

KI-gestützte Textproduktion in der Hochschule

Johanna Gröpler^a, Margret Mundorf^b
und Nicolaus Wilder^c

^a  <https://orcid.org/0000-0002-1546-9843>, ^b  <https://orcid.org/0000-0001-7992-6233>,

^c  <https://orcid.org/0000-0002-5884-6768>

Abstract Seit der Veröffentlichung von OpenAIs Chatbot ChatGPT sind auf künstlicher Intelligenz basierende Tools auch in der breiten Öffentlichkeit angekommen. Welche Möglichkeiten sich für ihren Einsatz an Hochschulen bieten, aber auch welche Herausforderungen sich ergeben, soll in diesem Artikel erörtert werden.

Keywords Künstliche Intelligenz, Textgenerierung

1. KI und die Disruption der akademischen Welt

Künstliche Intelligenz (KI) in Verbindung mit natürlicher Sprache wurde in der Öffentlichkeit lange Zeit nur mit Service-Chatbots in Verbindung gebracht, die auf Anfrage etwa Auskunft über die Öffnungszeiten einer Bibliothek oder die Allgemeinen Geschäftsbedingungen einer Telefongesellschaft geben. Diese waren bislang an ihrem begrenzten „Wissen“, verstanden als auf einen bestimmten Anwendungsbereich zugeschnittene Fähigkeiten der Informationsverarbeitung, zu erkennen (Adiwardana & Luong 2020). Dies änderte sich schlagartig im November 2022, als das US-Unternehmen OpenAI ChatGPT, ein KI-basiertes Tool für dialogisch geführte Gespräche, veröffentlichte. Die Anwendung ermöglicht es, mithilfe von Eingabeaufforderungen (*Prompts*) in natürlicher Sprache in einem Chatfenster mit einem auf Algorithmen basierenden System menschenähnliche Dialoge zu führen, etwa Fragen zu stellen, sich Sachverhalte erklären zu lassen oder komplexere Handlungen auszuführen. Da die Ausgaben des Systems plausibel und kohärent erscheinen, ist vielen Nutzer*innen nicht bewusst, dass diese nicht auf unstrittigen Daten beruhen, sondern lediglich das Ergebnis statistischer Berechnungen sind. Bereits die Chats mit der auf dem Sprachmodell GPT-3 und der verbesserten Version GPT-3.5 basierenden kostenlosen

Variante von ChatGPT¹ war in vielerlei Hinsicht nicht von der Interaktion mit einem Menschen zu unterscheiden. Der Turing-Test – entwickelt, um maschinelles von menschlichem Denken zu unterscheiden – schien bestanden (Borchers 2022).

Schon 2019 veröffentlichte Springer Nature unter dem Pseudonym *Beta Writer* die erste komplett von einer KI generierte Fachpublikation aus der Lithium-Ionen-Batterieforschung (Beta Writer 2019). Darin wurden sämtliche Artikel zum Thema Lithium Batterien zusammengefasst und referenziert. Dies gab einen Ausblick darauf, wie im wissenschaftlichen Arbeiten der Schritt des Literature Review künftig automatisiert werden könnte (Wilder et al. 2022, 217–218). Im Bildungsbereich löste der Hype um ChatGPT gleichermaßen dystopische Panik, utopisch-naive Verklärung und heftig geführte Debatten zu schriftlichen Prüfungsformen wie Haus- und Abschlussarbeiten aus: Wie kann ein Text den Erkenntnisprozess und Wissensstand einer Person dokumentieren, wenn er womöglich von einer Maschine erzeugt wurde? Während auf der einen Seite skeptische Stimmen laut wurden, dass Täuschungsversuchen Tür und Tor geöffnet sei und diese drastisch ansteigen könnten, stellten andere, darunter auch das Marketing der Anbieter, die Vorzüge des KI-Einsatzes in der Bildung, der nicht weniger als eine Bildungsrevolution entfachen sollte, heraus: Jede Person unabhängig von ihrem Lernstand, ihren Fähigkeiten, ihrer sozio-ökonomischen Bedingungen etc. stehe nun ein individueller Lernbegleiter zur Verfügung, der den Lernprozess in allen Phasen unterstützen könne und so einen erheblichen Beitrag zum Kompetenzerwerb der jeweiligen Person leiste. Die kontroverse Diskussion um Chancen und Risiken beim Einsatz KI-gestützter Anwendungen im akademischen Kontext nahm also unmittelbar mit der Veröffentlichung von ChatGPT Fahrt auf und wird den Bildungsbereich auch in Zukunft noch beschäftigen und herausfordern. Im Kern führen die im Bildungskontext entstandenen kontroversen Diskurse zu der wohl unumgänglichen Zukunftsaufgabe, Lehr-, Lern- und Prüfungskulturen grundlegend zu überdenken (Weßels et al. 2022).

Disruptive Auswirkungen zeichnen sich insbesondere auf den akademischen Schreibprozess, letztlich auf die Sprache und ihren Gebrauch sowie auf die sprachliche Aneignung und Vermittlung von Wissen ab. Die weitere Entwicklung und deren Bewertung ist kaum abzusehen und entscheidet sich wohl weniger an der Technologie der KI-Systeme an sich, sondern vielmehr an den Fragen, wann diese wie und wozu eingesetzt werden (Limburg et al. 2023).

Um den gegenwärtigen Diskussionsstand skizzieren und einordnen zu können, erläutert der vorliegende Artikel zunächst die Funktionsweise von *Large Language Models* (LLMs; große Sprachmodelle), auf denen die aktuellen Tools wie ChatGPT von

1 Am 14.03.2023 wurde das Nachfolgemodell GPT-4 als kostenpflichtige Version veröffentlicht und um multimodale Fähigkeiten sowie den Zugriff auf Daten aus dem Internet erweitert. Seit November 2023 gibt es die Möglichkeit, für bestimmte Aufgaben spezialisierte Plugins zu nutzen bzw. eigene GPTs mit eigenen Daten zu erstellen. Seit Mitte Mai 2024 ist GPT-4o (omni) für alle kostenlos zu nutzen.

OpenAI² oder das Open-Source-Projekt *Open Assistant*³ basieren. Generierte Outputs dieser Systeme lassen sich so mit einhergehenden Chancen und Risiken systematisch analysieren und künftige Entwicklungen und ihre Bedeutung für den Bildungsbereich einschätzen. Im Anschluss werden beispielhaft mögliche Einsatzszenarien in Lehre und Forschung dargestellt und bisherige Aktivitäten an Hochschulen beleuchtet, um Lehrenden und Studierenden eine Orientierung zu geben. Dabei werden Herausforderungen wie Datenschutz, Urheberrecht und Bias mitberücksichtigt.

2. Zur Funktionsweise von KI-Textgeneratoren

Beim Begriff *künstliche Intelligenz* handelt es sich um einen Oberbegriff für verschiedene Systeme, deren Algorithmen darauf ausgelegt sind, originär menschliche Aufgaben mehr oder weniger eigenständig zu übernehmen. Nach dem aktuellen Entwurfsstand des EU AI Act (Art. 3), dem Gesetzentwurf, der eine Regulierung von KI-Systemen innerhalb der Europäischen Union vorsieht, werden KI-Systeme definiert als

Software, die mit einer oder mehreren der in Anhang I aufgeführten Techniken und Konzepte [das sind im Wesentlichen maschinelles Lernen, Logik- und wissensgestützte Konzepte sowie statistische Ansätze] entwickelt worden ist und im Hinblick auf eine Reihe von Zielen, die vom Menschen festgelegt werden, Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das Umfeld beeinflussen, mit dem sie interagieren. (Europäische Kommission 2021, 46)

Ein Teilbereich mit besonderer Bedeutung für die Leistungsfähigkeit gegenwärtiger Systeme ist das *Machine Learning* (ML), bei dem das System selbstständig lernt: Dabei werden auf Grundlage eines vorgegebenen Algorithmus aus Trainingsdaten (Beispielen) Muster und Gesetzmäßigkeiten (ein Modell) extrahiert, um daraus Voraussagen abzuleiten (das Modell also auf unbekannte Beispiele anzuwenden). Eingesetzt wird ML in der Regel dann, wenn die auszuführende Aufgabe zu kompliziert ist, um sie in eindeutigen Regeln abzubilden (Döbel et al. 2018, 8). Exemplarische und typische Anwendungsbereiche von ML sind Bilderkennung oder Textgenerierung. Das Konzept „Hund“ beispielsweise intensional so genau zu bestimmen, dass es auf Bildern extensional exakt alle Hunde erkennt und keine anderen Objekte als Hunde identifiziert, ist, wenn überhaupt, nur mit sehr ausführlichen Beschreibungen, die alle Einzelfälle abdecken, möglich. Sehr viel effizienter ist es hingegen, einen Algorithmus

2 S. <https://chat.openai.com>, zuletzt aufgerufen am 18.06.2024.

3 S. <https://open-assistant.io/de>, zuletzt aufgerufen am 18.06.2024.

einen großen Datensatz an Hundebildern analysieren zu lassen. Dieser kann so selbst ein Modell des Konzeptes „Hund“ entwickeln, welches zur Bestimmung weiterer Bilder mit der Unterscheidung von Hunden vs. Nicht-Hunden eingesetzt werden kann. Gleichermaßen verhält es sich bei Sprache, deren Wesen sich nicht auf syntaktische Regeln reduzieren lässt.

Beim ML wird üblicherweise zwischen drei unterschiedlichen Verfahren, sogenannten Lernstilen, unterschieden: (a) *überwachtes*, (b) *unüberwachtes* und (c) *bestärkendes Lernen*. Beim *überwachten Lernen* (a) müssen im Vorfeld die Trainingsdaten von Menschen in einem aufwendigen Prozess mit Labels, einer Annotation mit der Zuordnung einer zugehörigen Kategorie für die gewünschte Ausgabe, versehen werden. Dazu werden z. B. unterschiedlichen Tierfotos die Labels Hund, Katze oder Maus – die als Ergebnis bei diesem Lernstil bereits bekannt sein müssen – zugeordnet. Die Maschine erstellt dann im Wesentlichen über die Beziehung zwischen Label und Bild, Text o. ä. ihr Modell. Da die Vorhersagen des Modells mit den bekannten richtigen Ergebnissen abgeglichen werden können, kann der Lernprozess „überwacht“ werden (Müller & Guido 2017, 2). Beim *unüberwachten Lernen* (b) hingegen entfällt die Vorbereitung der Daten und der Algorithmus sucht eigenständig nach Mustern in dem eingegebenen Datensatz. Unüberwachtes Lernen wird insbesondere dann eingesetzt, wenn keine klaren Zielwerte vorgegeben sind (Patel 2019, 7–8): z. B. zur Anomalieerkennung bei der Analyse von Finanztransaktionen, um Betrugsversuche zu identifizieren, oder bei Empfehlungssystemen, die basierend auf den Vorlieben der Nutzenden Vorschläge generieren für Musik, Filme oder Käufe, aber auch zur Textanalyse, um beispielsweise bestimmte Themen oder Trends in einer Sammlung zu identifizieren. Beim Lernstil des *bestärkenden Lernens* (c) ist die KI darauf programmiert, eine Strategie zu entwickeln, die Belohnungen zu erhöhen, die sie als Ergebnis ihrer Interaktion mit einer Umwelt erhält. Wird eine Strategie belohnt, steigt die Wahrscheinlichkeit, diese wiederzuverwenden; wird sie nicht belohnt oder negativ sanktioniert, sinkt die Wahrscheinlichkeit. Es handelt sich also im Wesentlichen um ein Trial-and-error-Verfahren (Sutton et al. 2018). Bestärkendes Lernen wird u. a. eingesetzt zum Erlernen von Spielen wie Go oder Schach, wo Siege als positive Verstärker und Niederlagen als negative in die Modellentwicklung einfließen. Aber auch die Bewertungsmöglichkeiten von Output z. B. bei ChatGPT fließen so in die Verbesserung der Modelle ein.

Maschinelles Lernen ist die fundamentale Grundlage für die Entwicklung von *Large Language Models*, KI-Modellen, die auf *Natural Language Processing* (NLP), der maschinellen Verarbeitung natürlicher Sprache, basieren. Es trainiert diese Modelle daraufhin, Textdaten zu analysieren und brauchbare Vorhersagen für den nächsten Textabschnitt zu treffen.

Large Language Models berechnen Korrelationen auf Basis ihrer Trainingsdaten und somit die wahrscheinlichste Textausgabe für die jeweilige Eingabe der User*innen. So entstehen häufig plausible, auf der sprachlichen Textoberfläche kohäsive Texte, bei denen – etwa aufgrund des begrenzten *common ground* und Weltwissens –

jedoch eine „Illusion von Kohärenz“ (Lotze 2016) entstehen kann, weshalb ihnen auch die viel diskutierte Bezeichnung als „stochastische Papageien“ (Bender et al. 2021) zuteil wurde.

Die Wahrscheinlichkeiten für den Output werden dabei jedoch nicht auf Wortebene, sondern auf Tokenebene berechnet, da so auch passende grammatische Endungen oder Interpunktionszeichen vorhergesagt werden können. Token sind bedeutungstragende Einheiten, die über den Prozess der Tokenisierung hergestellt werden (Michelbacher 2013, 8). Bei GPT-3.5 umfassen Tokens im Schnitt vier Buchstaben, was laut OpenAI ca. $\frac{3}{4}$ eines englischen Wortes entspricht (OpenAI 2023).

Exemplarisch für GPT-3 zeigt Abb. 1, dass nicht immer das Token mit der höchsten Wahrscheinlichkeit gewählt wird. Grund dafür ist der Parameter der *Temperatur*, die als Einstellungsvariable die Kreativität bzw. den Zufallsanteil des Modells festlegt. Je niedriger die Temperatur gesetzt wird, desto höher ist die Wahrscheinlichkeit für die Auswahl des wahrscheinlichsten Tokens. Eine Temperatur auf dem niedrigsten Wert 0 würde daher auf eine gegebene Eingabe immer mit der gleichen, eben wahrscheinlichsten, Ausgabe antworten. GPT-4, befragt nach einem Vorschlag für einen Namen für eine Disziplin, die sich mit den Auswirkungen von KI auf die Theologie beschäftigt, macht bei niedriger Temperatur den Vorschlag „Theologische KI-Studien“, bei hoher Temperatur hingegen „Divine Datology“.

Die Beispiele aus Open AIs GPT-Familie (das Akronym GPT steht für *Generative Pretrained Transformer*) und LaMDA (*Language Model for Dialogue Applications*) verwenden für das Generieren der Wahrscheinlichkeiten eine Transformerarchitektur, eine spezielle Form von neuronalem Netzwerk, das Google 2017 entwickelt hat (Vaswani et al. 2017; Collins & Ghahramani 2021). Diese Architektur ermöglicht eine schnellere und präzisere Analyse der Eingabe, indem durch die Eingabe des Prompts – der textbasierten Eingabe in ein KI-System – sowie der vorangegangenen Konversation im Falle von Dialogmodellen die Wahrscheinlichkeiten für die zu generierenden Tokens auf bestimmte semantische Räume gerichtet wird (sogenannter Aufmerksamkeitsmechanismus). Wird also bei einer Eingabe beispielsweise das Wort

KI-Technologien beeinflussen die Art und Weise, wie Menschen Informationen erhalten und nutzen.



Abb. 1 Wahrscheinlichkeitsverteilung für den folgenden Token im Playground von OpenAI

„Turm“ verwendet, ist für die Berechnung der Wahrscheinlichkeit für den Output entscheidend, ob in dem Input oder der vorangegangenen Diskussion von Schach, Burgen oder mittelalterlichen Straftaten die Rede ist: Die Aufmerksamkeit wird auf den jeweiligen semantischen Raum gerichtet, in dem nach dem wahrscheinlichsten Output gesucht wird. Das Anwendungsfeld dieser Architektur – entwickelt für die Verbesserung von maschinellen Übersetzungen, deren Manko in der fehlenden Berücksichtigung des Kontexts bestand, was nicht selten zu sinnverfälschenden Übersetzungen führte – wurde erheblich ausgeweitet und sorgt nun für den hochwertigeren und passgenaueren Output der aktuellen Systeme (Linde 2023).

Auch wenn exemplarisch Open AIs GPT-Sprachmodelle die zugrundeliegenden Eigenschaften eines modernen LLMs im Namen tragen (**Generative Pretrained Transformer**), basieren auch andere aktuelle Modelle auf dieser Grundlage. *Generative* steht hierbei dafür, dass etwas Neues entsteht. *Pretrained* verweist auf auf das Trainingsverfahren des LLM, also den Vorgang des zugrundeliegenden maschinellen Lernens. Die Trainingsdaten für die LLMs umfassen insgesamt etwa eine Billion Wörter (Saboo & Kublik 2023) und basieren im Falle von GPT-3.5 auf fünf Datensätzen (Common Crawl, WebText2, Books1, Books2 und Wikipedia) (Brown et al. 2020). LaMDA hingegen liegen englischsprachige Dialoge zugrunde (Collins & Ghahramani 2021). Hieraus ergibt sich bereits ein zentrales und vieldiskutiertes Problem für textgenerierende KIs, da den Trainingsdaten immer auch Werte und Anschauungen des Trainingsmaterials inhärent sind. Je häufiger Material zu einem Thema oder Konzept in den Trainingsdaten vorkommt, desto wahrscheinlicher ist es, dass dies im Output reproduziert wird, was Draelos (2023) zu dem Urteil bringt: „Large language models (LLMs) like ChatGPT are racist, sexist, homophobic, and in general packed full of all of the worst of society’s biases, because they are trained on biased data.“ Dass das in dem alltäglichen Gebrauch von ChatGPT nicht so sehr auffällt, liegt im Wesentlichen daran, dass nicht der rohe Output des LLMs präsentiert wird, sondern der durch eine *Moderation API* gefilterte Output, die unerwünschten Output herausfiltert (Markov et al. 2022). Dieser Umweg ist jedoch anfällig für sogenannte *prompt injections*, mit denen versucht wird, Systeme zu Aussagen zu bewegen, die das System eigentlich nicht veröffentlichen soll (Liu et al. 2023). Daher fordert Draelos den Schwerpunkt auf die Entwicklung von LLMs zu legen, die einen weniger starken Bias haben. Beispiele für Modelle, die diesen Weg einzuschlagen versuchen, sind z. B. das Sprachmodell Claude des US-Unternehmens Anthropic, das für sich beansprucht, beim Modelltraining die Werte der amerikanischen Verfassung zugrundegelegt zu haben, oder das Luminous-Modell des Heidelberger Start-ups Aleph Alpha.

3. Einsatz von KI-Textgeneratoren an Hochschulen

3.1 Auswirkungen

Dass ChatGPT frei zugänglich wurde, zog sehr unterschiedliche Reaktionen im Bildungsbereich nach sich – neben Verboten in den USA und einige Monate später auch in Italien gab es gegenläufige Bestrebungen, schnell Leitlinien zu erstellen, um den Lehrenden Orientierung für einen konstruktiven Einsatz von KI im Kontext von Lehre und Prüfungen zu geben. Inzwischen scheint sich ein weitgehender Konsens darüber gebildet zu haben, dass ein pauschales Verbot von KI-Tools in Bildungseinrichtungen und der Wissenschaft nicht zielführend ist und vielmehr Strategien verfolgt werden sollten, die Anwendungen sinnvoll in den Lehr- und Forschungsalltag einzubinden (Buck & Limburg 2023). Klar und notwendig ist, dass auch die Prüfungspraktiken an den technologischen Transformationsprozess angepasst werden und weniger die Informationsabfrage als die Kompetenz- und Persönlichkeitsentwicklung im Vordergrund stehen sollte (Budde et al. 2023).

Dies betrifft nicht nur, aber insbesondere die wahrscheinlich traditionellste Form von akademischer Erkenntnissicherung und Kompetenzfeststellung, den (schriftlichen) Text, und damit im Kern das grundlegende wissenschaftliche Selbstverständnis. Der Fokus bei der Erstellung und Bewertung schriftlicher Arbeiten wird sich ändern müssen, da in absehbarer Zeit – oder bereits zum Veröffentlichungszeitpunkt dieses Artikels – in vielerlei Hinsicht zumindest qualitativ gleichwertige Texte durch generative KI-Systeme erstellt werden können. Waren Haus- und Abschlussarbeiten bisher ein Medium, in dem Studierende ihren Wissenszuwachs dokumentierten und zeigten, dass sie wissenschaftlich arbeiten können, werden sie in Zukunft womöglich vielmehr Mittel zur Reflexion über den eigenen Lernprozess sein (Weßels 2022; Weßels et al. 2022; Klein 2023). Dabei ist die Entwicklung von Schreibkompetenzen allerdings Voraussetzung, um solche Textgenerierungstools überhaupt kompetent anzuwenden (Klein 2023). Dazu ist Schreiben stärker „in die Fachlehre [zu] integrieren und als Denkwerkzeug [zu] benutzen“, Studierende sind stärker für gegenseitiges Feedback einzubeziehen und beides auch durch Schreibtutor*innen zu unterstützen. Auch Studierendenkonferenzen sind probates Mittel, um den Austausch zwischen den Studierenden zu fördern (Klein 2023).

Mollick & Mollick (2022, 9) sehen in diesem Sinne eines konstruktiven Einsatzes von KI-Systemen im Schreibprozess „neue Möglichkeiten der Interaktion“ und illustrieren das anhand von drei konkreten Aufgaben, in denen generative Texttools helfen können, Kompetenzen im Umgang mit Wissen zu entwickeln: a) Wissenstransfer, indem eine Erklärung für neu zu erlernende Konzepte beim Sprachassistenten erfragt wird und dann die Ausgabe auf ihre Korrektheit überprüft wird; b) Überprüfung des Wissenszuwachses und Anregung des kritischen Denkens, indem die Lehrperson einen Prompt vorgibt, um z. B. einen Aufsatz zu erzeugen. Die Lernenden sollen nun das Ergebnis durch Interaktion mit der Anwendung den Output verbes-

sern. Die Dokumentation des Prozesses ermöglicht der Lehrperson, die Gedankengänge und Entscheidungen der Lernenden nachzuvollziehen. Anwendungsfall c) ist ein Vorschlag, um die kritische Informationsbewertung und das Hinterfragen eigener Vorannahmen anzuregen.

Weßels (2022) sieht in generativer KI insbesondere das Potenzial eines „persönliche[n] Lernbegleiter[s]“ und prognostiziert, dass Prüfungsgespräche, der Prozess der Texterstellung im Verhältnis zum Endprodukt und gegebenenfalls das Einbeziehen von Tools in den Schreibprozess wichtiger in der Bewertung werden. Dies bringt eine Rollenveränderung von Lehrenden hin zu Gestalter*innen des Lernraums mit sich. Für Studierende werden digitale Kompetenzen zunehmend wichtiger werden (Friedrich & Tobor 2023). In Bezug auf den konkreten Schreibprozess haben Meyer & Weßels (2023) erste Ergebnisse aus einer KI-Schreibwerkstatt veröffentlicht, die zeigen, dass die Studierenden die Herausforderungen dieser Anwendungen sehen und die Textoutputs nicht unbesehen in eine Hausarbeit übernehmen würden (Meyer & Weßels 2023, 244). Gleichzeitig halten sie generative Textgeneratoren für ein hilfreiches Mittel, um Schreibblockaden zu bearbeiten. Studierende der Hochschule der Medien sehen Potenzial sowohl in der Leistungsverbesserung, insbesondere durch Zeitersparnis, als auch in der Persönlichkeitsentwicklung, da sie den „Umgang mit Technologie“ lernen und dies als „Vorbereitung auf Beruf und Zukunft“ einschätzen (Bihlmeier et al. 2023, Folien 4–6).

Während also im akademischen Diskurs eine Vielzahl von Anwendungsszenarien diskutiert und in kleinen Pilotstudien erprobt werden und von einer Vielzahl von Expert*innen eine Integration in die akademische Lehre gefordert wird, zeichnet die Praxis zum Abfassungszeitpunkt dieses Artikels ein anderes Bild. Empirische Erhebungen aus Mai und Juni 2023 zeigen, dass zumindest der Großteil der dort befragten Studierenden noch nie KI für Prüfungszwecke verwendet hat (Humboldt-Universität zu Berlin 2023, 3). Unter den am häufigsten genannten Begriffen, die die befragten Studierenden im Zusammenhang mit KI im Hochschulbereich nennen, sind u. a. Betrug, Schummeln, Unsicherheit, Plagiat und Täuschung zu finden (ebd., 6). Dadurch erschließt sich, weshalb gleichzeitig ein großer Bedarf nach klaren Regeln und Anleitungen sowie Schulungen zum effektiven Umgang geäußert wird, um das Potenzial auszuschöpfen, das von einem großen Teil der Befragten trotz vieler negativer Assoziationen gesehen wird (ebd., 4–6). Ähnlich zeigt sich auch aufseiten der Lehrenden das Bedürfnis nach Orientierung, verbunden mit dem Wunsch nach Richtlinien und Weiterbildungen zum KI-Einsatz in der Bildung (Philologenverband NRW 2023; Surovell 2023).

Erste Orientierungshilfen wurden bereits seit dem Sommer 2022 an der FH Kiel und der Hochschule RheinMain in Form von Eigenständigkeitserklärungen erarbeitet (Limburg et al. 2022, 103; Hochschule RheinMain 2023). Sie bieten die Option, die Nutzung von textgenerierenden KI-Tools entweder mit oder ohne Kennzeichnung explizit zu erlauben oder auszuschließen. Die Anpassung ist notwendig, da der Begriff des Plagiats nicht greift, auch wenn er im Zusammenhang mit KI-Texttools unzutref-

fenderweise verwendet wird. Plagiate werden allerdings in Zukunft kaum mehr eine Rolle spielen, da LLM generierte Texte grundsätzlich einzigartig und nicht ohne Weiteres reproduzierbar sind. KI-generierte Texte stellen kein Plagiat dar, sondern sind vielmehr dem Ghostwriting zuzuordnen. Limburg et al. (2022, 101) sehen die Notwendigkeit einer „Revision des Plagiatsbegriffs“.

Um der Unsicherheit zu begegnen, ob ein vorliegender Text menschen- oder maschinengeschrieben ist, wurden Programme entwickelt, die das anhand von sprachlichen Merkmalen wie Stilbrüchen oder „KI-typischen“ Ausdrucksweisen unterscheiden sollen. Auch die führenden Anbieter von Plagiatserkennungssoftware arbeiten daran, eine Funktion zur Erkennung KI-generierter Texte in ihre Programme einzubauen (Hipold & Weber 2023). OpenAI hat seinen eigenen AI-Text-Identifizierer allerdings im Juli 2023 offline genommen, da die Erkennungsquote zu gering war (Sokolov 2023). Zu diesem Ergebnis kamen auch Weber-Wulff et al. (2023) in ihrer Studie, in der sie mehrere Programme dahingehend testeten, ob Texte von Maschinen und Menschen korrekt zugeordnet werden können. Es zeigte sich, dass es einen sehr hohen Anteil an falsch positiven und falsch negativen Ergebnissen gab (Weber-Wulff et al. 2023, 30). Das Problem der Fehlerhaftigkeit scheint dabei nicht auf die noch mangelhafte Qualität der Detektoren zurückzuführen, sondern vielmehr mathematischer Natur zu sein. Eine sichere Identifikation wäre damit dem Grunde nach nicht möglich (Wilkins 2023).

Klarheit zu rechtlichen Fragen – und damit ein großer Schritt in Richtung Orientierung – schaffte im März 2023 erstmals auch ein Rechtsgutachten, das vom Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen beauftragt wurde. KI-generierte Texte müssen derart gekennzeichnet werden, dass für Dritte ersichtlich ist, welche Teile in welchem Umfang mithilfe von KI-Tools erstellt wurden. Zu berücksichtigen sind die jeweilige Gesetzgebung und die Lizenz- und Nutzungsbedingungen der Software. Im Einzelfall kann die Kennzeichnung durch Selbstständigkeitserklärungen und in Prüfungsordnungen geregelt werden. Die Übernahme von KI-generiertem Text sei per se auch kein Verstoß gegen die gute wissenschaftliche Praxis, solange der hauptsächliche schöpferische Anteil bei dem Menschen, der das Tool bedient, liege. Generative KI könne also lediglich ein Hilfsmittel sein (Hoeren 2023, 23). Die Universität Hohenheim leitet aus ihrer Satzung zur guten wissenschaftlichen Praxis hingegen ab, dass die Verwendung von ChatGPT unredlich sei (Vogelgesang 2023, 8), auch wenn das nicht explizit darin formuliert ist.

Während im Zusammenhang mit Plagiaten die Sachlage klar ist, dass sich eine Person die Gedanken und/oder Ideen einer anderen Person – ob willentlich oder aus Unkenntnis der Regeln – zu eigen macht, und im Falle von Ghostwriting eine andere Person als der*die angegebene Autor*in den Text verfasst, gelten für KI-generierte Texte andere Regeln. Entsprechend dem Gesetz über Urheberrecht und verwandte Schutzrechte (UrhG) gilt eine KI nicht als Autor*in. Führende Wissenschaftsverlage haben daher künstliche Intelligenzen als Co-Autor*innen verboten (Nature 2023; Science 2023). Wenn im Rahmen einer Aufgabenstellung im Hochschulbereich doch

KI-generierte Textelemente übernommen werden, sollten diese gekennzeichnet werden. In welcher Form das geschehen kann, schlagen u. a. die *Modern Language Association* (MLA) und die *American Psychological Association* (APA) vor, die auch zwei der insbesondere in den Geisteswissenschaften besonders verbreiteten Zitierstile herausgeben (McAdoo 2023; MLA Style Center 2023). Allerdings wird dieses traditionelle Konzept von Autorschaft sich ändern müssen, da die Arbeit mit KI-Textgeneratoren zunehmen und es immer schwieriger werden wird, in den Texten klar zu markieren, welche Textteile vom Menschen und welche von der Maschine stammen (Kullmann 2023).

Erste Handreichungen zum Einsatz von KI-Textgeneratoren wurden im ersten Halbjahr 2023 bereits für und von Bildungseinrichtungen im deutschsprachigen Raum veröffentlicht (Mohr 2023). Der Grundtenor schließt sich dem des Rechtsgutachtens an. Ein Verbot von generativen KI-Tools und KI im Allgemeinen im Bildungsbereich wird nicht als sinnvoll erachtet. Vielmehr geht es darum, diese sinnvoll in die Lehre und Forschung einzubinden. Auch der Hochschullehrerbund fordert die Möglichkeit, KI-Tools in der Lehre zu nutzen und beruft sich bei der Ausgestaltung des Einsatzes auf die Freiheit der Lehre (Hochschullehrerbund 2023). Die Mitglieder fordern insbesondere aber einen Zugang für alle zu solchen Tools, da sie es als Bestandteil des Bildungsauftrags der Hochschulen ansehen, die Studierenden auf die Arbeitswelt vorzubereiten. Voraussetzung dafür ist allerdings auch ein grundlegendes Verständnis sowohl aufseiten der Lehrenden als auch der Lernenden. Erste Tutorials, um dieses Verständnis zu vermitteln und praktische Anwendungsszenarien vorzustellen, wurden bereits produziert (z. B. KI-Campus 2023).

3.2 Anwendungsbeispiele

Inzwischen können fast alle Aufgaben und jeder Arbeitsschritt des wissenschaftlichen Schreibprozesses von der Themenfindung bis zum Feedback mithilfe von KI-Tools bewältigt werden, zum Teil mit sehr hoher Qualität (Limburg et al. 2023, 4; Gröpler 2023, 367). Allerdings ist es zwingend erforderlich, die Ergebnisse kritisch zu überprüfen. Im Folgenden werden exemplarisch einige Anwendungsfälle gezeigt. Es sei darauf hingewiesen, dass die ausgewählten Tools jeweils als Stellvertreter für eine Reihe von Anwendungen mit gleichen oder ähnlichen Funktionen gewählt wurden, um die Vielfalt abzubilden.

Zu Beginn wird gezeigt, wie ein KI-gestütztes Texttool zur Recherche genutzt werden kann. Dafür wurde ChatGPT (GPT-3.5) gebeten, fünf wichtige deutschsprachige Publikationen zum Thema KI und Theologie auszugeben. Folgende drei Bücher waren samt der Beschreibung unter den Ergebnissen:

- „Gott im Netz: Ein Tagungsband zu Theologie und künstlicher Intelligenz“ von Andreas Mertin (Herausgeber) (Erscheinungsjahr: 2019)

- „Digitale Auferstehung: Von der Zukunft des menschlichen Geistes“ von Thomas Metzinger und Svenja Flaßpöhler (2018). Dieses Buch beleuchtet die philosophischen und theologischen Aspekte der KI und deren mögliche Auswirkungen auf das menschliche Bewusstsein und die Spiritualität.
- „Gott 4.0: Warum die Zukunft der Religion in der KI liegt“ von Thomas D. Williams (2017). Der Autor untersucht in diesem Werk die mögliche Rolle der KI in religiösen Kontexten und wie sie die Zukunft der Religion beeinflussen könnte.

Diese drei Einträge stammen aus zwei verschiedenen Anfragen im selben Tool. An der unterschiedlichen Formatierung erkennt man bereits, dass sich die Ausgaben jedes Mal unterscheiden. Die Ergebnisse hatten auch keinerlei Überschneidungen. Sie sehen auf den ersten Blick seriös aus, da sowohl Autor*in, Titel als auch Publikationsjahr vorhanden sind. Genau an dieser Stelle liegt die Gefahr, wenn die Existenz dieser Publikationen bzw. die Korrektheit der Angaben nicht überprüft wird. So gibt es nämlich eine Monografie mit dem Titel „Gott im Netz“, jedoch mit einem anderen Untertitel. Autor und Publikationsjahr stimmen mit der Ausgabe des Chatbots nicht überein. Es gibt allerdings einen Autor namens Andreas Mertin, der auch im Bereich Religion publiziert. Ebenso existieren Autor und Autorin des zweiten Buches, ein Buch dieses Namens ist allerdings nicht zu finden. Auch das dritte Buch hat keine reale Entsprechung, auch wenn es einen Autor dieses Namens gibt und eine Monographie „Gott Version 4.0“. Mögen die Ergebnisse zunächst der Beleg dafür sein, dass die Tools für eine fundierte Recherche unbrauchbar sind, können sie doch genutzt werden, um den kritischen Umgang mit Quellen zu üben oder auf bisher unbekannte Autoren aufmerksam zu werden. Zudem könnte die Ausgabe Ideen für weitere Suchbegriffe geben. Für vertrauenswürdige Quellen sollte aber immer auf die herkömmlichen Informationsinfrastrukturen zurückgegriffen werden. Aber auch in diesem Bereich gibt es mittlerweile eine Vielzahl KI-gestützter Alternativen und Recherchertools, die auf Basis semantischer Suchen wissenschaftliche Quellen erschließen.⁴

Dieses Beispiel illustriert anschaulich die irreführende Eigenschaft von LLMs zu halluzinieren und durch die stimmigen Formalia den Anschein hoher Plausibilität zu erwecken, was den stets kritischen Umgang mit KI-Textgeneratoren zwingend erforderlich macht.

Die Abb. 2 (nächste Seite) zeigt die Interaktion mit einem PDF mithilfe von PDFGear,⁵ das kostenlos heruntergeladen werden kann. Der integrierte PDFGear Copilot erweitert die üblichen Funktionen eines PDF-Anzeige- und Bearbeitungsprogramms durch einen Chatassistenten, der auf GPT-3.5 basiert und Interaktion mit

4 Eine strukturierte und regelmäßig aktualisierte Auswahl ist z. B. hier zu finden: <https://www.vkkiwa.de/ki-ressourcen>, zuletzt aufgerufen am 18.06.2024.

5 PDFGear steht hier stellvertretend für andere Tools, die über ähnliche Funktionen verfügen wie u. a. ChatGPT Pro mit Plug-in, BingChat oder ChatPDF.

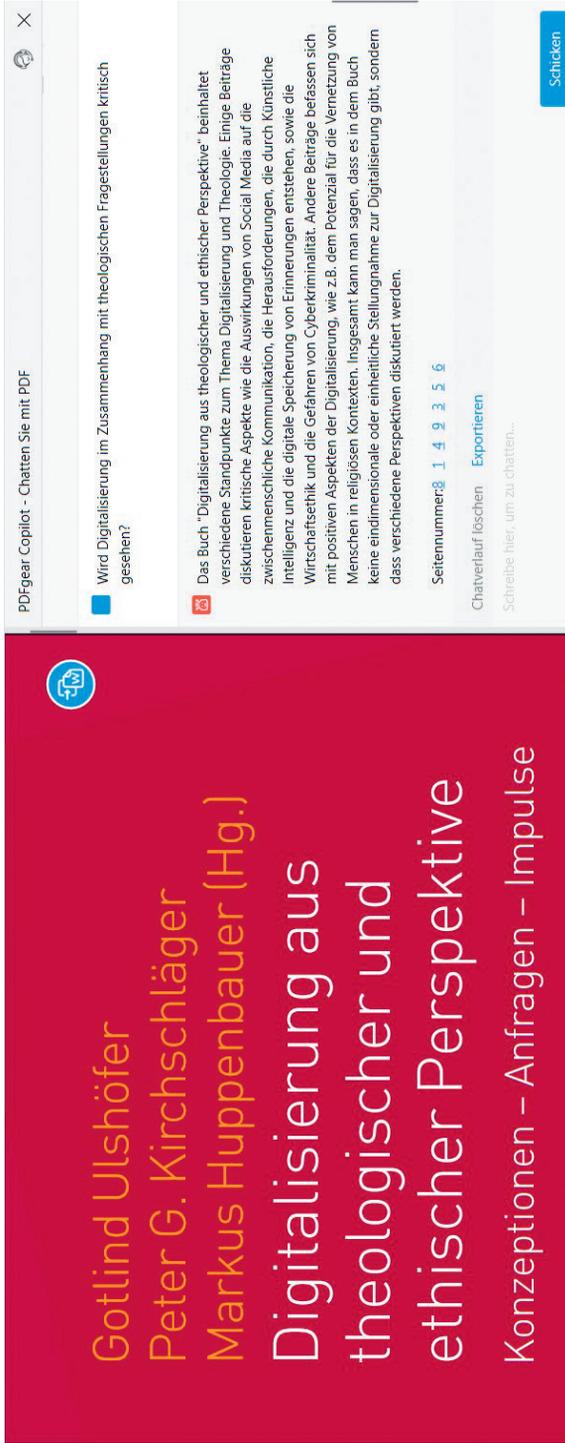


Abb. 2 Chatten mit einem PDF-Dokument

dem Text erlaubt. Eine Aufforderung könnte sein „Wird Digitalisierung im Zusammenhang mit theologischen Fragestellungen kritisch gesehen?“, um schnell herauszufinden, welche Positionen in den Artikeln vertreten werden.

Das Tool gibt nun eine knappe Antwort, die sich in Länge und auch Aufbau nach den Parametern der Eingabe richtet. So ist es auch möglich, sich eine mit Spiegelstrichen strukturierte Liste der Hauptaussagen des Gesamtbuches oder nur eines einzelnen Kapitels ausgeben zu lassen. Auch hier muss die Ausgabe kritisch überprüft werden, was wiederum in der Lehre eingesetzt werden kann. Zum Beispiel könnte eine Aufgabe darin bestehen, einen Text selbst zusammenfassen und dann eine Zusammenfassung automatisch erstellen lassen und beide daraufhin zu vergleichen, ob ähnliche Kernthesen herausgearbeitet wurden. Für Studierende, Lehrende und Forschende kann ein PDF-Auswertungstool darüber hinaus dabei helfen, die Erstsichtung elektronischer Literatur schneller durchzuführen oder herausfordernde Passagen zu erschließen. Diese Funktion ist auch in viele *AI Research Assistants* wie *Jenni.ai* integriert. Ähnlich wie bei Paraphrasierungstools wie z. B. *Quillbot* besteht bei dieser Anwendung die Gefahr von Plagiatsverschleierung, wenn Texte durch die KI neu formuliert werden und somit als Plagiat nicht erkennbar sind. Ein Gespräch oder eine Reflexion zu den Texten kann hier vorbeugen.

Wie bereits angesprochen können Textgeneratoren bei Reibungsverlusten helfen, wenn der Einstieg ins Schreiben oder die Themenfindung schwerfällt. *Open Assistant* erstellt auf Anfrage eine Auswahl an möglichen Forschungsfragen zum Thema „generative KI in der theologischen Forschung“ (Abb. 3).

Einige Fragestellungen, die zu dem Thema Generative KI und theologische Forschung zusammenhängen könnten sind:

- Wie kann man den Begriff "Generative KI" im Kontext der theologischen Wissenschaft verstehen?
- Welche Möglichkeiten bietet das Verwenden von Generativer KI für die Erforschung des menschlichen Geistes?
- Inwiefern kann die Generative KI dazu beitragen, unser Wissen über Gott zu erweitern oder zu vertiefen?
- Gibt es ethische Bedenken bezüglich der Anwendung von Generativer KI auf dem Gebiet der Theologie?
- Wie können wir sicherstellen, dass die durch Generativer KI erzeugten Ergebnisse mit den Prinzipien und Lehren der Theologie übereinstimmen?

Es ist wichtig zu beachten, dass diese Fragestellungen lediglich als Vorschlag dienen und nicht zwangsläufig alle relevanten Aspekte abdecken. Es ist empfehlenswert, eine breite Palette an Fragen zu erstellen, um ein möglichst detailliertes Verständnis des Themas zu gewinnen.

Abb. 3 Generierung von Forschungsfragen mithilfe eines KI-gestützten Chattools

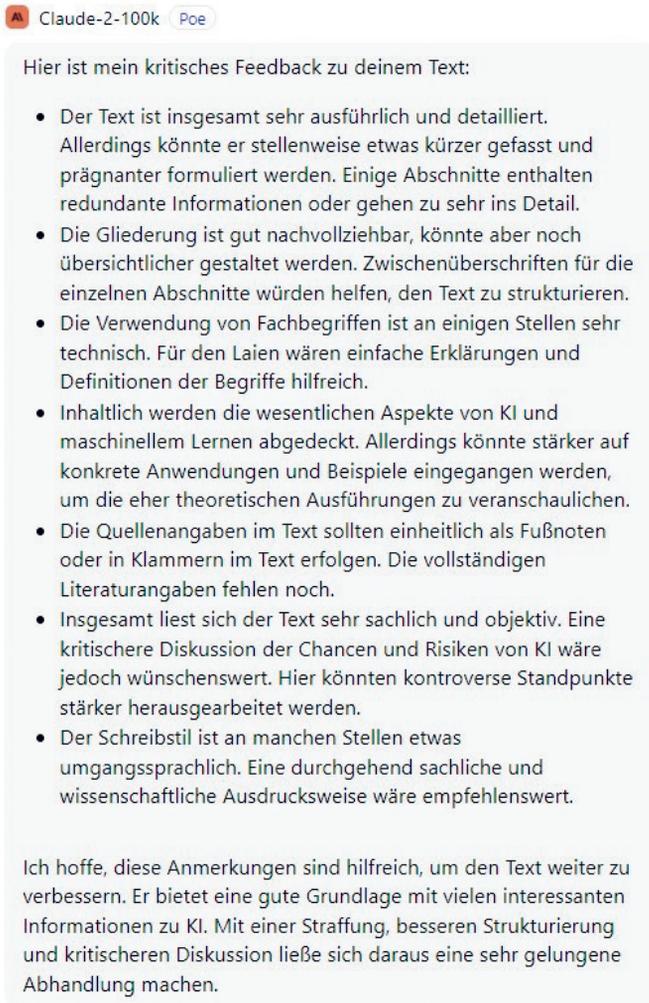


Abb. 4 Textkritik mithilfe eines KI-gestützten Chattools

Interessant ist der abschließende Hinweis, dass man sich nicht nur an dieses Ergebnis halten sollte. Diese werden keinesfalls immer ausgegeben. Bei wiederholter Eingabe desselben Prompts fallen die Antworten sehr verschieden und nicht immer entsprechend der Anweisung, kurz und aus drei Fragen bestehend, aus. Es ist nun möglich, auf Grundlage einer der ausgegebenen oder einer eigenen Frage einen Forschungsplan erstellen zu lassen, der entsprechend den realen Gegebenheiten angepasst werden kann.

Auf einem basalen Niveau sind die KI-Chatbots auch geeignet, um die eigenen Textentwürfe zu hinterfragen. Als Beispiel seien hier die Anmerkungen angeführt,

die das Modell Claude-2 auf die Bitte produziert hat, Feedback auf den Abschnitt „Zur Funktionsweise von KI-Textgeneratoren“ aus diesem Artikel zu geben (Abb. 4).⁶

Das Feedback bezieht sich sowohl auf inhaltliche als auch formale Aspekte, wenn auch die inhaltlichen Hinweise oberflächlich bleiben. Dennoch bietet es Anregungen, um an dem Text weiterzuarbeiten. Wichtig ist, dass dem Modell mit dem Prompt klar mitgeteilt wird, dass „kritisches Feedback“ gewünscht ist oder auf Schwachstellen oder Widersprüche hingewiesen werden soll, um möglichst konkrete Ergebnisse zu erhalten. Genaue Prompts legen, wie oben ausgeführt, fest, worauf die Aufmerksamkeit bei der Generierung des Outputs gerichtet werden soll, und sind somit der zentrale Schlüssel für eine brauchbare Ausgabe. Auch hier gibt es mittlerweile eine Reihe von Prompt-Leitfäden für den Einsatz im akademischen Kontext (z. B. Lenk-Ostendorf 2023). Für eine abschließende sprachliche Korrektur von Texten und auch Präsentationen können darauf spezialisierte Anwendungen wie *DeepLWrite*⁷ oder ebenfalls einer der Chattools auf Basis eines LLM verwendet werden.

4. Herausforderungen für Hochschulen, Forschung und Lehre

Der Einsatz generativer KI-Tools steht im Spannungsfeld von Chancen und Risiken. Er bietet viel Potenzial zur (Teil-)Automatisierung von Standardaufgaben an Hochschulen, in Lehre und Forschung, wie z. B. die Erstellung von Forschungsförderanträgen oder Literatur Reviews, sodass mehr Zeit und Fokus auf die eigentliche schöpferische Arbeit gelegt werden kann.

Allerdings stellen sich derzeit noch viele Herausforderungen. Neben dem Bias, der wie oben beschrieben durch die Trainingsdaten nicht komplett ausgeschlossen, aber durch eine möglichst sorgfältige Kuratierung wie bei Open-Source-Programmen (z. B. *Open Assistant* oder auf der US-Entwicklerplattform verfügbare Open Source-Modelle) minimiert werden kann, liegen drängende rechtliche Fragen im Bereich von Urheberrechten, Persönlichkeitsrechten und Datenschutz. Die am Markt bekanntesten und leistungsstärksten Anwendungen erfordern nicht nur die Eingabe von persönlichen Daten, wie mindestens einer E-Mail-Adresse, sondern nutzen auch die Prompts zum Trainieren des Sprachmodells. Inzwischen ist es zumindest bei ChatGPT auch in der kostenfreien Version möglich, dies optional in den Einstellungen auszuschließen. Erste Ansätze, um Studierenden einen geschützten Raum zu schaffen, gibt es bereits. So stellen deutsche Unternehmen, die die Schnittstellen zu den

6 Seit Mai 2024 steht in Europa regulär die Modellfamilie Claude-3 in drei Varianten Haiku, Sonnet und Opus zur Verfügung, seit Juni 2024 das leistungsfähige Modell Claude 3.5 Sonnet sowie das Feature Claude Artifacts, mit dem sich ohne Programmierkenntnisse z. B. interaktive Diagramme oder HTML-Webseiten erstellen und ausführen lassen.

7 S. <https://www.deepl.com/de/write>, zuletzt aufgerufen am 18.06.2024.

proprietären amerikanischen Sprachmodellen nutzen, eigene Nutzeroberflächen zur Verfügung. Die persönlichen Daten der User werden dadurch verwaschen. Bei *Fobizz* kann eine Lehrkraft einen Klassenraum einrichten, in dem die Lernenden KI-Tools austesten können (Fobizz 2023). Bei HAWKI von der Hochschule für angewandte Wissenschaft und Kunst Hildesheim/Holzminde/Göttingen können sich die Hochschulangehörigen mit ihren Hochschulaccounts anmelden und so geschützt auf die Schnittstelle zu OpenAI zugreifen (HAWK 2023).

Derzeit ist es noch so, dass LLMs, die für den Einsatz in sensiblen Bereichen u. a. wegen Intransparenz hinsichtlich der Trainingsdaten, schlechten Arbeitsbedingungen für die menschlichen Trainer*innen des Modells und des Datenschutzes kritisch zu sehen sind, am zuverlässigsten arbeiten. Das liegt v. a. an den umfangreichen Investitionen durch große Konzerne.

Eine europäische Lösung wird auch deswegen u. a. mit *OpenGPT-X* (Luber 2023), *Aleph Alpha* und *OpenAssistant* angestrebt, um konkurrenzfähig zu bleiben und nicht weiter in Abhängigkeit von den derzeit den Markt beherrschenden US-amerikanischen und chinesischen KI-Modellen zu geraten. *OpenAssistant* ist ein Projekt, das ein Gegengewicht zu den großen Akteuren setzen will. Gegründet in der Schweiz stehen sowohl Source Code als auch Trainingsdaten auf Github bereit (Köpf et al. 2023). Mit LEAM (*Large European AI Models*) wurde unter dem Dach des Bundesverbands der Unternehmen der Künstlichen Intelligenz in Deutschland e. V. (KI Bundesverband) ein Projekt auf den Weg gebracht, um in Zusammenarbeit mit europäischen Anbietern von LLM sowie Vertretern aus Wirtschaft und Wissenschaft die Entwicklung und den Ausbau von Modellen zu forcieren, die „europäischer Werte und [...] Standards und Regulierungen“ berücksichtigen (Bienert et al. 2023, 4. 12).

Die Transformationsprozesse, die durch die technologischen Entwicklungen im KI-Bereich gesamtgesellschaftlich angestoßen wurden, werden auch in Zukunft wahrscheinlich in einem Ausmaß, das bisher nicht einmal zu erahnen ist, einen erheblichen Einfluss auf die Praktiken des Systems Hochschule, beim Lehren, Lernen und Forschen mit dem zentralen Element der Textproduktion haben und uns möglicherweise dazu zwingen, zumindest in manchen Aspekten unser grundlegendes Selbstverständnis zu hinterfragen. Was heißt Autorschaft? Was heißt gute wissenschaftliche Praxis im Kontext des Einsatzes von KI? Und nicht zuletzt: Was ist Wissenschaft und welche Rolle spielt der Mensch darin? Nach der anfänglich erschütternden Disruption macht sich die Wissenschaft in bester Manier daran, erste Teilfragen zu beantworten und stiftet damit allmählich Orientierung im Chaos. Viele dieser Fragen dürften Gegenstand andauernder kontroverser Diskurse werden. Die Aufgabe aber von Forschung und Wissenschaft besteht nicht nur darin, Antworten zu finden, sondern wohl vor allem darin, die richtigen Fragen zu stellen. Und hier stehen wir ganz am Anfang, da niemand seriös auch nur kurzfristig die weiteren Entwicklungen voraussagen kann, wie die letzten Monate eindrücklich gezeigt haben. Zwischen den hier grob und exemplarisch skizzierten Chancen und Risiken eröffnet sich damit ein extrem großer Möglichkeitsraum für unsere Zukunft. Was ist technisch möglich? Was

wollen wir wie und wozu einsetzen? In welcher Zukunft wollen wir leben, lernen, forschen? Diesen Möglichkeitsraum gilt es nun wissenschaftlich zu erschließen und Positionierungen im gesamtgesellschaftlichen Diskurs auszuhandeln.

Literaturverzeichnis

- Adiwardana, D., & Luong, T. (2020). Towards a Conversational Agent that Can Chat About...Anything. In *Google Research Blog*. URL: <https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html> [zuletzt aufgerufen am 18.06.2024].
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots. In *ACM Digital Library, Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (S. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922> [zuletzt aufgerufen am 18.06.2024].
- Beta Writer. (2019). *Lithium-Ion batteries. A machine-generated summary of current research*. Cham: Springer. <https://doi.org/10.1007/978-3-030-16800-1> [zuletzt aufgerufen am 18.06.2024].
- Bienert, J., Broch, R., Bunk, P., Burchard, A., Cann, V., Deetjen, B., Dickmann, A., Flores-Herr, N., Friedmann, L., Götz, C., Haenel, F., Hillemann, D., Lambers, K., Lux, L., Otterbach, J., Paass, G., Röhl, C., Schmeier, S., Susallek, P., & Uszkoreit, H. (2023). *Große KI-Modelle für Deutschland. Machbarkeitsstudie*. Sankt Augustin: Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme (IAIS). <https://doi.org/10.24406/PUBLICA-1047> [zuletzt aufgerufen am 18.06.2024].
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, Ch., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, Ch., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). *Language Models are Few-Shot Learners*. Online: arXiv. <https://doi.org/10.48550/arXiv.2005.14165> [zuletzt aufgerufen am 18.06.2024].
- Buck, I., & Limburg, A. (2023). Hochschulbildung vor dem Hintergrund von Natural Language Processing (KI-Schreibtools). Ein Framework für eine zukunftsfähige Lehr- und Prüfungspraxis, *die hochschullehre*, 9(6), 70–84. <https://doi.org/10.3278/HSL2306W> [zuletzt aufgerufen am 18.06.2024].
- Budde, J., Tobor, J., & Beyermann, J. (2023). Blickpunkt Digitale Prüfungen. *Hochschulforum Digitalisierung*, 1–33. URL: https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_Blickpunkt_Digitale_Pruefungen.pdf [zuletzt aufgerufen am 18.06.2024].

- Collins, E., & Ghahramani, Z. (2021). LaMDA. Our breakthrough conversation technology. In *Google-Blog*. URL: <https://blog.google/technology/ai/lamda> [zuletzt aufgerufen am 18.06.2024].
- Döbel, I., Leis, M., Vogelsang, M. M., Neustroev, D., Petzka, H., Riemer, A., Rüping, S., Voss, A., Wegele, M., & Welz, J. (2018). *Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung*. Sankt Augustin/Leipzig/München: Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V. URL: https://www.bigdata-ai.fraunhofer.de/content/dam/bigdata/de/documents/Publicationen/Fraunhofer_Studie_ML_201809.pdf [zuletzt aufgerufen am 18.06.2024].
- Draelos, R. (2023). Debiasing ChatGPT, *Medium*, o. S. URL: <https://rachel-draelos.medium.com/from-chatgpt-to-puregpt-7124463a100d> [zuletzt aufgerufen am 18.06.2024].
- Friedrich, J.-D., & Tobor, J. (2023). Zur Bedeutung von ChatGPT & der Notwendigkeit eines progressiven Umgangs mit neuen KI-Technologien im Hochschulbereich. Ein Zwischenstand in 6 Thesen. In *Blog des Hochschulforum Digitalisierung*. URL: <https://hochschulforumdigitalisierung.de/de/blog/chat-gpt-6-thesen> [zuletzt aufgerufen am 18.06.2024].
- Gröpler, J. (2023). Schreiben oder schreiben lassen?, *Forum Bibliothek und Information*, 75(7), 366–368.
- Hoeren, T. (2023). Rechtsgutachten zum Umgang mit KI-Software im Hochschulkontext. In P. Salden & J. Leschke (Hrsg.), *Didaktische und rechtliche Perspektiven auf KI-gestütztes Schreiben in der Hochschulbildung* (S. 22–41). <https://doi.org/10.13154/294-9734> [zuletzt aufgerufen am 18.06.2024].
- Klein, A. (2023). „Die Hausarbeit ist tot, es lebe die Hausarbeit!“ Entwicklungsorientierung, wissenschaftliches Arbeiten und KI gemeinsam denken. In *Blog des Hochschulforum Digitalisierung*. URL: <https://hochschulforumdigitalisierung.de/de/blog/andrea-klein-KI-Hausarbeit> [zuletzt aufgerufen am 18.06.2024].
- Köpf, A., Kilcher, Y., Rütte, D. v., Anagnostidis, S., Tam, Z.-R., Stevens, K., Barhoum, A., Duc, N. M., Stanley, O., Nagyfi, R., ES, S., Suri, S., Glushkov, D., Dantuluri, A., Maguire, A., Schuhmann, C., Nguyen, H., & Mattick, A. (2023). *OpenAssistant Conversations. Democratizing Large Language Model Alignment*. Online: arXiv. URL: <https://arxiv.org/pdf/2304.07327> [zuletzt aufgerufen am 18.06.2024].
- Kullmann, S. (2023). Autorschaft im Schlaglicht. KI und die Regeln der guten wissenschaftlichen Praxis. *Deutsche Gesellschaft für Information & Wissen e. V.* URL: <https://dgi-info.de/autorschaft-im-schlaglicht-ki-und-die-regeln-der-guten-wissenschaftlichen-praxis> [zuletzt aufgerufen am 18.06.2024].
- Lenk-Ostendorf, B. *ChatGPT-4 Cookbook*. München: ProLehre | Medien und Didaktik. URL: https://www.prolehre.tum.de/fileadmin/woobtq/www/Angebote_Broschueren_Handreichungen/ChatGPT-4_Cookbook.pdf [zuletzt aufgerufen am 18.06.2024].
- Limburg, A., Bohle-Jurok, U., Buck, I., Grieshammer, E., Gröpler, J., Knorr, D., Lira Lorca, A., Mundorf, M., Schindler, K., & Wilder, N. (2023). *Zehn Thesen zur*

- Zukunft des Schreibens in der Wissenschaft*. Diskussionspapier. Berlin: Hochschulforum Digitalisierung [= Diskussionspapier, 23]. URL: https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_DP_23_Zukunft_Schreiben_Wissenschaft.pdf [zuletzt aufgerufen am 18.06.2024].
- Limburg, A., Salden, P., Mundorf, M., & Weßels, D. (2022). Plagiarismus in Zeiten Künstlicher Intelligenz, *Zeitschrift für Hochschulentwicklung*, 17(3), 91–106. <https://doi.org/10.3217/ZFHE-17-03/06> [zuletzt aufgerufen am 18.06.2024].
- Liu, Y., Deng, G., Li, Y., Wang, K., Zhang, T., Liu, Y., Wang, H., Zheng, Y., & Liu, Y. (2023). *Prompt Injection attack against LLM-integrated Applications*. Online: arXiv. URL: <https://arxiv.org/pdf/2306.05499> [zuletzt aufgerufen am 18.06.2024].
- Lotze, N. (2016). *Chatbots. Eine linguistische Analyse* [Diss.]. Berlin et al.: Peter Lang GmbH. [= *Sprache – Medien – Innovationen*, 9]. <https://doi.org/10.3726/b10402> [zuletzt aufgerufen am 18.06.2024].
- McAdoo, T. (2023). How to cite ChatGPT. In *Blog von APA Style*. URL: <https://apastyle.apa.org/blog/how-to-cite-chatgpt> [zuletzt aufgerufen am 18.06.2024].
- Meyer, E., & Weßels, D. (2023). Natural Language Processing im akademischen Schreibprozess – mehr Motivation durch Inspiration? Positionspapier basierend auf einer Fallstudie an der Fachhochschule Kiel. In T. Schmohl, A. Watanabe & K. Schelling (Hrsg.), *Künstliche Intelligenz in der Hochschulbildung. Chancen und Grenzen des KI-gestützten Lernens und Lehrens* (S. 227–251). Bielefeld: transcript Verlag [= *Hochschulbildung. Lehre und Forschung*, 4]. <https://doi.org/10.14361/9783839457696> [zuletzt aufgerufen am 18.06.2024].
- Michelbacher, L. (2013). *Multi-word tokenization for natural language processing* [Diss.]. Stuttgart: OPUS. Online Publikationen der Universität Stuttgart. <https://doi.org/10.18419/opus-3208> [zuletzt aufgerufen am 18.06.2024].
- Mohr, G., Reinmann, G., Blüthmann, N., Lübcke, E. & Kreinsen, M. (2023). Übersicht zu ChatGPT im Kontext Hochschullehre, *Hamburger Zentrum für universitäres Lehren und Lernen (HUL)*, 1–12. <https://www.hul.uni-hamburg.de/selbstlernmaterialien/dokumente/hul-chatgpt-im-kontext-lehre-2023-01-20.pdf> [zuletzt aufgerufen am 18.06.2024].
- Mollick, E. R., & Mollick, L. (2022). New Modes of Learning Enabled by AI Chatbots. Three Methods and Assignments, *SSRN Electronic Journal* [Preprint]. <https://doi.org/10.2139/ssrn.4300783> [zuletzt aufgerufen am 18.06.2024].
- Müller, A. C., & Guido, S. (2017). *Einführung in Machine Learning mit Python. Praxiswissen Data Science*. Übers. v. K. Rother. Heidelberg: dpunkt.verlag GmbH.
- Patel, A. A. (2019). *Hands-on unsupervised learning using Python. How to build applied machine learning solutions from unlabeled data*. Sebastopol: O'Reilly Media.
- Saboo, S., & Kublik, S. (2022). A Beginner's Guide to GPT-3. In *datacamp blog*. URL: <https://www.datacamp.com/blog/a-beginners-guide-to-gpt-3> [zuletzt aufgerufen am 18.06.2024].

- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning. An Introduction*. Adaptive Computation and Machine Learning series. 2. Ausgabe. Cambridge, Mass./London: The MIT Press [= *Adaptive Computation and Machine Learning series*].
- Ulshöfer, G., Kirchschräger, P. G., & Huppenbauer, M. (Hrsg.). (2021). *Digitalisierung aus theologischer und ethischer Perspektive. Konzeptionen – Anfragen – Impulse*. Baden-Baden: Nomos Verlagsgesellschaft [= *Religion – Wirtschaft – Politik*]. DOI: 10.5771/9783748924012 [zuletzt aufgerufen am 18.06.2024].
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. Online: arXiv. URL: <https://arxiv.org/pdf/1706.03762> [zuletzt aufgerufen am 18.06.2024].
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023, 21. Juni). *Testing of Detection Tools for AI-Generated Text*. Online: arXiv. URL: <https://arxiv.org/pdf/2306.15666> [zuletzt aufgerufen am 18.06.2024].
- Wilder, N., Weßels, D., Gröpler, J., Klein, A., & Mundorf, M. (2022). Forschungsintegrität und Künstliche Intelligenz mit Fokus auf den wissenschaftlichen Schreibprozess. Traditionelle Werte auf dem Prüfstand für eine neue Ära. In K. Miller, M. Valeva & J. Prieß-Buchheit (Hrsg.), *Verlässliche Wissenschaft – Bedingungen Analyse Reflexion* (S. 203–223). Darmstadt: wbg academic.
- Weßels, D., Mundorf, M., & Wilder, N. (2022). ChatGPT ist erst der Anfang. In *Blog des Hochschulforum Digitalisierung*. URL: <https://hochschulforumdigitalisierung.de/de/blog/ChatGPT-erst-der-anfang> [zuletzt aufgerufen am 18.06.2024].
- Weßels, D. (2022). ChatGPT. Ein Meilenstein der KI-Entwicklung, *Forschung & Lehre*, o. S. URL: <https://www.forschung-und-lehre.de/lehre/chatgpt-ein-meilenstein-der-ki-entwicklung-5271> [zuletzt aufgerufen am 18.06.2024].

Sonstige Referenzen

- Bihlmeier, T., Cicek, T., & Hoch, A. (2023). *Studie zur Nutzung von KI-Tools an der Hochschule der Medien [Vortragsfolien]*. URL: [https://filestore.hdm-stuttgart.de/s/k\]ILGELhFaHxhg9#pdfviewer](https://filestore.hdm-stuttgart.de/s/k]ILGELhFaHxhg9#pdfviewer) [zuletzt aufgerufen am 18.06.2024].
- Borchers, D. (2022). Zahlen, bitte! 10⁹ Speicher voller Modelle. Vom Turing-Test bis ChatGPT. *heise online*. URL: <https://www.heise.de/hintergrund/Zahlen-bitte-10-Speicher-voller-Modelle-vom-Turing-Test-bis-Chat-GPT-7392171.html> [zuletzt aufgerufen am 18.06.2024].
- Europäische Kommission. (2021). Vorschlag für eine Verordnung des europäischen Parlamentes und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union. (2021/0106 (COD)). URL: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0019.02/DOC_1&format=PDF [zuletzt aufgerufen am 18.06.2024].

- Fobizz (Hrsg.). (2023). Digitalen Unterricht einfach und sicher gestalten. Alle hilfreichen Tools & KI für die Planung und Durchführung von Unterricht an einem Platz | fobizz Tools [Website]. URL: <https://fobizz.com/ki-assistenz-fobizz-tools> [zuletzt aufgerufen am 18.06.2024].
- Hipold, P., & Weber, S. (2023). *Erste Plagiatssoftware launcht KI-Erkennung [Aktuelles – Meldung der Universität Innsbruck]*. URL: <https://www.uibk.ac.at/de/newsroom/2023/erste-plagiatssoftware-launcht-ki-erkennung> [zuletzt aufgerufen am 18.06.2024].
- Hochschule für angewandte Wissenschaft und Kunst Hildesheim/Holzwinden/Göttingen – HAWK (Hrsg.). (2023). HAWKI [KI-Chatbot]. URL: <https://ai.hawk.de/login.php> [zuletzt aufgerufen am 18.06.2024].
- Hochschule RheinMain. (2023). *Eigenständigkeitserklärung*. URL: https://www.hs-rm.de/fileadmin/Home/Services/Didaktik_und_Digitale_Lehre/Eigenstaendigkeitserklaerung_HSRM_6_23.pdf [zuletzt aufgerufen am 18.06.2024].
- Hochschullehrerbund (Hrsg.). (2023). *Wissenschaftsbasierte Lehre und generative KI. Positionspapier*. URL: https://www.hlb.de/fileadmin/hlb-global/downloads/Positionen/2023-06-12_hlb-Positionspapier_Wissenschaftsbasierte_Lehre_und_generative_KI-Systeme.pdf [zuletzt aufgerufen am 18.06.2024].
- Humboldt-Universität zu Berlin – HU (Hrsg.). (2023). *Kurzbefragung zu KI und Prüfungen*. URL: https://pages.cms.hu-berlin.de/doeringn/dashboard/ErgebnisseKurzbefragungKIPruefungen_230621.pdf [zuletzt aufgerufen am 18.06.2024].
- KI-Campus. (2023). *Sprachassistenzen als Chance für die Hochschullehre | KI-Campus [Lehrveranstaltungscommentar]*. URL: <https://ki-campus.org/courses/sprachassistenzen-hochschule> [zuletzt aufgerufen am 18.06.2024].
- Linde, H. (2023). Künstliche Intelligenz. So funktioniert ChatGPT. *golem.de*. URL: <https://www.golem.de/news/kuenstliche-intelligenz-so-funktioniert-chatgpt-2302-171644.html> [zuletzt aufgerufen am 18.06.2024].
- Luber, S. (2023). Was ist OpenGPT-X? *BigData-Insider*. URL: <https://www.bigdata-insider.de/was-ist-opengpt-x-a-42a8df567400674b7f94418837a80279> [zuletzt aufgerufen am 18.06.2024].
- Markov, T., Zhang, C., Agarwal, S., Eloundou, T., Lee, T., Adler, S., Jiang, A., & Weng, L. (2023). *New and improved content moderation tooling* [Ankündigung von OpenAI]. URL: <https://openai.com/blog/new-and-improved-content-moderation-tooling> [zuletzt aufgerufen am 18.06.2024].
- MLA Style Center. (2023). *How do I cite generative AI in MLA style?* URL: <https://style.mla.org/citing-generative-ai> [zuletzt aufgerufen am 18.06.2024].
- OpenAI. (2023). *Tokenizer*. URL: <https://platform.openai.com/tokenizer> [zuletzt aufgerufen am 18.06.2024].
- Philologenverband NRW (Hrsg.). (2023). *Lehrer-Umfrage zu ChatGPT. Eher Skepsis und Unsicherheit*. URL: <https://phv-nrw.de/2023/04/05/lehrer-umfrage-zu-chatgpt-eher-skepsis-und-unsicherheit> [zuletzt aufgerufen am 18.06.2024].

- Science (Hrsg.). (2023). *Science Journals. Editorial Policies*. URL: https://www.science.org/content/page/science-journals-editorial-policies?adobe_mc=MCMID%3D42627578122959750013493612926978492722%7CMCORID%3D242B6472541199F70A4C98A6%2540AdobeOrg%7CTS%3D1691159684#authorship [zuletzt aufgerufen am 18.06.2024].
- Sokolov, D. A. (2023). OpenAI zieht KI-Detektor zurück. *heise online*. URL: <https://www.heise.de/news/OpenAI-kann-ChatGPT-Texte-nicht-von-menschlichen-unterscheiden-9227827.html> [zuletzt aufgerufen am 18.06.2024].
- Surovell, E. (2023). Faculty Members Still Aren't Sure What to Make of ChatGPT. *The Chronicle of Higher Education*. URL: <https://www.chronicle.com/article/faculty-members-still-arent-sure-what-to-make-of-chatgpt> [zuletzt aufgerufen am 18.06.2024].
- Springer Nature (Hrsg.). (2023). Artificial Intelligence (AI). In *Nature Portfolio*. URL: <https://www.nature.com/nature-portfolio/editorial-policies/ai> [zuletzt aufgerufen am 18.06.2024].
- Virtuelles Kompetenzzentrum KI und wissenschaftliches Arbeiten. (2023). *Virtuelles Kompetenzzentrum KI und wissenschaftliches Arbeiten*. Schreiben lehren und lernen mit KI. Tools und Techniken für Bildung und Wissenschaft [Website]. URL: <https://www.vkkiwa.de> [zuletzt aufgerufen am 18.06.2024].
- Vogelgesang, J., Bleher, J., Krupitzer, C., Stein, A., & Jung, R. (2023). Nutzung von ChatGPT in Lehre und Forschung. Eine Einschätzung der AIDAHO-Projektgruppe [Positionspapier]. URL: https://aidaho.uni-hohenheim.de/fileadmin/einrichtungen/aidaho/Dokumente/AIDAHO_ChatGPT_Positionspapier_23-02-09.pdf [zuletzt aufgerufen am 18.06.2024].
- Wilkins, A. (2023). Reliably detecting AI-generated text is mathematically impossible. *New Scientist*. URL: <https://www.newscientist.com/article/2366824-reliably-detecting-ai-generated-text-is-mathematically-impossible> [zuletzt aufgerufen am 18.06.2024].

Bildnachweis

Abb. 1: Eigener Screenshot von <https://platform.openai.com/playground?mode=complete> [zuletzt aufgerufen am 18.06.2024].

Abb. 2: Eigener Screenshot vom Chat mit dem E-Book PDFgear.

Abb. 3: Eigener Screenshot von der Chatantwort von Open Assistant.

Abb. 4: Eigener Screenshot von der Chatantwort von Claude, bereitgestellt bei poe.com.