. Published 2022 Aug 16. [doi:10.3389/fdgth.2022.929508](https://doi.org/10.3389/fdgth.2022.929508)

There is a spectrum of intended uses of AI, not all of which merit overt consideration as part of ethics review. One way to distinguish those uses that do merit consideration from those that do not is to separate routine uses from non-routine uses. For example, routine uses might include AI that is integrated into pre-existing or administrative tools. Non-routine uses can also be considered along a continuum from automation to generation to intervention (in the context of either clinical or non-clinical research). What is defined as 'routine' and 'non-routine' will inevitably change over time.

This tool is intended for use when research involves non-routine uses of AI but is not intended for use for research that involves routine uses of AI.

AI use spectrum

Intended use	Use of assessment tool required
Routine	✗ No
Non-routine	
Automation	✓ Yes
Generation	✓ Yes
Intervention	✓ Yes

Non-routine uses of AI

The spectrum of non-routine AI use ranges from automation to intervention.

Examples of automation in research include but are not limited to selection and randomisation of participants; participant reminders; creation of and/or conduct of online surveys; data cleaning, screening and extraction; limited data analysis; data visualization; and research task repetition. It is likely that much of what is currently considered to be non-routine use of automation in research will rapidly be re-categorised as routine use over time.

Examples of generation include but are not limited to using AI to draft text (e.g. for development of participant information sheets, interview questions, policy briefs, or academic publications); for reviewing literature; for hypothesis generation; or for data analysis.

Examples of intervention include but are not limited to the development of AI tools to improve educational or social outcomes; or to prevent, diagnose or manage disease.