

Der digitale Dozent: ChatGPT als Co-Pilot in der Lehrpersonenbildung

Jens Damköhler*, Wolfgang Lutz*, Thomas Trefzger*

*Lehrstuhl für Physik und ihre Didaktik, Universität Würzburg
jensdamkoehler@uni-wuerzburg.de

Kurzfassung

Die rasante Entwicklung künstlicher Intelligenz im vergangenen Jahr beeinflusst mittlerweile auch Diskussionen im Bildungsbereich. Während Studierende wie Lehrende grundsätzlich offen gegenüber der Nutzung von KI scheinen, sind Überlegungen zum konkreten Einsatz in der universitären Lehre häufig geprägt von Spekulationen und Unsicherheit. Im nachfolgenden Artikel werden zunächst Annahmen diskutiert, die für die Integration künstlicher Intelligenz im Bildungsbereich relevant erscheinen. Anschließend wird beschrieben, wie aufbauend auf diesen Überlegungen ein KI-Chatbot basierend auf dem bekannten Modell ChatGPT in einem Lehr-Lern-Labor-Seminar Physik an der Universität Würzburg genutzt wird, um Studierende bei der strukturierten Reflexion ihrer ersten Unterrichtserfahrungen zu begleiten und zu unterstützen.

1. Einleitung

Auch wenn die Einsatzmöglichkeiten Künstlicher Intelligenz (KI) bereits seit mehreren Jahren immer wieder diskutiert wurden, war kaum vorherzusehen, mit welcher Macht das Thema mit der Einführung von ChatGPT im Herbst 2022 in die unterschiedlichsten Fachdiskurse drängen würde. Gerade die elaborierten Sprachfähigkeiten von ChatGPT ermöglichen menschlich anmutende Dialoge – und die Entwicklung scheint noch lange nicht am Ende (van den Berg & Du Plessis, 2023). Doch auf die erste Euphorie über die begeisternden Fähigkeiten der generativen Werkzeuge, z. B. bei der Text- und Bildproduktion, folgt oft eine Reihe von Enttäuschungen, wenn bestimmte Erwartungen doch nicht erfüllt werden können (Salvagno, Taccone & Gerli, 2023). So kommt z.B. Emsley (2023, S. 2) zu dem Schluss: „Just as I would not recommend collaborating with a colleague with pseudologia fantastica, I do not recommend ChatGPT as an aid to scientific writing.“. Dennoch bieten Werkzeuge der generativen künstlichen Intelligenz besondere Möglichkeiten, die im Folgenden näher vorgestellt werden.

Mit generativer KI wird eine Klasse von Werkzeugen bezeichnet, die Deep-Learning-Modelle nutzen, um menschenähnliche („human-like“) Inhalte als Reaktion auf komplexe und vielfältige Aufforderungen zu erzeugen (Lim, Gunasekara, Pallant, Pallant & Pechenkina, 2023). Unter ihnen bilden die Large-Language-Modelle eine Teilmenge, die sich mit dem Verstehen und Erzeugen menschlicher Sprache beschäftigen (Wang et al., 2024). Diese Modelle, zu denen auch der OpenAI Chatbot ChatGPT gehört, beeindruckt durch Dialoge, die denen mit Menschen verblüffend nahe kommen und wecken Hoffnungen und Erwartungen für den Einsatz in unterschiedlichen Bereichen – auch im Bildungswesen (Kasneci et al., 2023; van den Berg & Du Plessis, 2023). Inwieweit die Fähigkeiten des Systems gewinnbringend

eingesetzt werden können, gilt es durch gezielte Forschung zu untersuchen.

An der Universität Würzburg wurde ein Projekt ins Leben gerufen, in dem die Wirkung des Einsatzes von ChatGPT als Gesprächspartner bei der Reflexion von Unterrichtserfahrungen von Studierenden in einem Lehr-Lern-Labor-Seminar (LLL-Seminar) Physik untersucht wird. Die bei der Entwicklung zugrunde liegenden Annahmen sowie das Projekt werden im Folgenden erläutert.

2. Annahmen

2.1. Wir wissen vieles über KI... ..nicht

Eines der häufig in Verbindung mit KI genannten Probleme ist das der fehlenden Erklärbarkeit der Ergebnisse und Funktionsweise von Werkzeugen, die auf künstlicher Intelligenz beruhen. Diese fehlende Erklärbarkeit fußt zum Teil auf der Komplexität der zugrunde liegenden Algorithmen, die eine Kontrolle unmöglich machen, zum Teil aber auch auf einer nicht auflösbaren prinzipiellen Undurchsichtigkeit, die auf nichtlinear interagierenden Operationen und Rückkopplungsschleifen im Rechenprozess beruht (Peters, 2023). In diesem Fall ist es sogar bei vollständiger Offenlegung des Quellcodes nicht möglich, den Weg von einer Eingabe zur zugehörigen Ausgabe nachzuvollziehen.

Zusätzlich sind entsprechende Werkzeuge, sofern sie Teil einer kommerziellen Unternehmung sind, selbst innerhalb des Möglichen oft nur schwach dokumentiert. Diese schwache Dokumentation betrifft auch die häufigen Überarbeitungen der Modelle, die sich oft erst in der Nutzung offenbaren. Dazu ist das Verhalten der KI-Systeme stark von der zum Training verwendeten Datenbasis abhängig (Leavy, O'Sullivan & Siapera, 2020). Ist diese unbekannt, führt das zu einem weiteren Unsicherheitsfaktor. Dies führt dazu, dass viele der vermeintlichen Wissensbestände und Erkenntnisse über KI unsicher und vorläufiger Natur sind, was die folgenden Beispiele illustrieren.

In einzelnen Veröffentlichungen zum Einsatz von ChatGPT bei der Literaturrecherche existierten in der Anfangszeit ein Teil der von ChatGPT aufgeführten Quellen nicht, sondern wurden vom System frei erfunden (z. B. Vogelsang, Bleher, Krupitzer, Stein & Jung, 2023). Ein Jahr später ist nicht nur das Phänomen der Halluzination weithin bekannt, es wird auch so aktiv bearbeitet, dass die Literaturrecherche mit der Folgeversion gpt4 inzwischen stark verbessert wurde.

Neuere Versionen großer Sprachmodelle sind in der Lage, auch große Textmengen als Eingaben zu verarbeiten. Dabei zeigt sich allerdings, dass Informationen zu Beginn und Ende des Textes mit höherer Sicherheit korrekt verarbeitet werden, als Informationen, die eher in der Mitte des Textes zu finden sind (Liu et al., 2024). Dieses unter dem Namen Lost-in-the-Middle bekannte Phänomen lässt sich bislang allerdings nicht zuverlässig erklären.

In der Zeit um die Veröffentlichung von ChatGPT zeigten große Sprachmodelle emergente Fähigkeiten, also Fähigkeiten, die sich nicht aus der einfachen Skalierung herkömmlicher Sprachmodelle vorhersagen ließen (Wei, Tay et al., 2022). Diese Fähigkeiten, die heute als Teil des Funktionsumfangs großer Sprachmodelle akzeptiert werden, waren zu diesem Zeitpunkt auch für Spezialisten nicht erklärbar.

Das Wissen um unser Unwissen, die Vorläufigkeit von Erkenntnissen und die kontinuierliche Weiterentwicklung bilden wichtige Grundlagen der Überlegungen für das Würzburger KI-Projekt. Auch wenn wir die im Hintergrund ablaufenden Prozesse nicht direkt kennen und verstehen können, geht es uns darum, an bisher dokumentierten Erkenntnissen anzuknüpfen und diese durch eigene Erfahrungen zu ergänzen.

2.2. Pragmatisches Prompt-Engineering

Unter einem Prompt versteht man einzelne oder eine Sammlung von Anweisungen, z. B. an große Sprachmodelle, auf deren Eingabe das Modell mit einer bestimmten Ausgabe reagiert (White et al., 2023). Unter dem Begriff des Prompt Engineerings versteht man nun eine Zusammenstellung von Techniken und Methoden, um Prompts zu formulieren, die zu tendenziell höherwertigen Ausgaben führen (Velásquez-Henao, Franco-Cardona & Cadavid-Higuaita, 2023). Während einige der vorgeschlagenen Techniken sehr einfach sind und z. B. auf Klarheit und Präzision der Anweisung, kontextbezogene Informationen, Angaben zum Antwortformat und zur gewünschten Ausführlichkeit der Antwort abzielen (Lo, 2023), warten andere mit einem komplexen Regelwerk mit durchaus ungewöhnlich anmutenden Tipps („Add ,I’m going to tip \$xxx for a better solution!““ (Bsharat, Myrzakhan & Shen, 2023, S. 5)) auf.

Dabei konnte gezeigt werden, dass einige Techniken wie das Chain-of-thought prompting, bei dem das Modell zur Begründung einzelner Schritte aufgefordert wird (Wei, Wang et al., 2022), oder das One-/Few-Shot-Prompting, bei dem dem Modell ein oder

mehrere Ausgabebeispiele in der Anweisung präsentiert werden (Brown et al., 2020), in der Praxis Erfolg versprechen können.

Im Würzburger KI-Projekt werden verschiedene Techniken des Prompt Engineering angewendet. Zudem werden in iterativen Schleifen die Prompts in simulierten Szenarien getestet, um sie anschließend zu verbessern.

2.3. Die Verantwortung bleibt beim Menschen

Die neuen und teilweise überraschenden Fähigkeiten scheinen auch die Sprache der Menschen im Zusammenhang mit künstlicher Intelligenz zu beflügeln. So wird in einer Vortragsankündigung von der geöffneten Büchse der Pandora (Gauß, 2024) gesprochen, während in einem Presseartikel von der Befreiung des Geistes aus der Flasche zu lesen ist (Sachse, 2023). Für die Implementierung in das LLL-Sem wird künstliche Intelligenz hingegen zunächst nüchtern als Werkzeug aufgefasst. Dies impliziert die Notwendigkeit der Übernahme der Verantwortung durch einen menschlichen Akteur – im Fall des Projektes durch die Dozierenden. Hiervon ist direkt das Ausmaß an eigenständiger Handlung, das einem KI-Werkzeug zugestanden wird, betroffen. Im Geiste des oben beschriebenen defensiven Vorgehens übernimmt die KI keine direkten Feedback-Aufgaben. Dabei handelt es sich aber bei diesem Beschluss um eine Momentaufnahme. So sind künftig durchaus Szenarien denkbar, in denen auch größere Aufgaben an künstliche Intelligenz ausgelagert werden können. Zum jetzigen Zeitpunkt und mit den in diesem Projekt verwendeten Werkzeugen erscheint das aber nicht angebracht.

3. Das Würzburger Lehr-Lern-Labor Physik

Das Würzburger Lehr-Lern-Labor-Seminar (LLL-sem) im Fach Physik ist eine Pflichtveranstaltung für Studierende des Lehramts an Gymnasien und Realschulen und folgt einem iterativen Aufbau (Damköhler, Elsholz & Trefzger, 2024; Elsholz & Trefzger, 2017). Der prinzipielle Ablauf ist in Abbildung 1 dargestellt.

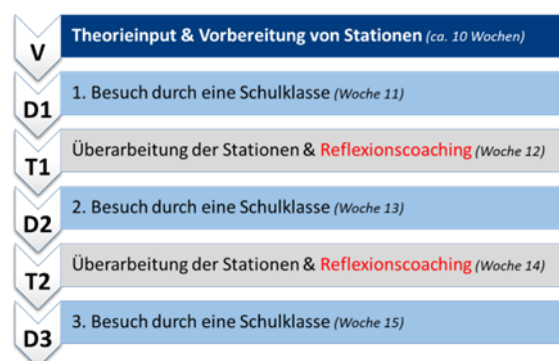


Abb. 1: Aufbau des LLL-Seminars an der Universität Würzburg

In einer mehrwöchigen Vorbereitungszeit (V) erhalten die Studierenden zunächst theoretischen Input zu relevanten Themen für das Lehren im LLL.

Daraufhin entwickeln sie unter der Betreuung von zwei Dozierenden Experimentierstationen, an denen die besuchenden Schülerinnen und Schüler während der späteren Durchführungstage (D1, D2, D3) selbstständig forschen dürfen. An den Stationen vermitteln die Studierenden innerhalb von 30 Minuten Gruppen von je rund vier bis sechs Schüler*innen physikalische Inhalte zu einem vorgegebenen Rahmenthema.

In den Wochen zwischen den Durchführungsterminen haben die Studierenden die Möglichkeit zur Überarbeitung und Weiterentwicklung ihrer Station. Dabei beziehen sie die Erfahrungen aus den vergangenen Durchführungen ein. In diesem Prozess werden sie von den Dozierenden durch Feedback und Ratschläge unterstützt. Darüber hinaus beschäftigen sich die Studierenden an diesen Tagen mit theoretischen und praktischen Inhalten, die den Umgang mit Erfahrungen, also Reflexionsprozesse selbst, in den Fokus nehmen (Damköhler, Elsholz & Trefzger, 2023). Zu diesen Inhalten gehören zum einen Reflexionsstrukturmodelle, wie das ALACT-Modell (Korthagen & Kessels, 1999) und das ERTO-Modell (Krieg & Kreis, 2014), oder das Zwiebelmodell (Korthagen, 2004), das verschiedene Determinanten des Lehrerhandelns in unterrichtlichen Situationen beleuchtet und dabei die Bedeutung der professionellen Identität der Lehrkraft betont. Zum anderen werden auch Fähigkeiten, die im Reflexionsprozess von Nutzen sind, wie z. B. die Noticingfähigkeit (van Es & Sherin, 2002) als Aspekt der professionellen Unterrichtswahrnehmung, thematisiert und eingeübt. Intensive Coaching-Gespräche zu diesen Themen können aus Zeitgründen jedoch nur in Ausnahmefällen geführt werden.

4. ChatGPT als Reflexionscoach

In der Adaption des LLL-Seminars unter Einbezug von ChatGPT soll ein Gesprächsagent (Chatbot) auf Basis von ChatGPT (gpt4) die Rolle eines Reflexionscoachs übernehmen. Für den ersten Versuch werden hierfür so genannte GPTs verwendet, Erweiterungen des ursprünglichen ChatGPT-Chatbots, die mithilfe von Prompts auf die eigenen Bedürfnisse angepasst werden können. Diese GPTs können Studierenden anschließend für eigene Chats zur Verfügung gestellt werden.

An den Überarbeitungs- und Reflexionsterminen (T1 und T2) erhalten die Studierenden zwar nach wie vor einen gewissen theoretischen Input, ein zentrales Element stellen aber nun Reflexionsgespräche mit ChatGPT dar, in denen der Chatbot die Studierenden durch einen strukturierten Reflexionsprozess begleitet. Dabei übernimmt der Chatbot die Gesprächsführung und strukturiert auf diese Weise den Prozess, bzw. wählt Betrachtungsschwerpunkte aus. Dabei sollen die Studierenden mit zwei verschiedenen Chatbots in Kontakt kommen, die, nach einer anfänglichen Phase der allgemeinen Einschätzungen, je einen unterschiedlichen Fokus haben.

Einer der Chatbots legt den Schwerpunkt auf die Analyse der vergangenen Stationsdurchführung hinsichtlich der Unterrichtsqualität mit einem besonderen Fokus auf der kognitiven Aktivierung. Als Grundlage für diese Betrachtungen wird das bereits in der Vorbereitungsphase (V) thematisierte Modell der Basisdimensionen guten Unterrichtens (Praetorius, Klieme, Herbert & Pinger, 2018) verwendet. Die Studierenden werden zu ihren Einschätzungen in Bezug auf die Erfüllung der drei Dimensionen Klassenführung, konstruktive Unterstützung und kognitive Aktivierung der Lernenden befragt und dabei im Sinne einer höheren Reflexionstiefe (vgl. hierzu Damköhler et al., 2024) um Erklärungen gebeten. Als adaptiv agierender Gesprächspartner reagiert der Chatbot hier flexibel auf Eingaben und fragt für mehr Details nach. Nach gründlicher Analyse des Verlaufs fordert der Chatbot die Studierenden auf, Vorschläge für alternative Handlungsmöglichkeiten zu nennen und die erhoffte Wirkung zu beschreiben und zu begründen. Das Gesprächsziel besteht in der Entwicklung eines konkreten Handlungsplans für die nächste Durchführung, der den identifizierten Schwächen gezielt begegnen und zur Erhöhung der Unterrichtsqualität führen soll.

Der zweite Chatbot verfolgt einen etwas anderen Ansatz. Nach einer allgemeinen Analysephase sollen die Studierenden wiederum Aspekte ihrer Stationsdurchführung identifizieren, für die sie sich Handlungsalternativen wünschen. In diesem Fall unterbreitet allerdings der Chatbot, nachdem er im Gespräch möglichst viele Details über die Stationsdurchführung und die empfundenen Schwächen dieser „eingesammelt“ hat, eine gewisse Zahl eigener Vorschläge für mögliche Handlungen bei der nächsten Durchführung. Im Sinne der Förderung eines kritischen Denkens im Umgang mit KI analysieren und bewerten die Studierenden anschließend – wiederum im Gespräch mit dem Chatbot die einzelnen Ideen und treffen eine Wahl für eine Handlungsalternative. Diese wird schließlich mit den übrigen Studierenden und den Dozierenden besprochen und das Urteil somit vor Umsetzung validiert. Im Gegensatz zum ersten Chatbot, der primär die analyse- bzw. reflexionsbezogenen Fähigkeiten der Studierenden adressiert, werden durch die Arbeit mit dem zweiten Chatbot gezielt Kompetenzen der Studierenden im Umgang mit KI-Werkzeugen angesprochen. Zudem wird eine Haltung der Studierenden gefördert, die durch ein KI-Werkzeug gelieferten Ergebnisse kritisch zu hinterfragen.

5. Erste Erkenntnisse und Ausblick

Mit dem beschriebenen Projekt wird das Ziel verfolgt, ein bestehendes Lehrveranstaltungsformat durch den Einsatz eines KI-Werkzeugs zu modifizieren und die Wirkung zu untersuchen. Der Einsatz des KI-Coaches und die gesammelten Erfahrungen sollen in einem nächsten Schritt genutzt werden, um neben der Reflexionsfähigkeit weitere unterrichtsrelevante Kom-

petenzen zu fördern, beispielsweise die experimentelle Kompetenz. Eine Anwendung von ChatGPT und die Übertragung der Erkenntnisse aus dem Projekt auf Lehrveranstaltungen anderer Fächer sind im interdisziplinär angelegten Didaktikzentrum M!ND-Center der Universität Würzburg (Völker & Trefzger, 2010) nicht nur denkbar, sondern auch gewünscht. Aus diesem Grund ist es wichtig, nicht nur Lehramtsstudierende der Physik, sondern auch aller anderen Lehrämter in Hinblick auf ihre Einstellungen zu künstlicher Intelligenz besser kennenzulernen.

Nach der Entwicklung eines ersten Konzepts wurden insgesamt 150 Studierende unterschiedlicher Fächer und Lehrämter an der Universität Würzburg zu ihren Einstellungen gegenüber künstlicher Intelligenz befragt. Hierbei wurden neben studienbezogenen Variablen auch Hoffnungen und Befürchtungen mit Bezug zum Einsatz von KI-Werkzeugen erhoben.

Dabei äußerten Studierende vor allem Sorgen, die Verluste betreffen (Abb. 2). Die Palette der Verluste reicht hierbei von Fähigkeitsverlusten – allen voran die Fähigkeit selbstständig zu denken – über Bedeutungsverluste hin zu Verlusten von Werten: Verlust von Fähigkeiten (Aufmerksamkeit, körperliche Fähigkeiten, Kreativität, Denken), Bedeutungsverlust von Menschen, Verlust von Tätigkeit, Kontrollverlust, Verlust von sozialen Beziehungen, Privatsphäre, Unabhängigkeit, Individualität, Werten, Glück, Kritikbereitschaft, Grenzen, Transparenz, Vertrauen, Persönlichkeit. Sorgen um einen missbräuchlichen Einsatz betrafen in der Mehrzahl die Prüfungsgerechtigkeit, aber auch die Befürchtung, Opfer von Desinformationskampagnen zu werden.

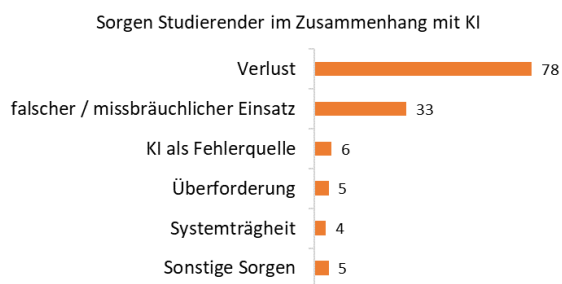


Abb. 2: Sorgen Studierender

Diesen Sorgen steht eine Reihe von genannten Hoffnungen (Abb. 3) gegenüber. So hofft die größte Zahl der Studierenden, durch den Einsatz von KI einen Produktivitätszuwachs, bzw. eine Effizienzsteigerung zu erfahren. Dieser Effizienzzuwachs betrifft vor allem Routineaufgaben aus dem Erfahrungsbereich der Studierenden. Hierzu zählen sie auch Rechercheaufgaben, schematisch ablaufende, sich wiederholende, Prozesse und sonstige Arbeitserleichterungen. Einige der Studierenden erhoffen sich allerdings auch direkte positive Auswirkungen auf ihren Unterricht, beispielsweise durch bessere Möglichkeiten zur Binnendifferenzierung oder eine abwechslungsreiche Unterrichtsgestaltung. Ein wiederum häufig genannter Aspekt betrifft die Nutzung von KI

zur Unterstützung kreativer Prozesse. Hier erkennen Studierende bereits das Potenzial der Technologie als „digitaler Kollege“ Ideen zu liefern, die sie selbst dann weiterverfolgen können. Einige Studierende benennen auch den (technologischen) Fortschritt an sich als eigenen Wert.

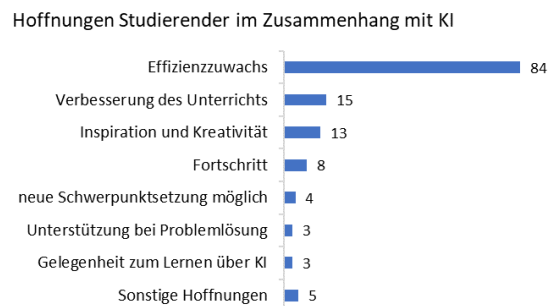


Abb. 3: Hoffnungen Studierender

Diese Studie wird durch eine zweite Erhebung mit der gleichen Zielgruppe an der Universität Würzburg ergänzt. In dieser sollen – neben Vorerfahrungen der Studierenden mit KI in Studium und Lehre – auch allgemeine Haltungen (Schepman & Rodway, 2023) und selbst eingeschätzte Fähigkeiten im Umgang mit KI (Carolus, Koch, Straka, Latoschik & Wienrich, 2023) abgefragt, sowie ein Fachtest zur Ermittlung dieser Fähigkeiten (Hornberger, Bewersdorff & Nerdel, 2023) durchgeführt werden. Die Ergebnisse dieser Erhebung können helfen, die untersuchte Gruppe der Teilnehmenden am LLL-Seminar Physik in Bezug zur Gesamtheit der Lehramtsstudierenden zu setzen.

Die Weiterentwicklung der Lehrveranstaltung knüpft bezüglich der Aspekte der Reflexionsförderung an ein Dissertationsprojekt im LLL-Seminar Physik, das sich mit der Untersuchung von Reflexionsprozessen beschäftigt, an (Damköhler et al., 2024). Daher wird, wie im Dissertationsprojekt, während der Durchführung dieses Projekts die Reflexionspraxis der Studierenden durch qualitative Erhebungen an den Durchführungstagen begleitend untersucht. Abschließend wird das Veranstaltungskonzept im Rahmen von Interviews unter den teilnehmenden Studierenden evaluiert.

Bei der Evaluation der Lehrveranstaltung, die im Sommersemester 2024 erstmalig in der beschriebenen Form stattfinden wird, stehen, neben prinzipiellen Betrachtungen zur Realisierbarkeit und konkreten Umsetzung und zur Entwicklung der Reflexionspraxis der Studierenden, vor allem Fragen zur Akzeptanz der Methode durch die Studierenden und zur Wirkung der digitalen Reflexionscoachings im Zusammenspiel mit klassischen Betreuungsgesprächen im Vordergrund. Unter anderem wird dabei untersucht, wie Studierende die Betreuung durch eine KI erleben und wie sie die mit dieser in Zusammenhang stehenden Veränderung der Betreuungsbeziehungen empfinden.

6. Literatur

- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P. et al. (2020, 28. Mai). *Language Models are Few-Shot Learners*. Verfügbar unter: <http://arxiv.org/pdf/2005.14165v4>
- Bsharat, S. M., Myrzakhan, A. & Shen, Z. (2023, 26. Dezember). *Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4*. Verfügbar unter: <http://arxiv.org/pdf/2312.16171v2>
- Carolus, A., Koch, M., Straka, S., Latoschik, M. E. & Wienrich, C. (2023). *MAILS -- Meta AI Literacy Scale: Development and Testing of an AI Literacy Questionnaire Based on Well-Founded Competency Models and Psychological Change-and Meta-Competencies*. <https://doi.org/10.48550/arXiv.2302.09319>
- Damköhler, J., Elsholz, M. & Trefzger, T. (2023). Förderung der Reflexionskompetenz im Lehr-Lern-Labor. Unterstützung der Entwicklung der Reflexionskompetenz Studierender in einem Lehr-Lern-Labor-Seminar. In H. Grötzebauch & S. Heinicke (Hrsg.), *PhyDid B, Didaktik der Physik. Beiträge zur DPG-Frühjahrstagung 2023*. Bad Honnef: Fachverband Didaktik der Physik.
- Damköhler, J., Elsholz, M. & Trefzger, T. (2024). Förderung der Reflexionskompetenz angehender Physiklehrkräfte im Lehr-Lern-Labor (im Druck). In M. Jungwirth, J. Haarmann, N. Harsch, F. Haupts, J. Marks & Y. Noltensmeier (Hrsg.), *Wegmarken für eine zeitgemäße Lehrkräftebildung – Konzeptionelle Ansätze im Fokus. Tagungsband des 16. Bundeskongresses der Zentren für Lehrer*innenbildung* (Schriften zur allgemeinen Hochschuldidaktik, Bd. 9). Münster.
- Elsholz, M. & Trefzger, T. (2017). Professionalisierung durch Praxisbezug – Begleitforschung zu den Würzburger Lehr-Lern-Laboren. In C. Maurer (Hrsg.), *Implementation fachdidaktischer Innovation im Spiegel von Forschung und Praxis*. (S. 488–491). Regensburg: Universität Regensburg.
- 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>
- Hochschule Ravensburg-Weingarten. (2024). *Künstliche Intelligenz - Die Büchse der Pandora ist geöffnet*. Weingarten.
- Hornberger, M., Bewersdorff, A. & Nerdel, C. (2023). What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence*, 5, 100165. <https://doi.org/10.1016/j.caeai.2023.100165>
- Kasneji, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F. et al. (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>
- Korthagen, F. (2004). In search of the essence of a good teacher: towards a more holistic approach in teacher education. *Teaching and Teacher Education*, 20(1), 77–97. <https://doi.org/10.1016/j.tate.2003.10.002>
- Korthagen, F. & Kessels, J. P. A. M. (1999). Linking Theory and Practice: Changing the Pedagogy of Teacher Education. *Educational Researcher*, 28(4), 4–17. <https://doi.org/10.3102/0013189X028004004>
- Krieg, M. & Kreis, A. (2014). Reflexion in Mentoringgesprächen - ein Mythos? *Zeitschrift für Hochschulentwicklung*, 9(1). <https://doi.org/10.3217/zfhe-9-01/11>
- Leavy, S., O'Sullivan, B. & Siaper, E. (2020, 28. Juli). *Data, Power and Bias in Artificial Intelligence*.
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I. & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), 100790. <https://doi.org/10.1016/j.ijme.2023.100790>
- Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F. et al. (2024). Lost in the Middle: How Language Models Use Long Contexts. *Transactions of the Association for Computational Linguistics*, 12, 157–173. https://doi.org/10.1162/tacl_a_00638
- Lo, L. S. (2023). The Art and Science of Prompt Engineering: A New Literacy in the Information Age. *Internet Reference Services Quarterly*, 27(4), 203–210. <https://doi.org/10.1080/10875301.2023.2227621>
- Peters, U. (2023). Explainable AI lacks regulative reasons: why AI and human decision-making are not equally opaque. *AI and Ethics*, 3(3), 963–974. <https://doi.org/10.1007/s43681-022-00217-w>
- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM – Mathematics Education*, 50(3), 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Sachse, M. (2023, 11. Mai). Der Geist ist aus der Flasche. *Frankfurter Allgemeine Zeitung online* (Kommentar). Verfügbar unter: <https://www.faz.net/aktuell/wirtschaft/kuenstliche-intelligenz-der-geist-ist-aus-der-flasche-18884740.html>
- Salvagno, M., Taccone, F. S. & Gerli, A. G. (2023). Artificial intelligence hallucinations. *Critical Care (London, England)*, 27(1), 180. <https://doi.org/10.1186/s13054-023-04473-y>
- Schepman, A. & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory Validation and Associations with Personality, Corporate Distrust, and General Trust. *International Journal of Human-Computer Interaction*, 39(13), 2724–2741. <https://doi.org/10.1080/10447318.2022.2085400>

- Van den Berg, G. & Du Plessis, E. (2023). ChatGPT and Generative AI: Possibilities for Its Contribution to Lesson Planning, Critical Thinking and Openness in Teacher Education. *Education Sciences*, 13(10), 998.
<https://doi.org/10.3390/educsci13100998>
- Van Es, E. & Sherin, M. (2002). Learning to Notice: Scaffolding New Teachers' Interpretations of Classroom Interactions. *Journal of Information Technology for Teacher Education*, 10(4), 571–596.
- Velásquez-Henao, J. D., Franco-Cardona, C. J. & Cadavid-Higuaita, L. (2023). Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering. *DYNA*, 90(230), 9–17.
<https://doi.org/10.15446/dyna.v90n230.111700>
- Vogelsang, J., Bleher, J., Krupitzer, C., Stein, A. & Jung, R. (2023, 7. Februar). *Nutzung von ChatGPT in Lehre und Forschung – eine Einschätzung der AIDAHO-Projektgruppe*. Hohenheim: Universität Hohenheim.
- Völker, M. & Trefzger, T. (2010). Lehr-Lern-Labore zur Stärkung der universitären Lehramtsausbildung. In *Phydid B, Didaktik der Physik, Beiträge zur DPG-Frühjahrstagung in Hannover*. Bad Honnef: Fachverband Didaktik der Physik.
- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J. et al. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6).
<https://doi.org/10.1007/s11704-024-40231-1>
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S. et al. (2022, 15. Juni). *Emergent Abilities of Large Language Models*. Verfügbar unter:
<http://arxiv.org/pdf/2206.07682v2>
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F. et al. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho & A. Oh (Hrsg.), *Advances in Neural Information Processing Systems* (Bd. 35, S. 24824–24837). Curran Associates, Inc. Verfügbar unter:
https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H. et al. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*.
<https://doi.org/10.48550/arXiv.2302.11382>

Danksagung

Das beschriebene Projekt wird im Rahmen von WueDive gefördert. WueDive ist ein von der „Stiftung Innovation in der Hochschullehre“ gefördertes Projekt zur Weiterentwicklung digitaler Lehre an der Universität Würzburg. WueDive initiiert und koordiniert mit der Einrichtung eines Digital Higher Education Hub die Auseinandersetzung mit neuen Lernanforderungen und Lerngelegenheiten.

Transparenzhinweis

Bei der Anfertigung dieses Artikels wurden Werkzeuge der generativen KI zur Literaturrecherche und dem Übersetzen von Artikeln verwendet. Sämtliche Aktivitäten wurden durch die Autoren überwacht und überprüft, so dass sie die Verantwortung entsprechend tragen.