

FI Guidelines: How to Discuss LLMs and GenAI in Your Courses

What is GenAI?

Generative AIs (GenAI) are tools trained on large datasets used primarily to create text, images, code, and synthetic data sets. ChatGPT (Chat Generative Pre-Trained Transformer) is such a text-based tool, belonging to the class of large language model (LLM) tools. Because they generate outputs based on inputted training data, these sorts of GenAIs record and reify existing biases, errors, mis- and disinformation present in their source texts. This is well documented in the technical, philosophical, and science studies literature. ChatGPT, the most well-known GenAI LLM, is trained on text available freely on the internet including reliable and unreliable sources such as books, web articles, Wikipedia, and Reddit, with a temporal lag between when information is published and when ChatGPT becomes adequately trained on the available information. The code of GenAI LLMs relies on a probabilistic approach to word prediction and processing, using both the language it was trained on and algorithms it depends on to arrive at the most likely string of words or symbols, not the most truthful. ChatGPT is an iteratively trained LLM, meaning that it uses the prompts given by users as training material to further improve its accuracy and scope.

UU policy on Gen AI in education

ChatGPT exploded in popularity in the academic year of 2022-2023, triggering a global need to reassess guidelines on how to understand the role of GenAI and LLMs in academia. At Utrecht University, preliminary guidelines regarding the use of ChatGPT were issued stating that it will be considered *academic fraud* if a student uses ChatGPT or a similar GenAI tool to create text, presentations, or preparation material and attempts to submit this as their own work.¹ Beyond these cases, it is the position of the University that it is up to the discretion of the teacher to determine how and to what extent students may use ChatGPT and other GenAI LLM (referred to as GenAI in the remainder of the text for simplicity) tools in the classroom.² As such, faculties and departments have been left to develop their own approaches to dealing with GenAI in education, of which this document is one of many such approaches. At the Freudenthal Institute, we would like to take this opportunity to discuss not only how students *ought* to use GenAI, or how instructors *ought* to redesign their curricula in response to GenAI but instead enter into a dialogue about the limitations, dangers and benefits of using these new tools as they are emerging and becoming instantiated in our society.

Audience and Goals

Audience: Staff and students of the Freudenthal Institute, including the MSc students, PhD candidates, and Postdoctoral researchers.

¹ [https://students.uu.nl/en/news/guidelines-chatgpt#:~:text=Guideline%3A,or%20parts%20thereof\)%2C%20cf.](https://students.uu.nl/en/news/guidelines-chatgpt#:~:text=Guideline%3A,or%20parts%20thereof)%2C%20cf.)

² <https://www.uu.nl/en/education/education-at-uu/teaching/generative-ai>

Goals: To develop a cursory understanding of how these LLMs function, what their limitations are, and how we can critically and consciously engage with these tools; enter into an exercise in self-reflection and co-creation with our colleagues and students in order to cultivate a robust and well-informed approach to advising on and using GenAI in education and research.

Approach: Provide prompts and space for reflection on typical approaches to using GenAI in academia, giving space for initial thoughts, collaborative dialogue, and reflection. This document can be used to guide discussions with students and peers and should be approached as a tool for dialogue rather than an advice report. That being said, I will endorse one bit of explicit advice wrt lecturers: it is the responsibility of lecturers to determine the extent to which they are going to permit the use of GenAI in their courses, within the boundaries of the official UU policy. It is advised to take a proactive approach to setting clear boundaries for use in the courses you lead, without waiting for the situation to emerge when a student has openly admitted to using GenAI to address the question.

In the following part, we want to encourage you to consider a recent assignment you have (had) to make while answering the questions in the boxes. The questions are meant to make you reflect on the use of GenAI in your own research, and having an assignment in mind allows you to go through the specific steps you would take. Perhaps you have never considered using GenAI for some parts of your research process; it's still useful to reflect on why this would (not) be a good choice and to discuss this with other students to consider different perspectives.

Should I use GenAI for ...

GenAI can be used as copy editor.

How reliable is GenAI to check grammar and syntax? And for generating synonyms, or academic and scientific language?

Are there more specific tools for that? If so, what are they and what are their comparative advantages?

Does GenAI introduce expressions and tones I would not use? Or that are inappropriate?

Does the text feel 'mine' after I copyedited with GenAI?

GenAI can be used to assist in the development of research, e.g.:

Generate research hypothesis

Does GenAI consider the relevant literature for my research topic?

Can GenAI formulate a novel and original research question?

Conduct literature review on work on sources

What are the temporal constraints the datasets GenAI is trained against? What is the oldest and most recent information it has been trained on?

What sort of academic texts has GenAI been trained on? Do you know if it has been trained on scanned texts, texts behind paywalls?

Has GenAI been trained to discriminate between sources? Will it assign the same weight to an openly available academic text as a reddit post?

Assess contents and evaluate arguments

Is GenAI able to identify argumentative structures in sources? For example, will it recognize that it recommends two papers for the same purpose even though the arguments from the papers (indirectly) attack or undermine each other?

Is GenAI able to contextualise knowledge claims with respect to historical, philosophical, cultural, or other factors?

GenAI can be used to help to create and organize notes more efficiently.

*What sort of assumptions will GenAI make when I give it a text to summarize?
What information will it consider as valuable?*

How will my insights be limited by relying on an LLM to summarize texts?

I inputted substantial text in GenAI, now what?

Substantial use of GenAI can have consequences for other people's work.

How does GenAI utilize the input it receives from users? Am I not breaching privacy or intellectual property rights?

What parts of my own work would I consider using as input for GenAI? What if this work contains content from others or confidential information?

In what way can your own work as a GenAI output be concerning?

Impact on authorship

Can GenAI be listed as an author? If not why?

What does being an author of a scientific paper entail? Consider ethical, legal, creative responsibilities and obligations?

Consider the model architecture and training data that GenAI is based on. In what ways do you think the different components of your answer to the previous question align or clash with how GenAI was trained and how it operates?

The following question is derived from one of the dilemmas that is being used in the GSNS master course Dilemmas of the Scientist:

Imagine that your co-author has added ChatGPT as one of the authors for your research paper. The reason they give for this is because ChatGPT has written a substantial amount of their part of the research paper. Why would you (not) list ChatGPT as an author in this situation? Give two reasons for either option.

Using GenAI correctly?

What is a prompt? And how is a good prompt formulated?

Reflect on your experience and competence on prompt writing and how this might affect the quality of the output.

Copy editor

1. ChatGPT is less reliable than other specialized grammar/spelling/syntax AI tools. ChatGPT's reliability decreases with text length and the volume of grammatical issues in the text. ChatGPT's predictive process can be 'tailored' to produce text in a specific manner (i.e. 'scientific' or 'academic' tones) however, this requires precise prompt engineering and is only effective when multiple trials are conducted.
 - a. General point: prompt engineering and iterative trials are essential when working with ChatGPT.
 - b. General point: ChatGPT is not veridically oriented and therefore is neutral wrt the truth of its responses
2. Yes, there are more specific tools. Grammarly is a great example. Grammarly is a specific tool and therefore was trained with a specific intention, to adjust grammar and spelling. N.B. Careless use of Grammarly can also produce text that is meaningless or otherwise incomprehensible. Reliance on AI tools for grammar is only effective when there is sufficient Human-AI interaction and the Human possess a high degree of expertise.
3. Yes, ChatGPT will generate text that you might find disagrees with your intentions, your values, your principles, etc. ChatGPT's probabilistic, predictive algorithm means that the program is only concerned with predicting likelihood. Biases and hallucinations are likely.
4. The text might feel yours, however there is an evident linguistic bias in ChatGPT. Even if the text feels like you, there is a high chance that ChatGPT has introduced a degree of linguistic standardization. Do we want to lose track of the individual voice?



Generate Research Hypothesis

1. You do not know which texts ChatGPT is relying on to generate responses. You know with certainty that ChatGPT is not trained exclusively on academic literature, and that it is not biased to weigh the suspected veridical nature of the texts it is trained on.
2. There is evidence to suggest that ChatGPT can deal surprisingly well with the notion of novelty in scientific research.²

Conduct Literature Review on Work on Sources

1. Chat GPT 3.5 - January 2022, Chat GPT 4 Turbo – April 2023, Chat GPT 4 Turbo + Browse with Bing – Allows users to receive prompts from search engines in real time.
2. We don't know – the datasets are proprietary. Chat GPT can extract text from PDFs, but we do not know if ChatGPT was trained on PDFs. We can assume that ChatGPT has not been trained on text that is not freely available, however, ChatGPT Browse with Bing had to be temporarily suspended because it was capable of bypassing paywalls, so it might be the case that in the training paywalled materials were used. The core issue is, we simply do not know.
3. No, ChatGPT is trained to predict the likelihood of a word occurring in the context of the preceding words. The neural network architecture processes sequential data and does not take into consideration veracity or virtue.

Assess Contents and Evaluate Arguments

1. ChatGPT is generally quite argumentatively weak – it is not very good at deciphering the patterns of logical argumentation or of repeating that logical argumentation in the texts it generates. ChatGPT will hallucinate, generate false information, and generate misleading information. ChatGPT is not veridically oriented, it has no awareness of whether the texts it generates are truthful, only if there is a high credence of conformation with sequentially probabilistic data.
2. ChatGPT can generate text that can have the appearance of taking into consideration these factors with delicate prompt engineering, however, the veracity or depth of this is not reliable. Expert Human-AI interaction is needed in order to determine the extent to which ChatGPT is reliably generating such texts. ChatGPT itself is ignorant to any of these domains, it is exclusively concerned with generating sequences of texts that conform to a probabilistically determined architecture.

ChatGPT can be Used to Help to Create and Organize Notes More Efficiently

1. ChatGPT can be used to summarize texts, however, you do necessarily run the risk of missing important information as ChatGPT does not discriminate contextually. Human-AI interaction is needed still in order to determine the extent to which you are pleased with the response that ChatGPT has generated. This also means that you as an individual need to develop summarizing skills so that you can adequately judge the output.
2. While you are developing these skills, it is important to remain cognizant of the modes of knowledge production you are alienating yourself from by offloading labor to an AI. Repetition is valued for its didactic contributions, engaging in the process of summarizing class notes can play as substantive a role in developing an understanding as participating in class or reading the

summaries. You are depriving yourself of something by offloading that labor. Making risk-benefit calculations to determine whether you value efficiency or depth is relevant to this level of decision making.

Substantial use of ChatGPT Can Have Consequences for Other People's Work

1. ChatGPT (OpenAI) reserves the right to use the information you provide in prompts to further train their AIs. It is entirely possible that you are participating in the violation of intellectual property rights/violating copyright agreements depending on the license of the text you are uploading to ChatGPT.
2. Notes from the GSNS SPOC course on Authorship:

The European Code of Conduct (ALLEA) discerns 4 principles:

- 1) Reliability in ensuring the quality of research, reflected in the design, the methodology, the analysis and the use of resources.
- 2) Honesty in developing, undertaking, reviewing, reporting and communicating research in a transparent, fair, full and unbiased way
- 3) Respect for colleagues, research participants, society, ecosystems, cultural heritage and the environment.
- 4) Accountability for the research from idea to publication, for its management and organisation, for training, supervision and mentoring, and for its wider impacts.

These principles apply to good publication practice and authorship, e.g. honesty is explicitly described as reporting research in a transparent, fair, full and unbiased way.

Writing prompts

While it seems very easy to use a Gen-AI interface for generating text contents, prompt writing is a sub-field of engineering on its own and the quality of prompt writing may fundamentally affect the output.

Open prompt sharing & community tools:

- <https://hero.page/ai-prompts>
- <https://www.ptsearch.info/tags/list/>
- <https://ignacio-velasquez.notion.site/2-500-ChatGPT-Prompt-Templates-d9541e901b2b4e8f800e819bdc0256da>
- <https://github.com/prompts-lab/Awesome-Prompt-Engineering>

Other useful resources:

Christine Fox from the GSLS has produced a number of Open Education Resources on LLMs and AI with the intention to increase AI literacy for the students of GSLS. These modules are for students and lecturers and can be accessed by anyone with a SOLIS-ID.

- **GSLS Master's Student Guidelines:** [student guidelines](#)
- **GSLS Master's Student Tutorials:** [open-access tutorials](#)
- **GSLS Teacher Guidelines:** [Guidelines for Incorporating Generative AI \(GenAI\) in GSLS Education](#).

- **GSLs Teacher tutorials:** [open access tutorials](#)
- **GSLs GenAI in Master's research project supervision:** [GenAI guidelines for research projects for students and supervisors.](#)

We would also like to draw attention to a GenAI LLM created by scholars at EUR:

- ErasminLM (<https://www.eur.nl/en/news/launch-erasmian-language-model>) GenAI developed at EUR for academic research.

While the general concerns related to the use of GenAI in education and research remain, ErasminLM is more transparent about the data sets they have used to train their LLM. We are not advising the use of this GenAI, but are drawing attention to it for the sake of harm reduction.

The Commission and the European Research Area countries and stakeholders have jointly put forward a set of **guidelines** to support the European research community in their responsible use of **generative Artificial Intelligence (AI)**.

- [Living guidelines on the responsible use of generative AI in research | Research and innovation \(europa.eu\)](#)

This document has been drafted by Maura Burke, Odette Fischer, Marilou Niedda, and Federica Russo, with comments and suggestions from several FI colleagues. For any questions, comments, or suggestions, don't hesitate to contact us directly.