UNSW
SYDNEY

AI, Large Language Models, and the Responsible Conduct of Research at UNSW

This document outlines UNSW's position on the use of AI and Large Language Models in Research. It should be read alongside the UNSW Research Code of Conduct, and advice on the use of Generative AI in Teaching and Assessment.

There has recently been a step change in the ability of, and access to, large language models which produce useful long-form written content. New applications and uses for this technology are emerging rapidly, with the potential for significant positive impact on the conduct of research across many areas of inquiry.

For researchers, the availability of large language models can have both positive and negative implications for their work. On one hand, these models can assist with data analysis, text summarization, and other tasks that can speed up the research process. On the other hand, there is a risk of researchers relying too heavily on these models, which can lead to unoriginal work and potential plagiarism. Additionally, the ease of generating realistic-looking text and data with language models raises concerns about the validity and reliability of research results.

Given the rapid pace of change, researchers, journals, and organisations are considering their position on the appropriate use of these tools, especially in the context of research integrity. Individual researchers have also indicated that the lack of clarity is concerning, and guidance on the appropriate use of these tools would be valued. This paper is intended to clarify UNSW's position on the use of these tools in research and the implications for our research integrity processes.

Ethical and Responsible Use

It is important for researchers to use language models in an ethical and responsible manner and to be transparent about their use in their methods and reporting.

The ethical implications of using large language models in research and other fields are complex and wide-ranging. Some of the key ethical concerns include:

- **Responsibility for content generated by the models:** The text generated by language models can contain biased, false, or harmful information, and it can be difficult to determine who is responsible for this content.
- **Impact on human work:** The increasing sophistication of language models has the potential to automate tasks previously performed by humans, leading to job loss and other economic consequences.
- **Privacy and data protection**: Large language models are trained on massive amounts of personal data, raising questions about the use and storage of this information and the potential for misuse. There are also concerns around unauthorised disclosure where confidential or sensitive data are provided to large language models.
- **Authenticity and originality:** The ease with which language models can generate text raises concerns about the authenticity and originality of work, especially in academic and creative fields.
- **Bias and discrimination:** Language models can perpetuate and amplify existing biases and stereotypes, leading to discrimination and harm.

CRICOS Provider Code 00098G

These ethical concerns highlight the need for responsible and transparent development, deployment, and regulation of language models, as well as ongoing efforts to address and mitigate their potential negative impacts.

Care should be taken when engaged in research activities with external organisations which may have specific requirements. By way of example, both the NHMRC and ARC have policies that restrict or preclude the use of AI tools in certain activities, including in the assessment of grants.

Authorship

Recently, manuscripts have begun to appear in preprint repositories and published works which include ChatGPT as an author[1,2]. Academic authorship is generally reserved for individuals who have made significant contributions to the design, conduct, or analysis of a study or other research project. While language models can assist researchers in generating text, they do not have the capability to independently design, conduct, or analyse research, and therefore should not be considered as authors.

However, it may be appropriate to acknowledge the use of language models in the methodology or materials and methods section of a study or research paper, to ensure transparency and to allow for accurate interpretation and replication of results.

Managing Risk

When using these tools in research, it is also important that risks associated with the following issues be identified and measures are in place for reducing them:

- **Large-scale knowledge:** Although LLM's represent large-scale knowledge bases that can be used to generate information and support research, the source of the knowledge can be opaque.
- **Need for accuracy:** The information generated by language models may not be accurate and reliable, and errors can have significant consequences for users and consumers of the information.
- **Vulnerability to bias:** Large language models are susceptible to bias and can perpetuate existing stereotypes and inaccuracies, which can negatively impact the quality of information generated.
- **Potential for abuse:** Large language models can be manipulated or used for malicious purposes, such as spreading false information or generating biased results.
- **Responsibility for content:** Large language models raise questions about responsibility for the content generated, and who should be held accountable for errors, inaccuracies, or malicious use of the information.

HDR Specific Issues

Whilst the general research considerations apply to HDR candidates, additional issues arise related to the integrity of examination. UNSW has clear Guidelines on editing in Research Theses. The reflect the position of the Institute for Professional Editors, endorsed by the Australian Council of Graduate Research, which outlines standards for editing research theses. HDR Candidates should review the

[1] Kung, T. H. et al. Preprint at medRxiv https://doi.org/10.1101/2022.12.19.22283643 (2022)

[2] Stokel-Walker, C. Nature 613, 620-621 https://doi.org/10.1038/d41586-023-00107-z (2023)

IPED guidelines and ensure that the use of the tool is in line with the editing advice they provide, and that the tool is acknowledged where this is appropriate. Supervisors should be consulted for advice if there are questions on application of these guidelines.

HDR candidates should also consider any impact that the use of AI tools may have on the criteria that theses are examined against, which are outlined in the thesis examination procedure.

Principles

Considering the above, UNSW has taken the following position regarding the use of generative AI in Research:

- Generative AI tools are likely to find many useful applications in research.
- Care should be taken when using these tools to ensure that research integrity is maintained.
- Appropriate training is to be provided to assist researchers in navigating integrity requirements.
- The use of generative AI tools should be disclosed and appropriately cited where it would be reasonably expected (there may be reasons for non-disclosure such as privacy concerns or that the tool is generally expected to be used). This will help researchers maintain the integrity of their work and the trust of the research community. Given the rapidly changing nature of this field, a prescriptive definition of appropriate is likely to change, and may differ between research areas.
- Large Language Models such as ChatGPT cannot ever meet the requirements for authorship or inventorship. As such, researchers must take responsibility of the output of these tools if they are to be relied on in research.
- UNSW, through the Division of Research and Enterprise, will keep a watching brief on emerging AI technology and its applications as well as its impact on the responsible conduct of research at UNSW.

Prof. Dane McCamey
Pro Vice-Chancellor Research

Prof. Jonathan Morris
Pro Vice-Chancellor Research Training and
Dean of Graduate Research

March 2024

This document was produced using input from ChatGPT.