Hochschulforum Digitalisierung

DISKUSSIONSPAPIER NR. 29 / FEBRUAR 2024

9 Mythen über generative KI in der Hochschulbildung

Rund um generative Künstliche Intelligenz in der Hochschulbildung gibt es in der öffentlichen Wahrnehmung einige Mythen. Neun dieser verbreiteten Mythen werden in diesem Diskussionspapier beleuchtet. Im Fazit zeigen die Autor:innen auf, was vor dem Hintergrund der aktuellen KI-Entwicklungen nun zu tun ist.

Autorinnen & Autoren

Julius-David Friedrich / Jens Tobor / Martin Wan

Einleitung

Seit dem Ausrollen von ChatGPT am 30.11.2022 erfährt das Thema generative KI in der Breite der deutschen Hochschullandschaft eine große Aufmerksamkeit. Kein KI-Tool zuvor hat die Hochschulwelt so aufgerüttelt. Das ist nicht überraschend, denn der Erstkontakt mit diesem Tool verläuft oft wie folgt: Ungläubiges Staunen, Verängstigung und Faszination. Daher ein kleiner Appell vorab: Falls noch nicht passiert, bitte einmal ausprobieren. Nur so bekommt man exemplarisch eine Idee davon, wie generative KI-Tools funktionieren.

Es gibt viele Ängste, Hoffnungen und Erwartungen, die an generative KI-Tools geknüpft sind. Teilweise nimmt die Debatte dabei sogar unreflektierte, teils religiöse Züge an: Von der Warnung der KI als größter Bedrohung der Menschheit (Musk et al., vgl. auch AI-Doomsday-Marketing) bis hin zur KI als „Erlöserfigur", die sämtliche menschlichen oder menschengemachten Probleme lösen soll. Eine nüchterne Herangehensweise und ein tiefgreifendes Verständnis der Technologie erscheinen uns angebrachter, um tatsächliche Einsatzgebiete mit Mehrwert für diese vielversprechende und faszinierende Technologie zu identifizieren. Wir wollen mit diesem Diskussionspapier entmystifizieren, um Platz zu schaffen für einen produktiven Austausch zu Chancen und Risiken von generativen KI-Tools in Studium und Lehre.

1 Generative KI-Sprachmodelle sind fehlerfrei

Generative KI-Systeme lernen, Inhalte zu erzeugen, die den Daten ähnlich sind, mit denen sie trainiert wurden. Dies spiegelt sich in einer Vielzahl möglicher medialer Outputs wider, darunter Texte, Bilder, Musik und sogar Videos. Generative KI-Sprachmodelle operieren dabei auf Basis von Wortwahrscheinlichkeiten in den Texten des Trainingsmaterials. Durch die automatisierte Verarbeitung dieses umfangreichen Materials und anschließenden menschlichen Feedbacks „eignen" sie sich an, wie Sprache üblicherweise genutzt wird. Dies ermöglicht es ihnen, Sprache auf eine möglichst natürliche Weise zu imitieren. Das bedeutet, KI-Sprachmodelle reihen Wörter nach Wahrscheinlichkeit aneinander (Salden, 2023, S. 7). Weil es in erster Linie um stochastische Operationen geht, neigen die Modelle dazu, zu „halluzinieren". Das bedeutet, dass die Antworten der KI möglicherweise sinnvoll erscheinen, jedoch inhaltlich falsch sein können. Sprachmodelle wie ChatGPT liefern entsprechend keine Wahrheiten, sondern wohlklingende, semantisch korrekte und plausibel wirkende Texte. Dies kann besonders für Lernende, die sich in ein neues Gebiet einarbeiten, problematisch sein. Sie können eventuell nicht beurteilen, ob die Ausgaben eines Chatbots inhaltlich korrekt sind. Gleiches gilt für andere generative KI-Tools.

2 KI versteht Bedeutungen

KI-Modelle eignen sich gut, um Korrelationen zu entdecken. In Kombination mit der Funktionalität großer Sprachmodelle können sie semantisch und syntaktisch sinnvolle Texte produzieren. Dabei entsteht leicht der Eindruck, die KI besitze so etwas wie „Verständnis" oder „common sense", was auch in der menschlichen Tendenz begründet liegt, Phänomene als „Gegenüber" wahrzunehmen, insbesondere, wenn sie wie solche erscheinen. Tatsächlich „versteht" KI jedoch keine Bedeutungen und Zusammenhänge, sondern arbeitet mit Korrelationen: Worte und Begriffe werden (als sog. Tokens) mithilfe von sog. Embeddings (vorstellbar als Achsen im Token-Koordinatensystem, auf welchem sinnverwandte Wörter in räumlicher Nähe zueinander stehen) in Bedeutungsdimensionen einsortiert: ChatGPT arbeitet hier mit über 10.000 solcher Dimensionen für jedes Wort. Ein vereinfachtes Beispiel: Soll das Wort „König" im Training eines Sprachmodells „embedded" werden, könnte das Modell es in den (fiktiven) Kategorien „Herrscher", „männlich", „monarchisch" jeweils mit einer *1* markieren, in den Kategorien „weiblich" oder „demokratisch" jedoch mit einer *0*. Löschte man nun die Markierung bei „männlich" und setzte stattdessen eine bei „weiblich", würde voraussichtlich das Wort „Königin" erscheinen.

Diese Kategorien erscheinen uns Menschen sinnvoll. Für das Sprachmodell haben die Dimensionen des Embeddings allerdings keinen Sinngehalt. Es erschließt sich diese Tabellen lediglich aus der Masse an Trainingsdaten und erzeugt so einen Vektor, mit dessen Hilfe es Tokens anhand ihrer Position im Verhältnis zu anderen Tokens sortieren kann.

Korrelationen können natürlich auch auf Kausalitäten hinweisen, tun es aber nicht zwangsläufig. Für uns sind Scheinkorrelationen teilweise schnell zu entlarven. Dass es zwischen der Anzahl verkaufter Eiskugeln und Todesfällen durch Ertrinken in Freibädern keinen direkten kausalen Zusammenhang gibt, sagt uns der gesunde Menschenverstand. Generativer KI fehlt diese Fähigkeit. Da KI-Systeme, wie gezeigt, keine Wahrheiten produzieren, sind die generierten Texte zwangsläufig kritisch zu hinterfragen.

3 KI wird immer klüger

Für die Erstellung von KI-Modellen mittels Machine Learning werden große Datenmengen benötigt: Bei GPT 3.5 ist die Rede von 570 GB an (gefiltertem) Textmaterial, das für das Training des Modells verwendet wurde. Darunter befinden sich frei zugängliche Daten wie Wikipedia-Artikel, Patentdatenbanken oder Nachrichtenseiten genauso wie kopiergeschütztes Material. Gleichzeitig stehen die Schöpfer neuer KI-Modelle vor dem Problem, dass das Training (exponentiell) größerer Modelle auch (exponentiell) größere Trainingsdaten benötigt, während ein Großteil der verfügbaren Trainingsdaten bereits „abgegrast" ist.

Im Juli vergangenen Jahres beobachteten Nutzer bei der Verwendung von ChatGPT übereinstimmend eine Verschlechterung der Antwortqualität. Während OpenAI angibt, nichts am System geändert zu haben, wurde in der Community über die möglichen Ursachen spekuliert. Eine spannende – und nicht unplausible – These besagte, dass die Verschlechterung der Ergebnisse damit zusammenhängen könnte, dass nun zunehmend KI-generierte Inhalte im Web zu finden sind – und sich die Trainingsdaten der Modelle wiederum aus dem Web speisen. Dass Rekursion, also das Füttern eines Modells mit KI-generierten Daten, mittelfristig zu schlechteren Ergebnissen führt, wurde bereits an verschiedenen KI-Modellen beobachtet. Forscher der Universitäten Stanford und Rice tauften das Phänomen passend MAD (Model Autophagy Disorder).

Immer mehr qualitativ hochwertige Trainingsdaten für immer größer werdende Sprachmodelle zu finden, ist eine der größten Herausforderungen in der KI-Entwicklung. Einige Forscher:innen sind der Ansicht, dass wir bereits 2026 bei frischen Daten hoher Qualität an eine Grenze stoßen könnten.

4 KI ist eine Blackbox

Wie bereits beschrieben, speisen sich generative KI-Modelle aus einer großen Menge an Daten. Gerade weil für den Laien in der Regel nicht nachvollziehbar ist, wie die KI zu ihrem Ergebnis kommt, d.h. mit welchen Daten wie operiert wurde, entsteht schnell der Eindruck einer Blackbox. Der Blackbox-Charakter trainierter KI-Modelle liegt zum einen in der unüberschaubaren Anzahl ihrer Parameter begründet, denn die Auswirkungen dieser auf das Ergebnis sind nicht ohne weiteres nachvollziehbar. Zum anderen bleibt unklar, auf welcher Datenbasis diese Operationen überhaupt durchgeführt werden und ob diese Basis beispielsweise frei von Verzerrungen ist. Denn KI-Unternehmen wie OpenAI bleiben in der Regel intransparent, wenn es darum geht, welche Daten genau für das Training verwendet wurden. Deshalb hat es sich die Disziplin der „erklärbaren KI" (XAI) zur Aufgabe gemacht, mittels verschiedener Methoden (etwa nachträgliche Sichtbarmachung von Heatmaps, gezielte Manipulation der Input-Daten, Vereinfachung) das Verhalten einzelner KI-Modelle nachträglich transparenter und erklärbarer zu machen. Obwohl dies zum generellen Verständnis der Funktionsweise der Systeme beiträgt, steht diese Möglichkeit der 'Nachvollziehbarkeit' individueller KI-Outputs für Laien bisweilen nicht zur Verfügung. Hier bleibt generative KI vorerst eine Blackbox.

5 KI ist kreativ

Generative KI mag den Anschein erwecken, kreativ zu sein; in Sekundenschnelle können neue Bilder erstellt oder Texte generiert werden. Bei den erzeugten Werken handelt es sich allerdings immer um Remixe z.B. bestehender Kunstwerke. Nüchtern betrachtet ist unsere Kreativität selbst also die Ressource, aus der die generative KI ihre „Kreativität" schöpft; KI-generierte „Kunst" spiegelt lediglich menschlicher Kreativleistungen

Neben diesem Fundus menschlicher Kreativität, das von den generativen KI-Tools gemäß ihrer Operationsweise rekombiniert wird, ist aber auch die Kreativität derer nicht zu übersehen, die das Tool „befähigen". Kann ich als Toolnutzer:in eine Anweisung (Prompt) formulieren, die das Tool dazu veranlasst, einen möglichst originellen Output zu generieren? Die Befähigung der Tools selbst setzt bei komplexen – und damit komplex bleibenden – Herausforderungen menschliche Kreativität weiter voraus. Nicht ohne Grund wird diskutiert, ob nicht auch Prompts, also die Anweisungen, die zum Output der KI geführt haben und besonders originell sind, urheberrechtlich geschützt werden können.

Ein konstruktiver Umgang mit generativen Sprachmodellen besteht darin, das Tool als „Inspirationsmaschine" zu nutzen. Es werden Ideen, Lösungsansätze, Sprachbilder und Formulierungen vorgeschlagen, auf die wir aus dem Stegreif vielleicht nicht unmittelbar gekommen wären. Das bahnbrechend Neue wird KI aber nicht erschaffen können: Menschliche Kreativität lebt von Emotion, Lebenserfahrung, Intuition und Improvisation. All das sind Aspekte, die für die KI nicht zugänglich sind und damit bleibt auch das kreative Schaffen der KI begrenzt – begrenzt auf mathematische Operationen.

6 KI hat ein Bewusstsein

Im Juni 2022, wenige Monate vor dem Erscheinen von ChatGPT, behauptete der damalige Google-Ingenieur Blake Lemoine, dass der von ihm getestete Google-Chatbot LaMDA ein Bewusstsein entwickelt habe und veröffentlichte sein Gespräch mit der KI auf seinem Blog. Derselbe Lemoine bezeichnet sich in einem Blogbeitrag von 2019 als „gnostischer Christ", also als Vertreter eines dualistischen Weltbilds. Es ist frappierend, wie sehr jahrtausendealte dualistische Topoi bei Vertretern der starken KI (also der Vorstellung, dass KI ein Bewusstsein und eigene Intentionalität entwickeln kann) zum Vorschein kommen – ist die Vorstellung einer bewussten KI nichts anderes als die Vorstellung eines körperlosen Geistes.

Diese Vorstellung basiert jedoch auf Glaubenssätzen, die nicht zwangsläufig geteilt werden müssen. So bezeichnet etwa der Heidelberger Psychiater Thomas Fuchs die vermeintliche Alternative zwischen einem subjektiven Ich und dem Gehirn als Urheber von Handlungen, als verengt. Das Gehirn vermöge nämlich als Organ überhaupt keine Entscheidungen zu treffen – Begriffe wie Fühlen, Wollen und Entscheiden seien auf physiologischer Ebene gar nicht anwendbar.[1] Ein vom Körper unabhängig in einer Nährstofflösung lebendes Gehirn wäre demnach genauso wenig denkbar wie eine bewusste KI.

Es gibt in der intensiv geführten Mind-Brain-Debatte eine Vielzahl ungelöster Probleme: Neben dem Qualia-Problem (Wie können aus neuronalen Prozessen überhaupt wahrnehmbare

[1] „Das Gehirn verfügt nicht über geistige Zustände oder über Bewusstsein, denn das Gehirn lebt nicht – es ist nur das Organ eines Lebewesens, einer lebendigen Person. Nicht Neuronenverbände, nicht Gehirne, sondern nur Personen fühlen, denken, nehmen wahr und handeln." (Thomas Fuchs: Das Gehirn – ein Beziehungsorgan. Eine phänomenologisch-ökologische Konzeption, Stuttgart 22009, 283.).

Bewusstseinserlebnisse entstehen?), der Frage nach der Intentionalität, dem Bindungsproblem (Warum gibt es ein subjektiv erfahrbares Ich?) wird vor allem die Frage nach der Willensfreiheit diskutiert. Bislang gibt es für keine dieser Fragen einen Konsens, alle möglichen Antworten basieren zu einem großen Teil auf Glaubenssätzen, die andere Diskussionsteilnehmer:innen teilen können, aber nicht müssen. Festzuhalten bleibt: Was ein „Bewusstsein" ist oder wie es entsteht, haben wir nicht ansatzweise verstanden.

Ungeachtet dessen erwecken Techno-Optimisten wie Ray Kurzweil oder sog. „Singularians" den Eindruck, wir stünden kurz davor, mithilfe von KI ein künstliches Bewusstsein zu erschaffen. In eine ähnliche Richtung gehen auch Vorstellungen einer „Artificial General Intelligence", einer Art Allzweck-KI, die eigenständig sämtliche Probleme zu lösen imstande wäre. Aber auch Ray Kurzweil fällt in seinem Buch „How to Create a Mind" auf Glaubenssätze zurück, wenn er erklären möchte, wie aus einer (scheinbar) perfekten Simulation ein Bewusstsein werden soll.[2]

Zur Mystifizierung von KI trägt auch die Behauptung von „emergenten Fähigkeiten" von Large Language Models (LLM) durch die aktuelle KI-Forschung bei. Unter „Emergenz" in diesem Kontext wird verstanden, dass größere Modelle in der Lage seien, Probleme zu lösen, auf welche sie nicht explizit trainiert worden sind. Das erweckt den Eindruck einer Intentionalität bzw. Selbstwirksamkeit von KI-Modellen, für die es bei näherer Betrachtung keinerlei Beleg gibt. Liest man derartige Paper dann genauer, bleibt von der behaupteten Emergenz auch oft nicht mehr als die banale Feststellung, dass länger trainierte Modelle komplexere Aufgaben zu lösen imstande sind als kürzer trainierte Modelle.

Interessanter scheint uns ein Phänomen, das der Computerpionier Joseph Weizenbaum bereits 1966 beobachtet hat: Nämlich, dass Menschen offenbar bereit sind, sich von vergleichsweise simpler Technologie täuschen zu lassen, wenn sie den Eindruck eines Gegenübers erweckt. Die Erfahrung, dass seine eigenen Mitarbeiter bei der Nutzung des von ihm programmierten, sehr rudimentären Chatbot ELIZA den Eindruck eines intimen Gesprächs mit der Software erweckten, verstörte ihn damals so sehr, dass er zu einem frühen Skeptiker der IT-Technologie wurde.

7 Wissenschaftliches Arbeiten wird obsolet

Wir leben in einem Zeitalter der ubiquitären Information. Doch nicht alle Informationen, auf die wir stoßen, sind notwendigerweise zuverlässig. Wie oben beschrieben, trifft das in besonderer Weise auf durch Sprachmodelle generierte Texte zu. In diesem Informationsfluss kommt der Fähigkeit der Beurteilung, welche Information zuverlässig und relevant ist, eine besondere Bedeutung zu. Das persönliche Erlangen und Aufbauen von Fachwissen bleibt hierbei wesentliche Grundlage wissenschaftlichen Arbeitens, um Informationen zu bewerten und einzuordnen (Friedrich & Tobor, 2023). Wissenschaftliches Arbeiten umfasst jedoch mehr als das Anhäufen und die Reproduktion von Wissen. Die Rekombination bestehender Materialien ist entsprechend nur der erste Schritt, denn Wissenschaft bedeutet, daraus neue Erkenntnisse zu gewinnen, eigene Ideen zu entwickeln und genau das kann generative KI nicht (siehe hierzu auch Mythos 5 "KI ist kreativ"). Selbst für die bloße Reproduktion bestehenden Wissens sind LLMs auch weiterhin im akademischen Kontext nicht geeignet: Weist man sie an, aktuelle Wissensstände reproduzierende Essays unter Angabe von

[2] „My objective prediction is that machines in the future will appear to be conscious and that they will be convincing to biological people when they speak of their qualia. […] We will come to accept that they are conscious persons. My own leap of faith is this: Once machines do succeed in being convincing when they speak of their qualia and conscious experiences, they will indeed constitute conscious persons." (Ray Kurzweil: How to Create a Mind. The Secret of Human Thought Revealed, New York 2013, 209f., Hervorhebung durch Verf.).

Quellen zu generieren, sind diese Quellen weiterhin oft frei erfunden und/oder belegen gerade nicht den vom LLM generierten Textabschnitt (vgl. Martin Wan, „Warum ChatGPT nicht das Ende des akademischen Schreibens bedeutet“ / „Warum ChatGPT (auch weiterhin) nicht das Ende des akademischen Schreibens bedeutet“).[3]

KI-Sprachmodelle werden voraussichtlich wissenschaftliche Schreibpraktiken verändern. KI-Tools produzieren kostenfrei bzw. kostengünstige Texte, die sprachlich gut sind. Diese Texte werden sich zunehmend von menschlichen Texten nicht mehr unterscheiden lassen (Limburg et al., 2023).[4] Es ist wahrscheinlich, dass KI-Schreibtools immer stärker im wissenschaftlichen Schreibprozess genutzt werden.[5] Umso wichtiger wird es sein, die genannten Grenzen im Blick zu behalten.

8 KI führt zu mehr Arbeitslosigkeit

Sobald eine neue Technologie in der Öffentlichkeit wahrgenommen wird, die in Teilbereichen menschliche Fähigkeiten erreichen bzw. übersteigen soll, folgt mit großer Zuverlässigkeit eine Debatte über die Vernichtung von Arbeitsplätzen. Dieses Muster lässt sich medial gut zurückverfolgen.[6] Auch mit dem Release von ChatGPT findet diese Debatte wieder Einkehr in die Berichterstattung.[7]

Die Geschichte zeigt bisher aber immer wieder, dass die jeweilige Technologie nicht unmittelbar zu mehr Arbeitslosigkeit führt, sondern die Arbeitspraxis verändert. Damit geht natürlich einher, dass bestimmte Jobprofile aus der Zeit fallen, modifiziert oder neu interpretiert werden müssen. Völlig neue Jobs entstehen. Dass jedoch menschliche Arbeitskraft im Zuge technologischer Entwicklung obsolet werden könnte, hat sich bislang weder für die Industrie noch für die Kreativ- oder Finanzbranche gezeigt.8 Von allen weiteren Vermutungen über rückläufige Zahlen in verschiedenen Tätigkeitsbereichen wollen wir an dieser Stelle Abstand nehmen. Stattdessen wird anhand der bereits entkräfteten Mythen sichtbar, dass generative KI zum jetzigen Stand viele menschliche Eigenschaften, die in Arbeitskontexten erforderlich sind, nicht gänzlich in der Lage sind zu substituieren. Ob dies in Zukunft der Fall sein wird und in der Summe dazu führen kann, dass

---

[3] Auch ChatGPT-Plugins wie ScholarAI oder angepasste GPTs wie Scispace (vormals ResearchGPT) helfen hier nur bedingt weiter: Zwar können sie in vielen Fällen relevante Publikationen zu bestimmten Themen finden eine wirkliche „Rezeption“ dieser Arbeiten findet durch das Sprachmodell allerdings nicht statt: Weist man es an, auf Grundlage der gefunden Publikationen einen Essay zu schreiben, geht das Ergebnis über eine Reformulierung und Aneinanderreihung der jeweiligen Abstracts nicht hinaus. Insofern werden auch einzelne Artikel nicht auf andere Artikel bezogen, sondern lediglich im Sinne einer Inhaltsangabe aneinandergereiht.

[4] Fest steht, KI-Sprachmodelle werden nicht mehr verschwinden und ihr Einsatz lässt sich insbesondere bei unüberwachten schriftlichen Prüfungen, wie z.B. der Hausarbeit, nicht immer zweifelsfrei nachweisen. Auch KI-Detektoren können nicht verlässlich KI-generierte Texte erkennen (Weber-Wulff et al., 2023). Ein einfaches Verbot stößt daher schnell an seine Grenzen.

[5] Welche Implikationen das für Forschende, Lehrende und Studieren hat, kann in “Wissenschaftliches Schreiben im Zeitalter von KI gemeinsam verantworten. Diskussionspapier Nr. 27. Berlin: Hochschulforum Digitalisierung” nachgelesen werden.

[6] So titelte beispielsweise der Spiegel am 17. April 1978: “Die Computer-Revolution - Fortschritt macht arbeitslos”. Das Cover schmückt ein riesiger Spielzeugroboter, dessen Hand den Blaumann eines hilflosen Arbeiters am Latz packt und diesen wegschafft. Damals standen die noch neuartigen Mikroprozessoren, die Einkehr in die industrielle Produktion fanden, im Fokus des Artikels. Eine weitere Titelseite des Spiegels vom 3. September 2016 weist gewisse Parallelen auf. Auch hier eine Roboterhand, die einen Angestellten greift und seines Arbeitsplatzes verweist. Diesmal trifft es den Büroangestellten. Der Titel dieser Ausgabe: “Sie sind entlassen! Wie uns Computer und Roboter die Arbeit wegnehmen – und welche Berufe morgen noch sicher sind”.

[7] So fragt die NZZ: ”Die KI kommt: Treibt uns Chat-GPT in die Arbeitslosigkeit?” (Pressl & Rueegg, 02.06.2023). Die FAZ titelt: “In welchen Berufen kann ChatGPT übernehmen?” (Bös & Diemand, 28.02.2023).

[8] Wie bei jeder Extrapolation ist allerdings auch hier davor zu warnen, den historischen Wandel der Arbeitswelt nun 1:1 auf die Implikationen von KI fortzuschreiben. Fakt ist jedoch, generative KI-Systeme haben aktuell die in diesem Papier genannten Limitationen. Die Gefahr einer Massenarbeitslosigkeit lässt sich daraus zum jetzigen Stand nicht ableiten.

tatsächlich irgendwann gesagt werden kann 'KI vernichtete Arbeitsplätze', ist reine Spekulation und überblendet die viel konstruktivere Perspektive: Die der wechselseitigen Interaktion zwischen Mensch und Maschine, in der dem Menschen seine Autonomie erhalten bleibt.[9]

9 KI kann in die Zukunft blicken

Eine grundlegende Limitation jeglicher datenbasierten Vorhersagen: Sie stützen sich zwangsweise auf Daten aus der Vergangenheit. In vielen Szenarien können so wahrscheinliche Entwicklungen antizipiert werden (z.B. beim Kundenverhalten im Einzelhandel über historische Verkaufsdaten). Da es allerdings auch unwahrscheinliche, aber folgenschwere Ereignisse gibt, mit denen niemand rechnen kann, stößt die Fortschreibung der Zukunft mit Daten aus der Vergangenheit nicht selten an ihre Grenzen. Obwohl der Verlauf der Coronapandemie immer präziser vorhergesagt werden konnte (z.B. über das sog. Abwassermonitoring), erwischte uns ihr Eintreten im Jahr 2020 völlig unvorbereitet. Während also in enggeführten Szenarien, die über quantifizierbare und begrenzte Variablen verfügen, Vorhersagen über Eintrittswahrscheinlichkeiten gut möglich sind, sinkt die Vorhersagequalität mit der Breite und der Quantifizierbarkeit der zu berücksichtigen Variablen (z.B. Berechnungsmodelle zum Eintritt einer Pandemie). Egal ob Vorhersagen nun als zuverlässig oder als vage gelten, sie bleiben immer mit Unsicherheiten behaftet, unabhängig davon, wie gut und vollständig die Datenbasis erscheint und wie durchdacht das Berechnungsmodell ist.10 Die Zukunft bleibt offen und sie überrascht.

Auch der Einsatz von KI-Systemen, die in den oben aufgeführten Beispielen bereits zum Einsatz kommen, ändert nichts daran. Trotz der Tatsache, dass sie mit riesigen Datenmengen umgehen können und Muster darin identifizieren, die uns potenziell verschlossen bleiben, bleiben sie von den oben genannten Limitationen nicht unberührt.[11] Dennoch scheinen diese Grenzen oft ignoriert zu werden, was nicht selten zu einer Überschätzung von KI-basierten Vorhersagen führt. Wie die Wissenschafts- und Technikforscherin Helga Nowotny (2023) schreibt, prägt der in der Moderne verankerte Glaube, dass Technik die Probleme der Menschheit lösen kann (siehe auch Mythos 6), auch den Umgang mit KI-basierten Vorhersagen.[12]

Diese könnten zwar unseren Blick erweitern, aber sie bergen auch das Risiko, dass wir andere Möglichkeiten, die 'außerhalb' der Vorhersage liegen, übersehen. Die Zukunft allein der KI zu überlassen, ist verführerisch, bedroht aber unsere Autonomie.

---

[9] Warum KI nicht ohne uns auskommt, lässt sich z.B. mit dem Problem des 'Verantwortungsdefizits' aufgreifen. Weder KI noch Technik im Allgemeinen sind fehlerfrei (siehe hierzu Mythos 1). Das wirft die Frage auf, was passiert, wenn der Fehler eines autonomen KI-Systems zum (juristisch relevanten) Schadensfall führt. Wer übernimmt die Verantwortung? Da Vertrauen an Vernunft, Freiheit und damit an Autonomie gebunden ist (Staudacher & Nida-Rümelin, 2022), was der KI nicht innewohnt, ist auch ihre Verantwortungsübernahme problematisch. Daraus folgt: Ohne menschliche Kompetenzträger:innen lassen sich 'KI-induzierte' Schaffensprozesse nicht verantworten. Dies spricht nicht für eine Substitution menschlicher Arbeitskraft, sondern für einen Einsatz von KI als geschätztes Hilfsmittel.

[10] Trotz etlicher Faktoren und stetig verbesserter Modelle und Messtechniken, die bei der Vorhersage des Wetters Anwendung finden, kommt es oft genug vor, dass sich über 'Falschmeldungen' echauffiert wird. Zu komplex, dynamisch und damit fehleranfällig sind die Faktoren, die zum Wetter führen, um gutes oder schlechtes Wetter garantieren zu können. Dennoch gilt die Wettervorhersage im Feld der Vorhersage als relativ zuverlässig.

[11] Bereits die Bereitstellung der notwendigen Referenzdaten ist enorm voraussetzungsvoll. Beim Release von ChatGPT im November 2022 durften wir z.B. feststellen, dass das LLM mit Daten bis September 2021 trainiert wurde. Als wir das System aufforderten, Prognosen zu erstellen, erschienen diese oft bereits unplausibel.

[12] Gerade weil für uns Laien nicht nachvollziehbar ist, wie KI funktioniert (siehe hierzu Mythos 4), haftet dem Ganzen eine mystische Komponente an, die über die Grenzen datenbasierter Vorhersagen hinwegtäuscht.

Mit Blick auf die Hochschulbildung, in der KI als Vorhersageinstrument zunehmend in den Fokus zu rücken scheint, gibt es daher auch einige Einsatzszenarien, die kritisch reflektiert werden sollten. Während der Einsatz im Verwaltungsbetrieb, z.B. bei der Prognose von Raumauslastungen, eher unproblematisch erscheint, ist dies z.B. bei KI-Empfehlungssystemen weniger der Fall. So sind Systeme denkbar, die auf Basis eines Datenprofils der Studierenden den optimalen Studienverlauf prognostizieren und entsprechende Ratschläge geben, wie man sich diesem annähern kann. Das kann bis zu einem gewissen Maß hilfreich sein. Vor dem Hintergrund der Grenzen der KI-Vorhersage ist es jedoch wichtig, einen nüchternen Blick auf solche potenziell folgenreiche Entscheidungsempfehlungen zu werfen, um sich möglichen Alternativen, die außerhalb der Empfehlungen liegen, nicht zu verschließen.[13]

Fazit: Bedeutung von generativen KI-Anwendungen für Hochschulen

Generative KI-Tools werden nicht wieder verschwinden. Die Hochschulen sollten also neue Wege finden, um mit Tools wie ChatGPT umzugehen – sie gar zum Ausgangspunkt für Überlegungen machen, wie mit ihnen ein besseres Hochschulstudium ermöglicht werden kann. Dies kann nur dann funktionieren, wenn es sich dabei um einen entmystifizierten Blickwinkel handelt. Erst dann wird erkennbar, inwieweit die Modelle einen positiven Einfluss auf Studium und Lehre nehmen können: In welchem Rahmen können sie z.B. zur Personalisierung des Lernprozesses beitragen, bei der Gestaltung von Lehrveranstaltungen und Prüfungen unterstützen oder die administrative Verwaltung entlasten? Auf all diesen Ebenen gibt es zugleich Einsatzmöglichkeiten als auch Grenzen des Einsatzes. Der nüchterne Blick auf die Technologie, der hier zum Austarieren von Möglichkeiten und Grenzen einlädt, zeigt uns also, worauf es jetzt ankommt:

1. "KI-Literacy", d.h. die Fähigkeit, KI-Systeme, deren Outputs und Wirkungsdimensionen zu verstehen und zu interpretieren, muss ebenso wie konkrete Anwendungskompetenzen, z.B. das Prompting, verstärkt in alle Studiengänge integriert werden.
2. Es ist und bleibt wichtig, Fachwissen und Kompetenzen zu entwickeln, die in spezifischen Anwendungskontexten zur reflexiv-kritischen Grundlage jeder verantwortungsvollen Mensch-KI-Interaktion werden. Die Entwicklung dieser Kompetenzen muss in der Lehre erhalten und überprüfbar bleiben.
3. Der Umgang mit KI, eine Urteilsfähigkeit in Bezug auf Stärken und Schwächen von KI-Systeme sollte als Teil der Persönlichkeitsbildung an Hochschulen angesehen werden: Zur Mündigkeit in einer von KI immer stärker geprägten Welt gehört der souveräne Umgang mit der Technologie und die Urteilsfähigkeit über ihren Einsatz. Aus diesem Grund hat das Hochschulforum Digitalisierung den Think Tank „Künstliche Intelligenz: Essenzielle Kompetenzen an Hochschulen“ ins Leben gerufen, der einen fachunabhängigen Kompetenzrahmen entwickelt.

[13] So macht auch der Ethikrat darauf aufmerksam, zu berücksichtigen, dass KI-Systeme als Empfehlungstools lediglich Empfehlungen aussprechen, nicht aber Entscheidungen über ihre Anwender:innen treffen dürfen und problematisiert dabei den Automation Bias: „Selbst wenn ein KI-System normativ strikt auf die Rolle der Entscheidungsunterstützung begrenzt wird, kann Automation Bias dazu führen, dass ein KI System allmählich in die Rolle des eigentlichen ´Entscheiders´ gerät und menschliche Autorschaft und Verantwortung ausgehöhlt werden“ (Deutscher Ethikrat, 2023, S. 26).

4. Es sind hochschulweite Rahmenbedingungen für den Umgang mit KI erforderlich. Diese Rahmenbedingungen leiten den Umgang der Hochschulangehörigen mit der Technologie unter Wahrung der akademischen Integrität. Hochschulweite KI-Leitlinien können ein Instrument darstellen, welches diese Orientierungsleistung stützt. Der HFD-Blickpunkt "Leitlinien für den Umgang mit generativer KI" bietet Hilfestellungen bei diesem Prozess.
5. KI-Anwendungen werden sich beständig weiterentwickeln. Es sollte verstärkt Experimentierräume geben, in denen die Hochschulangehörigen statusgruppenübergreifend erproben können, wie KI-Tools typische Praktiken (Lernen, Lehren, Prüfen) verändern und beispielsweise den dahinter steckenden Arbeitsaufwand (z. B. Prüfungsgestaltung) reduzieren können.
6. Der freie Zugang zu datenschutzkonformen und leistungsstarken KI-Modellen ist eine wichtige Voraussetzung dafür, dass alle Hochschulangehörigen diese möglichst risikofrei und unabhängig von individuellen ökonomischen Ausgangsbedingungen nutzen können. Im besten Fall sollte es europäische Open-Source-Alternativen geben. Darauf können Hochschulen jedoch nicht warten, weshalb faire Education-Lizenzmodelle seitens der Anbieter von KI-Systemen nötig sind.
7. Lehrende müssen zu den informierten Expert:innen im Umgang mit KI gehören, weswegen es spezielle Weiterbildungsformate braucht, die auf die Entwicklung von KI-Kompetenzen einzahlen.

Literatur

Brommer, S., Berendes, J., Bohle-Jurok, U., Buck, I., Girgensohn, K., Grieshammer, E., Gröner, C., Gürtl, F., Hollosi-Boiger, C., Klamm, C., Knorr, D., Limburg, A., Mundorf, M., Stahlberg, N., Unterpertinger, E. (2023). *Wissenschaftliches Schreiben im Zeitalter von KI gemeinsam verantworten.* Diskussionspapier Nr. 27. Berlin: Hochschulforum Digitalisierung.

Deutscher Ethikrat (2023). *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz* (Stellungnahme – Kurzfassung). Berlin. https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine-kurzfassung.pdf.

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.* Berlin: Hochschulforum Digitalisierung.

Limburg, A., Bohle-Jurok, U. , Buck, I., Grieshammer, E., Gröpler, J., Knorr, D., Mundorf, M., Schindler, K., Wilder, N. (2023). *Zehn Thesen zur Zukunft des wissenschaftlichen Schreibens.* Diskussionspapier Nr. 23. Berlin: Hochschulforum Digitalisierung. https://hochschulforumdigitalisierung.de/wp-content/uploads/2023/09/HFD_DP_23_Zukunft_Schreiben_Wissenschaft.pdf.

Nowotny, H. (2023). *Die KI sei mit euch. Macht, Illusion und Kontrolle algorithmischer Vorhersage.* Aus dem Englischen von Sabine Wolf. Berlin: Matthes & Seitz.

Salden, P. (2023). *Didaktische und rechtliche Perspektiven auf KI-gestütztes Schreiben in der Hochschulbildung.* Zentrum für Wissenschaftsdidaktik der Ruhr-Universität Bochum. https://doi.org/10.13154/294-9734.

Staudacher, K., Nida-Rümelin, J. (2022). *Philosophische Überlegungen zur Verantwortung von KI: Eine Ablehnung des Konzepts der E-Person.* Bidt DE. https://doi.org/10.35067/bv16-2z28.

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O. (2023). *Testing of Detection Tools for AI-Generated Text.* https://arxiv.org/abs/2306.15666.

IMPRESSUM

Diskussionspapiere des HFD spiegeln die Meinung der jeweiligen Autor:innen wider.
Das HFD macht sich die in diesem Papier getätigten Aussagen daher nicht zu Eigen.

ISSN (Online) 2365-7081; 5. Jahrgang

Zitierhinweis
Friedrich, J.-D., Tobor, J., Wan, M. (2024). 9 Mythen über generative KI in der Hochschulbildung. Diskussionspapier Nr. 29. Berlin: Hochschulforum Digitalisierung.

Herausgeber
Geschäftsstelle Hochschulforum Digitalisierung beim Stifterverband für die Deutsche Wissenschaft e.V.
Hauptstadtbüro • Pariser Platz 6 • 10117 Berlin • T 030 322982-520
info@hochschulforumdigitalisierung.de

Korrektorat
Katja Engelhaus

Verlag
Edition Stifterverband – Verwaltungsgesellschaft für Wissenschaftspflege mbH
Barkhovenallee 1 • 45239 Essen • T 0201 8401-0 • mail@stifterverband.de

Layout
Satz: Emily Fröse. Katja Engelhaus
Vorlage: TAU GmbH • Köpenicker Straße 154a • 10997 Berlin

Das Hochschulforum Digitalisierung ist ein gemeinsames Projekt des Stifterverbandes, des CHE Centrums für Hochschulentwicklung und der Hochschulrektorenkonferenz. Förderer ist das Bundesministerium für Bildung und Forschung.

www.hochschulforumdigitalisierung.de

STIFTERVERBAND

CHE Centrum für Hochschulentwicklung

HRK Hochschulrektorenkonferenz

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung