Senate Ethics in Research Committee (EiRC) Guidelines and recommendations for the use of generative artificial intelligence (AI) tools in research

Title	Guidelines and recommendations for the use of generative artificial intelligence (AI) tools in research
Approval date	31 October 2023 (via EiRC Chair's Circular)
Person(s) responsible for drafting, review and revision	Research Integrity Manager Research Integrity Director
Next review date	October 2024
Status	Guideline
Approval route	Office of Research Integrity (ORI) > Senate Ethics in Research Committee (EiRC)
Approving body	EiRC
Relevant related policies, procedures and guidelines	• UCT Policy for Responsible Conduct of Research • UCT Research Ethics Code for Research Involving Human Participants • UCT Research Ethics Code for Use of Animals in Research and Teaching • Authorship Practices Policy • Conflict of Interest Policy • UCT Policy and Procedures for Breach of Research Ethics Codes and Allegations of Misconduct in Research (colloquially known as the 'Research Misconduct' policy)

Table of Contents

1. Purpose and preamble

The purpose of this document is to provide practical approaches for researchers[1] based at the University of Cape Town to responsibly use generative AI tools in their research practice.

On the one hand, researchers working directly in the field of generative AI have acknowledged a multitude of potential positives of these supportive tools. This includes but is not limited to:

- Automation of repetitive tasks such as summarising or rephrasing of text;
- Identification of potential new avenues or topics for research through analysing and assessing large volumes of published work in a field;
- Creation of templates which researchers can build on and further develop;
- Translation of languages, therefore enabling non-First-language English speakers to compose more accurate and natural texts;
- Testing of hypotheses through running simulations and exploring more or alternative hypotheses;
- Natural language processing and helping to edit, refine or correct sample texts and,
- Preliminary analysis of data sets.

While on the other hand a range of potential challenges are also presented by generative AI tools. This includes but is not limited to:

- Lack of clarity on precisely which data are used to train the models or tools (commonly referred to as a 'black-box'), therefore potentially infringing on copyrighted material;
- Lack of clarity regarding the tool's approach to decision-making;
- End or expiry dates of training data sets, which limit the tool's ability to generate content that takes into account the most recent and emerging information on a given topic;
- Data privacy when used in the context of analysing data that contains sensitive or personal information;
- Systemic biases in material used to train the tools;
- Potential for malicious actors to use the tools to facilitate irresponsible research practices linked to fabrication, falsification and plagiarism and to make obfuscation easier;
- The speed at which the tools are being developed and being made publicly and openly available to users coupled with the lack of a coherent regulatory environment;
- Potential to exacerbate issues around colonialism and Global North/South power dynamics;
- Potential for homogeneity of responses, and therefore the homogenisation of research outputs, as a result of the inherent (Global North) bias of training data and lack of representation of other (Global South) ways of being/doing/creating;
- New opportunities to commit research misconduct and questionable research practices and practice academic dishonesty;
- Errors and inaccuracies, including generation of fictitious references or fake/false information (sometimes optimistically referred to as 'hallucinations') and,
- As yet unknown or unimagined safety issues.

These examples are not comprehensive, and the field of generative AI is a growing and dynamic one, meaning that the positives and negatives are likely to grow and adapt as the technology does.[2]

It is the view of the EiRC that researchers should be guided in how to approach use of these tools with an ethos of ethical responsibility, responsibly and with accountability. The tools are already being widely used and this trend will continue and therefore researchers should receive sufficient training and guidance.

Finally, it must be acknowledged that there is a limit to the guidance that an institution or committee can produce, given the dynamic nature of the topic and the inherent nature of AI generation tools where questions of training black-boxes and potential copyright infringement already exist.

---

[1] 'Researchers' in this case is meant to be understood to include postgraduate students, staff (part-time, full-time, adjunct, emeritus and otherwise) working directly as researchers or in research management, administration, and support.

[2] Section 1 was developed out of participant responses in the 2023 EiRC workshop and out of the work that Maria Keet, presented at this workshop, noted in the references below.

Guidance, developed by the Centre for Innovation in Learning and Teaching is available for the Teaching and Learning context, on their website.

2. Relationship with institutional policies

This guidance document should be read in conjunction with institutional research policies, not limited to but including:

- UCT Policy for Responsible Conduct of Research
- UCT Research Ethics Code for Research Involving Human Participants
- UCT Research Ethics Code for Use of Animals in Research and Teaching
- Authorship Practices Policy
- Conflict of Interest Policy
- UCT Policy and Procedures for Breach of Research Ethics Codes and Allegations of Misconduct in Research (colloquially known as the 'Research Misconduct' policy)

3. Understanding generative AI tools

Disclaimer: Generative AI technology is evolving rapidly and may include models beyond Large Language Models (LLMs) described below. As the technology evolves, the EiRC will consider the ethical implications and endeavour to incorporate these into this guideline.

Large Language Models (LLMs) are language models[3] that try to complete an input piece of text in a human readable way. The LLM uses statistics of letters, sequences of letters, words and sequences of words to randomly generate human readable text. The tool therefore generates one word at a time without a global or overarching plan. As a result of this approach it is possible that a generative AI tool will generate incorrect or nonsensical answers to questions and prompts.

LLMs (for example, but not limited to OpenAI's GPT3.5, commonly referred to as ChatGPT) are capable of summarising, responding to questions and completing basic programming tasks. These capabilities are increasing exponentially, and it is known that forthcoming iterations of LLMs are trained on larger datasets.

Outputs are only as good as inputs

In addition to the concerns already expressed regarding the data that the tools are trained on it must also be noted that the researcher's use of the tool is also going to determine the output that is generated. Questions and prompts which are inadequately or inappropriately phrased or lack detailed explanation of the expected output will result in responses which could be inadequate, inappropriate or nonsensical. Even with adequate and appropriate inputs, researchers should always be cautious of the outputs, applying a critical approach to using any texts that are generated. [4]

4. Values

"*When you invent a new technology, you uncover a new class of responsibilities*" (Harris and Raskin, 2023)

The EiRC advises that researchers using generative AI tools should uphold long standing values already practiced in the responsible conduct of research and outlined in national and international statements, guidelines, and regulations. These values include but are not limited to:

---

[3] In this context 'language' is represented by a wide range of inputs, including (but not limited to) images, MRI data, music, written text, computer code, videos and others.
[4] Section 3 was developed, in large part, out of the work Kyle Abrahams presented at an EiRC Workshop, noted in the references below.

1. Honesty in all aspects of research
2. Professional courtesy and fairness in working with others
3. Good stewardship of research on behalf of others[5]
4. Transparency in conducting research and dissemination of findings[6]
5. Fair practice from conception to implementation of research
6. Shared accountability in the conduct of research
7. Indigenous knowledge recognition and epistemic justice[7]

5. Operationalising values

The EiRC recognises that values can be abstract and not clear in how they should or could be implemented in concrete research practice. Therefore the following operationalising practices are also provided.

1. Researchers should be open and honest when using generative AI in their research writing (whether for research publications, conference papers, grant proposals or translation into teaching material), this may mean providing a disclaimer or explanatory note identifying the tool that has been used and describing where and how the tool was used in the final written product.
2. Researchers, especially those who are supervising students or junior colleagues, should create an environment that encourages transparency and responsible use of AI tools, including providing access to resources to enable upskilling, development and education in this emerging practice.
3. Researchers should take great care in designing the prompts or input texts used when engaging generative AI tools. Writing detailed prompts that take a 'step-by-step' approach or breakdown a query into discrete parts are more likely to get meaningful outputs as a result. Approaches include inclusion of the '5Ws' (who, what, where, when, why) in a prompt or asking the tool to take on a particular role.
4. Researchers should use generative AI tools in the context of UCT policies. The Authorship Practices Policy requires that authors (i) make a substantive intellectual contribution to a piece of work and (ii) are able to defend it against criticism. In the context of the use of generative AI use, this means that listed authors should be (and will be) held accountable for the work that they produce, and any errors, inaccuracies or hallucinations that are included in the text. Depending on the extent of any problematic text and the intent of the author this could mean irresponsible research practices that are identified as (at best) questionable research practices (QRPs) or (at worst) research misconduct. This could trigger the university to implement the Research Misconduct Policy if QRPs or research misconduct is identified.
5. Researchers are encouraged to distinguish between using AI tools for editorial improvements (for example using Grammarly to improve the wording, flow and coherence of a piece of work) and for generation of new text and should let readers of their writing know about it. The use of AI tools for editorial assistance is likely to represent a low-risk to the researcher and the research output, while the use of an AI tool to generate sections of text or analyse data represents a higher-risk activity to the research enterprise. In the spirit of shared accountability, this detail should be declared by the researcher.

Leveraging the values identified above in the use of generative AI tools allows for people involved in the research enterprise to have a shared understanding of texts and approaches and to collaboratively develop and understand what responsible approaches look like to a community of researchers.

---

[5] Items 1 through 3 are taken from the Singapore Statement on Research Integrity (2010) (https://www.wcrif.org/guidance/singapore-statement)

[6] Item 4 is taken from the Montreal Statement on Research Integrity in Cross-Boundary Research Collaborations (2013) (https://www.wcrif.org/guidance/montreal-statement)

[7] Items 5 through 7 are taken from the Cape Town Statement on Fostering Research Integrity through Fairness and Equity (2023) (https://www.wcrif.org/guidance/cape-town-statement)

6. Duties of the institution

The institution (University of Cape Town) has a duty to create an environment which encourages and enables responsible use of generative AI tools in the research context. This mandate can be realised through a variety of practices:

1. Development of clear and enabling guidance, which is regularly updated;
2. Provision of training and educational resources to upskill and enable responsible use,
3. Mechanisms to address irresponsible use of generative AI tools (such as leveraging the research misconduct policy, if a situation requires it) and,
4. Monitoring and evaluation of this guideline on a regular basis (annually) to incorporate up-to-date information and guidance.

The EiRC recognises that generative AI tools are quickly being integrated into a researcher's tool kit and it is therefore necessary to upskill researchers in the responsible use of the tools.

7. Concluding remarks

The EiRC acknowledges the fast pace at which generative AI tools are being developed. It also notes that enabling researchers to integrate their use in a responsible, honest and transparent way must be the basis of the approach for using these tools.

8. References and resources

- Kyle Abrahams, *Generative AI Basics – What is it and how does it work?*, presentation delivered to the EiRC workshop, 02 August 2023
- Tristan Harris and Aza Raskin, *The AI Dilemma*, Centre for Humane Technology presentation, 9 March 2023 (https://www.youtube.com/watch?v=xoVJKj8lcNQ).
- Maria Keet, *Ethical issues regarding the us of AI authoring tools in research,* presentation delivered to the EiRC workshop, 02 August 2023
- *The Singapore Statement on Research Integrity* (2010) (https://www.wcrif.org/guidance/singapore-statement)
- *The Montreal Statement on Research Integrity in Cross-Boundary Research Collaborations* (2013) (https://www.wcrif.org/guidance/montreal-statement)
- *The Cape Town Statement on Fostering Research Integrity through Fairness and Equity* (2023) (https://www.wcrif.org/guidance/cape-town-statement)

9. Authorship and attribution

This document was authored by Mrs Paula Saner, on behalf of the Senate Ethics in Research Committee (EiRC). The EiRC were instrumental in reviewing and adapting the document to make it fit-for-purpose.

The contributions of the participants at the 2023 EiRC Annual Workshop, held on 02 August 2023 are gratefully acknowledged as a source of inspiration for this work. The formal presentations are appropriately cited where they are used.