GÖTEBORGS UNIVERSITET

Guidance for ethical use of generative AI in research

This document is an English translation of the Guidance for ethical use of generative AI in research. In the event of ambiguity, the Swedish version always applies.

Introduction

This guidance is intended to support all research staff in making well-informed decisions regarding the use of generative AI in the research process, including whether, when, and how they choose to use generative AI. The guidance focuses on principles for the ethical and responsible use of AI. At the same time, researchers are expected to have a basic understanding of what generative AI is, how it works, and why it may raise concerns from a research ethics perspective. As generative AI is evolving rapidly and its applications vary across research disciplines, these principles need to be formulated in broad terms.

The European Commission has developed guidelines on the responsible use of generative AI in research ("Living Guidelines on the Responsible Use of Generative AI in Research"[1]). These guidelines build, among other sources, on ALLEA's "The European Code of Conduct for Research Integrity"[2] and interpret how AI can be used in an ethically responsible manner within research based on the principles set out in that code. The Commission's guidelines constitute an important source for this guidance and have been developed in light of the current state of technological development and the surrounding policy framework. Both are dynamic in nature, meaning that the guidelines and, consequently, this guidance will need to be adapted, further developed, and updated on a regular basis. The Research Board's Committee for Ethics in Research, in consultation with the Research Board, is responsible for conducting annual reviews of this guidance.

Responsibility

The individual researcher is always responsible for the quality, accuracy, and integrity of their research, including when generative AI tools are used. This requires a critical approach to AI-generated material and an awareness of the limitations of such tools, including systematic biases, fabricated content (hallucinations), unreliability resulting from a lack of reproducibility (i.e. varying outputs in response to identical inputs), low specificity in generated outputs, and factual errors. AI must never act as the driving force in the research process; its role must be limited to that of a supporting tool. While generative AI can be a valuable aid in various stages of the research process, AI tools lack the capacity to replace the researcher's human judgement, accountability, and contextual understanding in complex situations.

The University has a responsibility to ensure that all research staff are offered training in the ethical use of AI. With regard to doctoral students, the faculties are responsible for ensuring that they receive the necessary training in the ethical use of AI in research. Supervisors should support doctoral students in developing the ability to reflect critically and ethically on the use of AI, both in research generally and within their specific field of research.

Openness and transparency

When generative AI tools have been used in the research process, this should be disclosed openly and in a manner consistent with established practice within the relevant research field. More extensive uses, such as for data analysis, literature reviews, writing, identification of knowledge gaps, formulation of research objectives, or development of hypotheses, should be described. Where relevant, prompts and AI-generated material should also be made available in accordance with the principles of Open Science. However, views on which parts of the research process are appropriate for AI support may vary across research disciplines.

Researchers should take into account that generative AI tools may produce different outputs from the same input and should strive to enable scrutiny and reproducibility of their results and conclusions. They should also disclose the limitations of the tools used, including potential biases and any measures taken to mitigate them.

Sensitive information and sensitive contexts

When using generative AI, it is important to be aware that submitted or generated material may be reused, for example in the development and training of AI models. Unpublished, confidential, or otherwise sensitive material should therefore not be shared with external AI services unless there are clear legal guarantees that the information will neither be stored nor reused. Certain AI tools provided by the University of Gothenburg do not use uploaded information to train AI models.[3] This means that some types of unpublished material (provided they do not contain the kinds of information described below) may be used in these tools. At the same time, the University recommends that information containing sensitive personal data and other personal data requiring enhanced protection, information subject to confidentiality provisions, or other sensitive material (for example, information that must be handled in a particular manner under contractual obligations) should not be uploaded.[4] This is because such information is transferred to the companies providing the tools, which could constitute an unauthorised disclosure of information or otherwise be in breach of contractual obligations.

Output generated by generative AI may be particularly sensitive with regard to the protection of intellectual property rights and personal data. AI-generated material may contain elements derived from the work of others, which requires proper citation and respect for the authorship of others. The risk of plagiarism, whether involving text, code, or visual material, must therefore be carefully considered. Where personal data appear in AI-generated content, the researcher is responsible for ensuring that such data are processed in a lawful and ethically appropriate manner.

In processes where the use of generative AI may have implications for other researchers or organisations, such as peer review, the assessment of applications, or other evaluation processes, AI

may be used as a tool in certain parts of the process, but not in the researcher's own synthesis, evaluation, or judgement. This is intended to reduce the risk of unfair or biased assessments resulting from the limitations of such tools and to avoid unclear lines of responsibility arising from the use of generative AI in these contexts. For example, the Swedish Research Council states in its guidelines that generative AI tools must not be used in the scientific assessment of applications, but only to improve the language of written evaluations.[5]

Competence development and ethical reflection

As generative AI tools are evolving rapidly, researchers need to continuously update their knowledge of how these tools can be used in a responsible and effective manner. This can be achieved through training, exchange of experiences, and sharing of good practices.

Research may be affected by AI in different ways depending on the field of study, and the focus of ethical reflection on the use of AI may therefore vary across disciplines. In cases of co-authorship, joint reflection on the project's use of AI is required so that all authors can agree on the framework for the use of AI and take responsibility for its use. For doctoral students, an open dialogue and ongoing reflection together with supervisors are essential to ensure that generative AI is used in a thoughtful and ethical manner. In addition, uncritical or poorly considered use of AI may undermine the development of the skills and knowledge required for conducting scientific research.

The development and use of generative AI may have negative impacts on people and the environment, for example through resource consumption, working conditions associated with data processing, and the handling of data. Opportunities to mitigate such impacts should be taken into account by selecting appropriate tools and using them in a thoughtful, sustainable, and resource-efficient manner (for example, by minimising unnecessary runs, limiting the volume of generated output, and choosing models that are proportionate to the task at hand).

References

1. European Commission. (2024). *Living Guidelines on the responsible use of generative AI in research.* https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf
2. All European Academies (ALLEA). (2023). *The European Code of Conduct for Research Integrity.* https://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf
3. University of Gothenburg. *Guidance on the use of generative AI tools.* https://gunet.sharepoint.com/sites/mp-stod-och-service/SitePages/en/V%C3%A4gledning-vid-anv%C3%A4ndning-av-AI-verktyg.aspx
4. University of Gothenburg. *AI for staff.* https://gunet.sharepoint.com/sites/mp-stod-och-service/SitePages/en/AI-verktyg-f%C3%B6r-medarbetare.aspx
5. Swedish Research Council. *Guidelines for the use of AI tools.* https://www.vr.se/english/applying-for-funding/applying-for-a-grant/guidelines-for-the-use-of-ai-tools.html