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

[Version 2.1, June 2025]

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) 17 June 2025 (via EiRC round robin, post committee meeting edits)
Person(s) responsible for drafting, review and revision	Research Integrity Manager Research Integrity Director
Next review date	June 2027
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 guidance to researchers[1] based at the University of Cape Town for responsible use of 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 potentially positive use-cases for these supportive tools. These include but are 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, thereby 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.

A range of potential challenges are also presented by generative AI tools. These include but are 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'[2]), therefore potentially infringing on copyrighted or ethically problematic material;
- Lack of clarity regarding the tool's approach to decision-making;
- End or expiry dates of training data sets, which may 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;
- Possible 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, with the tools then also allowing for easier obfuscation of such;
- 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 areas of concern related to issues such as 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; including a possible future tendency to dismiss the validity ad relevance of Diversity, Equity and Inclusivity considerations in a research context.
- New opportunities to commit research misconduct and engage in questionable research practices, and to practice academic dishonesty;
- Errors and inaccuracies, including generation of fictitious references or fake/false information (sometimes 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 change in nature and impact as the technology does.[3]

It is the view of the EiRC that researchers can benefit from guidance in how to approach use of these tools with an ethos of ethical responsibility, responsibly, and with accountability, given that they are being used at an ever-

[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] A 'black-box' refers to an artificial intelligence system whose internal decision-making processes are opaque and not fully understandable, even to its creators, despite observable inputs and outputs. For further information see: Tredinnick, L., & Laybats, C. (2023). Black-box creativity and generative artifical intelligence. *Business Information Review*, *40*(3), 98-102. https://doi.org/10.1177/02663821231195131

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

increasing rate. When using generative AI tools, it is critical that researchers take moral and legal responsibility for their use of these tools. Additionally, researchers must uphold the values described in Section 4 (below), and remain accountable for the outputs of generative AI tools.

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 generative AI tools where questions of training black-boxes and potential copyright infringement already exist.

Guidance on this topic for the Teaching and Learning context has been developed by the Centre for Innovation in Learning and Teaching, and is available 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[4] 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. As a result of this approach it is possible that a generative AI tool will generate incorrect or nonsensical answers to questions and prompts.

Large Language Models (LLMs)[5] try to complete tasks assigned to it by users, whether human users or other AI tools. An LLM uses statistics related to the frequency of letters; and sequences of letters and words; to generate human readable text. Given the current limitations of AI meta-cognition regarding aspects such as intent or strategy, it is possible that generative AI tools could generate incorrect or nonsensical answers to questions and prompts.

LLMs (for example, but not limited to OpenAI's ChatGPT) are capable of summarising, responding to questions, and completing basic programming tasks. These capabilities are increasing exponentially, and future Generative AI tools will benefit from training by datasets of ever-increasing size.

[4] 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.
[5] 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.

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. [6]

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:

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[7]
4. Transparency in conducting research and dissemination of findings[8]
5. Fair practice from conception to implementation of research
6. Shared accountability in the conduct of research
7. Indigenous knowledge recognition and epistemic justice[9]

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 transparent 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 result in meaningful outputs. Possible approaches are the inclusion of the '5Ws' (who, what, where, when, why) in a prompt; or asking the tool to take on a particular role such as "journal editor" or "lab assistant".
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

[6] Section 3 was developed, in large part, out of the work Kyle Abrahams presented at an EiRC Workshop, noted in the references below.
[7] Items 1 through 3 are taken from the *Singapore Statement on Research Integrity* (2010) (https://www.wcrif.org/guidance/singapore-statement)
[8] Item 4 is taken from the *Montreal Statement on Research Integrity in Cross-Boundary Research Collaborations* (2013) (https://www.wcrif.org/guidance/montreal-statement)
[9] 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)

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 entail research practices that are identified as (at best) questionable research practices (QRPs), or (at worst) research misconduct. Where identified, QRPs or research misconduct would invoke the UCT Research Misconduct Policy.

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), versus for generation of new text, and should inform readers of which applies to the research output. The use of AI tools for editorial assistance is likely to represent a low-risk to the researcher little or no compromising of 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. Declaration of details regarding AI use affirms and promotes a spirit of shared accountability, and supports an ethical culture in research generally.
6. Researchers should be aware of potential biases in the generative AI tools used (for reasons that could include being developed by Western-/Euro-centric technology companies, or not intended specifically for research purposes), and should be mindful of risking the perpetuation of homogenous Western-/Eurocentric worldviews, or contributing to epistemic injustice by excluding marginalised or indigenous knowledge sources. Indigenous and marginalised communities should be consulted before researchers incorporate their contributions into generative AI models, to ensure that their wishes are taken into account.

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.

6. Duties of the institution

The institution (the 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. Regular monitoring and annual evaluation of this guideline on a regular basis, 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 that 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. It is critical that the human researcher who makes use of generative AI tools reviews its outputs, and takes accountability for any and all use of the tools in the research process.

Students and researchers are encouraged to develop and demonstrate essential research skills independently. The responsible use of generative AI tools may be permitted, under certain circumstances, to streamline research practices, provided such use supports, rather than replaces, the acquisition of core research competencies and adheres to principles of research integrity outlined in this guideline.

8. References and resources

- Abrahams, K., *Generative AI Basics – What is it and how does it work?*, presentation delivered to the EiRC workshop, 02 August 2023
- Bekker, M. (2024). Large language models and academic writing: Five tiers of engagement. *South African Journal of Science*, 1-5. doi:https://doi.org/10.17159/sajs.2024/17147
- Harris, T., and Raskin, A., *The AI Dilemma*, Centre for Humane Technology presentation, 9 March 2023 (https://www.youtube.com/watch?v=xoVJKj8lcNQ).
- Keet, M., *Ethical issues regarding the use of AI authoring tools in research,* presentation delivered to the EiRC workshop, 02 August 2023
- Resnick, D. B., & Hosseini, M. (2025). Disclosing artificial intelligence use in scientific research and publication: When disclosure should be mandatory, optional or unnecessary? *Accountability in Research*, 1-13. doi:https://doi.org/10.1080/08989621.2025.2481949
- *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.

Version 1: October 2023; Version 2.1: June 2025 (current)

Appendix 1: Use case examples

This section should be read in conjunction with the Center for Innovation in Learning and Teaching (CILT) *Researcher's Guide: Ethical use of Generative AI for research purposes*. The guide considers the use of generative AI tools at various stages of the research life cycle. This appendix seeks to add further clarification by applying a value- and risk-based approach to the different use cases in the research life cycle. These stages are described as follows:

1. Conceptualisation and research idea generation
2. Development of ethics applications
3. Literature Review
4. Data collection/transcription
5. Data analysis
6. Write-up

Resnick and Hosseini (2025) suggest that use of generative AI tools should be disclosed if the use is "intentional and substantial" (p. 1), and that substantial is defined as:

> "(a) The AI tool ***makes decisions that directly affect research results***. For example, using AI to extract data from articles to conduct a systematic review would be substantial and intentional because data extraction decisions affect the outcomes of the review.
> (b) The AI tool ***generates or synthesizes content, data or images***. For example, using AI to write sections of a paper, integrate notes or other pieces of information, translate language in the paper, or create images or synthetic data would be substantial and intentional because AI has generated or synthesized new content that directly affect research outcomes.
> (c) The AI tool ***analyzes content, data or images***. For example, using AI to analyze genomic data, text, or radiologic images would be both substantial and intentional because it produces analyses that supports findings and conclusions and affects the content of a publication." (p. 6) (emphasis added)

It is critical, when using generative AI tools that the researcher (i.e., the human in the process) takes moral and legal responsibility for the use of the tool. Additionally, researchers must uphold the values described in Section 4 (above), and remain accountable for the outputs of generative AI tools, and their use in the research process.

Key

High risk activities	Stop, do not proceed. Recommended that generative AI tools are not used, or that they only be used under very specific conditions, with adequate security measures in place.
Medium risk activities	Proceed with caution. Generative AI tools may be used in limited cases, provided that risks are mitigated, and the values-based approach is leveraged in the use of generative AI tools.
Low risk activities	Go, slowly. Use of generative AI is permissible when the risks are minimal or limited, where no private, personally identifiable or confidential information is being used, and where researchers cautiously review and assess the outputs for validity and reliability before proceeding.

Stage	Risk-based consideration of generative AI use
Conceptualisation	High risk • Action: Inputting proprietary, confidential or novel ideas into an open tool to generate further ideas, uncritical reliance on AI output. • Foreseeable risks: Leakage of confidential information or loss of intellectual property rights, policy violations.

	Medium risk • Action: Drafting research proposal outlines, identifying methodologies. • Foreseeable risks: Possible bias, inaccurate or incomplete methodologies or literature reviews, privacy issues.
	Low risk • Action: Brainstorming general topics, generating research questions, clarifying concepts, summarizing public literature. • Foreseeable risks: Lowest privacy risk, bias in the generation of research questions based on publicly available information, over-generalisation of concepts and literature.
Development of ethics applications	High risk: • Action: AI-generated informed consent forms, information brochures, project summaries or descriptions of novel methodologies. • Foreseeable risk: Without human oversight AI-generated content may not be contextualised to the research environment and may not comply with legal or ethical norms, it may also put proprietary, sensitive or confidential information at risk. Bias in training data may not adequately summarise research in a context-specific, culturally sensitive manner, over reliance on genAI may undermine skills development and erode researcher accountability.
	Medium risk: • Action: Drafting or refining correspondence with ethics committees, summarising research protocols for comparative analysis, cross checking applications with institutional policies or national legislation. • Foreseeable risk: Confidential information contained in ethics committee correspondence breaches confidential processes of ethics committees, summaries using only publicly available information are likely to be incomplete as the tool cannot access studies behind publisher paywalls, use of fabricated or 'hallucinated' sources in summaries.
	Low risk: • Action: Identifying potential ethical issues associated with a particular research topic; standardising references/citations • Foreseeable risk: Bias in training data may not adequately identify potential ethical risks or may identify ethical risks that are not context relevant, cultural insensitivity in the context of the research, over reliance on genAI may undermine skills development and erode researcher accountability.
Literature Review	High risk • Action: AI-generated reviews without verifying references or review by the researcher, identifying gaps in the literature based on incomplete literature sources. • Foreseeable risks: Fabricated or 'hallucinated' sources used, plagiarism or copyright infringement, privacy breaches, misinformation (especially when using freely accessible internet content).
	Medium risk • Action: Drafting literature reviews with AI-generated citations to be reviewed and edited by the researcher. • Foreseeable risks: Fabricated or 'hallucinated' references, bias, incomplete review.
	Low risk • Action: Summarise and organise public literature with human oversight. • Foreseeable risks: Lacks criticality of using research published in journals.
Data collection/ transcription	High risk • Action: Processing personally identifiable information (PII)/confidential data using public generative AI tools without participant consent, undisclosed AI use.

	• Foreseeable risks: Data breaches, privacy breaches, ethical breaches, breaches of informed consent, legal violations of the Protection of Personal Information Act (POPIA).
	Medium risk • Action: Transcribing unpublished but non-sensitive data with or without participant consent, synthetic datasets. • Foreseeable risks: Data retention (depending on the generative AI tool settings), participant consent, re-identification, bias in generating synthetic datasets based on existing biases in data.
	Low risk • Action: Transcribing publicly available interviews or lectures, organising anonymised data. • Foreseeable risks: Minimal privacy risk as the information is in the public domain, fully anonymised data present minimal risk based on confidentiality and sensitivity.
Data analysis	High risk • Action: Processing PII/confidential data using public generative AI tools, unverified AI-only analysis. • Foreseeable risks: Data breaches, privacy breaches, ethical breaches, breaches of informed consent, biased data analysis, legal violations of the Protection of Personal Information Act (POPIA).
	Medium risk • Action: Analysing proprietary, sensitive, confidential or pseudonymized data, AI-generated insights • Foreseeable risks: Input of confidential data into public generative AI tools may create privacy breaches, especially if re-identification of data is possible; 'hallucinations' and biases may surface in the analysis
	Low risk • Action: Analysing anonymised public data, automating basic statistics/visuals • Foreseeable risks: Minimal or low risk based on using publicly available anonymised data; human oversight and understanding used to review statistical outputs before they are incorporated into the research process
Write-up	High risk • Action: Complete manuscripts drafted by generative AI tools without disclosure, input of confidential/sensitive/PII data to develop manuscripts. • Foreseeable risks: Plagiarism, privacy breaches, unreliable presentation of results, potential fabrication or falsification.
	Medium risk • Action: Drafting sections, AI-suggested citations, describing results • Foreseeable risks: Disguising poor research in well-written content, fabricated or 'hallucinated' references, bias, inaccurate reflection of research.
	Low risk • Action: Language editing, summarising public information, paraphrasing • Foreseeable risks: Provided there is sufficient human oversight, this presents low or minimal risk.

Risks and benefits in using generative AI tools exist at all stages of research, therefore careful use and the application of critical thinking as well as values- and risk-based reflection is recommended. Bekker (2024) has provided further guidance for the use of generative AI tools in academic writing, using a five-tiered system, summarised below.

Tier	Effect / type of tool	Place in the writing process	Most obvious benefit	Most obvious risk
1	Ban	n/a	Ensures 100% human authorship and does not compromise academic integrity	Inevitably flouted except under exam conditions
2	Proofing	After	Increases efficiency, reduces cost	Might subtly alter meaning or obscure intentions
3	Editing	During	Produces well-organised, word-perfect writing	Language may be bland, may foster laziness
4	Co-creating	Start	Offers an alternative to human partnership–making interpretative and instructive suggestions, error checks	Authorship is opaque, high risk of introducing hallucinations and biases, inexplicability, loss of autonomy, loss of critical reasoning, outsourcing of thought
5	All	All	Enables high-speed academic writing, maximum support for inexperienced academics	As per Tier 4, but more extreme

Figure 1: Summary of large language model (LLM) permission tiers, where they come into the writing process, and their most obvious benefits and risks (Bekker, 2024, p. 3)

Tiers 2 and 3 are aligned with low-risk use of generative AI tools; tier 4 is aligned with a medium-risk use of generative AI tools and, tier 5 is aligned with a high-risk use of generative AI tools.