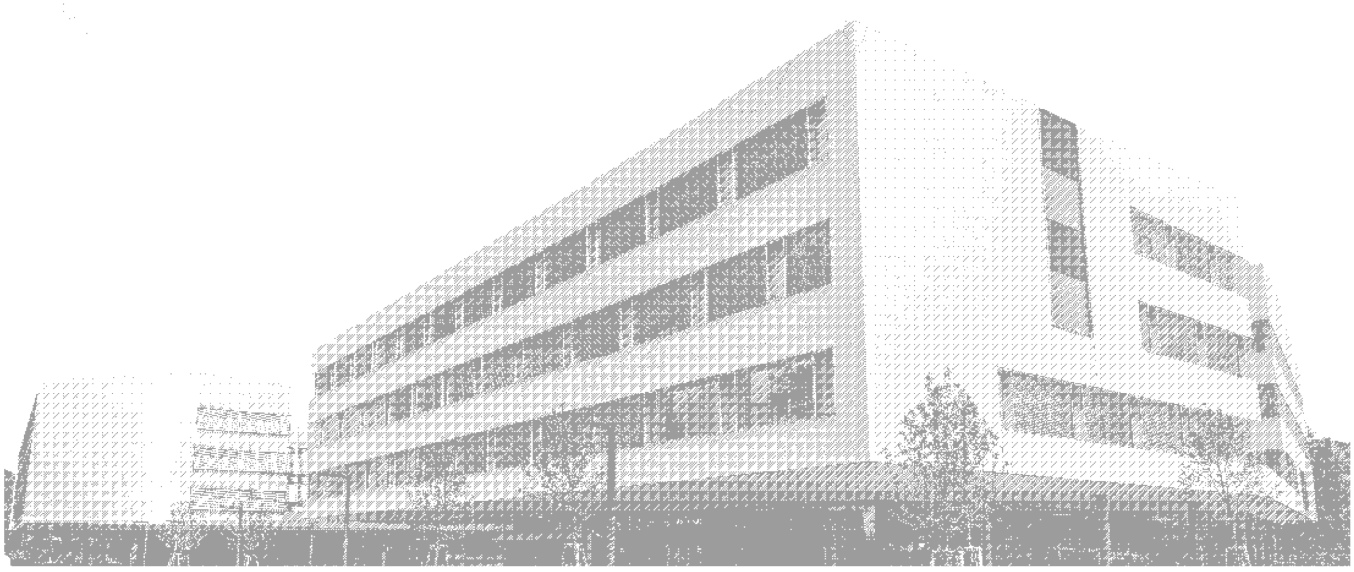
WE ARE EUDRES
EUROPEAN UNIVERSITY

fh st.pölten
UNIVERSITY OF APPLIED SCIENCES



Richtlinien für den Umgang mit generativen KI-Anwendungen an der FH St. Pölten

Lisa David, Marlies Temper, Simon Tjoa, Lukas Richter

1. Fassung vom 08.10.2024

Vorwort/Präambel

Das vorliegende Dokument vereint die im Juni 2023 vom Kollegium der Fachhochschule St. Pölten beschlossenen Empfehlungen im Umgang mit generativen KI-Anwendungen wie ChatGPT mit konkreten Richtlinien für den Einsatz generativer KI. Die Richtlinien sind angelehnt an den Artificial Intelligence Act (AIA) der Europäischen Union. AIA zielt darauf ab, die Regulierung von künstlicher Intelligenz zu stärken und den Einsatz von KI-Systemen innerhalb der EU zu harmonisieren. Das Gesetz ist mit 01. August 2024 in Kraft getreten und umfasst verschiedene Regelungen und Anforderungen für die Entwicklung, Bereitstellung und Nutzung von KI-Technologien.

Der AIA gliedert KI-Systeme in vier Hauptkategorien basierend auf ihrem Risiko und ihrer Anwendung. Diese Kategorien sind:

- **Verbotene KI-Systeme** (Unacceptable Risk) bezeichnet KI-Systeme, die als besonders gefährlich eingestuft werden und daher verboten sind. Beispiele sind Systeme, die die Würde der Menschen beeinträchtigen oder Entscheidungen treffen, die die Menschenrechte verletzen.
- Diese Kategorie der **hochriskanten KI-Systeme** (High Risk) bezieht sich auf KI-Systeme, die ein hohes Risiko für die Sicherheit, die Grundrechte oder die Gesundheit der Bürger*innen der EU darstellen. Beispiele hierfür sind KI in kritischen Infrastrukturen, im Verkehrswesen oder in der Gesundheitsversorgung.
- Bei **KI-Systemen mit begrenztem Risiko** (Limited Risk) handelt es sich um KI-Systeme, die ein gewisses Risiko darstellen, aber weniger als hochriskante Systeme. Dies können beispielsweise KI-Anwendungen im Bereich des Kund*innenenmanagements oder der Personalbeschaffung sein.
- **KI-Systeme mit minimalem Risiko** (Minimal Risk) umfassen KI-Systeme, die als sicher gelten und daher weniger regulierungsbedürftig sind. Hierzu gehören beispielsweise einfache Chatbots oder Spracherkennungssysteme.

Mithilfe konkreter Anwendungsfälle aus dem Lehr- und Lernalltag ist die Logik des AIA für den Einsatz in der Hochschulbildung von den Autor*innen adaptiert worden (Higher Education Act for AI – HEAT-AI). Das Ziel besteht in der Gewährleistung des geregelten Einsatzes von generativen KI-Tools in der Lehre der Fachhochschule St. Pölten.

Allgemeine Rahmenbedingungen

KI-basierte, generative Sprachmodelle (z.B. ChatGPT, Llama, DeepL, Microsoft CoPilot, Elicit) verwenden maschinelles Lernen und künstliche Intelligenz, um Texte zu generieren. Dazu werden Wortwahrscheinlichkeiten berechnet, um z.B. menschenähnliche Antworten auf Fragen zu erstellen.

Datenschutz
Das Eingeben von vertraulichen sowie personenbezogenen Daten (z.B. aus Interviews) verstößt ohne schriftliche Einwilligung der betroffenen Personen gegen die Datenschutz-Grundverordnung (DSGVO). Generell gilt im Umgang mit personenbezogenen Daten: Bei der Verwendung von KI-Anwendungen und anderen digitalen Diensten ist deren Umgang mit Datenschutz in jedem Fall zu prüfen, und muss den europäischen und nationalen Bestimmungen vollumfassend entsprechen! Dies bedeutet, dass personenbezogene Daten nur in KI-Systeme eingegeben werden dürfen, wenn die betroffenen Personen a) vorab über die Datenverarbeitung genau informiert wurden und dieser schriftlich zugestimmt haben und b) das KI-System den europäischen und nationalen Bestimmungen (etwa DSGVO) unterliegt.

D.h. Systeme, die kein transparentes Datenschutzsystem angeben und Dritten möglicherweise Zugriff auf Daten gewähren bzw. nicht den europäischen und nationalen Bestimmungen genügen, dürfen nicht genutzt werden. Das Eingeben von vertraulichen sowie personenbezogenen Daten verstößt daher unter anderem gegen die DSGVO. Bei Unsicherheit bzgl. des Datenschutzes darf das entsprechende KI-System nicht für den Umgang mit personenbezogenen Daten genutzt werden.

Transparenz
Generative KI-Anwendungen werden an der FHStP als Hilfsmittel gewertet und müssen somit ausgewiesen werden. Ausnahmen befinden sich in der Kategorie „Minimal Risk of Usage“. Im Rahmen von Prüfungen und anderen Leistungserbringungen in einer Lehrveranstaltung fällt eine unerlaubte Nutzung solcher Hilfsmittel unter Erschleichung einer Leistung (siehe FHG §20 und der Verweis auf §2a HS-QSG). Der unerlaubte Einsatz im Rahmen von Abschlussarbeiten wird grundsätzlich als Vortäuschen eigener wissenschaftlicher Leistung (siehe den FHSTP Leitfaden zum wissenschaftlichen Arbeiten) interpretiert. Ausnahmen müssen vorab mit dem*der Betreuer*in schriftlich vereinbart und in der Arbeit in der eidesstattlichen Erklärung explizit kundgemacht werden.

Quellenkritik
KI-Anwendungen wie ChatGPT sind Sprachmodelle und (noch) keine Experten*innensysteme. Häufig produzieren KI-Anwendungen erfundene bzw. plagiierte Ergebnisse. Genauso wie beim Umgang mit Literatur und mit Ergebnissen aus Internetsuchmaschinen ist eine korrekte wissenschaftliche Recherche sowie Quellenkritik unabdingbar.

Achtsamer Umgang mit den Tools
- Die Nutzung von ChatGPT und vergleichbaren Tools braucht ein Konto, für das persönliche Daten inkl. Telefonnummer eingegeben werden müssen. Hier ist abzuklären, inwiefern eine Kontoerstellung für den Kompetenzerwerb in einer Lehrveranstaltung notwendig ist.

- Anwendungen wie ChatGPT benötigen große Mengen an Energie[1]. Auch sind die Arbeitsbedingungen von Mitarbeiter*innen, die das Modell mit Daten versorgen, fraglich[2]. Daher ist ein bewusster Umgang mit derartigen KI-Anwendungen unbedingt nötig.
- In vielen entstehenden Texten werden bestimmte gesellschaftliche Normen und Sichtweisen (Bias) reproduziert bzw. verdichtet. Ergebnisse sollten entsprechend gemeinsam in der Lehrveranstaltung besprochen werden.

Wissenschaftliche Integrität

Wie oben erwähnt sind generative KI-Anwendungen wie das Sprachmodell ChatGPT keine Experten*innensysteme. Da Ergebnisse manchmal aus anderen Quellen kopiert wurden oder gänzlich erfunden sind, ist eine fundierte wissenschaftliche Recherche inkl. Quellenkritik im Umgang mit diesen Anwendungen besonders wichtig. Nur wer zuvor Wissen und Kompetenzen erworben hat, kann mit diesen Systemen adäquat umgehen und deren Ergebnisse korrekt einschätzen. Dementsprechend muss Kompetenzerwerb trotz der Existenz von Anwendungen wie ChatGPT sichergestellt werden. Zugleich benötigt es Anwendungskompetenz mit KI-Systemen, um einen sachgerechten, datenschutzkonformen und qualitätvollen Umgang zu gewährleisten.

Für Studierende: Verantwortungsbewusster Einsatz

Hochschulbildung zielt darauf ab, die Aneignung forschungsbasierten Wissens, berufspraktischer Kompetenzen sowie gesellschaftlichen Verantwortungsbewusstseins und Reflexionsfähigkeit zu ermöglichen. Obwohl das Verwenden von generativer KI z.B. bei Ideen-Brainstorming unterstützen kann, verführen solche Anwendungen dazu sich das Studierendenleben besonders leicht zu machen. Dies kann dazu führen die besagten Hochschulbildungsziele nicht zu erreichen und Kompetenzaneignung leichtsinnig zu überspringen. Dadurch besteht die Gefahr dem im Curriculum vorgesehen Qualifikationsprofil als Absolvent*in nicht zu entsprechen.

Für Lehrende: Kompetenzziele und Lernaktivitäten überprüfen

Damit Studierende bei gewissen Aufgaben nicht Gefahr laufen sich das Leben mit generativer KI so zu erleichtern, dass ein Kompetenzerwerb verhindert wird, ist es notwendig über Kompetenzziele und adäquate Prüfungsformate nachzudenken. Es ist daher zu überlegen, welche Lernergebnisse im Rahmen von Lehrveranstaltungen zu erreichen sind und welche Methoden trotz und/oder ggf. mit Hilfe von KI-Anwendungen zu diesen Ergebnissen führen. Die Leistungsfeststellung muss auf eine Art erfolgen, dass die eigene Leistung sichtbar wird. Hier müssen ggf. Prüfungsformate sowie Aufgabenstellungen angepasst werden. Ein Beispiel ist ein Abgabegespräch, das bei der Abgabe einer Programmieraufgabe, eines Projektes, eines Textes, eines Falles, eines Forschungsberichts einer Reflexion, u.ä., mehr oder weniger elaboriert geführt wird.

[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

Anwendungsfälle

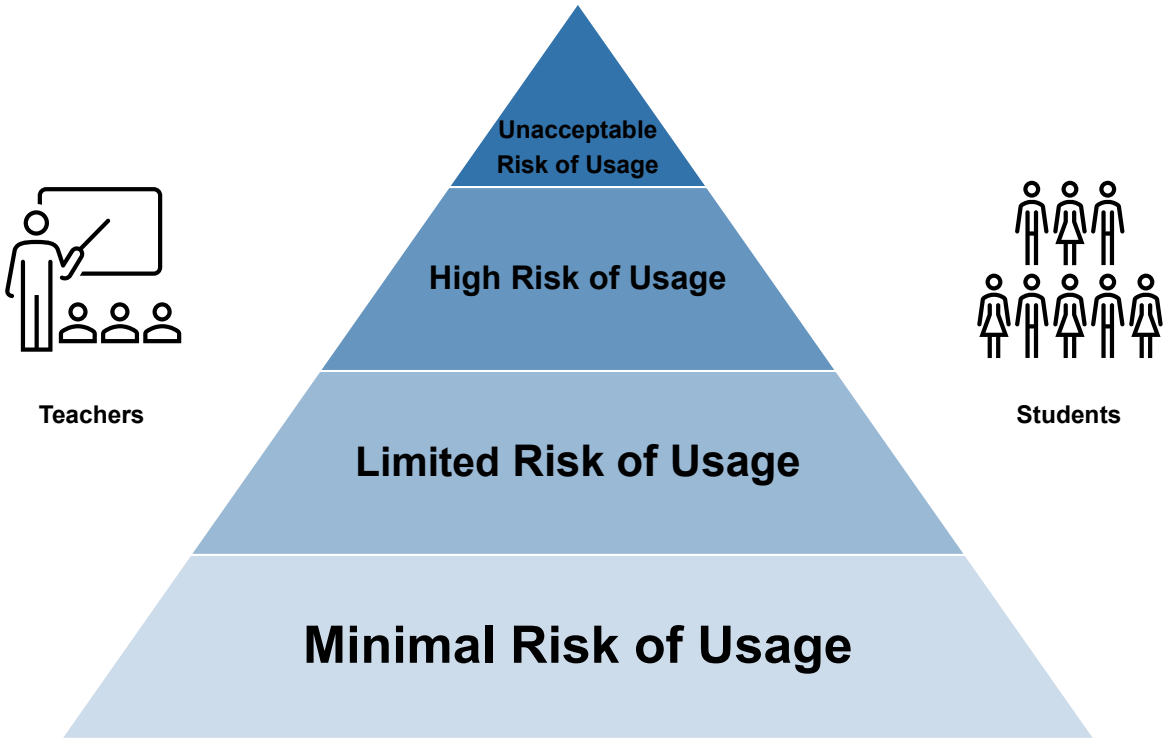


Abbildung 1: HEAT - AI

Aus der Übertragung der soeben beschriebenen Anwendungsfälle auf den AI Act der Europäischen Union ist der Higher Education Act for Artificial Intelligence (HEAT-AI) von den Autor*innen an der Fachhochschule St. Pölten entstanden. Hier werden die vier Risikokategorien in Bezug auf Lehren und Lernen genauer beschrieben. Daraus ergibt sich eine Tabelle, die als Orientierung für den Einsatz von generativen KI-Tools dient.

Unacceptable Risk of Usage – Inakzeptabler Einsatz

Bereiche, die ein inakzeptables Risiko bei der Nutzung darstellen, sind sowohl für Lehrkräfte als auch für Studierende untersagt, da der Einsatz von generativer KI in bestimmten Fällen sogar zu einer Verletzung von rechtlichen Rahmenbedingungen führt. Untersagt ist somit

- personenbezogene Daten in ein KI-Tool eingeben, ohne (schriftliche) Einverständniserklärung der betroffenen Personen,
- personenbezogene Daten in ein KI-Tool eingeben, das einen Verstoß gegen die Datenschutzgrundverordnung mit sich zieht,
- KI-generierte Inhalte (Texte, Bilder, Programmcode, …) als Eigenleistung auszugeben,

- Aufgaben rein mithilfe eines KI-Tools zu lösen (z.B. Literaturrecherche: Die KI sucht und fasst Publikationen zusammen, außer das ist explizit die Aufgabenstellung),
- studentische Leistungen durch KI-Systeme benoten (fehlende Nachvollziehbarkeit).

Inakzeptable KI-Nutzung von Studierenden wird als Erschleichung einer Leistung bzw. Plagiat (siehe dazu Leitfaden zum wissenschaftlichen Arbeiten) gewertet und entsprechende Maßnahmen werden eingeleitet. Lehrende riskieren den Entzug von Lehrveranstaltungen oder Verwarnungen. Sollte es zu Rechtsverletzungen kommen, werden diese gemeldet.

High Risk of Usage – Hochrisikobereich

Der Einsatz von KI in der Lehre, der als Hochrisikobereich betrachtet wird, ist streng reglementiert. Diese Kategorie besteht vorrangig aus Anwendungsfällen, die die Integrität von Wissenschaft und Wissensvermittlung gefährden.

KI-generierte Inhalte, die in Lehr-Lernsituationen eingesetzt werden, müssen sorgfältig überprüft werden, z.B. auf Vertrauenswürdigkeit, Validität, Bias und Verzerrungen. Werden KI-generierte Inhalte verwendet, muss durch eine entsprechende Markierung im Text konkret dargestellt werden, welcher Prompt in welchem Tool zu welchem Ergebnis geführt hat.

Besondere Sorgfalt gilt zudem bei der Erstellung von Prüfungen und Prüfungsfragen, bei der Entwicklung von Unterrichtsmaterialien sowie bei der Formulierung von Feedback für Studierende. Das Transkribieren von Interviews mithilfe von generativer KI ist zudem als hochriskanter Einsatz eingestuft, da hier ganz besonders auf Datenschutz geachtet werden muss.

Limited Risk of Usage – Begrenztes Risiko

Das Konzept des begrenzten Risikos bei der Nutzung von KI in der Lehre bezieht sich auf die potenziellen Risiken, die mit unzureichender Transparenz bei der Verwendung von KI verbunden sind. Dies ist z.B. der Fall, wenn Studierende KI-Tools nutzen, um Inhalte zu generieren, die dabei helfen ein anderes Lernergebnis zu erreichen (z.B. Erstellen einer Webseite) oder den selbst entwickelten Programmcode zu optimieren. Ist das Editieren bzw. Übersetzen von Textpassagen Teil der Prüfungsleistung (z.B. bei wissenschaftlichen Abschlussarbeiten) muss dies kenntlich gemacht werden. Für Lehrende fällt die Erstellung von Szenarien, Simulationen, Beispielen und Anwendungsfällen in die Kategorie des begrenzten Risikos. Eine Deklaration wie „KI generiert“ oder „mithilfe von KI erstellt“ ist ausreichend, um Transparenz herzustellen.

Minimal Risk of Usage – Freie Nutzung

Fallen die Nutzung von KI in die Minimal Risk of Usage Kategorie, wird die freie Nutzung von KI erlaubt. Das ist der Fall, wenn generative KI als Unterstützung dient und kein Teil der Prüfungsmodalität ist bzw. die Ergebnisse nicht direkt zur Benotung beitragen. Auch hier darf der Einsatz keine konkreten Kompetenzziele beeinträchtigen. Beispiele sind das Brainstormen von Ideen, aus denen dann eigene Ergebnisse entstehen.

Anwendungen für Lehren und Lernen auf einen Blick

Folgende Tabelle stellt mögliche Einsatzgebiete zutreffend für Lehrende und/oder Studierende dar, und zeigt die Einordnung in die vier Kategorien:

	Lehrende	Studierende
Anwendungsfälle – Unacceptable Risk of Usage		
Übergabe von personenbezogenen Daten an die KI a) ohne Einwilligungserklärung und/oder b) in KI-Systeme, die nicht der DSGVO entsprechen	●	●
Ausgabe von KI-generierten Inhalten als Eigenleistung	●	●
Automatisierte Benotung von Studienarbeiten, Prüfungen und ähnlichen Leistungen mittels KI	●	
Aufgaben rein mithilfe eine KI-Tools zu lösen (z.B. Literaturrecherche und -synthese)	●	●
Anwendungsfälle – High Risk of Usage		
Transkribieren von Interviews (ohne Übergabe personenbezogener Daten an die KI)		●
Erstellung von Prüfungen	●	
Entwicklung von Unterrichtsmaterialien	●	
Unterstützende inhaltliche Ausformulierung von Feedback zu Aufgaben und Prüfungen	●	
Übernahme von KI erstellten Inhalten (Texte, Bilder, Programmcode) in Berichten, Übungen, Abgaben, Abschlussarbeiten, usw.		●
Anwendungsfälle – Limited Risk of Usage		
Erstellung von Texten, Bildern und Videos unter Angabe, dass generative KI verwendet wurde, sofern der Inhalt nicht unmittelbar mit dem Lernziel verbunden ist (z.B. bei dem Lernziel, eine Webseite eigenständig zu erstellen, können KI-generierte Bilder verwendet werden)	●	●
Texte in andere Sprachen übersetzen (wenn Teil der Prüfungsleistung)		●
Texte editieren: kürzen, erweitern, umformulieren oder sprachlich korrigieren lassen (wenn Teil der Prüfungsleistung)		●
Erstellung von komplexen Szenarien oder Simulationen, um Studierenden theoretische Konzepte näher zu bringen und Problemlösungen zu forcieren	●	

	Lehrende	Studierende
Erstellung von Anwendungsfällen oder Beispielfirmen	●	
Optimierung von eigenen Programmcodes		●
Anwendungsfälle – Minimum Risk of Usage		
Texte in andere Sprachen übersetzen (wenn nicht Teil der Prüfungsleistung)	●	●
Texte editieren: kürzen, erweitern, umformulieren oder sprachlich korrigieren lassen (wenn nicht Teil der Prüfungsleistung)	●	●
Nutzung, um Personen inklusiven Unterricht zu ermöglichen (Live-Untertitelung für Menschen mit eingeschränktem Hörvermögen oder Audiobeschreibungen für Menschen mit geringem Sehvermögen)	●	
Nutzung von KI als Innovationstool, um auf Ideen zu kommen (Werden die Ideen weiterentwickelt und diente die KI nur als Sparringpartner, müssen die eigenen und weiterentwickelten Ideen nicht gekennzeichnet werden.)	●	●
Erstellung interaktiver Folien aus vertrauenswürdigen (eigenen) Dokumenten	●	●
Strukturierung und Gliederung von Berichten, Arbeiten, usw.	●	●
Erstellung von Lehrplänen und Lernzielen	●	
Inspiration von Studierenden, kreative Schreibkompetenz zu fördern (z.B. KI beginnt Geschichte, die Studierende dann fortsetzen und bearbeiten)	●	
Einsatz von KI, um Lernmaterialien wie Zusammenfassungen, Mindmaps oder Flashcards zu generieren, um den eigenen Lernprozess zu unterstützen		●
Verwendung von KI-gestützten Tutor*innen für individuelle und personalisierte Lernunterstützung	●	●