

Stellungnahme zu ChatGPT

für den Bundestagsausschuss
Bildung, Forschung und
Technikfolgenabschätzung

Deutscher Bundestag

Ausschuss für Bildung, Forschung
und Technikfolgenabschätzung

Ausschussdrucksache
20(18)108c

24.04.2023

24.04.2023

Sehr geehrte Mitglieder des Ausschusses für Bildung, Forschung und
Technikfolgenabschätzung,

hiermit möchte ich Ihnen meine aktuellen Beobachtungen und Erfahrungen beim Einsatz
generativer KI-Systeme an Schulen wie auch Hochschulen darlegen und darauf basierend
Handlungsempfehlungen **zum Einsatz und zur Regulierung von ChatGPT in den
Bereichen Bildung und Forschung unter Berücksichtigung der Förderung von
Innovationen in Wirtschaft und Gesellschaft** formulieren.

Vorbemerkungen

Der Bereich Bildung und Forschung ist im Zeitalter generativer KI-Systeme einem
Umbruch ausgesetzt, den es in diesem Ausmaß bisher nicht gegeben hat. Mit der
Veröffentlichung von ChatGPT am 30. November 2022 ist eine Zäsur zu beobachten, die
in ihrem disruptiven Potenzial und ihren Folgewirkungen selbst heute nur schemenhaft zu
erkennen ist. Bill Gates hat in einem Beitrag vom 21.03.2023 zur Bedeutung der KI im
Bildungsbereich von einer revolutionären Entwicklung gesprochen: „*Education
Computers haven't had the effect on education that many of us in the industry have hoped.
There have been some good developments, including educational games and online
sources of information like Wikipedia, but they haven't had a meaningful effect on any of
the measures of students' achievement. But I think in the next five to 10 years, AI-driven
software will finally deliver on the promise of revolutionizing the way people teach and
learn*“ (Gates 2023).

Die gesellschaftliche und politische Diskussion ist in der deutschen Bildungslandschaft
stark auf das Werkzeug **ChatGPT** vom Anbieter OpenAI aus San Francisco fokussiert,
verkennt damit aber den schnell wachsenden Gesamtmarkt vergleichbarer KI-
Chatlösungen, die wiederum auf anderen großen KI-Sprachmodellen („Large Language
Model“ - LLM) basieren. Laut dem AI Index 2023 des Institute for Human-Centered AI der
Stanford University dominieren heute die privaten Tech-Giganten die Entwicklung der

großen KI-Modelle, während bis 2014 die meisten bedeutenden Modelle von öffentlichen Forschungseinrichtungen entwickelt wurden. Diese Entwicklung spiegelt wider, dass wegen des immensen Kapital- und Ressourceneinsatzes, der für die KI-Forschung in diesem Bereich notwendig ist, **öffentlich finanzierte Forschungseinrichtungen nicht mehr wettbewerbsfähig** scheinen. Auf einen weiteren Problempunkt für den KI-Forschungsstandort in Deutschland weist der AI Index (Maslej 2023, S. 44) hin, demzufolge sich die ersten neun Plätze der zehn besten Forschungseinrichtungen, bewertet nach der Anzahl der KI-relevanten Publikationen zwischen 2010 und 2021, in China befinden. Das amerikanische MIT liegt auf Platz 10. **Europa ist in der Top-Liga überhaupt nicht vertreten.**

Bei der Bewertung der Technologie generativer KI-Systeme ist in der aktuellen Diskussion der Fokus bei ChatGPT zudem einseitig auf die **Textgenerierung** gerichtet. Mit den dahinter liegenden KI-Sprachmodellen GTP-3.5 bzw. der neuen Version GTP-4 handelt es sich um sogenannte „General purpose AI“, d.h. KI-Systeme, die eine große Bandbreite an Funktionen für den Umgang mit multimodalen Content-Formen aufweisen, die deutlich über die Textgenerierung und -bearbeitung hinausgehen. Anzumerken ist, dass sich bereits bei dem Sprachmodell GTP-3, das im Jahr 2020 veröffentlicht wurde, **emergente Fähigkeiten** zeigten, die über die intendierten Trainingsziele des Modells hinausgingen, so z.B. mathematische Fähigkeiten oder die Interpretation und die Generierung von Software-Code. Diese Phänomene zeigen sich sehr deutlich auch bei neueren Modellversionen wie GPT-4 und führen zu Diskussionen über eine „starke KI“ bzw. „artificial general intelligence“ (Bubeck et al. 2023).

Bisher noch weitgehend unbekannt, zeigt sich der nächste Entwicklungsschritt generativer KI-Systeme hin zu **generativen KI-Agenten**. Die Verbindung von großen KI-Sprachmodellen mit interaktiven Agenten führt zu neuen Architektur- und Interaktionsmustern, die glaubwürdige Simulationen menschlichen Verhaltens ermöglichen (Park et al. 2023). Durch diese neuen Szenarien entstehen aber auch gänzlich neue Herausforderungen für die menschliche Kontrolle und Steuerung. Es erscheint wie eine neue Ära der KI, wenn KI-Agenten weitgehend autonom eine komplexe Aufgabenstellung bestehend aus eigenständig identifizierten Einzelaufgaben bearbeiten. Die experimentelle Open-Source-Anwendung AutoGPT (<https://autogpt.net/>) ist eines der ersten Beispiele für die vollständige Autonomie des Einsatzes von GPT-4.

Ein weiterer Trend in der nahen Zukunft dürfte die Integration von GPT-4 und vergleichbaren KI-Systemen mit humanoiden Robotern in Richtung **"Embodied AI"** sein, wodurch die räumliche Dimension hinzukäme.

Diese Dynamik zeigt, dass ein engmaschiges und permanentes Monitoring der technologischen Entwicklung erforderlich ist. **Die Reaktionsgeschwindigkeit und die Anpassungsfähigkeit** von Organisationen im Bildungssektor werden zukünftig in einem bisher nicht gekannten Ausmaß gefordert sein. Die bisherige Entwicklungsgeschwindigkeit der KI-Systeme (innerhalb weniger Monate von GPT-3.5 über GPT-4 bis hin zu KI-Agenten) spiegelt bereits eine Überforderung der Bildungseinrichtungen wider, deren Entwicklungszyklen eher in Jahren und Jahrzehnten gemessen werden.

Zielsetzung und Vorgehen

Die durch generative KI-Lösungen ausgelösten Herausforderungen benötigen zur Bewältigung neben dem permanenten Monitoring dieser kontinuierlich fortschreitenden Entwicklung einen adäquaten Lösungsraum, um zukunftsweisende Konzepte für den Bildungs- und Forschungsstandort Deutschland zu entwickeln. Dabei müssen sowohl die technischen wie auch rechtlichen, ethischen und sozialen Auswirkungen von ChatGPT & Co. auf Bildung und Wissenschaft berücksichtigt werden.

Eine besondere Herausforderung stellt die anzustrebende **Balance** dar, um die Eckpfeiler für notwendige Regulierungsmaßnahmen im Lebenszyklus generativer KI-Systeme zu definieren, ohne die deutsche Wirtschaft in ihrer Innovationsfähigkeit und internationalen Wettbewerbsfähigkeit unnötig einzuschränken.

Aufgrund der aktuellen nationalen und europäischen Bemühungen und des offenbar bevorstehenden Kompromissvorschlages für den EU AI Act möchte ich mich nachfolgend auf den **Bildungsbereich** konzentrieren. Hilfreich ist der **UNESCO-Leitfaden für „ChatGPT und Artificial Intelligence in higher education“**, der mit Bezug zu den ethischen Zielsetzungen des KI-Einsatzes politische Empfehlungen für diese sieben Bereiche formuliert (Sabzalieva und Valentini 2023, S. 12):

1. Eine systemweite Vision und strategische Prioritäten
2. Übergreifendes Prinzip für KI und Bildungspolitik
3. Interdisziplinäre Planung und sektorübergreifende Steuerung
4. Politiken und Vorschriften für eine gerechte, integrative und ethische Nutzung von AI
5. Masterpläne für den Einsatz von KI im Bildungsmanagement, Lehren, Lernen und Bewertung
6. Pilotversuche, Überwachung und Bewertung sowie Aufbau einer Evidenzbasis
7. Förderung lokaler KI-Innovationen für die Bildung

Ich möchte nachfolgend mit Schwerpunktsetzung auf den ersten Punkt **der systemweiten Vision und strategischen Prioritäten** - und unter Berücksichtigung relevanter Interdependenzen zu den anderen sechs Punkten - die Bildung einer Task-Force „KI-Bildung“ für den Bildungsbereich anregen und diesen Ansatz nachfolgend ausführen, dabei einzelne Facetten herausgreifen und weitergehend erläutern. Die Implikationen für den Bereich der Forschung werde ich nur punktuell aufgreifen, um Redundanzen zu erwartbaren Stellungnahmen anderer Expert*innen in dieser Diskussion zu vermeiden. Abschließend werde ich ausgewählte rechtliche Aspekte thematisieren, die für die weitere Nutzung und Ausgestaltung der Lehre und Forschung im KI-Zeitalter relevant erscheinen.

Task-Force „KI-Bildung“ für alle deutschen Bildungseinrichtungen

Deutschland benötigt eine konzertierte Aktion, eine **Task-Force „KI-Bildung“**, die in der Lage ist, auf Basis der aktuellen Erkenntnislage wegweisende Lösungsansätze umzusetzen - schnell und mit der erforderlichen Intensität. Der Chefredakteur des Handelsblatts, Sebastian Matthes, hat am 21. April 2023 in einem Kommentar zum Thema KI sehr pointiert das „KI-Bootcamp“ für Deutschland gefordert - in Anlehnung an die Initiative der Stanford University, die amerikanische Parlamentsangehörige in einem KI-Bootcamp qualifiziert hat (Matthes 2023).

Diese Forderung ist nachvollziehbar: Ungewöhnliche Herausforderungen erfordern auch neuartige Lösungsansätze. Eine Task-Force „KI-Bildung“ erscheint dringend notwendig, da wir derzeit eine **Wissens- und Kompetenzasymmetrie** erleben, die unsere bereits bestehende digitale Spaltung der Gesellschaft noch weiter zu verstärken droht. Während die Lernenden an Schulen wie auch Hochschulen die „Power-User“¹ bei der Nutzung generativer KI-Systeme wie z.B. ChatGPT sind, hinken die Lehrenden und die Leitungsebenen der Bildungseinrichtungen der Entwicklung vielfach hinterher. Diese Divergenz stellt den Anspruch unseres Bildungssystems mit den Rollenmodellen von Lehrenden und Lernenden ganz grundsätzlich in Frage. Aus diesem Grund benötigen wir dringend eine schnell umsetzbare und auch wirksame KI-Qualifizierungsoffensive für Lehrende und Verantwortliche im gesamten Bildungsbereich, mit dem besonderen Fokus auf staatliche Bildungsträger wie Schulen und Hochschulen. Es bedarf hierzu aktueller Qualifizierungsangebote, die von einer zentralen Organisationseinheit gesteuert und bundesweit zum Einsatz kommen sollten. Es sollte bewusst ein zentralistischer Ansatz mit maximaler Reichweite gewählt und zur Vermeidung von Redundanzen auf förderalistische Strukturen verzichtet werden. Vorhandene Lösungsansätze, die bereits in einigen Bundesländern erfolgreich praktiziert werden, sind einzubeziehen.

Analog dazu sind die Nutzung vorhandener Strukturen und die Einbindung von Netzwerken, die gleichermaßen das Ziel der KI-Qualifizierung verfolgen, zwingend notwendig. Hierzu gehören u.a. Initiativen wie der **KI-Campus** mit seiner Lernplattform (<https://ki-campus.org/>) und das **Hochschulforum Digitalisierung** (<https://hochschulforumdigitalisierung.de/>) für die Hochschulbildung. Bei der Ausgestaltung der Task-Force KI-Bildung kann zum Vergleich das **European Center for Algorithmic Transparency** (ECAT) (https://algorithmic-transparency.ec.europa.eu/index_en) herangezogen werden. Dieses neue Zentrum soll als interdisziplinäres Team von Wissenschaftlerinnen und Wissenschaftlern die EU-Kommission bei der Durchsetzung der Vorschriften des **Digital Services Acts** unterstützen. Aus Ressourcengründen sollte das Zentrum bei der Analyse von Systemen und der Durchführung von Feldstudien die Expertise anderer Forschungsprojekte und Forschungseinrichtungen einbinden und nutzen. Mit Blick auf die gesamte Bildungslandschaft ist zu betonen, dass der Fokus stark auf Hochschulen gerichtet ist und es kaum analoge Initiativen für die Schulen gibt.

Qualifizierung der Lehrenden an Hochschulen

Die Erhebung des Hochschulforums Digitalisierung aus dem Jahr 2022 zeigt auf, dass Lehrende häufig motiviert sind sich weiterzubilden, es ihnen aber an Zeit fehlt. Erschwerend kommt hinzu, dass Anreizsysteme für Lehrende fehlen und Qualifizierungsangebote nicht optimal auf die Zielgruppe ausgerichtet sind. Lehrende werden zudem durch enge Rahmenbedingungen und Vorgaben (z.B. Lehrverpflichtungsverordnungen und unflexible curriculare Festlegungen) limitiert, die sie in der Ausgestaltung der Lehre einschränken (Hense und Goertz 2023, S. 8).

¹ Aussage Hendrik Haverkamp, Vorsitzender des Instituts für zeitgemäße Prüfungskultur und Koordinator Digitalität am Evangelisch Stiftische Gymnasium in Gütersloh bei der Sitzung der AG Bildung der Initiative D21 e.V. zum Thema CHATGPT IN DER SCHULISCHEN UND UNIVERSITÄREN PRAXIS am 29.03.2023.

Diese Ein- und Beschränkungen gilt es aufzulösen. In den Bildungseinrichtungen sollten **regelmäßig und in kurzer zeitlicher Taktung KI -Qualifizierungsangebote** für alle Lehrenden angeboten werden. Dazu müssen entsprechende Freiräume in den Stundenplänen der Lehrenden geschaffen werden, um sich gemeinsam mit Kolleginnen und Kollegen fortzubilden.

Die Qualifizierung der Lehrenden ist Voraussetzung für die **Nutzung der Potenziale**, die generative KI-Systeme bieten. Systeme wie ChatGPT und Co. können das Lehren, Lernen und Bewerten in einer neuen Qualität durch vielfältige Rollenmuster wie ein Lerntutor unterstützen, siehe Abbildung 1. Daraus ergeben sich für die Lernenden neue Wege des selbstbestimmten und individualisierten Lernens. Der Zugang zu Informationen wird erleichtert, Bildung wird demokratisiert, wenn es einen chancengleichen Zugang zu diesen Systemen gibt.

Eine besondere Bedeutung und neue Potenziale bieten KI-gestützte Lehr- und Lernsettings angesichts des **Fachkräftemangels bei Lehrenden** in Schulen und Hochschulen. Diese Potenziale gilt es intensiv zu analysieren und auszuschöpfen.

Role ⁶	Description	Example of implementation
Possibility engine	AI generates alternative ways of expressing an idea	Students write queries in ChatGPT and use the Regenerate response function to examine alternative responses.
Socratic opponent	AI acts as an opponent to develop and argument	Students enter prompts into ChatGPT following the structure of a conversation or debate. Teachers can ask students to use ChatGPT to prepare for discussions.
Collaboration coach	AI helps groups to research and solve problems together	Working in groups, students use ChatGPT to find out information to complete tasks and assignments.
Guide on the side	AI acts as a guide to navigate physical and conceptual spaces	Teachers use ChatGPT to generate content for classes/courses (e.g., discussion questions) and advice on how to support students in learning specific concepts.
Personal tutor	AI tutors each student and gives immediate feedback on progress	ChatGPT provides personalized feedback to students based on information provided by students or teachers (e.g., test scores).
Co-designer	AI assists throughout the design process	Teachers ask ChatGPT for ideas about designing or updating a curriculum (e.g., rubrics for assessment) and/or focus on specific goals (e.g., how to make the curriculum more accessible).
Exploratorium	AI provides tools to play with, explore and interpret data	Teachers provide basic information to students who write different queries in ChatGPT to find out more. ChatGPT can be used to support language learning.
Study buddy	AI helps the student reflect on learning material	Students explain their current level of understanding to ChatGPT and ask for ways to help them study the material. ChatGPT could also be used to help students prepare for other tasks (e.g., job interviews).
Motivator	AI offers games and challenges to extend learning	Teachers or students ask ChatGPT for ideas about how to extend students' learning after providing a summary of the current level of knowledge (e.g., quizzes, exercises).
Dynamic assessor	AI provides educators with a profile of each student's current knowledge	Students interact with ChatGPT in a tutorial-type dialogue and then ask ChatGPT to produce a summary of their current state of knowledge to share with their teacher/for assessment.

Abbildung 1: Rollen von ChatGPT - Quelle Grafik: (Sabzalieva und Valentini 2023, S. 9)

Die in der Abbildung 1 dargestellten zehn Rollen der KI benötigen eine didaktische Begleitung durch die Lehrperson, die wiederum eine für sie **neue Rolle** erlernen und übernehmen muss. Im Falle von ChatGPT geht es somit nicht um ein Gegeneinander von KI-Chatbot und menschlichen Lehrkräften, sondern um eine neue Form der Kollaboration im Sinne eines konstruktiven Miteinanders von Mensch und Maschine, natürlich unter bestmöglicher Nutzung vorhandener Synergien. Das bedeutet, dass Lehrende diesen neuen Lehr- und Lernraum inhaltlich gestalten müssen. Sie schlüpfen in die Rolle von Architekten und Gestaltern des virtuellen und analogen Lernraums.

Auch die **Leitungsebenen in Schulen sowie Hochschulen** müssen in die Qualifizierungsmaßnahmen einbezogen werden, um die notwendigen Umsetzungsmaßnahmen auf operativer Ebene in ihrem Umfeld vorzunehmen. Dabei kommt es vor allem auf Schnelligkeit an, denn ein Auseinanderdriften der jeweiligen Entwicklungen im Bildungssystem versus Wirtschaftssystem gefährdet den Innovationsstandort Deutschland. Zu betonen ist, dass Weiterbildungs- und Qualifizierungsmaßnahmen an Schulen und Hochschulen sehr unterschiedlich gehandhabt werden. Diese Besonderheiten sind zu berücksichtigen.

Für beide Bereiche, d.h. für den Einsatz in Schulen und Hochschulen, müssen gleichermaßen **datenschutz- und rechtskonforme Einsatzszenarien** entwickelt und ethische Grundsätze eingehalten werden. Eine besondere Bedeutung erlangen DSGVO-konforme **deutsche bzw. europäische KI-Sprachmodelle** für den Einsatz in der Bildung mit Lizenz- bzw. Nutzungsmodellen, die im Bildungsbereich umsetzbar erscheinen. Die Bemühungen zum Aufbau derartiger Lösungen sollten von der **Bundesregierung mit Nachdruck** verfolgt und unterstützt werden, siehe die Initiative „Large European AI Models“ (LEAM) (<https://leam.ai/>) des KI-Bundesverbands.

Unterstützung der Leitungsebenen in Schulen und Hochschulen

Die Task-Force „KI-Bildung“ adressiert nicht nur die Lehrenden, sondern auch die Leitungsebenen in den jeweiligen Einrichtungen. Seit der Veröffentlichung von ChatGPT am 30. November letzten Jahres hat sich deren Überforderung im Umgang mit dieser digitalen Disruption sehr deutlich gezeigt. Einige Hochschulleitungen haben ChatGPT in Gänze verboten, andere Präsidien haben den fallweisen Einsatz und die Entscheidung darüber in die Fakultäten, Institute bis hin zu den einzelnen Lehrenden verlagert. Andere Präsidien haben bis heute überhaupt nicht reagiert. In den Justitiariaten der Hochschulen zeigt sich eine ähnliche Bandbreite an Reaktionen.

In der Folge haben sowohl Lehrende wie auch Studierende geäußert, sich orientierungslos und alleingelassen zu fühlen. Letztere werden im Falle eines Verbots sogar indirekt kriminalisiert, da ChatGPT inzwischen als ubiquitäres Tool zu bewerten ist. Mit der Freigabe der Schnittstelle durch OpenAI am 1. März 2023 ist ChatGPT in eine große Fülle von Software-Anwendungen eingeflossen, ohne dass den Anwender dieser Systeme der im Hintergrund liegende Systemaufruf von ChatGPT bewusst ist.

Eine besondere Bedeutung haben bereits in den letzten **Jahren KI-gestützte Schreibwerkzeuge und Digital Research Assistant Tools** erlangt, die im Bildungsbereich auch schon vor ChatGPT eine zunehmende Verbreitung gefunden haben. Gleichzeitig ist die Fülle der Angebote drastisch gestiegen, wie

Übersichtsplattformen am Beispiel von Futurepedia (<https://www.futurepedia.io/>) eindrucksvoll zeigen. Es ist offensichtlich, dass dieser Umstand bei vielen deutschen Schul- und Hochschulleitungen unbekannt ist, was wiederum als Indiz für den Bedarf an Aufklärung und Qualifizierung aller Stakeholder-Gruppen im Bildungsbereich gewertet werden muss (siehe hierzu auch die Zielsetzung des Virtuellen KI-Kompetenzzentrums „Schreiben lehren und lernen mit KI“ an der FH Kiel <https://www.vkkiwa.de/>).

Darüber hinaus haben Anbieter wie die Firma Microsoft bereits angekündigt, dass KI-gestützte Werkzeuge zur Texterstellung und -überarbeitung bald in deren Office-Produkte wie Textverarbeitungsprogramme integriert werden (Stallbaumer 2023). Damit werden generative KI-Werkzeuge omnipräsent als integrierte Lösungen in Softwareprodukte der Tech-Giganten, die von Milliarden Anwendern weltweit genutzt werden.

Der Stellenwert von IT in der Selbstverwaltung der Bildungsorganisationen

In Bezug auf die erforderlichen Qualifizierungsmaßnahmen der **Task-Force „KI-Bildung“** werden die Führungsebenen an Schulen wie auch Hochschulen eine zentrale Rolle übernehmen. Sie müssen diesen Rahmen schaffen und erkennen, dass nicht nur große Bereiche der Wirtschaft, sondern auch die Organisationen im Bildungsbereich in vielen Bereichen „IT-driven organization“ sind und ihren Organisationsaufbau anpassen müssen. Die Verantwortung für IT gehört analog zu Wirtschaftsunternehmen auch bei Schulen und Hochschulen auf die **Top-Management-Ebene** und muss wegen ihrer strategischen Bedeutung dort fest verankert sein. Sie erfordert im KI-Zeitalter dezidierte Verantwortlichkeiten und qualifiziertes Personal mit hoher IT-/KI-Expertise. Die finanziellen Mittel für diesen organisatorischen Anpassungsprozess müssen hoch priorisiert bereitgestellt werden.

Im weiteren Prozessverlauf sind regelmäßige interne und externe **AI Audits** für Hochschulen (Sabzalieva und Valentini 2023, S. 14) zu empfehlen, die alle relevanten Stakeholder einbeziehen und klare Handlungsempfehlungen für Hochschulleitungen formulieren. Die **Einbeziehung der Studierenden und ihrer Interessenvertreter** wie z.B. der Fachschaften und ASten und der damit verbundene partizipative Ansatz sind wichtige Erfolgsfaktoren für den „Change“ im KI-Zeitalter.

Rechtliche Aspekte

Rechtliche Bewertung des KI-gestützten Schreibens in der Hochschulbildung

Das Projektteam von KI:edu.nrw an der Ruhr-Universität Bochum um Dr. Peter Salden hat in Zusammenarbeit mit Prof. Dr. Thomas Hoeren von der Westfälischen Wilhelms-Universität Münster (WWU) eine umfassende juristische Bewertung grundlegender Rechtsfragen im Zusammenhang mit dem **Einsatz von KI-basierten Schreibwerkzeugen an Hochschulen** (Salden 2023) vorgelegt.

Bei der Frage der **Urheberschaft** kommt das Gutachten zu dem Schluss, dass eine Software wie ChatGPT nicht Urheber eines Textes sein kann. Die Urheberschaft gilt aber für Studierende beim Einsatz KI-gestützter Schreibwerkzeuge, wenn sie in erheblichem Maße ihre geistige Eigenleistung bei der Textproduktion zeigen. Wegen der fehlenden Trennschärfe verweisen die Gutachter aber auch hier auf eine Einzelfallentscheidung.

Das Gutachten betont, dass ein generelles **Verbot von Anwendungen wie ChatGPT an Hochschulen weder zielführend noch tatsächlich umsetzbar** sein dürfte. Vielmehr verweist es darauf, dass Hochschulen auf Basis ihrer individuellen prüfungsrechtlichen Regelungen definieren müssten, wann und unter welchen Voraussetzungen KI-Schreibwerkzeuge von Studierenden eingesetzt werden können (Salden 2023, S. 5).

Die Entwicklungen der Corona-Pandemie haben gezeigt, dass beispielsweise der Einsatz von Tools wie Zoom oder MS Teams erhebliche praktische Erleichterungen brachte, die sich zum Teil auch über die Pandemie hinaus erhalten und in der Lehre nachhaltig etabliert haben. Während der „Corona-Zeit“ waren Schulen und Hochschulen aber zugleich gezwungen, die mit diesen Tools verbundenen datenschutzrechtlichen Implikationen und die darauf aufbauenden Einschätzungen der Datenschutzbehörden zu beachten, was zum Teil - schon aufgrund teilweise divergierender Einschätzungen - für große Rechtsunsicherheit im Umgang mit diesen Werkzeugen sorgte. Es muss vermieden werden, dass sich ähnliche Umstände auch im Zusammenhang mit generativen KI-Lösungen wiederholen, die zu Hemmungen führen könnten, derartige KI-Tools von Anbietern außerhalb der EU überhaupt einzusetzen. Hier gilt es, pragmatische Lösungsansätze in Form eines Spagats zwischen weitestgehender rechtlicher Sicherheit und den Nutzungspotenzialen zu finden. Die praktische Erfahrung zeigt, dass es ohnehin bereits jetzt **an der Lebenswirklichkeit vorbeiginge, die Nutzung von Anwendungen wie ChatGPT zu untersagen**. Das Ziel der digitalen Bildung sollte vielmehr darin bestehen, die Digital- und Medienkompetenz der Lernenden im Umgang mit KI-Werkzeugen zu stärken und damit auch die Grundlagen für einen kompetenten und reflektierten Umgang mit den weiteren, bereits absehbaren schnelllebigen Entwicklungen zu legen.

KI-Fakes, Kennzeichnungspflichten und Transparenz

Die in der Presse in den letzten Wochen intensiv diskutierte KI-Fakes wie die vermeintliche Trump-Verhaftung oder der Papst im Designer-Daunenmantel haben uns als Gesellschaft stark für die Risiken generativer KI-Systeme sensibilisiert: Wir müssen als Medienrezipienten erkennen können, wo wir auf KI-generierten Content (sei es Text, Bild, Bewegtbild, Musik o.ä.) stoßen, der ohne Zuordnung zu einem Menschen und ohne dessen erkennbare Verantwortung (mit nachfolgender Haftung) für die Inhalte verbreitet wird. Das bedeutet eine **Kennzeichnungspflicht** für diese Form von KI-generierten Inhalten. Fakten und Fiktion müssen für die Mediennutzer unterscheidbar und einfach erkennbar sein. Deren Frage lautet: Was darf ich glauben, was ist „wahr“ bzw. „falsch“?

Der Content-Anbieter muss sich seiner Verantwortung stellen und KI-generierten Content vor der Veröffentlichung gewissenhaft prüfen. **Aufklärungsbedarf** herrscht bei einem Grundproblem generativer KI-Systeme, die wie ChatGPT auf KI-Sprachmodellen der Transformer-Architektur basieren. **Systeminhärente Merkmale** sind sogenannte Halluzinationen (frei erfundene Aussagen), Verzerrungen in Form vielfältiger Bias-Facetten durch die Trainingsdaten und das einseitige Ausrichten an menschlichen Präferenzen (Alignment). Für diesen Problemkomplex gibt es derzeit keine umfassenden Lösungsansätze, da derartige KI-Modelle mit probabilistischen Algorithmen systembedingt nicht auf Faktentreue trainiert werden können. Die Kopplung mit Suchmaschinen oder auch die Einbindung nachgelagerter Sekundärsysteme bieten vielversprechende Perspektiven. Vielversprechend sind neue Methoden wie AtMan (Deb

et al. 2023), das Lösungsansätze für „Explainable AI“ im Sinne von mehr Transparenz und Nachvollziehbarkeit für generative KI-Sprachmodelle auf Basis der Transformer-Architektur bietet.

KI-Governance und Zuverlässigkeit als wissenschaftlicher Wert im KI-Zeitalter

Für die Wissenschaft und Forschung ist die Zuverlässigkeit im Zeitalter Künstlicher Intelligenz im Zusammenspiel von Mensch und Maschine zu analysieren. Für die Beurteilung der Zuverlässigkeit im Sinne der Vertrauenswürdigkeit KI-gestützter Systeme sind die Datenqualität, die Transparenz, Erklärbarkeit und Widerspruchsfreiheit sowie die Verantwortung, Haftung und Systemvertrauen von besonderer Relevanz. Der gesellschaftliche Anspruch an KI-Systeme erfordert ein **neues Regelungssystem der KI-Governance** (Wilder et al. 2022). Eine KI-Governance wird nach aktueller Einschätzung zu einer Kennzeichnungspflicht der von Wissenschaftler*innen verwendeten KI-Systeme in ihren Publikationen führen, die bereits vielfach gefordert und praktiziert wird, und bei großen KI-Sprachmodellen zu **qualitätsgeprüften KI-Sprachmodellen und darauf basierenden Werkzeugen**. Wie die UNESCO-Empfehlung (Deutsche UNESCO-Kommission e.V. 2023) betont, sind auch die **Trainingsdaten** ein relevanter Aspekt der Analyse. Sie sollten im Idealfall öffentlich verfügbar und durch unabhängige Organe überprüfbar sein. Der TÜV-Verband hat bereits 2020 in einem Positionspapier zur Sicherheit KI-gestützter Anwendungen (TÜV-Verband 2020) seine Mitwirkung signalisiert und die Bedeutung unabhängiger Prüforganisationen betont.

Zusammenfassung

Abschließend möchte ich meine zentralen Empfehlungen zusammenfassen:

- Wir benötigen eine **Task-Force „KI-Bildung“**, die sich mit zielgruppengerechten Qualifizierungsangeboten an alle Lehrenden und Verantwortlichen im deutschen Bildungsbereich wendet.
- Diese Task-Force muss **zeitnah gebildet werden und zentral verankert** sein, um die erforderliche Schnelligkeit, Schlagkraft und Reichweite in der deutschen Bildungslandschaft zu erzielen.
- Wir benötigen dringend **DSGVO-konforme (deutsche bzw. europäische) KI-Sprachmodelle** für den Einsatz in der Bildung mit Lizenz- bzw. Nutzungsmodellen, die zum Bildungsbereich passen und praktikabel sind.
- Der an vielen Stellen noch **„rechtsfreie“ Raum** beim Einsatz generativer KI-Systeme und der Nutzung des KI-generierten Contents in dem kollaborativen Prozess von Mensch und Maschine muss dringend geklärt werden. Wir benötigen gesellschaftlich akzeptierte Eckpfeiler für ein **rechtskonformes und ethisch verantwortliches Handeln**.
- Last, but not least, sind diese Maßnahmen als **Teil einer systemweiten Vision und strategischen Prioritätensetzung von Digitalisierung** in unserer Gesellschaft zu verstehen. Deutschland muss ein **„chancen- und werteorientiertes digitales Mindset“** entwickeln. Bürokratische Hemmnisse müssen überwunden und eine Aufholjagd gestartet werden, um in Zeiten des beschleunigten Wandels und generativer KI-Systeme die Zukunftsfähigkeit und damit auch die Wettbewerbsfähigkeit Deutschlands sicherzustellen.

Literaturverzeichnis

- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T. & Zhang, Y. (2023, 22. März). *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. <https://arxiv.org/pdf/2303.12712>.
- Deb, M., Deiseroth, B., Weinbach, S., Schramowski, P. & Kersting, K. (2023, 19. Januar). *AtMan: Understanding Transformer Predictions Through Memory Efficient Attention Manipulation*. <https://arxiv.org/pdf/2301.08110>.
- Deutsche UNESCO-Kommission e.V. (Hrsg.). (2023). *UNESCO-Empfehlung zur Ethik Künstlicher Intelligenz. Handlungsansätze zur Implementierung in Deutschland*, Bonn. https://www.bundestag.de/resource/blob/942696/993ee22a6d04a87778e06f3c66a49dc8/20-18-107a-Dt_Unesco-data.pdf. Zugegriffen: 22. April 2023.
- Gates, B. (2023, 21. März). The Age of AI has begun. <https://www.gatesnotes.com/The-Age-of-AI-Has-Begun>. Zugegriffen: 23. April 2023.
- Hense, J. & Goertz, L. (März 2023). *Monitor Digitalisierung 360 Grad. Wo stehen die deutschen Hochschulen?* (Hochschulforum Digitalisierung, Hrsg.) (Arbeitspapier Nr. 68). https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_AP_68_Monitor_Digitalisierung.pdf. Zugegriffen: 21. April 2023.
- Maslej, N. e. a. (April 2023). *The AI Index 2023 Annual Report* (AI Index Steering Committee, Hrsg.). Stanford, CA: Institute for Human-Centered AI, Stanford University. https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf. Zugegriffen: 21. April 2023.
- Matthes, S. (2023, 21. April). Kommentar: Wir führen die falsche KI-Debatte. *Handelsblatt*. <https://www.handelsblatt.com/meinung/kommentare/kommentar-wir-fuehren-die-falsche-ki-debatte/29106118.html>. Zugegriffen: 22. April 2023.
- Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P. & Bernstein, M. S. (2023, 07. April). *Generative Agents: Interactive Simulacra of Human Behavior*. <https://arxiv.org/pdf/2304.03442>.
- Sabzalieva, E. & Valentini, A. (April 2023). *ChatGPT and Artificial Intelligence in higher education. Quick start guide* (UNESCO, Hrsg.) (Document code: ED/HE/IESALC/IP/2023/12). https://www.iesalc.unesco.org/wp-content/uploads/2023/04/ChatGPT-and-Artificial-Intelligence-in-higher-education-Quick-Start-guide_EN_FINAL.pdf. Zugegriffen: 21. April 2023.
- Salden, P. (2023). *Didaktische und rechtliche Perspektiven auf KI-gestütztes Schreiben in der Hochschulbildung*.
- Stallbaumer, C. (Microsoft, Hrsg.). (2023, 16. März). Introducing Microsoft 365 Copilot—A whole new way to work. <https://www.microsoft.com/en-us/microsoft-365/blog/2023/03/16/introducing-microsoft-365-copilot-a-whole-new-way-to-work/>. Zugegriffen: 23. April 2023.
- TÜV-Verband (Hrsg.). (August 2020). *Zur Sicherheit KI-gestützter Anwendungen. Eckpunkte eines europäischen Rechtsrahmens für den sicheren Einsatz von KI*, Berlin. <https://www.tuev-verband.de/positionspapiere/tuev-verband-zur-sicherheit-ki>. Zugegriffen: 22. April 2023.
- 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, Analysen, Reflexionen* (1. Auflage, S. 203–223). Darmstadt: wbg Academic in Wissenschaftliche Buchgesellschaft (wbg). https://files.wbg-wissenverbindet.de/Files/Article/ARTK_ZOA_1025976_0001.pdf. Zugegriffen: 25. Juni 2022.