Australian Government
National Health and Medical Research Council

NHMRC

BUILDING A HEALTHY AUSTRALIA

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)
2. Human agency/oversight
3. Transparency/explainability
4. Bias/representativeness
5. Technical robustness/safety
6. Data reliability
7. Data governance/security

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)
- Literature review
- Planning and development
- Ethics and/or scientific review
- Recruitment
- Intervention (if relevant)
- Data collection
- Data management
- Data analysis
- 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

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

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.

Australian Government
National Health and Medical Research Council

NHMRC BUILDING A HEALTHY AUSTRALIA

Questionnaire: Guide for assessing research involving artificial intelligence, machine learning, large language model technology ("ai")

Researchers are asked to respond to the following questions in completing their application for ethics review. HRECs or other ethics review bodies will then consider the information provided in assessing the research proposal.

Instructions for researchers

- *In providing your responses, consider each stage of the research in which you intend to use AI and each tool that you intend to use.*
- *Where applicable, include references to the relevant project document/s (with page numbers) that include more complete information that is relevant to your response to the question.*

Is AI used for non-routine purposes in the research project?

- (•) no, further responses required
- [x] yes, continue

1. **Identify AI use stage** (check all that apply)

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

2. **How will the AI tool be used at each relevant stage of the research project?**

3. Will participant data be entered into any AI system operated by a third party?[4]

☐ no *(continue to next question 4)* ☐ yes

If yes describe:

a. if the data will be retained by the AI system

b. if there are any restrictions, licensing conditions, or ethical considerations in the tool's creation that may affect participants' rights or data privacy

c. whether the data will be used to further train the same model or other models

d. how you will ensure that the data will not be disclosed, exchanged or sold by the third party without participant consent.

4. What is known about the provenance of the AI tool's training data?

a. Do the data sets that are used to train the AI algorithms adequately represent the population/s impacted by or relevant to the research?

☐ yes ☐ no

b. If not, justify why it is appropriate for this tool to be used on your research cohort.

4 Note that submission to and retention of human data to a third party (e.g. OpenAI via ChatGPT, Microsoft via Copilot, Google via Gemini, etc.) where data are very unlikely to remain in Australia may constitute a cross-border disclosure under the Privacy Act 1988. If this is the case, researchers would incur a burden under the Act to ensure sufficient information is given to people from whom data have been collected and that consent exists for such use and disclosure. Use of local AI running on researchers' hardware or mechanisms such as Google NotebookLM may mitigate these concerns.

5. What are the characteristics of the AI tool that you intend to use?

a. Has the AI system been validated for its intended use/s? If so, how? If not, why not?

b. Is the AI tool non-adaptive or adaptive (involving continuous learning algorithms while the researcher is using the tool)?

c. If the AI tool is adaptive, what are the limits to permitted adaptations, if any?

d. Is reproducibility and/or replicability important for your AI tool. If so, explain why and how this has been adequately addressed.

e. Will the AI interact with research participants and/or research staff? If so, how?

6. Describe your plan for oversight of the AI and the outputs of the tools or systems used.

a. What safeguards are in place to include human oversight, especially when critical decisions are necessary or unexpected developments arise?

b. How will you know if the tool has produced biased or unfair outputs? What steps will you take to detect and address the risk of this occurring?

c. Who on the research team has the skills to critically assess the outputs of the AI tools/ systems and how will these team members review these outputs before they can have an impact on participants or others?

d. Does your plan include the capacity to monitor and assess the performance of the AI over time? Explain how this will be done.

e. If relevant, will training be offered to members of the research team or organisational staff who interact with the AI during the research?

7. **Describe the risks associated with your use of the AI tool in your research and your plan for mitigation and management of those risks.**

   a. What are the potential harms that could result from the use of the AI tool and to whom?

   b. How likely are these harms?

   c. How will these risks be mitigated and managed, and their successful mitigation and management be evaluated?

   d. Will risk assessment be conducted on an ongoing basis during (and potentially after) the project? If so, will it be conducted in accordance with a defined schedule? If so, what is the proposed schedule?

8. **What information about the use of AI in the research and the projected impact of any AI-assisted or AI-generated outputs of the research will be made available to participants, and the public (if relevant) and how will it be communicated?**