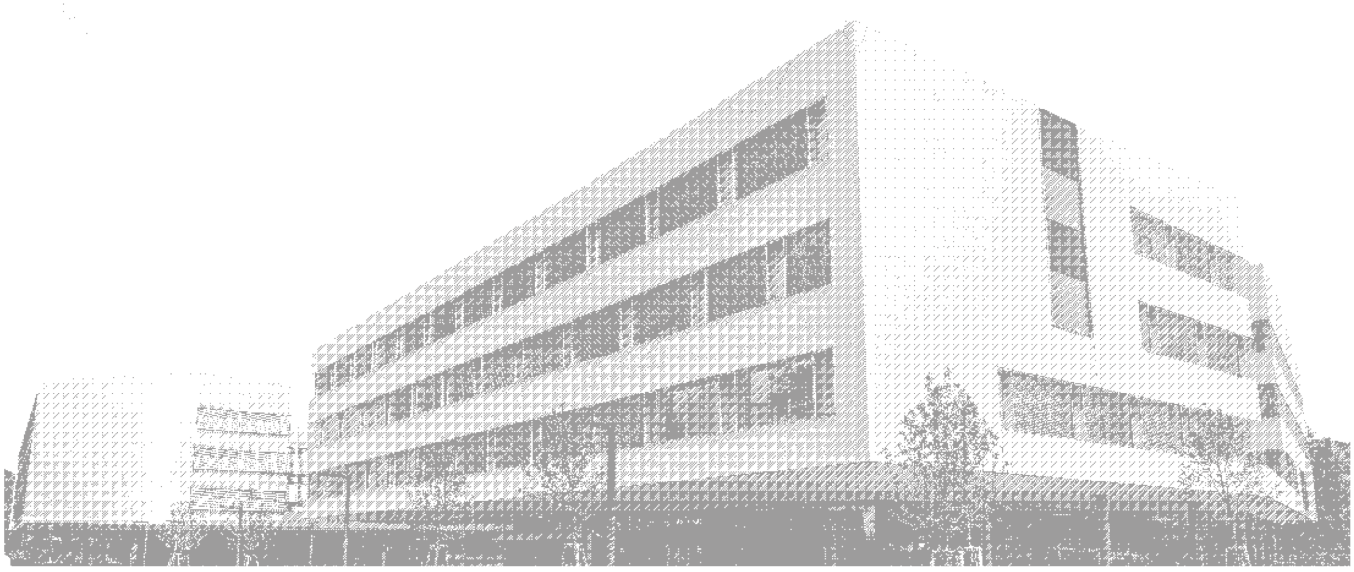
WE ARE EUDRES
EUROPEAN UNIVERSITY

fh st.pölten
UNIVERSITY OF APPLIED SCIENCES



Guidelines for the Approach to Generative AI Applications such as ChatGPT

Lisa David, Marlies Temper, Simon Tjoa, Lukas Richter

Version 1: 07.10.2024

Fachhochschule St. Pölten GmbH, Campus-Platz 1, 3100 St. Pölten, T: +43 (2742) 313 228, E: csc@fhstp.ac.at, I: www.fhstp.ac.at

Preface/Preamble

This document merges the “Recommendations for the Approach to AI Applications such as ChatGPT” enacted by the University of Applied Sciences St. Pölten Board with specific guidelines for the use of generative AI (i.e., applications). These guidelines have been strongly inspired by the Artificial Intelligence Act (AIA) of the European Union. The EU’s proposal for an AI Act aims at regulating the emerging developments in the AI sector, and the EU is one of the first large economies to establish harmonised rules for the development and use of AI. The legal framework became effective on 01 August 2024 and includes the classification of AI systems according to their risks, thereby establishing obligations and responsibilities for providers and users of AI.

The AI Act uses the four risk categories of unacceptable risk, high risk, limited risk, and minimal risk:

- “Unacceptable risk” refers to AI systems that violate fundamental rights or values of the European Union. Examples could be systems that compromise human dignity or make decisions that violate human rights.
- The category of “high-risk” AI systems refers to systems that pose a high risk to the safety, fundamental rights, or health of EU citizens. Examples include AI that is used in critical infrastructure, transportation, or healthcare.
- AI systems with “limited risk” are systems that do pose a certain risk, but less than high-risk systems. These can be AI applications in customer management or recruitment, for example.
- AI systems with “minimal risk” include AI systems that are considered safe and therefore require less regulation. These include, for example, simple chat bots or voice recognition systems.

Based on concrete use cases from everyday teaching and learning, the logic of the AIA was adapted by the authors for use in higher education (Higher Education Act for AI, HEAT-AI). The objective is to ensure the regulated use of generative AI tools in teaching at the St. Pölten University of Applied Sciences.

General Framework Conditions

AI-based generative language models (e.g., ChatGPT, Llama, DeepL, Micorosft CoPilot, Elicit) use machine learning and artificial intelligence to generate texts. They do this by calculating the probability of words in order to, for example, give human-like answers to questions.

Data Privacy

It is a violation of the General Data Protection Regulation (GDPR) to enter any confidential or personal data (e.g., from interviews) into these tools without the written consent of the affected person. A general principle for dealing with personal data is: When using AI applications or digital services in general, it is important to carefully check their approach to data privacy, which always needs to be fully in line with the European and national data protection regulations! This means that personal data may be processed with the help of AI systems only if the affected persons a) were accurately informed about the data processing beforehand and agreed to it, and b) the AI system is governed by the European and national regulations (such as the GDPR).

In other words, systems that do not indicate a transparent data protection system and might enable third parties to access the data, or that do not comply with the European and national regulations, must not be used. To enter confidential or personal data into such systems means an infringement of the General Data Protection Regulation, among other provisions. In case of uncertainty in terms of data privacy, the AI system in question must not be used to process any personal data.

Transparency

AI applications are considered as writing aids, which is why their output must be clearly declared as “generated by AI”. Exceptions are all use cases in the category “Minimal Risk of Usage”. When it comes to exams or other assessments, the use of such aids constitutes a fraudulent acquisition of achievements (see § 20 FHG and the referral to § 2a HS-QSG). The use of such aids in writing final theses is generally considered a pretence of one’s own scientific performance (see the St. Pölten UAS’ Guidelines for Scientific Work). Any exceptions to this rule are to be negotiated with the thesis supervisor beforehand and put down in writing. Additionally, the Declaration of Honour in the thesis is to make explicit reference to the use of any such aids.

Source Criticism

AI applications including ChatGPT are language models and not (yet) expert systems. They frequently produce made-up or plagiarised results. Just like with literature study and results from Internet search engines, it is imperative to carry out correct scientific research and critically examine any sources used.

Careful Use of AI Tools

- The use of ChatGPT and similar tools requires an account and, therefore, the disclosure of personal data including a telephone number. It needs to be clarified whether it is necessary to create an account for the acquisition of competencies in a course.

- Applications such as ChatGPT require great amounts of energy[1]. Furthermore, the working conditions of the people supplying the model with data are questionable[2]. Greater awareness in approaching such AI applications is definitely called for.
- Many of the resulting texts reproduce or consolidate certain societal norms and views (in other words: bias). Results should, therefore, be discussed together in class.

Scientific Integrity

As mentioned above, AI applications such as ChatGPT are language models and not (yet) expert systems. As results are sometimes copied from other sources or made up altogether, it is particularly important to carry out a sound scientific study including the verification of sources in dealing with these applications. Only persons who have previously acquired knowledge and competencies can make adequate use of these systems and correctly assess their results. This means that the acquisition of competencies needs to be ensured despite the existence of AI applications.

For Students: Responsible Use

Higher education is designed to enable the acquisition of research-based knowledge, professional and practical competencies, an awareness of social responsibility, and reflection capability. While the use of ChatGPT can, e.g., support the brainstorming of ideas, these applications also tend to hold out the promise of making student life easier. This, in turn, might mean that the above-mentioned goals of higher education are not achieved, and that the acquisition of actual competencies is carelessly skipped. Students are thus at risk of not living up to the qualification profile outlined in the curriculum after graduation.

For Lecturers: Review of and Reflection on Competency Goals

In order to prevent students from being tempted to use ChatGPT to make life easier for themselves and from failing to acquire the necessary competencies, lecturers need to consider competency goals and adequate examination formats. They should reflect on which learning outcomes can be attained within the framework of a course, and which methods may lead to these outcomes despite and/or with the aid of AI applications. The performance needs to be assessed in such a way that students' own achievements become visible. Examination methods and assignments have to be adapted accordingly, one example being a more or less elaborate interview accompanying the submission of a programming task, project, text, case study, research report, reflection, etc.

[1] Landwehr, Tobias (2023). Der Energiehunger von KIs. In: Süddeutsche Zeitung. Online: https://www.sueddeutsche.de/wissen/chat-gpt-energieverbrauch-ki-1.5780744?reduced=true [05.2023]

[2] Wolfangel, Eva (2023): Ausgebeutet, um die KI zu zähmen. In: Zeit Online. Online: https://www.zeit.de/digital/2023-01/chatgpt-ki-training-arbeitsbedingungen-kenia?utm_referrer=https%3A%2F%2Fwww.google.com%2F [05.2023]

HEAT-AI: Higher Education Act for Artificial Intelligence

Use Cases

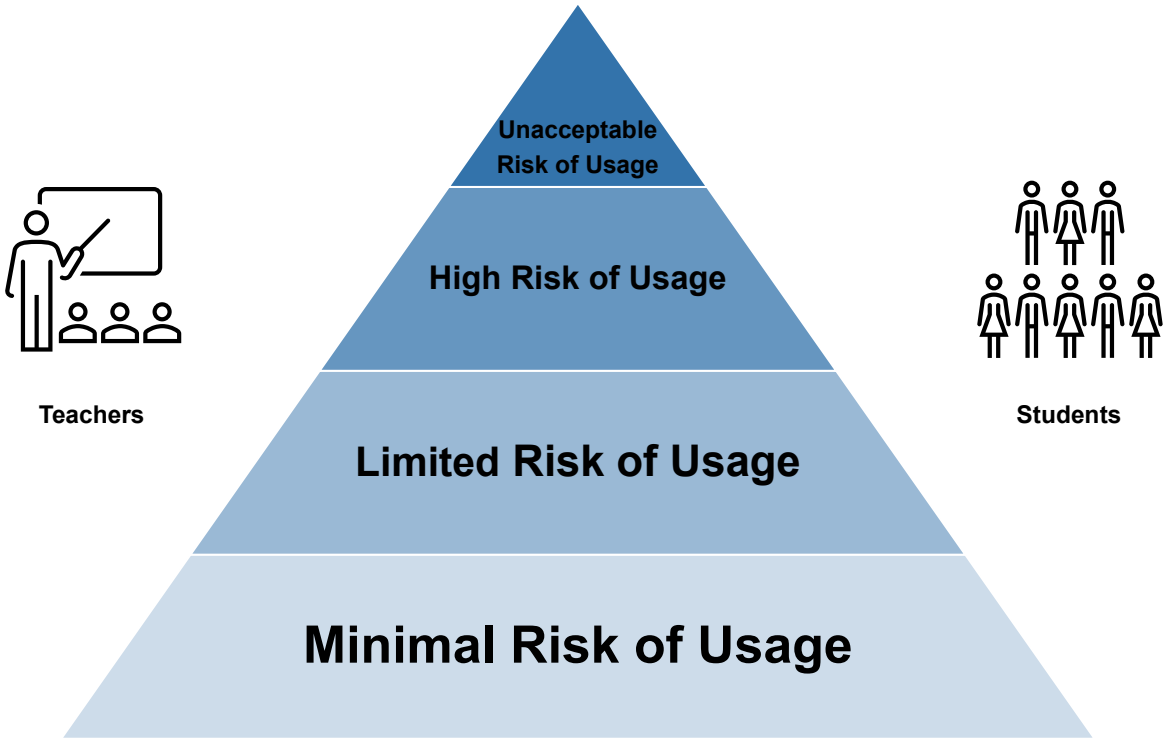


Figure 1: HEAT - AI

By transferring the use cases described above to the AI Act of the European Union, the authors from the St. Pölten University of Applied Sciences developed the Higher Education Act for Artificial Intelligence (HEAT-AI). Here, the four risk categories are described in more detail in the context of teaching and learning. The result is a table that serves as orientation for the use of generative AI tools.

Unacceptable Risk of Use

Areas whose use constitutes an unacceptable risk are prohibited for both teachers and students because in some cases, the use of generative AI even results in a violation of the legal framework conditions. It is thus forbidden to

- enter personal data into an AI tool without the explicit (written) consent of the affected person.
- enter personal data into an AI tool that constitutes a violation of the General Data Protection Regulation.
- claim AI-generated contents (texts, images, programme code, etc.) as one's own work.
- solve tasks by means of an AI tool alone (e.g., literature research where the AI tool searches for and summarises publications, unless explicitly demanded in the task description).
- grade students' performances using AI systems (lack of transparency).

Unacceptable AI use on the part of students is classified as a fraudulent acquisition of achievements, or plagiarism (see the St. Pölten UAS' Guidelines for Scientific Work), and measures are taken accordingly. Members of the teaching staff risk losing their teaching assignments or receiving a warning. Any legal infringements are reported.

High Risk of Use

The use of AI in teaching, which is considered a high-risk area, is strictly regulated. This category includes all areas of application where the integrity of science and knowledge transfer is at risk, or a violation of the above-mentioned principles might occur.

AI-generated content to be used in teaching/learning situations must be carefully examined and disclosed as such. More specifically, AI-generated content must be checked with regard to trustworthiness, validity, bias, and distortions. If these are used, it must be specifically marked in the text which prompt and which tool has led to this result.

Moreover, special care needs to be taken in the preparation of exams and exam questions, in the development of teaching materials, and in the formulation of feedback for students. In addition, the transcription of interviews using generative AI has been classified as a high-risk use of the technology because special attention must be paid to data protection here.

Limited Risk of Use

The concept of limited risk in the use of AI in teaching refers to the potential risks associated with insufficient transparency in the use of AI.

For example, this is the case when students use AI tools to generate content that helps them to achieve a different learning outcome (e.g., designing a website) or to optimise their self-developed programme code. Furthermore editing or translating text passages in final theses needs to be made transparent.

For lecturers, the creation of scenarios, simulations, sample companies, and application scenarios falls into the category of limited risk.

A declaration such as "AI-generated" or "created with the aid of AI" is sufficient in order to ensure transparency.

Minimal Risk of Use / Free Use

If the use of AI falls into the "minimal risk of use" category, the free use of AI is permitted. This is the case when generative AI serves as support only, constitutes no part of the examination modalities, and its results do not directly contribute to grading. Moreover, its use must not compromise any concrete competency goals. Examples include the brainstorming of ideas that are then used to develop own results.

Applications for Teaching and Learning at a Glance

The following table lists potential application scenarios for teachers and/or students including their classification in the four categories:

	Teacher	Student
Use Cases – Unacceptable Risk of Usage		
Disclosure of personal data to an AI tool a) without the persons' declaration of consent and/or b) in case the AI systems that do not comply with the GDPR	●	●
Disguise of AI-generated content as own work that is graded or reviewed	●	●
Assessment of course work, exams, and similar achievements using AI	●	
Purely AI-based literature research: The AI searches for and summarises publications	●	●
Use Cases – High Risk of Usage		
Transcription of interviews (without disclosing personal data to the AI)	●	●
Generation of exams and exam questions	●	
Development of teaching materials	●	
Supporting formulation of feedback on tasks and exams	●	
Use of AI-generated content (texts, images, programme code) in reports, exercises, assignments, theses, etc.		●
Use Cases – Limited Risk of Usage		
Creation of texts, images, and videos indicating that generative AI has been used, unless the content is directly related to the learning objective: For example, AI-generated images can be used to achieve the learning objective of creating a website independently.	●	●
Translation of texts into different languages (if the texts are part of the assessment)		●
Editing of texts: shortening, expanding, rephrasing, or linguistically correcting (if the texts are part of the assessment)		●
Creation of complex scenarios or simulations to familiarise students with theoretical concepts and promote problem-solving	●	

	Teacher	Student
Creation of use cases or example companies	●	
Optimisation of one‘s own programme codes		●
Use Cases – Minimum Risk of Usage		
Translation of texts into different languages (if the texts are not part of the assessment)	●	●
Editing of texts: shortening, expanding, rephrasing, or linguistically correcting (if the texts are not part of the assessment)	●	●
Use of AI to enable inclusive teaching (live subtitling for people with impaired hearing or audio descriptions for people with impaired vision)	●	
Use of AI as an innovation tool to come up with ideas: If the ideas are further developed, and the AI only served as a sparring partner, the author’s own and further developed ideas do not have to labelled as AI-generated.	●	●
Creation of interactive slides from trusted documents	●	●
Structuring and organisation of reports, papers, etc.	●	●
Creation of curricula and learning objectives	●	
Teachers can use generative AI to inspire students and encourage creative writing projects: For example, they could start a story that students then continue and edit.	●	
Use of AI to generate learning materials such as summaries, mind maps, or flashcards to support one’s own learning process		●
Use of suitable generative AI as a tutor to foster individual and personalised learning	●	●