

Large Language Models und ihre Potenziale im Bildungssystem

Impulspapier der Ständigen Wissenschaftlichen Kommission
der Kultusministerkonferenz





Über die Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK)

Die Ständige Wissenschaftliche Kommission (SWK) ist ein unabhängiges wissenschaftliches Beratungsgremium der Kultusministerkonferenz. Ihr gehören 16 Bildungsforscher:innen aus unterschiedlichen Disziplinen an. Die SWK berät die Länder zu bildungspolitischen Fragen. Sie identifiziert bestehende Herausforderungen und gibt evidenzbasierte Empfehlungen für deren Lösung. Dabei nimmt die Kommission eine interdisziplinäre, längerfristige und systemische Perspektive ein. Die SWK bindet externe Sachverständige in ihre Arbeit ein und hört Vertreter:innen aus Politik, Verwaltung, Bildungspraxis und Zivilgesellschaft an. Eine Geschäftsstelle unterstützt die Kommission bei ihrer Arbeit.

Mitglieder

Prof. Dr. Olaf Köller (Co-Vorsitzender der SWK)

Geschäftsführender Wissenschaftlicher Direktor des IPN

IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik

Prof. Dr. Felicitas Thiel (Co-Vorsitzende der SWK)

Professorin für Schulpädagogik und Schulentwicklungsforschung

Freie Universität Berlin

Prof. Dr. Isabell van Ackeren-Mindl

Professorin für Bildungssystem- und Schulentwicklungsforschung

Universität Duisburg-Essen

Prof. Dr. Yvonne Anders

Professorin für frühkindliche Bildung und Erziehung

Universität Bamberg

Prof. Dr. Michael Becker-Mrotzek

Professor für deutsche Sprache und ihre Didaktik

Direktor des Mercator-Instituts für Sprachförderung und Deutsch als Zweitsprache

Universität zu Köln

Prof. Dr. Ulrike Cress

Direktorin des IWM

Leibniz-Institut für Wissensmedien (IWM), Tübingen

Prof. Dr. Claudia Diehl

Professorin für Mikrosoziologie

Universität Konstanz



Prof. Dr. Thilo Kleickmann

Leitung der Abteilung Schulpädagogik
Christian-Albrechts Universität zu Kiel

Prof. Dr. Birgit Lütje-Klose

Professorin für Schulische Inklusion und sonderpädagogische Professionalität
Universität Bielefeld

Prof. Dr. Susanne Prediger

Professorin für Mathematikdidaktik und fachbezogene Professionalisierungsforschung
Technische Universität Dortmund

Prof. Dr. Susan Seeber

Professorin für Wirtschaftspädagogik und Personalentwicklung
Georg-August-Universität Göttingen

Prof. Dr. Birgit Ziegler

Professorin für Berufspädagogik und Berufsbildungsforschung
Technische Universität Darmstadt

Ständige Mitglieder:

Prof. Dr. Doris Lewalter

Vorstandsvorsitzende Zentrum für internationale Vergleichsstudien (ZIB)
Technische Universität München

Prof. Dr. Kai Maaz

Sprecher Autorengruppe Bildungsberichterstattung
DIPF – Leibniz-Institut für Bildungsforschung und Bildungsinformation

Prof. Dr. Christian Reintjes

Vorsitzender wissenschaftlicher Beirat der Steuerungsgruppe "Feststellung der Leistungsfähigkeit des Bildungswesens im internationalen Vergleich"
Universität Osnabrück

Prof. Dr. Petra Stanat

Wissenschaftlicher Vorstand Institut zur Qualitätsentwicklung im Bildungswesen (IOB)
Humboldt-Universität Berlin



Zentrale Diskussionspunkte

Das Bildungssystem in Deutschland steht aktuell vor der Aufgabe, das Potenzial generativer KI-Technologien wie Large Language Models (LLM) zu nutzen, dabei ihre Grenzen klar zu erkennen und verantwortungsbewusst mit ihren Beschränkungen umzugehen. Die Ständige Wissenschaftliche Kommission möchte mit diesem Impulspapier einen Diskussionsbeitrag leisten und Anregungen geben für den schulischen Einsatz sowie für Forschungs- und Entwicklungsaufgaben, um den lernförderlichen Einsatz von LLM sicherzustellen.

- **Übergangsphase zur systematischen Erprobung anlegen:** Kurzfristig ist eine Übergangsphase notwendig, in der eine systematische Erprobung von LLM bei offener Fehlerkultur stattfindet, in enger Zusammenarbeit von Praxis und Wissenschaft, insbesondere der Fachdidaktiken, Medienpsychologie, Medienpädagogik und Landesinstitute. Zudem bedarf es eines raschen, systematischen und wissenschaftsunterstützten Ausbaus der Fortbildungsangebote für Lehrkräfte und Lehrende an Hochschulen.
- **Domänenspezifische Tools entwickeln:** LLM müssen mit qualitativ hochwertigen, fachspezifischen Daten trainiert werden, um für den Bildungskontext zuverlässig und valide zu funktionieren. Befunde aus der Lehr-Lernforschung, z. B. zu Intelligenten Tutoriellen Systemen (ITS), zu Feedback, Schreib- und Kollaborationsprozessen, müssen auf die Entwicklung spezifischer LLM für den Bildungskontext adaptiert werden.
- **LLM ab der Sekundarstufe einsetzen:** Der Einsatz von LLM sollte angepasst an die jeweiligen Bildungsetappen erfolgen. In der Grundschule und zu Beginn der Sekundarstufe I scheint ein weitgehender Verzicht auf LLM und stattdessen ein Fokus auf den systematischen Aufbau von basalen Lese- und Schreibkompetenzen angemessen. Bis zum Ende der Sekundarstufe I sollte eine zunehmende Nutzung von LLM-Tools zur Erstellung und Überarbeitung von Texten erfolgen, während weiterhin Texte auch ohne LLM (hilfsmittelfrei) erstellt werden.
- **Prüfungskultur verändern:** In prozessorientierten Prüfungsformaten sollte die versierte Koaktivität mit LLM als Lernziel geübt und geprüft werden. Bei der Bewertung können KI-basierte Tools Lehrkräfte unterstützen, die Verantwortung für die Bewertung muss in menschlicher Hand bleiben. Ergänzend sind hilfsmittelfreie Prüfungsformate beizubehalten sowie Prüfungsformate zu entwickeln, bei denen das *Prompting* im Vordergrund steht.
- **Rahmenbedingungen schaffen:** Staatliche Organisationen sollten sicherstellen, dass alle Lernenden und Lehrenden in Bildungseinrichtungen einen (kostengünstigen oder kostenfreien) Zugriff auf KI-Systeme haben. Neben aktuellen LLM-Ansätzen sollte auch die Weiterentwicklung und Nutzung KI-basierter Ansätze zur individuellen Förderung systematisch gefördert werden, deren Lernwirksamkeit bereits belegt werden konnte.
- **LLM in Lernplattformen integrieren und Commons-Lösungen entwickeln:** Technologische, ethische und rechtliche Probleme können die Rechtmäßigkeit eines unterrichtlichen Einsatzes von LLM in Frage stellen. Aufgabe der Bildungspolitik muss es daher sein, LLM in geeignete Lernplattformen zu integrieren oder Commons-Alternativen zu entwickeln, deren Funktionsumfang und Verzerrungen bekannt sind und frei beforscht werden können. Für rechtliche und ethische Grauzonen sind klare Regulierungen zu treffen.



Inhaltsverzeichnis

1. Einführung	7
2. Charakteristika und Entwicklungstrends von Large Language Models	8
3. Potenziale und Grenzen der aktuellen LLM-Tools für Lehren und Lernen	10
3.1 Nutzung von LLM bei Texterstellung und Recherchen	10
3.2 Binnen- und Fachdifferenzierung, adaptives Lernen und individuelles Feedback	11
3.3 Erstellung und Nutzung von Lernmaterial	12
4. Notwendige Voraussetzungen zur kompetenten Nutzung von LLM	13
4.1 Technologiebezogenes Wissen und Reflexion	13
4.2 Quellenbewertung, -prüfung und -reflexion	13
4.3 Inhaltliches Wissen und <i>Prompt-Tuning</i>	14
4.4 Selbstregulationskompetenzen	14
4.5 Mediendidaktische Kompetenzen bei Lehrenden	14
5. Herausforderungen	15
5.1 Rechtliche Aspekte – Urheberrecht und Datenschutz	15
5.2 Auswirkungen auf die Prüfungskultur	16
5.3 Leistungsscheren zwischen Gruppen	17
5.4 Qualitätssicherung	18
6. Schlussfolgerungen	18





1. Einführung

Mit der kostenfreien Bereitstellung von GPT-3 hat das Unternehmen OpenAI im November 2022 weltweit für Furore gesorgt und die Diskussionen über Chancen und Risiken von künstlicher Intelligenz (KI) und insbesondere von Large Language Models (LLM) im Bildungskontext beflügelt. Innerhalb kürzester Zeit haben sich Millionen von Nutzer:innen registriert und das System u. a. für die Textproduktion oder die Generierung von Programmiercode genutzt. Aktuelle Schätzungen gehen davon aus, dass mindestens 20 Prozent der Schüler:innen in Deutschland ChatGPT für die Textproduktion und -übersetzung nutzen (Füller, 2023). Valide Zahlen, in welchem Umfang Lehrkräfte KI-Tools nutzen, liegen noch nicht vor.

Für kaum einen anderen Bereich wurden Potenziale und Risiken dieser Tools so massiv diskutiert wie für den Bildungsbereich (Fütterer et al., 2023). Während die Diskussion zunächst von Fragen zur Einschränkung von LLM in Schulen und anderen Bildungseinrichtungen dominiert war, ist sie inzwischen differenzierter und widmet sich auch den Potenzialen und Veränderungen im Bildungswesen. Gegenüber anderen KI-Tools, welche die Bildungspläne spezifischer Fächer abbilden, scheint das Potenzial von LLM wie ChatGPT, die auf einer breiten, unspezifischen Datenbasis trainiert wurden, besonders groß und disruptiv zu sein. Ersten Umfragen zufolge schreiben Lehrkräfte LLM in allen Unterrichtsfächern Einsatzpotenzial zu (bildung.digital, 2023).

Die Länder haben schnell auf die neuen Tools reagiert und bieten Handreichungen und Fortbildungen im Unterrichtszusammenhang an (z. B. Hessisches Kultusministerium, 2023; Landesinstitut für Schulqualität und Lehrerbildung - Sachsen-Anhalt, 2023; Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2023). Erste Länder ermöglichen Lehrkräften einen datenschutzkonformen Zugang zu ChatGPT (z. B. Sachsen-Anhalt, Mecklenburg-Vorpommern, Rheinland-Pfalz). Ebenso stellen erste Hochschulen ihren Studierenden datenschutzkonforme Zugänge zur Verfügung (z. B. HAW München). Außerdem hat die Kultusministerkonferenz im Januar 2023 eine Arbeitsgruppe zur Künstlichen Intelligenz eingerichtet, um relevante Themen für schulische und hochschulische Kontexte zu identifizieren und Vorhaben untereinander abzustimmen.

Im deutschen Bildungs- und Wissenschaftssystem werden bislang die Chancen und Risiken von ChatGPT und anderen LLM vor allem im Zusammenhang mit Textproduktionen diskutiert (z. B. Salden & Leschke, 2023). Im Kontext Schule steht darüber hinaus im Fokus, ob sich Unterrichtsziele (z. B. Verschiebung des Fokus vom Schriftlichen hin zum Mündlichen) verändern müssen und die informationsbezogene Grundbildung (*ICT-Literacy*) bzw. die Medienkompetenz um KI-Aspekte angereichert werden sollte, damit Schüler:innen kompetent mit LLM umgehen können. Unsicherheiten herrschen auch im Hinblick auf Prüfungsformate und wie dringend Prüfungsanforderungen durch neue Rahmenvorgaben aktualisiert werden müssen. Darüber hinaus werden lizenzrechtliche und datenschutzrelevante Aspekte erörtert. In all diesen Diskussionen kommt zu kurz, dass digitale Medien bereits seit langem Unterstützungspotenziale für sprachproduktive Leistungen bieten. Textverarbeitungsprogramme beispielsweise vereinfachen das nicht-lineare Schreiben und die Revision von Texten per *Copy and Paste* erheblich. Der Aspekt, dass LLM nicht nur Lernende unterstützen, sondern auch Lehrende, wird in Diskussionen unterschätzt. Für Lehrende bieten LLM Poten-



ziale in der Entwicklung und Anpassung von Unterrichtsmaterial sowie bei der summativen und formativen Beurteilung der Schüler:innen.

LLM stellen keineswegs die erste Technologie dar, welche die Bildungspraxis verändert. Die Einführung von Taschenrechnern, Computern oder des Internets hatte Auswirkungen auf Lehr-Lernprozesse, ohne dass grundlegende Ziele des Fachunterrichts in Schule und Hochschule aufgegeben wurden. Bildungsforschung und Lernpsychologie haben Modelle entwickelt, die beschreiben, ob und unter welchen Umständen technologische Innovationen lernförderlich wirken. Diese Modelle sind auf LLM und deren zukünftigen Weiterentwicklungen übertragbar.

Auch wenn es noch keine gesicherten empirischen Befunde darüber gibt, wie sich LLM auf Lehr-Lernprozesse auswirken werden, braucht es eine Orientierung, wie LLM in der Bildungspraxis zu verorten sind: Welche Potenziale bieten sie? Welche Beschränkungen sind zu berücksichtigen und welche Aufgaben stellen sich mittel- bis langfristig für Praxis und Wissenschaft?

Die Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK) möchte mit diesem Impulspapier einen Diskussionsbeitrag leisten und Anregungen geben, wie LLM zukünftig in Schulen genutzt werden können, wobei ein Großteil der Anregungen auf den Hochschulkontext übertragbar ist. Ein besonderes Augenmerk soll dabei auf den weiteren Forschungs- und Entwicklungsanstrengungen liegen, um den gewinnbringenden Einsatz der LLM in Lehr-Lernsituationen sicherzustellen. Damit knüpft die SWK an ihre Empfehlungen zur Digitalisierung im Bildungssystem und der systematischen Integration und Erforschung digitalisierungsbezogener Inhalte in Schule und Lehrkräftebildung an (Ständige Wissenschaftliche Kommission der Kultusministerkonferenz [SWK], 2022). Ein sinnvoller Umgang wird voraussichtlich sehr fachspezifisch geprägt sein. Damit adressiert dieses Impulspapier insbesondere die Fachdidaktiken, denn es wird vor allem die Erprobung in Schulen mit fachdidaktischer Begleitung und Evaluierung sein, über die schnell gesicherte Erkenntnisse zu Einsatzmöglichkeiten und Risiken erreicht werden können.

2. Charakteristika und Entwicklungstrends von Large Language Models

Generative KI kann auf Basis ihres Trainings alle Arten von Texten erstellen, die menschengemachten Inhalten ähneln (vgl. Albrecht, 2023, S. 19–26). Prominentes Beispiel hierfür sind große Sprachmodelle (LLM). Sie werden auf Basis immenser Datenmengen daraufhin trainiert, das nächste Wort vorherzusagen (Brown et al., 2020). Sie kodieren somit wahrscheinliche Wortreihenfolgen im Kontext, haben aber kein über die sprachliche Wahrscheinlichkeit von Wortabfolgen hinausgehendes Modell von *Bedeutung* oder *Weltwissen*.

Die Entwicklung der bekannten Modelle ist vor allem durch große private Unternehmen erfolgt (Ouyang et al., 2022). Zu den aktuell (Stand Dezember 2023) populärsten Modellen gehören die GPT-Reihe von OpenAI, die LaMDA/PaLM/Bard-Modelle von Google Research und die LLaMA-Serie von Meta. *Chatbots* wie z. B. ChatGPT basieren auf den Sprachmodellen. Sie sind auf Dialoge optimiert und können große Mengen an



Kontextdaten verarbeiten. Sie reagieren auf Spracheingabe (*Prompts*) und ermöglichen eine kohärente Kommunikation mit den Nutzer:innen. Über diese *Prompts* können Hinweise zur Art der Antwort oder einzubeziehenden Informationsquellen gegeben werden. Die Qualität der Antworten hängt mit der Qualität der *Prompts* zusammen, die möglichst exakte Anforderungen an das Textprodukt definieren sollen (*Prompt-Tuning*).

Während in den ersten Versionen natürliche Sprache als Auslöser diente, können multimodale Modelle (*Large Multimodal Models*) auch andere Datentypen verarbeiten und z. B. Texte, Bilder oder Programmcode integrieren (S. Wu et al., 2023). Der Entwicklungstrend geht zu generativen (Multi-)KI-Agentensystemen, bei denen Software-Agenten individuell konfiguriert werden, so dass sie untereinander interagieren bzw. kommunizieren, um gemeinsam und ohne menschliche Steuerung an komplexen Aufgabenstellungen zu arbeiten (Q. Wu et al., 2023). Seit November 2023 ermöglicht Open AI die Erstellung und Veröffentlichung individuell erstellter GPTs für spezifische Zwecke, die über *Prompts* konfiguriert und dann gespeichert werden können (OpenAI, 2023a). Damit eröffnen sich neue Möglichkeiten für die Nutzung in Bildungskontexten.

Eine Beschränkung der aktuellen LLM-Tools stellen so genannte Halluzinationen dar (Rawte et al., 2023; Y. Zhang et al., 2023). Generierte Inhalte klingen plausibel, enthalten aber erfundene Sachverhalte, Zitationen (Frederking, 2023; Kühl, 2022) und Fehler bzw. widersprechen dem konventionalisierten Weltwissen (Altmann, 2022; Athaluri et al., 2023). Zudem erinnert sich eine LLM nicht inhaltlich an die von ihr erstellten Inhalte (wohl aber an die eingegebenen *Prompts*). Das bedeutet zum einen, dass einmal kreierte Inhalte nicht durch das LLM weiterbearbeitet werden können, sondern mit jedem *Prompt* neue Inhalte generiert werden, zum anderen, dass neu generierte Inhalte zuvor erstellte Inhalte in Frage stellen können. LLM sagen wahrscheinliche Sprachsequenzen vorher, sie haben keinen anderen Bezug zur Welt oder semantisches Wissen. Zudem sind LLM nur so gut wie ihre Trainingsdaten. Sie weisen erhebliche Schwächen auf, wenn ihr Anwendungsbereich sich über die Domäne der Trainingsdatensätze hinweg ausdehnt (Yadlowsky et al., 2023). Je nach Anbieter sind keine aktuellen Informationen in den zum Training genutzten Texten enthalten und es kann nicht auf spezifische fachliche Inhalte zugegriffen werden. Das Training der aktuellen LLM beinhaltet inzwischen allerdings einen Großteil der digital verfügbaren Textmenge, also Internetseiten und Nachrichtentexte ebenso wie Literatur und andere fiktive Texte, Sachtexte unterschiedlichster Qualität und Meinungsäußerungen bis zu absichtlicher Desinformation. Im Bildungskontext ist die Frage nach der Datengrundlage und deren transparenter Veröffentlichung besonders relevant. Trotz intensiver Forschungsanstrengungen zu Erklärbarkeit und Zuverlässigkeit (Y. Wu et al., 2023) bleibt eine zentrale Herausforderung, dass LLM auf die Produktion wohlgeformter Sprache optimiert sind, so dass letztlich die Nutzer:innen die Kompetenz zur Bewertung, Korrektheit und Vertrauenswürdigkeit der Inhalte besitzen müssen.

Ein mit den Halluzinationen verwandtes Resultat der Funktionsweise von LLM sind *Biases*, die systematische Verzerrung von Informationen (Navigli et al., 2023), die sich aus der Auswahl der Trainingsdaten ergibt (Gombert et al., 2023). Sie können eine Vielzahl von Dimensionen betreffen, wie z. B. die moralische Ausrichtung (Schramowski et al., 2022), Religion, Gender, Ethnie, Beruf (Nadeem et al., 2021) oder die politische Ausrichtung (Feng et al., 2023; Rozado, 2023). Unterschiedliche LLM können etwa verschiedene politische Grundhaltungen einnehmen (Feng et al., 2023) oder Vorurteile reproduzieren (z. B. Bergener et al., 2023). Eine zusätzliche problematische Dimension ergibt sich aus der Tatsache, dass die hinter den LLM



stehenden Unternehmen die Möglichkeit haben, die Kriterien für die Generierung und Freigabe von Inhalten zu bestimmen. Diese Kontrolle birgt das Risiko einer selektiven Informationsfilterung, die durch die geschäftlichen oder ideologischen Präferenzen dieser Unternehmen geleitet wird. Eine derartige Kontrolle kann zu einer subtilen Form der Informationssteuerung führen, die bestimmte Perspektiven bevorzugt oder unterdrückt. Dies kann die Glaubwürdigkeit und Neutralität von LLMs wesentlich in Frage stellen und die Gefahr einer einseitigen Beeinflussung des öffentlichen Diskurses bergen (Nuxoll, 2023).

3. Potenziale und Grenzen der aktuellen LLM-Tools für Lehren und Lernen

Die Lehr- und Lernforschung zeigt, dass Lernen nur dann erfolgreich stattfindet, wenn Lernende kognitiv aktiviert und tiefe Informationsverarbeitungsprozesse (Elaborationen) gefördert werden. Rezeptive Verarbeitung allein ist nicht zielführend, sondern muss durch produktive und kollaborative Verarbeitung ergänzt werden (Chi & Wylie, 2014). LLM bieten besonderes Potenzial, kognitive Aktivierung, Elaboration und Kollaboration zu fördern und damit lernförderlich zu wirken (Kasneci et al., 2023). Gleichzeitig bergen sie durch ihre leichte Bedienung und die (zumindest oberflächlich) hohe sprachliche Qualität der Ergebnisse die Gefahr, dass Lernende fertige Produkte erstellen lassen, deren inhaltliche Qualität sie nicht einschätzen können, und dabei die eigentlich lernförderlichen Aktivitäten auslagern. Motivationsverluste sind zu befürchten, wenn Lernende den Eindruck haben, Aktivitäten ausüben zu müssen, die eine KI übernehmen könnte.

3.1 Nutzung von LLM bei Texterstellung und Recherchen

LLM dienen als Textgeneratoren und Informationsquelle und können Schreibtätigkeiten sowie rudimentäre Recherchen unterstützen und beschleunigen. Sie bringen eine neue Qualität technologischer Unterstützung im Schreibprozess, da sie im Vergleich zu bisherigen digitalen Tools vollautomatisiert Texte produzieren können (vgl. Steinhoff, 2023). In kreativen Lernszenarien können sie genutzt werden, um Ideen zu generieren und die Gestaltung literarischer oder künstlerischer Projekte zu unterstützen. Allerdings ist die eigene Informationssuche und –bewertung sowie die Produktion von Text für Lernende ein wichtiger elaborativer Prozess. Bei Nutzung von LLM zur Informationssuche ist die Bewertung systematisch erschwert oder unmöglich, da die Quelle der Informationen oft nicht bekannt oder, sofern eine Quelle angegeben ist, diese oft frei erfunden ist (z. B. Weiser, 2023). Um lernförderlich zu sein, sollte daher dieser Prozess nicht vollständig auf LLM ausgelagert werden. Sinnvoll erscheint eine versierte Koaktivität, in der die Lernenden die KI als Denkwerkzeug zielgerichtet nutzen (Cress & Kimmerle, 2023). Neben LLM können dabei auch andere KI-basierte Ansätze eine effektive Form der Unterstützung bieten (Weiss et al., 2023). Im Gegensatz zu den aktuellen LLM Ansätzen existieren für den Einsatz Intelligenter Tutorieller Systeme (ITS) im Schulkontext bereits Belege für Lernwirksamkeit (Meurers et al., 2019).

Mit Blick auf den schulischen Einsatz von LLM werden primär Fragen der Unterstützung bei der Textproduktion erörtert. Um LLM tatsächlich gewinnbringend im Schreibprozess zu nutzen, müssen die fachdi-



daktischen und lernpsychologischen Forschungsergebnisse zu Schreibkompetenz und ihrem Erwerb berücksichtigt werden (Becker-Mrotzek, 2022). In der Forschung besteht Einigkeit, dass die Schreibkompetenz von Schüler:innen primär in der Schule entwickelt wird, beginnend mit dem Schriftspracherwerb in der Grundschule, gemeinsam mit dem Lesen. Im Verlauf der weiteren Schulkarriere ist eine kontinuierliche Weiterentwicklung der Schreibkompetenz zu beobachten: Während Schüler:innen zu Beginn der Sekundarstufe I große Schwierigkeiten haben, den Schreibprozess mit seinen vier Komponenten *Planen*, *Formulieren*, *Verschreiben* und *Revidieren* zielführend in ein konzises Produkt zu übertragen, gelingt dies bei den meisten Schüler:innen gegen Ende der Sekundarstufe I deutlich besser. Das Arbeitsgedächtnis wird dann durch die Entwicklung von Schreibroutinen entlastet, so dass hierarchiehöhere Fertigkeiten und Fähigkeiten (Planen und Revidieren) eingesetzt werden können (Becker-Mrotzek & Böttcher, 2020). Diese hierarchiehöheren Fertigkeiten und Fähigkeiten sind nötig, um Texte (auch LLM generiert) beurteilen zu können. Übernehmen LLM frühzeitig die Textproduktion, wird diese Kompetenzentwicklung behindert und im ungünstigsten Fall unterbunden.¹

Mit dem Schreiben (z. B. Zusammenfassungen von Texten) sind Transformationsprozesse verbunden, die das tiefere Verständnis fachlicher Inhalte und damit erfolgreicherer fachliches Lernen ermöglichen. Deshalb können durch (epistemisches) Schreiben fachliches Wissen und fachliche Kompetenzen in den Sachfächern und sprachlichen Fächern aufgebaut werden (Graham et al., 2020). Metaanalysen zeigen mittlere Effekte von Schreibaktivitäten auf den Wissens- und Kompetenzerwerb in unterschiedlichen Fächern und Jahrgangsstufen (Graham et al., 2020). Wird hier Textproduktion durch LLM ersetzt, unterminiert dies möglicherweise fachbezogene Lernprozesse.

3.2 Binnen- und Fachdifferenzierung, adaptives Lernen und individuelles Feedback

LLM können Lehrkräfte bei der Erstellung adaptiver Unterrichtskonzepte zur Binnendifferenzierung und Individualisierung unterstützen (Pöler, 2023). Durch spezifische Anfragen (*Prompt-Tuning*) lassen sich individuelle Anpassungen von Lehr-Lerngelegenheiten an die Interessen von Lernenden vornehmen. Auf LLM basierende *Chatbots* können Schüler:innen interaktiv Rückmeldungen geben und auch außerhalb des regulären Unterrichtsrahmens Lehr-Lernprozesse unterstützen (Beck & Levine, 2023; Seßler et al., 2023). Bei Verständnisproblemen können Folgeprompts für Erklärungen für leichtere Sprache genutzt werden. Zur Unterstützung der Produktion von Texten können Prompts Gliederungs-, Aufbau- und Ausgestaltungsentwürfe generieren. Schüler:innen können Feedback zu ihren Texten hinsichtlich orthografischer Korrektheit und Textkohärenz erhalten. Allerdings zeigen aktuelle Analysen der Feedbacktypen und der Qualität des Feedbacks, dass die Rückmeldungen allgemein gehalten sind, oftmals nur wenig konkrete Verbesserungsvorschläge enthalten (Yoon et al., 2023) und die Qualität des Feedbacks erheblich variiert (Wang & Demczyk, 2023). Es gibt bislang nur wenige empirische Studien dazu, wie Lernende mit KI-generiertem Feedback

¹ Ausgehend von der Stellungnahme des schwedischen Karolinska-Instituts zur Digitalisierungsstrategie der nationalen schwedischen Bildungsagentur werden die Effekte digitaler Medien auf das Lernen teilweise kritisch diskutiert (Schmoll, 2023).



umgehen (Woitt et al., 2023). In diesem Zusammenhang wird auch zu untersuchen sein, wann eine inhaltliche oder sprachliche Unterstützung durch LLM Lernprozesse unterstützt.

Eine adaptive, individuelle Förderung setzt prinzipiell voraus, dass, basierend auf einer Diagnose der aktuellen Leistungen und mit Blick auf konkrete Lernziele, passende Lernangebote gemacht werden, bei deren Bearbeitung Lernende unterstützt werden können. Zu diesem Zweck beruhen ITS auf expliziten Modellen von Lernenden, der Lerndomäne und der pädagogischen und fachdidaktischen Optionen (Meurers et al., 2018). Dies setzt bislang eine aufwändige Modellierung der zu erwerbenden Kompetenzdomäne voraus, ebenso wie die Entwicklung adaptiver Lernpfade und die Formulierung potenzieller Rückmeldungen, die für die intendierte Zielgruppe verständlich und lernförderlich sind. Auch wenn LLM in einer Interaktion mit entsprechenden *Prompts* scheinbar in die Rolle eines Lernbegleiters versetzt werden können, ist zu bedenken, dass sie dabei bisher weder individuelle Lernende, ihre Kompetenzen und Misskonzepte modellieren und die curricularen Lernziele eines Schulfachs kennen noch die pädagogischen und fachdidaktischen Schritte zum Erreichen dieser Ziele.

Derzeit ist noch offen, wie die Stärken von LLM und anderen KI-basierten Ansätzen in ITS-Anwendungen effektiv kombiniert werden können. Einerseits sollte für einen breiten Einsatz in vielen Schulfächern der explizite Modellierungsaufwand reduziert und Interaktion anhand natürlicher Sprache durch LLM unterstützt werden. Andererseits ist die Bereitstellung von Lerngelegenheiten mit individualisierter Diagnostik und konkreten Zielen aus den Bildungsplänen unabdingbar, um das Lernen auf individuell adaptiven Lernpfade effektiv unterstützen zu können.

3.3 Erstellung und Nutzung von Lernmaterial

LLM bieten pädagogischen Fachkräften zahlreiche Unterstützungsmöglichkeiten bei der Konzeption von Lehr-Lernmaterialien (Küchemann et al., 2023) oder Unterrichtsplanungen (Wang & Demszyk, 2023). Zeitaufwendige Prozesse der Inhaltserstellung, wie die Bewertung relevanter Informationen bei Erstellung von Aufgaben (z. B. Xiao et al., 2023), können durch den Einsatz von LLM vereinfacht werden. Lehrkräfte können auf Basis von Aufgaben Wissenstests für spezifische Lerneinheiten (z. B. mit unterschiedlichem Schwierigkeitsgrad) erstellen und angepasst an Lernfortschritte modifizieren. Weiterhin eröffnen LLM Potenziale für innovative Lehransätze, wie etwa auf Lehrinhalten basierende Lernspiele (Chu & Liu, 2023). LLM bieten auch die Möglichkeit, Unterricht jenseits der Ebene von geschriebenen Texten weiterzuentwickeln, z. B. Einsätze in der Musik (Agostinelli et al., 2023) oder der Kunst (OpenAI, 2023b). Darüber hinaus haben generative Modelle für Videos (Stengel et al., 2023) das Potenzial, benachteiligte Gruppen stärker zu berücksichtigen – z. B. durch Avatare zur Übersetzung in Gebärdensprache oder zur Sichtbarmachung marginalisierter Gruppen. Bei all diesen Einsatzbereichen bleibt es angesichts der Fehleranfälligkeit der Ergebnisse essentiell, dass Lehrkräfte eine regulative Funktion ausüben (Ausat et al., 2023).

Schließlich können LLM Lehrkräfte bei der summativen und formativen Rückmeldung entlasten, indem sie Feedback zu Rechtschreibung, Grammatik, Textkohärenz, Argumentationsstruktur und inhaltlicher Qualität von Texten geben (z. B. Bewersdorff et al., 2023; Fang et al., 2023; Seßler et al., 2023; H. Wu et al., 2023;). Dabei ist eine Ko-Konstruktion zwischen LLM und Lehrkraft unerlässlich, um Fehler der LLM zu



erkennen und zu korrigieren. Die Bereitstellung von Musterlösungen für das genutzte LLM durch die Lehrkraft kann die Qualität des Feedbacks verbessern. Das Potenzial bei der Unterstützung der Bewertung bzw. der Generierung von Rückmeldungen zu Aufgaben bzw. Prüfungsleistungen ist hoch (Zesch et al., 2023). Allerdings ist zusätzliche Forschung und Entwicklung nötig, da die bisher entwickelten Systeme noch zu Überkorrekturen neigen und bestimmte Fehlertypen nicht erkennen.

Zusammenfassend können LLM die didaktischen Fachkenntnisse der Lehrkräfte und das fachdidaktisch fundierte Design von Lernumgebungen nicht ersetzen. Die KI kann bei der Bearbeitung didaktischer Herausforderungen unterstützen, sie bedarf aber der didaktisch treffsicheren Einbettung in die Lehr-Lernsituation.

4. Notwendige Voraussetzungen zur kompetenten Nutzung von LLM

4.1 Technologiebezogenes Wissen und Reflexion

Erst wenn Lernende bereichsspezifisches Vorwissen mitbringen, verstanden haben, wie LLM trainiert wurden, wie die automatisierte Textproduktion funktioniert, zu welchem Zweck LLM aufgesetzt wurden und welche Rolle *Prompts* für das Antwortverhalten von LLM spielen, können sie eine kompetente Einordnung der durch LLM gelieferten Informationen vornehmen und diese für das eigene Lernen produktiv nutzen. Auch Lehrkräfte benötigen diese Kompetenzen, weil sie diese Fähigkeit zur kritischen Reflexion sowohl vermitteln als auch selbst einsetzen müssen. So ist es wichtig zu verstehen, dass LLM zwar menschähnliche Texte als Antworten auf Fragen generieren können, dass sie aber weder Suchmaschinen (zur Suche nach vorhandenen Antworten auf diese Fragen) noch Maschinen mit menschlichen Eigenschaften (z. B. einem inneren mentalen Modell des Inhalts von Texten oder Empathie) sind. Diese Unterscheidungen werden zunehmend herausfordernder, wenn LLM-Systeme sich suchmaschinen- und menschenähnlicher verhalten. Studien zum Umgang mit Suchmaschinen zeigen, dass oftmals keine ausreichende Überprüfung der Ergebnisse stattfindet (Bräten et al., 2018; Kammerer & Brand-Gruwel, 2020). Bei der Nutzung von LLM ist anders als bei Suchmaschinenresultaten die Quelle oft unbekannt oder es werden erfundene, aber realistisch klingende Referenzen angegeben (Emsley, 2023). Daraus folgt eine zunehmende Bedeutung von Informationskompetenzen und *ICT-Literacy*.

4.2 Quellenbewertung, -prüfung und -reflexion

Spätestens seit der Verbreitung von *Social Media* sind Lernende daran gewöhnt, im Netz Informationen unterschiedlicher Güte zu finden, zunehmend auch Fake News und Propaganda. Die Forschung zeigt, dass Lernende dennoch häufig über unzureichende Fähigkeiten zur Informationsbewertung verfügen (vgl. Bräten et al., 2018; Schultheiß & Lewandowski, 2023). Sie überschätzen ihre Kompetenzen und können Quellen kaum nach Expertise, Motivation oder verborgenen Interessen beurteilen (Kammerer et al., 2015).



Solche Informations- und Medienkompetenzen werden durch LLM noch stärker gefordert: LLM-generierte Texte sind von menschengemachten Texten schwer unterscheidbar, sie erscheinen hochplausibel und sprachlich einwandfrei. Erste Studienergebnisse deuten darauf hin, dass KI-generierten Inhalten die gleiche Glaubwürdigkeit zugeschrieben wird wie menschengemachten (Huschens et al., 2023). LLM, die mit wenig Trainingsmaterial Stimmen klonen und authentisch wirkende Videos erzeugen können, verstärken diesen Eindruck. Diese generativen Bilderstellungsprogramme stellen Erziehungsberechtigte und Lehrkräfte vor neue, hochsensible Herausforderungen (Schneider, 2023). Kritisches Prüfen und Reflexion, mithin generell Medien- und Nachrichtenkompetenzen, sind damit Fähigkeiten, die nicht nur für den individuellen Lernprozess essenziell sind, sondern auch für die Gesellschaft als Ganzes.

4.3 Inhaltliches Wissen und *Prompt-Tuning*

Um mit LLM-Systemen sinnvoll lernen zu können, wird es zukünftig nicht ausreichen, präzise Fragen nach Inhalten stellen zu können, wie im Umgang mit Suchmaschinen. Vielmehr müssen in interaktiven Szenarien sinnvoll und an das Lernziel angepasste Sequenzen von *Prompts* eingesetzt werden. Dazu braucht es sowohl metakognitives Wissen über eigene Ziele, Verständnisprobleme und Kenntnislücken, als auch ein erhebliches Wissen über den Lerninhalt. Wissen wird durch LLM keinesfalls obsolet. Interaktives *Prompting* muss mit kritischer Informationsprüfung kombiniert werden, z. B. durch die Prüfung auf logische Schlüssigkeit, Objektivität und Validität von Informationen. *Prompt-Tuning* setzt damit kritisches und analytisches Denken voraus und Lernende benötigen breites fachliches Wissen, um die Glaub- und Vertrauenswürdigkeit der LLM-generierten Inhalte einschätzen können. Dies ist gerade im Bildungskontext eine besondere Herausforderung, wo bei schwächeren Lernenden diese Kompetenzen eher nicht vorausgesetzt werden können.

4.4 Selbstregulationskompetenzen

Beim Einsatz von Technologien, die Lernende selbstbestimmt nutzen können, nehmen Selbstregulationskompetenzen eine zentrale Rolle ein (Hoch & Fütterer, 2023). Für eine lernwirksame Nutzung von Technologien ist es entscheidend, dass diese Technologien konstruktiv und interaktiv genutzt werden (vgl. Chi & Wylie, 2014). Gleichzeitig sollte die lernrelevante kognitive Anstrengung erhöht und lernirrelevante Belastung (z. B. Ablenkung oder Passivität) vermieden werden (vgl. Paas & Sweller, 2019).

4.5 Mediendidaktische Kompetenzen bei Lehrenden

Wenn die oben genannten Kompetenzen für Lernende gelten, müssen Lehrende diese ebenso beherrschen. Im Bildungskontext (z. B. Lehrkräftefortbildungen) wird häufig auf das europäische Rahmenkompetenzmodell *DigCompEdu* (Redecker, 2017) Bezug genommen, das grundlegende Kompetenzen von Lehrenden zur Nutzung digitaler Medien zusammenfasst. Diese sind auch für die Nutzung von KI-Technologie und insbesondere LLM zentral. Es sind jedoch Anpassungen des Modells an die spezifischen Charakteristika dieser Technologie nötig. Zum Erwerb dieser Kompetenzen müssen Lehrende die Tools selbst ausprobie-



ren können. Darüber hinaus brauchen Lehrende allgemeine und fachspezifische mediendidaktische Kompetenzen. Sie müssen nicht nur bei Lernenden die allgemeine Medienkompetenz fördern, sondern wissen, inwiefern und unter welchen Bedingungen LLM sich für die Inhalte ihres Fachs lernwirksam einsetzen lassen und als Vorbild für die Lernenden fungieren. Die dynamische Entwicklung der Tools fordert die Lehrkräfte besonders. Dabei werden Lehrende je nach eigenem Vorwissen bzw. Interesse unterschiedlichen Unterstützungsbedarf haben, so dass differenzierte Angebote nötig sind. Hier besteht schneller Handlungsbedarf der Bildungsforschung, insbesondere der Fachdidaktiken, das Lehrpersonal in enger Kooperation mit den Landesinstituten für Lehrkräftebildung zu schulen und in einer realistisch-reflektierten Haltung zu stärken. Auch bereits bei weniger elaborierten digitalen Tools (z. B. Tabellenkalkulation), deren Nutzung seit 20 Jahren im Lehrplan vorgeschrieben ist, ist eine didaktisch angemessene Integration in den Unterricht bislang nicht flächendeckend von allen Lehrkräften realisiert. Dies zeigt die hohen Aus- und Fortbildungsbedarfe zu digitalen Anwendungen generell, die sich nun um LLM-Tools erweitern.

5. Herausforderungen

Mit der Nutzung von LLM-Tools sind neben den Anforderungen an die Kompetenzen bei Lehrenden und Lernenden weitere Herausforderungen verbunden. Diese Herausforderungen sind vor allem rechtlicher und ethischer Natur, betreffen aber auch Prüfungskultur und Qualitätssicherung. Darüber hinaus sind weitere Fragen mit der Nutzung von LLM verbunden, die in diesem Impulspapier angesichts des Fokus auf Lehr-Lernprozesse nicht weiter behandelt werden sollen: Politische und ethische Fragen, die sich im Kontext der Programmierung stellen (z. B. der Umgang mit *Clickworkern*), aber auch Fragen des Umweltschutzes und Energieverbrauchs. Eine flächenendeckende Einführung von LLM-Systemen in den Unterricht geht nach aktuellen Schätzungen mit erhöhten Anforderungen an Serverleistung, Netzkapazität und Energiebedarf einher.

5.1 Rechtliche Aspekte – Urheberrecht und Datenschutz

Ungeklärt sind bisher Fragen des Urheberrechtes. Insbesondere bei der Nutzung von literarischen und anderen künstlerischen Werken für das Training ist unklar, ob die Zustimmung der Urheber:in nötig ist (z. B. Alter & Harris, 2023). Aktuelle Beurteilungen sehen das Urheberrecht auch verletzt, wenn eine Lehrkraft Prüfungsleistungen direkt zur Bewertung in die KI einspeist und diese zum Training der KI weiterverwendet werden (Hoeren, 2023). Bei der Interaktion werden zudem Informationen vom Nutzer bzw. von der Nutzerin weitergegeben. Auch hier stellt sich die Frage des Urheberrechts, wenn diese Informationen zum weiteren Training genutzt werden. Ebenso ungeklärt ist der rechtliche Status KI-generierter Texte und Bilder: Damit das Urheberrecht greift, muss eine menschlich-gestalterische Tätigkeit vorliegen. Folglich kann eine KI keine Urheberschaft beanspruchen. Ein KI-Systeme nutzender Mensch kann ein Urheberrecht nur bei wesentlichen eigenen gestalterischen Entscheidungen beanspruchen (Salden & Leschke, 2023, S. 25-26).

Die Verarbeitung der durch Nutzer:innen eingegebenen Daten durch die KI berührt auch Fragen des Datenschutzes. Nutzt eine Lehrkraft z. B. die KI zur Unterstützung bei Bewertungen, könnten sensible Schülerda-



ten in die Datensammlung der KI eingespeist werden. An welchen Orten diese Daten gespeichert werden und welche Standards gelten, ist bei kommerziellen Anbietern oft nicht transparent bzw. entspricht häufig nicht den deutschen Datenschutzvorgaben. Darüber hinaus ist aktuell eine Bezahlung durch Datenerfassung etabliert: Kommerzielle Anbieter bieten kostengünstig Zugang zu ihren Produkten gegen die Preisgabe persönlicher Daten, die weiterverkauft oder anderweitig monetarisiert werden (D. Zhang, 2023). Eine Umsetzung solcher Verfahren bei KI/LLM-Tools schränkt eine schulische Nutzung weiter ein. Wichtig ist, dass bei der Nutzung von KI/LLM ethische und datenschutzrechtliche Grundsätze eingehalten werden (vgl. Hutson, 2023). In einem aktuellen systematischen Literaturüberblick zu KI im Unterricht wird deutlich, dass bei vielen KI-Systemen Datenschutzaspekte gegenwärtig keine Rolle spielen (Fütterer et al., in Vorbereitung).

5.2 Auswirkungen auf die Prüfungskultur

Die Nutzung von LLM beeinflusst die Prüfungskultur hinsichtlich der Lernziele und Kompetenzen der Lernenden, der Validität der Leistungsüberprüfung und der Kompetenzen seitens der Prüfenden. Angesichts der Schwierigkeit, KI-generierte Texte von menschlich erstellten Texten zu unterscheiden, argumentieren manche Autor:innen (z. B. Susnjak, 2022), dass traditionelle Prüfungsformate wie Hausarbeiten oder Hausaufgaben ihre Aussagekraft verlieren. Außerdem ist der rechtliche Status von mithilfe der KI erstellten Texten unklar: Aufgrund der fehlenden Autorenschaft und der Intransparenz der Datenbasis handelt es sich weder um ein Werk im urheberrechtlichen Sinne noch um eine zitierfähige Quelle (Nehlsen & Fleck, 2023). Diese Probleme bestehen jedoch nicht erst seit der Einführung von LLM. So konnten auch vorher durch Kopieren von im Internet verfügbaren Texten rasch neue Texte produziert werden. Zunehmend versierte Software, die Texte bei Plagiatsverdacht überprüfen können, bieten dagegen einen wirksamen Schutz. Aufgrund der Funktionsweise der LLM versagt Plagiatssoftware allerdings bei der Identifikation von durch LLM erstellte Texte. Die Deutsche Forschungsgemeinschaft (DFG, 2023) hat auf die Problematik reagiert und fordert ihre antragstellenden Wissenschaftler:innen auf, die Textteile, die mit LLM geschrieben wurden, anzugeben. In diesem Zusammenhang ist es plausibel, LLM auch bei Bachelor- und Masterarbeiten zuzulassen mit der Auflage, dass auf die Verwendung von LLM hingewiesen wird.

In Bildungskontexten könnten prozessorientiertere Prüfungsformate an Bedeutung gewinnen. Lehrkräfte können ihren Unterricht z. B. so organisieren, dass die eigentlichen Lern- und Verstehensleistungen im Unterricht erbracht werden und zuhause LLM-unterstützt Texte vorbereitet werden. Handlungsleitend für die Erstellung von Prüfungsformaten sollen fachdidaktische Erwägungen bleiben, unabhängig vom Einsatz der Hilfsmittel. Zudem kann (z. B. analog zur Nutzung des Taschenrechners im Fach Mathematik) in Prüfungen unterschieden werden zwischen hilfsmittelfreien Prüfungsteilen, in denen auf LLM nicht zurückgegriffen werden darf und Teilen, in denen sie genutzt werden können.

Der produktive und sichere Umgang mit KI bringt gleichzeitig neue Lernziele mit sich. Die LLM können Schreibprozesse übernehmen, aber die Steuerung im Prozess durch *Prompts* und die Interpretation und Reflexion der Schreibprodukte sind Aufgaben der Nutzer:innen. Da eine versierte Koaktivität mit LLM eine wichtige Zukunftskompetenz darstellen wird, müssen in den Fächern Prüfungsformate entwickelt werden, in denen Lernende diese Kompetenz unter Beweis stellen können und Aufgaben bereitgestellt werden, die sie darauf vorbereiten. Solche Aufgabenformate könnten projektartig und prozessorientiert angelegt sein



und besondere Anforderungen an das *Prompting* oder die kritische Reflexion des LLM-Outputs stellen. Sofern dies technisch und datenschutzrechtlich möglich ist, bietet sich z. B. für die Beurteilung der Bearbeitung solcher Aufgaben eine Auswertung der Chatverläufe an.

Diese Weiterentwicklungen in der Prüfungskultur betreffen auch die Kompetenzen der Prüfenden. LLM können Lehrkräfte bei der Erstellung von (binnendifferenzierenden) Prüfungsmaterialien und bei der Bewertung von Prüfungsleistungen unterstützen. Im ungünstigsten Fall, wenn Lernende und Prüfende KI nutzen, vergleichen zwei KI-Systeme ihre Ergebnisse miteinander. Lehrkräfte müssen daher sicherstellen, dass die Überprüfung der Leistungserbringung valide und reliabel die Leistung der Lernenden misst, nicht die Qualität des Produkts (Biggs, 1996). LLM können in der Prüfungsumgebung wertvolle Werkzeuge sein, um die Effizienz und Objektivität zu erhöhen. Sie können z. B. eine konstante und objektive Bewertung bieten und so das Lehrpersonal im Bereich der Korrektur von Aufgaben und Prüfungen ergänzen (vgl. Hackl et al., 2023; Karademir et al., eingereicht; Menzel et al., 2023). Den Lehrkräften müssen dabei aber die Grenzen der Systeme bewusst sein (Gombert et al., 2023). Daher ist entscheidend, LLM als Ergänzung und nicht als Ersatz für menschliche Expertise zu sehen. Bewertungen müssen immer in der Verantwortung einer natürlichen Person liegen (DFG, 2023).

5.3 Leistungsscheren zwischen Gruppen

Ob durch den Einsatz von LLM leistungsstärkere oder leistungsschwächere Lernende profitieren, ist eine offene Frage, die empirisch beantwortet werden muss. Einerseits können Lernschwächen (z. B. Rechtschreibschwächen) ausgeglichen werden, so dass die Inhalte eines Textes bedeutsamer werden und Schwächere von der elaborierteren Auseinandersetzung mit dem Inhalt profitieren. LLM können benachteiligten Lernenden auch temporär (z. B. neu zugewanderten Schüler:innen) oder dauerhaft (z. B. Schüler:innen mit bestimmten Förderschwerpunkten) die Unterrichtspartizipation erleichtern. Andererseits können besonders die stärkeren Lernenden LLM auf sinnvolle und effektive Weise für Lernaktivitäten nutzen. Schüler:innen, die hohe sprach- und fachbezogene Kompetenzen besitzen, und sich daher durch *Prompts* verständlich machen, den Output von LLM verstehen und ihn gekonnt mit der eigenen Sprachproduktion verbinden können, sind bei der Nutzung im Vorteil. Leistungsstärkere Lernende verfügen außerdem über einen erhöhten fachlichen Wissensstand und ein ausgeprägteres kritisches Bewusstsein und können somit metakognitive Strategien und kritische Überprüfungen bei LLM-generierten Texten erfolgreich anwenden.

Problematisch ist, dass LLM nicht für alle Lernenden gleichermaßen zugänglich sind. Eingeschränkte Zugänge können durch z. B. mangelnde Kompetenz in der Sprache, Dyslexie oder Einschränkungen der Sehfähigkeit bestehen. Für Lernende mit niedrigem sozio-ökonomischem Hintergrund stellen finanzielle Hürden durch private Anbieter und mangelnde Unterstützungsmöglichkeiten durch die Eltern eine Zugangsbarriere dar.



5.4 Qualitätssicherung

Halluzinationen und Verzerrungen (*Biases*) stellen besondere Anforderungen an die Qualitätssicherung. Ähnlich wie Schulbücher in einzelnen Ländern vor der Nutzung geprüft werden bzw. wurden, ist eine Prüfung der LLM-Systeme vor Einsatz in der Schule vorzusehen. Dafür sind Qualitätsmerkmale mit Berücksichtigung aktueller Forschungsergebnisse zu definieren. Der entstehende Markt von LLM-basierten Anwendungen macht Rahmenvorgaben nötig, um ein Mindestmaß an Transparenz (z. B. über Trainingsdaten und deren Nachjustierung) sicherstellen zu können. Bei der Entwicklung spezifischer Tools (fachspezifisch oder bzgl. bestimmter Unterstützungsbedarfe) ist die Einbindung insbesondere fachdidaktischer Expertise und weitere Forschung wichtig.

Die Entwicklung der leistungsstarken LLM-Systeme liegt zur Zeit vor allem in den Händen großer US-amerikanischer Konzerne. Damit einher gehen Abhängigkeiten im Zusammenhang mit Datenschutz, Transparenzforderungen und allgemein Möglichkeiten des rechtlichen Zugriffs. Aktuelle Entwicklungen wie z. B. die Investitionen in Aleph Alpha zeigen, dass hier auch auf Bundesseite nach Alternativen gesucht wird (Álvarez & Menn, 2023).

6. Schlussfolgerungen

Das Bildungssystem in Deutschland steht aktuell vor der Aufgabe, das Potenzial generativer KI-Technologien wie LLM zu nutzen, dabei ihre Grenzen klar zu erkennen und verantwortungsbewusst mit ihren Beschränkungen umzugehen. Lernende können im ungünstigen Fall durch die Nutzung von LLM lernförderliche Arbeitsaufträge auslagern. Ein Verbot von LLM ist dennoch weder angemessen noch realistisch. Wegen ihrer hohen lernförderlichen Potenziale und ihrer gesellschaftlichen und wirtschaftlichen Bedeutung sollte stattdessen eine produktive Nutzung etabliert werden. LLM können vor allem dann unterstützen, wenn Lernende über hohe fachliche, Schreib-, Lese- und digitale Kompetenzen (Grundlagenwissen über KI und LLM, hohes Reflexionsvermögen und Beurteilungskompetenz) verfügen. In der Forschung ist die Nutzung von LLM keineswegs unerwünscht oder gar verboten (vgl. DFG, 2023). Vor diesem Hintergrund sollten Texte in Schulen und Hochschulen auch unter Zuhilfenahme von LLM geschrieben werden können. Gleichzeitig plädiert die SWK dafür, dass auch zukünftig die Schreibkompetenz unter Verzicht auf LLM im schulischen Kontext systematisch aufgebaut wird und LLM erst in höheren Klassenstufen zum Einsatz kommen. Dafür sind lizenzrechtliche und datenschutzrechtliche Barrieren abzubauen.

KI kann und sollte Lehr-Lernprozesse unterstützen, die finale Entscheidung bzw. Bewertung und die Verantwortung für das Endprodukt muss beim Menschen liegen. Eine unreflektierte Übergabe von Aufgaben z. B. an LLM kann zu Fehlern, Ungerechtigkeiten und Kompetenzverlust führen. Lerninhalte und Prüfungen müssen den Kriterien von Wahrheit, Fairness, Objektivität, Validität und Reliabilität unterliegen.



Übergangsphase zur systematischen Erprobung bei offener Fehlerkultur

Kurzfristig ist eine Übergangsphase nötig, in der eine systematische Erprobung von LLM bei offener Fehlerkultur stattfindet. Dazu müssen zeitnah der rechtssichere Zugang ermöglicht, lernförderliche Nutzungsszenarien geschaffen und Fortbildungen für Lehrkräfte forciert werden. Dabei ist eine enge Zusammenarbeit von Praxis und Wissenschaft nötig, um fachspezifische und altersgerechte Nutzungs- und Aufgabenszenarien zu entwickeln, die den Potenzialen von LLM gerecht werden. Fachdidaktiken, Medienpsychologie, Medienpädagogik und Landesinstitute können hier ko-konstruktiv Impulse geben.

Sowohl Lernende als auch Lehrkräfte benötigen umfassende Kompetenzen zum lernförderlichen Umgang mit LLM: Dazu gehört das Verstehen der Funktionsweise der entsprechenden Tools, ihrer Stärken und Schwächen sowie *Prompt-Tuning*. Es bedarf daher eines raschen und umfassenden, systematischen und wissenschaftsunterstützten Ausbaus der Fortbildungsangebote für Lehrkräfte und Lehrende an Hochschulen zur Nutzung von LLM, um bei ihnen entsprechende Kompetenzen aufzubauen, damit sie KI-bezogene *ICT-Kompetenzen* erwerben und LLM in ihren Fachunterricht didaktisch treffsicher einsetzen können. Drei Fragen sollten zum Einsatz im Fachunterricht leitend sein und systematisch empirisch erforscht werden: Wie können LLM bei sprachproduktiven Aufgaben eingesetzt werden? Welche Potenziale bieten LLM für die summative und formative Bewertung von Schülerleistungen? Welche Potenziale haben LLM für die Erstellung und Anpassung von Unterrichtsmaterialien (insbesondere auch Prüfungsaufgaben)?

Bei Lernenden steigen die Notwendigkeit und Relevanz der Fähigkeit zur Selbststeuerung und zum kritischen Prüfen drastisch an. Diese Kompetenzziele sind bereits in vielen Curricula verankert, müssen aber um KI-spezifische Inhalte ergänzt und zügig in die Lehrpraxis integriert werden. Hier ist ein Aufbau digitaler Kompetenzen und *ICT-Literacy* (inkl. Grundkenntnisse über KI) als fächerübergreifende Aufgabe in der Primarstufe und der Sekundarstufe I nötig, ergänzt um entsprechende Angebote im Informatikunterricht der Sekundarstufe I. Ein starker Fokus sollte dabei auf Kompetenzen im Bereich der Bewertung der Glaubwürdigkeit von Texten bzw. Quellen liegen.

Domänenspezifische Tools entwickeln

Mittel- und langfristig ist die Entwicklung domänenspezifischer Tools notwendig. Derzeitige LLMs sind generisch und nicht spezifisch für den Bildungskontext und für einzelne Fächer entwickelt. Sie müssen mit qualitativ hochwertigen, fachspezifischen Daten trainiert werden, um zuverlässig und valide zu funktionieren. Eine systematische Variation von Trainings- und Stimulusmaterialien der LLM, inspiriert von experimentalpsychologischer Forschung, kann helfen, diesen Prozess voranzutreiben und die Reliabilität und Validität der generierten Informationen zu erhöhen. Aktuelle technische Weiterentwicklungen lassen hier ein großes Potenzial vermuten. Befunde aus der Lehr-Lernforschung, z. B. zu ITS, zu Feedback, Schreib- und Kollaborationsprozessen, müssen auf die Entwicklung spezifischer LLM übertragen und systematischer als bisher mit anderen Tools in die Entwicklung von fachdidaktisch fundierten digitalen Lernumgebungen integriert werden.



LLM ab der Sekundarstufe einsetzen

Der Einsatz von LLM sollte angepasst an die Bildungsetappen erfolgen. In der Grundschule und zu Beginn der Sekundarstufe I sollte weitgehend auf LLM verzichtet werden. Stattdessen sollte der systematische Aufbau von basalen Lese- und Schreibkompetenzen fokussiert werden, die den gesamten Schreibprozess, d. h. die rekursiv-iterative Planung, Erstellung und Revision von Texten systematisch berücksichtigen. Analoges Schreiben ist Aufgabe aller Fächer, um den fachspezifischen Aufbau von Wissen und Kompetenzen zu unterstützen. Im Laufe der Sekundarstufe I können zunehmend digitale Tools zur Erstellung von Texten eingesetzt werden, die Verwendung von LLM sollte in den ersten Jahren nur sehr eng begleitet erfolgen. Ab der achten Jahrgangsstufe ist ein regelmäßiger Einsatz von LLM als Schreibunterstützung in allen schulischen Fächern, in denen schriftlichen Leistungen erbracht werden müssen, zu erwägen. Ab der Sekundarstufe II können LLM vielfältig zur Erstellung von Texten verschiedener Genres eingesetzt werden (hilfsmittelunterstützt), während weiterhin auch Texte ohne LLM (hilfsmittelfrei) erstellt werden. An den Hochschulen schließlich sollten LLM gemäß den Richtlinien der DFG verwendet werden können.

Prüfungskultur verändern

Die Prüfungskultur sollte angepasst werden. Prüfungsformate mit und ohne Hilfsmittel können sich ergänzen. In prozessorientierten Formaten sollte die versierte Koaktivität mit LLM als Lernziel geübt und geprüft werden. Bei der Bewertung können KI-basierte Tools Lehrkräfte unterstützen, es ist jedoch zentral, dass die Kontrolle und Verantwortung für die Bewertung in menschlicher Hand bleiben. An Hochschulen können schriftliche Hausarbeiten (z. B. Bachelor- oder Masterarbeiten) unter Nutzung von LLM als Prüfungsformate genutzt werden. Texte, die mit Unterstützung von LLM erstellt wurden, müssen einen Hinweis enthalten. Ergänzend sind hilfsmittelfreie Prüfungsformate beizubehalten sowie Prüfungsformate zu entwickeln, bei denen das *Prompting* im Vordergrund steht.

Rahmenbedingungen schaffen

Lernende und Lehrkräfte benötigen einen (kostengünstigen oder kostenfreien) Zugang zu Tools. Staatliche Organisationen sollten sicherstellen, dass alle Lernenden in Bildungseinrichtungen gleichermaßen Zugriff auf hochwertige KI-Systeme haben, die für das Lernen genutzt werden können. Dabei sollte neben aktuellen LLM-Ansätzen auch die Weiterentwicklung und Nutzung weiterer KI-basierter Ansätze zur individuellen Förderung systematisch gefördert werden, deren Lernwirksamkeit belegt werden konnte. Die bereits heute bestehenden Ungleichheiten bei der Nutzung kostenpflichtiger Lernprogramme oder Nachhilfe könnten sich durch die Verbreitung weiterer KI-Anwendungen verschärfen. Die Entwicklung von Lizenzmodellen beispielsweise in Form von Landeslizenzen wäre ein möglicher Weg.



LLM in Lernplattformen integrieren und Commons-Lösungen entwickeln

Technologische, ethische und rechtliche Probleme können die Rechtmäßigkeit eines unterrichtlichen Einsatzes von LLM in Frage stellen. Der Einsatz kommerzieller LLM ist marktwirtschaftlichen Interessen unterworfen. Bildungspolitisch sollte das zum Anlass genommen werden, LLM in geeignete Lernplattformen zu integrieren oder Commons-Alternativen zu entwickeln. So kann sichergestellt werden, dass Bildungsinteressen gegenüber wirtschaftlichen Interessen gewahrt werden und keine Informations-Monopole entstehen. Für rechtliche und ethische Grauzonen sollten klare Regulierungen getroffen werden. Perspektivisch sollten LLM verwendet werden, deren Funktionsweise und Verzerrungen bekannt sind und frei beforscht werden können.



Literatur

- Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., Huang, Q., Jansen, A., Roberts, A., Tagliasacchi, M., Sharifi, M., Zeghidour, N. & Frank, C. (2023). *MusicLM: Generating Music From Text*. <https://doi.org/10.48550/ARXIV.2301.11325>
- Albrecht, S. (2023). *ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen* (TAB-Hintergrundpapier Nr. 26). <https://doi.org/10.5445/IR/1000158070>
- Alter, A. & Harris, E. A. (2023, 20. September). *Franzen, Grisham and Other Prominent Authors Sue OpenAI*. <https://www.nytimes.com/2023/09/20/books/authors-openai-lawsuit-chatgpt-copyright.html> (aufgerufen am 09.01.2024)
- Altmann, S. (2022). *Introducing ChatGPT*. <https://openai.com/blog/chatgpt/> (aufgerufen am 09.01.2024)
- Álvarez, S. & Menn, A. (2023, 07. November). *Millionen-Deal für Aleph Alpha: So will das Start-up Deutschland voranbringen*. <https://www.wiwo.de/politik/deutschland/ki-standort-millionen-deal-fuer-aleph-alpha-so-will-das-start-up-deutschland-voranbringen/29484770.html> (aufgerufen am 09.01.2024)
- Athaluri, S. A., Manthena, S. V., Kesapragada, V. S. R. K. M., Yarlagadda, V., Dave, T. & Duddumpudi, R. T. S. (2023). Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus*, 15(4), e37432. <https://doi.org/10.7759/cureus.37432>
- Ausat, A. M. A., Massang, B., Efendi, M., Nofirman, N. & Riady, Y. (2023). Can Chat GPT Replace the Role of the Teacher in the Classroom: A Fundamental Analysis. *Journal on Education*, 5(4), 16100–16106. <https://jonedu.org/index.php/joe/article/view/2745> (aufgerufen am 09.01.2024)
- Beck, S. W. & Levine, S. R. (2023). Backtalk: ChatGPT: A powerful technology tool for writing instruction. *Phi Delta Kappan*, 105(1), 66–67. <https://doi.org/10.1177/00317217231197487>
- Becker-Mrotzek, M. (2022). Schreibkompetenz: Überlegungen zu einem didaktischen Konstrukt. In M. Becker-Mrotzek & J. Grabowski (Hrsg.), *Schreibkompetenz in der Sekundarstufe: Theorie, Diagnose und Förderung* (S. 9–27). Waxmann.
- Becker-Mrotzek, M. & Böttcher, I. (2020). *Schreibkompetenz entwickeln und beurteilen* (9. Auflage). *Sekundarstufe I + II*. Cornelsen.
- Bergener, J., Gossen, M., Hoffmann, M. L., Bießmann, F., Veneny, M. & Korenke, R. (2023). Evaluating the Quality of ChatGPT's Climate-related Responses. *Wirtschaften - Fachzeitschrift*, 38(3), 46–50. <https://doi.org/10.14512/OEW380346>
- Bewersdorff, A., Seßler, K., Baur, A., Kasneci, E. & Nerdel, C. (2023, 11. August). *Assessing Student Errors in Experimentation Using Artificial Intelligence and Large Language Models: A Comparative Study with Human Raters*. <https://arxiv.org/pdf/2308.06088.pdf>



- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364. <https://doi.org/10.1007/BF00138871>
- bildung.digital (2023). *KI im Klassenzimmer: Zentrale Erkenntnisse aus dem Pilotversuch*. <https://www.bildung.digital/artikel/erkenntnisse-aus-dem-pilotversuch> (aufgerufen am 09.01.2024)
- Bråten, I., Stadtler, M. & Salmerón, L. (2018). The role of sourcing in discourse comprehension. In M. Schorber, D. N. Rapp & M. A. Britt (Hrsg.), *Handbook of discourse processes* (S. 141–166). Taylor & Francis.
- 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 [Jeffrey], Winter, C., ... Amodei, D. (2020). Language Models are Few-Shot Learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan & H. Lin (Hrsg.), *Advances in neural information processing systems: Bd. 33. Advances in neural information processing systems: 34th conference on neural information processing systems (neurips 2020)*. NEURAL INFO PROCESS SYS F.
- Chi, M. T. H. & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Chu, Y. & Liu, P. (2023). Public aversion against ChatGPT in creative fields? *Innovation (Cambridge (Mass.))*, 4(4), 100449. <https://doi.org/10.1016/j.xinn.2023.100449>
- Cress, U. & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s11412-023-09409-w>
- Deutsche Forschungsgemeinschaft [DFG] (2023). *Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG*. https://www.dfg.de/download/pdf/dfg_im_profil/geschaefsstelle/publikationen/stellungnahmen_papiere/2023/230921_stellungnahme_praesidium_ki_ai.pdf (aufgerufen am 09.01.2024)
- Emsley, R. (2023). ChatGPT: these are not hallucinations - they're fabrications and falsifications. *Schizophrenia (Heidelberg, Germany)*, 9(1), 52. <https://doi.org/10.1038/s41537-023-00379-4>
- Fang, T., Yang, S., Lan, K., Wong, D. F., Hu, J., Chao, L. S. & Zhang, Y [Yue]. (2023). *Is ChatGPT a Highly Fluent Grammatical Error Correction System? A Comprehensive Evaluation*. <https://doi.org/10.48550/ARXIV.2304.01746>
- Feng, S., Park, C. Y., Liu, Y. & Tsvetkov, Y. (2023). *From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models*. <https://doi.org/10.48550/ARXIV.2305.08283>
- fobizz (2023, 27. November). *Ein Jahr ChatGPT. KI in der Schule: So nutzen Lehrkräfte die Technologie im Unterricht* [Pressemitteilung]. <https://drive.google.com/file/d/1O4XIX1-ZIArQdsdY7V6JFeQXOudey-Loo/view>



- Frederking, V. (2023). Von Fake News bis ChatGPT. Digitale Textsouveränität als ethisch-politische Bildungsaufgabe für Deutschdidaktik und Deutschunterricht in der digitalen Welt. *Medien im Deutschunterricht*, 1–27. <https://doi.org/10.18716/OJS/MIDU/2023.2.4>
- Füller, C. (2023, 10. November). *Schüler glauben, mit ChatGPT gut zu lernen*. <https://www.zeit.de/gesellschaft/schule/2023-11/chatgpt-schueler-lernen-umfrage> (aufgerufen am 09.01.2024)
- Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M. & Gerjets, P. (2023). ChatGPT in education: global reactions to AI innovations. *Scientific reports*, 13(1), 15310. <https://doi.org/10.1038/s41598-023-42227-6>
- Fütterer, T., Goldberg, P., Bühler, B., Sikimic, V., Trautwein, U., Gerjets, P., Stürmer, K. & Kasneci, E. (in Vorbereitung). *Artificial Intelligence in Classroom Management: A Systematic Review on Technical Implementations, Educational Purposes and Ethical Considerations*.
- Gombert, S., Di Mitri, D., Karademir, O., Kubsch, M., Kolbe, H., Tautz, S., Grimm, A., Bohm, I., Neumann, K. & Drachsler, H. (2023). Coding energy knowledge in constructed responses with explainable NLP models. *Journal of Computer Assisted Learning*, 39(3), 767–786. <https://doi.org/10.1111/jcal.12767>
- Graham, S., Kihara, S. A. & MacKay, M. (2020). The Effects of Writing on Learning in Science, Social Studies, and Mathematics: A Meta-Analysis. *Review of Educational Research*, 90(2), 179–226. <https://doi.org/10.3102/0034654320914744>
- Hackl, V., Müller, A. E., Granitzer, M. & Sailer, M [Maximilian]. (2023, 3. August). *Is GPT-4 a reliable rater? Evaluating Consistency in GPT-4 Text Ratings*. <https://arxiv.org/pdf/2308.02575.pdf>
- Hessisches Kultusministerium. (Juli 2023). *Künstliche Intelligenz (KI) in Schule und Unterricht: Eine Handreichung für Lehrkräfte zum Umgang mit KI-basierten Anwendungen*. https://digitale-schule.hessen.de/sites/digitale-schule.hessen.de/files/2023-07/ki_handreichung.pdf (aufgerufen am 09.01.2024)
- Hoch, E. & Fütterer, T. (2023). Kompetenzen für das Lernen mit digitalen Medien: Eine konzeptuelle Analyse. In K. Scheiter & I. Gogolin (Hrsg.), *Edition ZfE: Band 15. Bildung für eine digitale Zukunft* (Bd. 15, S. 81–101). Springer VS, Springer Fachmedien Wiesbaden GmbH. https://doi.org/10.1007/978-3-658-37895-0_4
- 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. 22–40): Zentrum für Wissenschaftsdidaktik der Ruhr-Universität Bochum.
- Huschens, M., Briesch, M., Sobania, D. & Rothlauf, F. (2023, 5. September). *Do You Trust ChatGPT? -- Perceived Credibility of Human and AI-Generated Content*. <http://arxiv.org/pdf/2309.02524v1>
- Hutson, M. (2023). Rules to keep AI in check: nations carve different paths for tech regulation. *Nature*, 620(7973), 260–263. <https://doi.org/10.1038/d41586-023-02491-y>
- Kammerer, Y. & Brand-Gruwel, S. (2020). Trainings and Tools to Foster Source Credibility Evaluation During Web Search. In W. T. Fu & H. van Oostendorp (Hrsg.), *Understanding and Improving*



- Information Search* (S. 213–243). Springer International Publishing. https://doi.org/10.1007/978-3-030-38825-6_11
- Kammerer, Y., Oloff, C., Gerjets, P. & Mayer, A.-K. (2015). Erfassung von Fertigkeiten im Umgang mit dem Social Web. In A.-K. Mayer (Hrsg.), *Informationskompetenz im Hochschulkontext: Interdisziplinäre Forschungsperspektiven ; [Beiträge des interdisziplinären Symposiums „Informationskompetenz im Hochschulkontext - Interdisziplinäre Forschungsperspektiven“, das im Mai 2014 vom Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) in Trier veranstaltet wurde* (S. 121–150). Pabst Science Publishers.
- Karademir, O., Di Mitri, D., Schneider, J., Jivet, I., Allmang, J., Gombert, S., Kubsch, M., Neumann, K. & Drachsler, H. (eingereicht). I don't have time! But keep me in the loop: Co-Designing Requirements for a Learning Analytics Cockpit with teachers. *Journal of Computer Assisted Learning*.
- Kasneji, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M [Michael], Schmidt, A., Seidel, T., ... Kasneji, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Küchemann, S., Steinert, S., Revenga, N., Schweinberger, M., Dinc, Y., Avila, K. E. & Kuhn, J. (2023). Can ChatGPT support prospective teachers in physics task development? *Physical Review Physics Education Research*, 19(2). <https://doi.org/10.1103/PhysRevPhysEducRes.19.020128>
- Kühl, E. (6. Dezember 2022). ChatGPT: Gut erfunden ist halb geglaubt. *Zeit Online*. <https://www.zeit.de/digital/internet/2022-12/chatgpt-kuenstliche-intelligenz-openai-chatbot> (aufgerufen am 09.01.2024)
- Landesinstitut für Schulqualität und Lehrerbildung - Sachsen-Anhalt (September 2023). *Sprachmodelle im Schulunterricht: Chancen und Herausforderungen: Handreichung zum Umgang mit Künstlicher Intelligenz in der schulischen Bildung*. https://www.bildung-lsa.de/files/5c5e6e7a4f362d6ab1bfd03a000ac17d/Sprachmodelle_im_Schulunterricht.pdf (aufgerufen am 09.01.2024)
- Menzel, L., Gombert, S., Weidlich, J., Fink, A., Frey, A. & Drachsler, H. (2023). Why You Should Give Your Students Automatic Process Feedback on Their Collaboration: Evidence from a Randomized Experiment. In O. Viberg, I. Jivet, P. J. Munoz-Merino, M. Perifanou & T. Papathoma (Vorsitz), *18th European Conference on Technology Enhanced Learning*, Aveiro, Portugal.
- Meurers, D., Kuthy, K. D. e., Möller, V., Nuxoll, F., Rudzewitz, B. & Ziai, R. (2018). Digitale Differenzierung benötigt Informationen zu Sprache, Aufgabe und Lerner: Zur Generierung von individuellem Feedback in einem interaktiven Arbeitsheft. In C. Gnutzmann, L. Küster & K. Schramm (Hrsg.), *Fremdsprachen Lehren und Lernen* (S. 64–82).
- Meurers, D., Kuthy, K. de, Nuxoll, F., Rudzewitz, B. & Ziai, R. (2019). Scaling Up Intervention Studies to Investigate Real-Life Foreign Language Learning in School. *Annual Review of Applied Linguistics*, 39, 161–188. <https://doi.org/10.1017/S0267190519000126>



- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (Februar 2023). *Umgang mit textgenerierenden KI-Systemen: Ein Handlungsleitfaden*. https://www.schulministerium.nrw/system/files/media/document/file/handlungsleitfaden_ki_msb_nrw_230223.pdf (aufgerufen am 09.01.2024)
- Nadeem, M., Bethke, A. & Reddy, S. (2021). StereoSet: Measuring stereotypical bias in pretrained language models. In C. Zong, F. Xia, W. Li & R. Navigli (Vorsitz), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Symposium im Rahmen der Tagung von Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.acl-long.416>
- Navigli, R., Conia, S. & Ross, B. (2023). Biases in Large Language Models: Origins, Inventory, and Discussion. *Journal of Data and Information Quality*, 15(2), 1–21. <https://doi.org/10.1145/3597307>
- Nehlsen, J. & Fleck, T. (31. März 2023). Ist ChatGPT ein zulässiges Hilfsmittel in Prüfungen? *Forschung & Lehre*.
- Nuxoll, F. (2023). KI in der Schule. *Aus Politik und Zeitgeschichte*, 73(42), 41–46.
- OpenAI (2023a, 6. November). *Introducing GPTs*. <https://openai.com/blog/introducing-gpts> (aufgerufen am 09.01.2024)
- OpenAI (2023b). *DALL-E 3*. <https://openai.com/dall-e-3> (aufgerufen am 09.01.2024)
- Ouyang, L., Wu, J [Jeff], Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J. & Lowe, R. (2022). Training language models to follow instructions with human feedback. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho & A. Oh (Hrsg.), *Advances in neural information processing systems: Bd. 35, Advances in Neural Information Processing Systems: 36th Conference on Neural Information Processing Systems (NeurIPS 2022)*. Curran Associates Inc.
- Paas, F. & Sweller, J. (2019). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer (Hrsg.), *Cambridge Handbooks in Psychology. The Cambridge handbook of multimedia learning* (Second edition, 5th printing, S. 27–42). Cambridge University Press. <https://doi.org/10.1017/9781108894333.009>
- Pöler, H. (15. März 2023). KI-Tools & ChatGPT – die “Demokratisierung” adaptiver/intelligenter Tutorssysteme? *unterrichten.digital*. <https://unterrichten.digital/2023/03/15/chatgpt-ki-tools-demokratisierung-tutorsysteme/> (aufgerufen am 09.01.2024)
- Rawte, V., Sheth, A. & Das, A. (2023). *A Survey of Hallucination in “Large” Foundation Models*. <https://arxiv.org/pdf/2309.05922.pdf>
- Redecker, C. (2017). *European framework for the digital competence of educators*. <https://op.europa.eu/en/publication-detail/-/publication/fcc33b68-d581-11e7-a5b9-01aa75ed71a1/language-en>
- Rozado, D. (2023). The Political Biases of ChatGPT. *Social Sciences*, 12(3), 148. <https://doi.org/10.3390/socsci12030148>



- Salden, P. & Leschke, J. (März 2023). *Didaktische und rechtliche Perspektiven auf KI-gestütztes Schreiben in der Hochschulbildung*. <https://doi.org/10.13154/294-9734>
- Schmoll, H. (2023, 11. Juli). *Zurück zu den gedruckten Büchern!* <https://www.faz.net/aktuell/politik/ausland/digitalisierung-an-schulen-karolinska-institut-ueber-negative-konsequenzen-19023795.html> (aufgerufen am 09.01.2024)
- Schneider, J. (2023). *Schülerinnen mit KI-Nacktbildern gemobbt. Deepnudes in Spanien*. <https://www.zdf.de/nachrichten/panorama/spanien-schuelerinnen-deepnudes-nacktbilder-100.html> (aufgerufen am 09.01.2024)
- Schramowski, P., Turan, C., Andersen, N., Rothkopf, C. A. & Kersting, K. (2022). Large pre-trained language models contain human-like biases of what is right and wrong to do. *Nature Machine Intelligence*, 4(3), 258–268. <https://doi.org/10.1038/s42256-022-00458-8>
- Schultheiß, S. & Lewandowski, D. (2023). Misplaced trust? The relationship between trust, ability to identify commercially influenced results and search engine preference. *Journal of Information Science*, 49(3), 609–623. <https://doi.org/10.1177/01655515211014157>
- Seßler, K., Xiang, T., Bogenrieder, L. & Kasneci, E. (2023). PEER: Empowering Writing with Large Language Models. In O. Viberg, I. Jivet, P. J. Muñoz-Merino, M. Perifanou & T. Papathoma (Hrsg.), *Lecture Notes in Computer Science. RESPONSIVE AND SUSTAINABLE EDUCATIONAL FUTURES: 18th european conference on* (Bd. 14200, S. 755–761). SPRINGER INTERNATIONAL PU. https://doi.org/10.1007/978-3-031-42682-7_73
- Ständige Wissenschaftliche Kommission der Kultusministerkonferenz [SWK] (September 2022). *Digitalisierung im Bildungssystem: Handlungsempfehlungen von der Kita bis zur Hochschule: Gutachten der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz (SWK)*. <https://doi.org/10.25656/01:25273>
- Steinhoff, T. (2023). Der Computer schreibt (mit). Digitales Schreiben mit Word, WhatsApp, ChatGPT & Co. als Koaktivität von Mensch und Maschine. *MiDU – Medien im Deutschunterricht*, 5(1), 1–16. <https://doi.org/10.18716/ojs/midu/2023.1.4>
- Stengel, M., Nagano, K., Liu, C., Chan, M., Trevithick, A., Mello, S. de, Kim, J. & Luebke, D. (2023). AI-Mediated 3D Video Conferencing. In *SIGGRAPH 2023*. Symposium im Rahmen der Tagung von Association for Computing Machinery, Los Angeles. <https://research.nvidia.com/labs/nxp/3dvc-siggraph-etech/> (aufgerufen am 09.01.2024)
- Susnjak, T. (2022, 19. Dezember). *ChatGPT: The End of Online Exam Integrity?* <https://arxiv.org/pdf/2212.09292.pdf>
- Wang, R. E. & Demszky, D. (2023). *Is ChatGPT a Good Teacher Coach? Measuring Zero-Shot Performance For Scoring and Providing Actionable Insights on Classroom Instruction*. <https://arxiv.org/abs/2306.03090>
- Weiser, B. (2023, 27. Mai). *Here's What Happens When Your Lawyer Uses ChatGPT*. <https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html> (aufgerufen am 09.01.2024)



- Weiss, Z., Woerfel, T. & Meurers, W. D. (2023). Intelligente digitale Werkzeuge in der sprachlichen Bildung. In M. Becker-Mrotzek, I. Gogolin, H.-J. Roth & P. Stanat (Hrsg.), *Sprachliche Bildung: Band 10. Grundlagen der sprachlichen Bildung* (S. 185–197). Waxmann.
- Woitt, S., Weidlich, J., Jivet, I., Orhan Göksün, D., Drachsler, H. & Kalz, M. (2023). Students' feedback literacy in higher education: an initial scale validation study. *Teaching in Higher Education*, 1–20. <https://doi.org/10.1080/13562517.2023.2263838>
- Wu, H., Wang, W., Wan, Y., Jiao, W. & Lyu, M. (2023). *ChatGPT or Grammarly? Evaluating ChatGPT on grammatical error correction benchmark*. <https://doi.org/10.48550/arXiv.2303.13648>
- Wu, Q., Bansal, G., Zhang, J., Wu, Y., Zhang, S., Zhu, E., Li, B., Jiang, L., Zhang, X. & Wang, C. (2023, 15. September). *AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation*. Microsoft. <https://www.microsoft.com/en-us/research/publication/autogen-enabling-next-gen-llm-applications-via-multi-agent-conversation-framework/> (aufgerufen am 09.01.2024)
- Wu, S., Fei, H., Qu, L., Ji, W. & Chua, T.-S. (2023). *NExT-GPT: Any-to-Any Multimodal LLM*. <https://doi.org/10.48550/arXiv.2309.05519>
- Wu, Y., Henriksson, A., Duneld, M. & Nouri, J. (2023). Towards Improving the Reliability and Transparency of ChatGPT for Educational Question Answering. In *European Conference on Technology Enhanced Learning* (S. 475–488).
- Xiao, C., Xu, S. X., Zhang, K., Wang, Y. & Xia, L. (2023). Evaluating Reading Comprehension Exercises Generated by LLMs: A Showcase of ChatGPT in Education Applications. In *BEA 2023*.
- Yadlowsky, S., Doshi, L. & Tripuraneni, N. (2023). *Pretraining Data Mixtures Enable Narrow Model Selection Capabilities in Transformer Models*. <https://doi.org/10.48550/arXiv.2311.00871>
- Yoon, S.-Y., Miszoglad, E. & Pierce, L. R. (2023). *Evaluation of ChatGPT Feedback on ELL Writers' Coherence and Cohesion*. <https://doi.org/10.48550/arXiv.2310.06505>
- Zesch, T., Horbach, A. & Zehner, F. (2023). To Score or Not to Score: Factors Influencing Performance and Feasibility of Automatic Content Scoring of Text Responses. *Educational Measurement: Issues and Practice*, 42(1), 44–58. <https://doi.org/10.1111/emip.12544>
- Zhang, D. (2023, 4. Mai). *Should ChatGPT and Bard Share Revenue with Their Data Providers? A New Business Model for the AI Era*. <https://arxiv.org/pdf/2305.02555.pdf>
- Zhang, Y [Yue], Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y [Yu], Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F. & Shi, S. (2023). *Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models*. <https://doi.org/10.48550/arXiv.2309.01219>





Das vorliegende Impulspapier wurde von den Mitgliedern der Ständigen Wissenschaftlichen Kommission im Januar 2024 verabschiedet.

Impulspapiere adressieren fachliche Diskussionen bezogen auf eine konkrete Maßnahme, Situation oder ein Vorhaben mit Handlungsaufforderung und entwickeln strategische Positionen.

Weitere Informationen zu den Formaten der SWK finden Sie hier:

www.kmk.org/kmk/staendige-wissenschaftliche-kommission/aufgaben.html

Mitwirkende

Unter Mitarbeit von

Prof. Dr. Hendrik Drachler (DIPF), Prof. Dr. Volker Frederking (Universität Nürnberg), Dr. Tim Fütterer (Universität Tübingen), Prof. Dr. Peter Gerjets (IWM Tübingen), Prof. Dr. Andrea Horbach (Universität Hildesheim), Prof. Dr. Enkelejda Kasneci (Technische Universität München), Prof. Dr. Andreas Lachner (Universität Tübingen), Prof. Dr. Marlit Lindner (IWM Tübingen), Dr. Christina Matschke (IWM Tübingen), Prof. Dr. Detmar Meurers (Universität Tübingen), Florian Nuxoll (Universität Tübingen), Prof. Dr. Torsten Steinhoff (Universität Siegen), Prof. Dr. Leo van Waveren (RPTU Kaiserslautern), Prof. Dr. Doris Weißels (Fachhochschule Kiel), Prof. Dr. Torsten Zesch (Fernuniversität Hagen)



Impressum

Herausgeber

Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK)

Geschäftsstelle

Graurheindorfer Str. 157
53117 Bonn

Tel. 0228 501 700

info@swk.kmk.org

www.swk-bildung.org

Verantwortlich

Dr. Isabelle Sieh (Leitung der Geschäftsstelle)

Die Veröffentlichungen der SWK stehen auf den Seiten der SWK zum [Download](#) zur Verfügung.



Zitationsvorschlag:

Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK) (2023): *Large Language Models und ihre Potenziale im Bildungssystem*. Impulspapier der Ständigen Wissenschaftlichen Kommission (SWK) der Kultusministerkonferenz. <http://dx.doi.org/10.25656/01:28303>

© Geschäftsstelle der SWK, Bonn 2024



Die Ständige Wissenschaftliche Kommission (SWK) ist ein unabhängiges wissenschaftliches Beratungsgremium der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK).

www.swk-bildung.org

