Guidelines for responsible use of GenAI for UU employees

1. Introduction

The accelerating pace of technological advancements in Artificial Intelligence (AI), particularly Generative AI or GenAI, presents both unprecedented opportunities and challenges. Although GenAI has numerous applications that can enhance and complement the work of university staff, its use comes with risks and pitfalls that are essential to consider. So how can we make the most effective use of GenAI tools, while also working safely and responsibly.

These guidelines aim to help UU staff make informed decisions on the use of GenAI technologies while mitigating potential risks. While this document offers a broad framework, it is not exhaustive. Given the rapidly evolving nature of GenAI, guidelines will regularly need to be updated as new technology and insights arise. Specific guidelines for the use of GenAI in education are currently under development.

2. What is Generative AI?

Generative AI is a type of artificial intelligence technology that can generate new content, such as text, images, audio, or video, based on patterns it has learned from existing data. This class of models includes Large Language Models (LLMs), among others. The most widespread use case for current GenAI applications involves users interacting with a generative model by providing it with an input "prompt", whereby the model generates text, images, etc. The model's abilities are acquired as a result of training on very large datasets. These datasets could be a collection of research papers, books, images, or any other form of data. The model identifies patterns, structures, and correlations within this data during its training phase.

Once the model is trained, it can generate content that mimics the structure and style of its training data and input. For instance, if a model is trained on a dataset which includes funding proposals, it can in principle generate a draft proposal based on the patterns it has learned. Since the model is probabilistic (i.e., produces the most statistically likely output), the output can change if the prompt is worded differently or even if the same prompt is used multiple times. While GenAI can produce impressive results, it does not "understand" the content it produces or is trained on in a way humans do. It is more accurate to consider such models as reacting to user input based on recurrent patterns in data. Therefore, a model does not recognize if its output is factually (in)correct, misleading, manipulative, offensive or in other ways legally or ethically objectionable. Therefore, while it can be a powerful tool, any GenAI-generated content should always be reviewed and edited by a human.

3. Benefits and uses of GenAI in research (support)

GenAI offers potential benefits for researchers and support staff. Keep in mind that GenAI is not a substitute for expertise and human insight. It should therefore be used as a supplementary tool, when relevant and appropriate. Do not overly rely on GenAI in your work. A few examples of benefits and use cases of GenAI in research (support) include:

The list of these examples does not mean that GenAI can always be used in those contexts without reservations or is risk-free. The use of GenAI, if appropriate at all, should always come with caution, taking into account its various risks and limitations as outlined in section 4.

1. **Drafting and improving text:** GenAI can quickly generate first drafts for sections of a document, such as an article or proposal. Since writing a first draft can be a barrier, GenAI can

provide a starting point to build on and save researchers and research support staff valuable time. It can also help to polish text or make it more concise (for example to fit a word limit).

2. **Data analysis and pattern recognition:** GenAI excels at identifying complex patterns in large datasets, which can reveal insights that might be missed by human researchers. Always be very critical about the results as they may be false or biased.
3. **Cross-linguistics inclusiveness:** A good example of GenAI use is using automatic translation to overcome language barriers for non-native speakers (for example, translating emails). Since research papers and proposals tend to be written in English, GenAI is particularly helpful for non-native English speakers.
4. **Task Automation for research projects:** GenAI can help drafting progress reports, meeting minutes, or create summaries of meetings or communication chains.
5. **Summarize extensive/complex information:** GenAI can summarize lengthy or complex documents, making it easier to extract key insights and communicate important points to colleagues. For example, funding support staff can use GenAI to summarize and distill dense funding calls or eligibility criteria, ensuring that researchers are presented only with the information most relevant to them.
6. **Experimental design optimization:** AI models can help design more efficient experiments by suggesting optimal parameters, sample sizes, and methodologies based on previous studies.
7. **Literature review assistance:** GenAI may facilitate literature reviews by synthesising or summarising literature, but users should always fact-check the results. GenAI may not be suited to identify relevant literature or summarise it correctly. Outcomes depend on what the model was trained on, and/or whether the model is able to generate based on live web search. More reliable tools exist; an example is ASReview, developed at UU.
8. **Code Generation:** Increasingly, researchers find GenAI models useful to either generate programming code based on a linguistic prompt, to translate code from one language into the other, or to get useful hints on code continuation within their programming environment, as well as creating documentation and assisting with de-bugging of code. Code-specific models also exist.

4. Risk overview and mitigation: how to use AI safely and responsibly?

To use GenAI safely and responsibly in research (support), numerous legal, policy and ethical considerations and risks must be taken into account. Below, we provide key principles for the use of AI[1][2] for research in the UU, as well as the main risks of using AI and mitigation tips.

Guiding Principles

1. **Human accountability and human oversight:** Always remain the 'human in the loop' when working with GenAI. Ensure content generated by AI is factually accurate.
2. **Data protection and security:** Familiarize yourself with the UU data security policies and know how to find expert help on data security in the organization.
3. **Transparency; show and tell:** Proactively learn about the benefits and dangers of the GenAI tools you use or provide, and proactively communicate how you used it.
4. **Diversity, non-discrimination and fairness**: Be aware of bias in GenAI data and output. Make sure output is in line with the UU values regarding Equality, Diversity and Inclusion.
5. **Environmental and societal well-being**: Whenever possible, choose GenAI tools and instruments designed with climate and social considerations in mind. Do not use GenAI if

[1] Inspired by the Assessment List for Trustworthy AI (ALTAI) by the European Commission; via https://digital-strategy.ec.europa.eu/nl/node/806

[2] See also https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

your purpose can be achieved as effectively by other means not involving the use of GenAI.

4.1 Research integrity

4.1.1 Plagiarism and authorship

Utrecht University's position on research integrity is described in the UU Code of Conduct for Scrupulous Academic Practice and Integrity[3] and the Netherlands Code of Conduct for Research integrity (2018)[4]. In the latter, plagiarism is defined as "the use of another person's ideas, work methods, results or texts without appropriate acknowledgement". Whether and to what extent AI-generated text counts as plagiarism is currently still a grey area. It therefore remains your responsibility to ensure that submitted work reflects your own effort, also to reflect rules on authorship described in the Netherlands Code of Conduct for Research integrity.

4.1.2 Originality

When your work is expected to be original, it's important that the audience can reasonably assume that you created the content yourself. This expectation is lower in certain cases, like writing a standard instruction manual or a user guide. Even in those situations, if AI is used to help create the content, it's still important for a person to review it to make sure it's accurate.

4.1.3 Transparency

If you use GenAI in your workflow or research in a 'substantial' way (for example, use that involves more than basic text editing support), be transparent about your use of GenAI tools. Although 'substantial' as a term cannot be simply defined, the EU Living Guidelines on the Use of Generative AI in Research provide a starting point to consider this.

As research support staff, you may sometimes work with material (for example text) that was written by a researcher to whom you are providing support. If you enter their work in a GenAI tool, make sure to ask whether the researcher you are supporting feels comfortable with you using a GenAI tool and specify which tool you intend to use (and how).

As a researcher, appropriately and transparently acknowledge the use of the source/platform as you would any other piece of evidence/material in your submission. This will vary widely depending on how individuals may use AI tools in specific instances and/or the conventions in different disciplines. For more information on citing GenAI in your work, see Appendix 2.

Research Integrity'

- Ensure that any work you submit remains a reflection of your own effort.
- In case of substantial use, be transparent (inside and outside the organisation) about your use of GenAI.

4.2 Data protection and governance

sing GenAI involves processing large amounts of data, which can in many cases contain personal data. Depending on the AI tool you use, the data with which you prompt the model could eventually be used by the model developers to train and improve the models on which the tool is based. As a result, that

[3] https://www.uu.nl/sites/default/files/code_of_conduct_for_scrupulous_academic_conduct_and_integrity_-_en_def_0.pdf

[4] https://www.universiteitenvannederland.nl/files/publications/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf

data can 'leak', e.g. be reproduced as output for another user and become publicly available although that was not the intention beforehand.

Anything done with personal data with the use of any computing tools, including AI, is subject to strict regulations, and the legal meaning of personal data is much broader than the lay understanding of the term. For further information, please consult the intranet pages on personal data[5] and privacy[6]. The General Data Protection Regulation (GDPR, or AVG in Dutch) sets rules for handling personal data. If you use GenAI with personal data, you must follow these rules. In academic research, there are some exceptions, but they don't remove most of the key requirements.

Before you input data into any AI tool, consult a relevant officer of your department if the data you intend to input into the AI tool is personal data and what conditions apply: every organizational unit in faculties and corporate offices of the UU has experts on data governance and protection that can help you[7].

Protecting privacy and data security

- When using GenAI, treat the information you enter as if you were posting it on a public site (e.g., a social network or a public blog).
- Do not use any personal or sensitive data as input when using GenAI tools, unless you are certain that the data protection law is respected.
- Many risks pertaining to GenAI are not immediately visible. In many cases, it is necessary to assess the impact of your use of GenAI on data protection and privacy. You are not on your own: UU has experts that are available to assist and advise you with this process (see footnote below).

4.3 Intellectual Property (IP) Issues and GenAI

When using GenAI tools, it is crucial to be aware of potential IP issues. In the context of any AI-generated content, questions may arise about who owns the IP rights of the generated content. Current laws do not provide clear answers to AI-related IP issues yet. In case of substantial use of GenAI (see 4.1.3), keep a track record of your workflow (e.g. earlier text drafts and revisions, literature summaries, etc.) to be able to demonstrate that your work is original if needed (e.g. if there is doubt as to the origin of the work).

Similar to personal data, works of others entered into an AI tool can also accidentally 'leak'. The works of others you input into an AI tool are protected by copyright of the authors and/ or publishers. Inputting those works into AI tools may constitute a violation of copyright. Read the licences under which those works are published. While older licenses do not explicitly account for use of works in AI tools, they will still contain relevant general provisions, and more recent licenses may already contain AI-specific clauses.

Avoiding legal risks

Ensure that the training data of your GenAI model does not infringe on any third-party IP rights and until clear laws are in place, err on the side of caution. To mitigate these risks:

[5] See UU intranet page: https://intranet.uu.nl/en/knowledgebase/what-is-personal-data
[6] See UU intranet page: https://intranet.uu.nl/en/knowledge-base/privacy-at-uu
[7] See for example the UU Intranet page: https://intranet.uu.nl/en/knowledgebase/privacyofficers
Here you will find contact information of all privacy officers – their expertise includes but is not limited to privacy. In the course of 2024, most of them will also receive education and training on AI.

- When used 'substantially' (see 4.1.3) in for example publications or grant proposals, make sure to be transparent about the way GenAI was used.
- Keep a track record of interactions with GenAI.
- Carefully select your GenAI tool, taking into account IP and data security risks (see also section 5.1).
- Some tools let you opt *out* of using your data for model training (under user settings). Others let you opt *in* and don't use your data for model training unless you've provided explicit consent. When possible, opt out of model training or use tools that don't use input data for model training.

4.4 Quality of Training Data and Bias

GenAI models are only as good as the data they are trained on. If the training data is biased or incomplete, the GenAI's output will also be biased or incomplete. For example, if a model is only trained on text coming from western countries, it will include the wording, cultural biases, and ideas prevalent in these countries. If, for instance, you happen to be writing a proposal about mental health conditions that are more prevalent in non-western countries, this could be an issue. It is crucial to use GenAI tools that use diverse and representative training data, and/or to screen the output of such tools carefully for potential biases.

Avoiding bias

- Understand the GenAI's training data as much as possible. Diverse and representative data lead to fairer outputs.
- Stay critical and continuously assess the GenAI's output for biased patterns.

4.5 AI Hallucinations and output variation

GenAI recognizes patterns and produces statistically likely output without understanding the content the way a human does. It can therefore generate output that seems plausible but is factually inaccurate or even nonsensical - a phenomenon often referred to as "hallucination". The AI might for example confidently describe a non-existent psychological theory called "Cognitive Resonance Theory" and attribute it to a fictional psychologist. Or, when asked for references, it might cite publications that do not actually exist. Since GenAI models seldom express uncertainty, they will tend to state such things baldly as facts, making it harder to detect inaccuracies.

The output of a GenAI model can also vary substantially depending on the way you phrase your prompt, instruction, or question. Thus, effective use may sometimes require modifying a prompt to obtain more relevant or precise output. Note that since these models are probabilistic, the exact same output will never be obtained, even with the same prompt, over multiple queries.

Fact checking GenAI outputs

- Finetune your prompts to obtain relevant and accurate output and keep a record of your prompts.
- Always critically evaluate GenAI outputs and check for hallucinations.

4.6 Over-reliance on GenAI

While GenAI can speed up processes and help generate novel ideas, over-reliance can lead to a lack of oversight and critical thinking. Some GenAI tools are marketed as time saving 'AI Research Assistants' (e.g., Elicit, Scite, Scholarcy). If you choose to use them, use them with caution and be aware of their limitations. Be the human-in-the-loop (or even better; in the driver's seat) at all times and evaluate the accuracy in the output of the AI.

Maintaining balance with GenAI

- Use (and promote) a balanced approach, keeping in mind and emphasizing that GenAI is a tool to complement human expertise rather than replace it entirely.
- Form your own GenAI peer review community: for longer and/or more important output, apply the many pairs of eyes principles that you would also apply for scientific publications.

4.7 Environmental impact of GenAI

Running, and especially training, GenAI models require significant computational power, leading to high energy consumption and a substantial carbon footprint. Data centers use nearly 1-2% of the world's energy, and 1 average Chat GPT-3 prompt uses around 10 times more electricity than an average browser search, roughly the equivalent of lighting a LED light bulb for an hour[8]. Aside from energy use, GenAI indirectly requires large amounts of fresh water for server cooling. When considering the larger picture, the global demand for AI could result in 4.2 to 6.6 billion cubic meters of water withdrawal by 2027[9]. This figure is greater than the annual water withdrawal of 4 to 6 times that of Denmark or half of the United Kingdom.

Though using GenAI can have many benefits for your work, it is important to balance maximising the benefits of AI, while taking the environmental impact into account. Learning where GenAI adds most value in your specific workflow and being well informed about its resource consumption can help to find a middle ground between these conflicting values.

Reducing your environmental impact

- Ask yourself whether GenAI is really needed for the task, or whether a less energy-intensive solution (like a browser search or -in a research context- standard machine learning approach) would be sufficient or even better suited.
- Large, general-purpose models are powerful, but also use a lot of energy. Use smaller, task-specific models when possible for simple tasks (e.g., grammar checks or translation).

[8] Vries, The growing energy footprint of artificial intelligence, Joule (2023), https://doi.org/10.1016/j.joule.2023.09.004

[9] Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making ai less" thirsty": Uncovering and addressing the secret water footprint of ai models.

5. Effective Use of GenAI

Having addressed the potential benefits and the risks of using GenAI in research, how can you make the most of its potential?

5.1 Selecting the right GenAI tool

With the enormous choice of GenAI tools, each designed for specific tasks, it is important to know what part of your work could benefit from GenAI use, and what risks the use of each tool poses. Which factors do you need to consider?

1. **Assess the specific needs** of your research or research support tasks and choose a tool/model that aligns with those requirements. Different tools excel in various domains, so select one that best suits your project.
2. **Review data handling policies:** Ensure that the AI tool you use or develop adheres to UU/UMCU privacy, data management, and security protocols. Before choosing a tool, check how the provider handles your data. Does the tool use your input to improve its model? Look for tools that clearly state they will not use your data for training and that have strong protections for sensitive information. Prioritize open-source tools with transparent data practices and clear privacy commitments. If in doubt, ask for expert advice[10].
3. **Transparency**: Choose a tool that provides transparency about how it works and how it was trained. The provider should be able to explain this in understandable terms.
4. **Legal and ethical considerations**: Consider the tool's implications in light of the legal and ethical considerations, such as respecting privacy and IP rights and setting guardrails against toxic (e.g. racist or sexist) output. The provider should have guidelines on this.
5. **Environmental impact**: Your model choice will affect the environmental impact (like carbon footprint, energy use and water use) of your GenAI use. Models may for example be hosted in countries that predominantly rely on fossil fuels or have very limited fresh water supplies, which increases the environmental impact. See also section 4.7.
6. **Stay informed on model updates:** Regularly check for updates. This ensures that you benefit from the latest security features and improvements while staying informed about changes in data handling policies.

In short, before adopting a new GenAI tool, take some time for a “background check”: is the tool safe to use for your work? And how can you use it most efficiently? Exchanging experiences and best practices with colleagues can be helpful in the process.

6. Conclusion

There is still a lot to learn about the effective, safe, and responsible use of GenAI in research and research support. Technology as well as laws and guidelines are evolving rapidly. Guidelines will therefore continuously need to be updated based on the latest developments, and new knowledge will have to be shared within our Utrecht University community. This will allow us to make the most of the technology’s potential while adhering to policies and regulations (as well as values) on an international, national, and local level. By doing so, we will ensure that we contribute significantly to the ethical, and effective integration and development of GenAI in our work at Utrecht University.

[10] See for example the UU Intranet page: https://intranet.uu.nl/en/knowledgebase/privacyofficers
There you will find contact information of all privacy officers – their expertise includes but is not limited to privacy. In the course of 2024, most of them will also receive education and training on AI.

Appendix 1: Glossary of AI terms

AI: Artificial Intelligence (AI) refers to computer systems designed to perform tasks that typically require human intelligence, such as visual perception, decision-making, and language processing. These systems use algorithms and data analysis to learn, adapt, and make predictions or decisions, with applications spanning a broad range of fields.

GenAI: Generative AI or 'GenAI' refers to artificial intelligence systems capable of generating content probabilistically, based on patterns learned from vast amounts of existing data. These systems, when given prompts or instructions, can produce outputs that often mimic human-created work in style and quality.

GDPR (AVG in Dutch): The General Data Protection Regulation[11]. Legal framework in the EU, laying down rules that relate to the protection of natural persons, their personal data and their fundamental rights and freedoms. Note that the GDPR is a legal instrument to control aspects pertaining to privacy. That concept is in itself not mentioned in the GDPR.

GPT: Generative pre-trained transformer. A family of generative AI models developed by OpenAI, which underlie the widely-used ChatGPT system, among others. All are based on the 'Transformer' architecture, a type of neural network which was been in use since around 2017.

LLM: Large language model: A language model predicts continuations of a prompt, based on their probability. The probabilities come from the analysis of very large datasets. The term 'large' refers both to the size of the model (how many parameters it uses to model probabilities) and to the size of the data it is trained on. Generative AI tools are based on LLMs, though they often extend beyond language itself, to include images, video and audio.

Algorithm: a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.

Personal data: Any information that relates, either by means if content, purpose or impact, to an identified or identifiable living natural person, as defined in the GDPR.[12]

Privacy: The ability of an individual to control the disclosure and use of his personal information and the right to protection of personal data expressed, a.o. in the EU General Data Protection Regulation (GDPR). [13][14]

[11] Find the complete regulation here: https://eur-lex.europa.eu/legal-content/NL/TXT/?qid=1465452422595&uri=CELEX:32016R0679

[12] See intranet page on personal data: https://intranet.uu.nl/en/knowledgebase/what-is-personal-data

[14] Also see intranet page on what is meant by privacy: https://intranet.uu.nl/en/knowledgebase/what-is-meant-by-privacy

Appendix 2. How to cite/acknowledge GenAI in your work

Citation[15]

Quoting a text that was generated by AI (for example ChatGPT) can be compared to sharing the output of an algorithm. Therefore, you should mention the source, both in the text and in your reference list. The maker of the AI tool is considered the author, the date is the year in which the version of the AI tool you used has been released and as title you mention the name of the AI tool: **Name of Company/creator of generative AI Tool. (Year). *Name of the generative AI tool* (version release date) [Large language model]. URL.**

For example:

- In the bibliography: OpenAI. (2023). *ChatGPT* (Mar 14 version) [Large language model]. https://chat.openai.com/chat.
- In-text citation: When given a follow-up prompt of "What is a more accurate representation?" the ChatGPT-generated text indicated that "different brain regions work together to support various cognitive processes" and "the functional specialization of different regions can change in response to experience and environmental factors" (OpenAI, 2023; see Appendix A for the full transcript).[16]

Acknowledgement and documentation

It may be more appropriate to acknowledge and document the use of AI tools rather than to cite them, e.g. depending on the guidance for submitting your assessment or the guidance provided by your publisher.

A basic acknowledgement should include:

- Name and version of the GenAI system used, e.g. ChatGPT-3.5
- the company that made the AI system, e.g. OpenAI
- URL of the AI system.
- Brief description of how the tool was used
- Date the content/output was generated

For example:

I acknowledge the use of ChatGPT 3.5 (Open AI, https://chat.openai.com) as a tool to proofread the final version of this work.

You may also wish, depending on the circumstances, to include prompts that were used, copies of outputs that were generated or how you used or edited the content generated.

[15] Guidelines UU Library. See https://libguides.library.uu.nl/citing/APA#s-lg-box-wrapper-19102609
[16] Source, including more information: https://apastyle.apa.org/blog/how-to-cite-chatgpt