

## **LLMs wie ChatGPT als individualisierte Tutoren im Mathematikunterricht: Eine explorative Annäherung**

Wir untersuchen den potenziellen Einsatz von textgenerierenden Large Language Models (LLMs) als maßgeschneiderte Tutoren im Mathematikunterricht. Trotz des wachsenden Interesses und der scheinbar vielfältigen Verwendungsmöglichkeiten im Unterricht identifizieren wir erhebliche Herausforderungen sowohl in Form inhärenter Schwächen solcher Modelle als auch hinsichtlich verbreiteter Fehlvorstellungen über deren Funktionsweise. Unsere Erhebung mit 54 Mathematik-Lehramtsstudierenden zeigt eine ambivalente Haltung bezüglich einer solchen Anwendung. Anhand dieser Erhebung und eines von uns entwickelten prototypischen LLM-Tutors diskutieren wir Grenzen, Möglichkeiten und Herausforderungen im Mathematikunterricht.

### **Einleitung: Illusion der Intelligenz**

Large Language Models (LLMs) wie ChatGPT sind autoregressiv arbeitende mathematische Modelle aus dem Bereich der Künstlichen Intelligenz (KI), die zur Textgenerierung eingesetzt werden können. Die massive Erhöhung der Parameteranzahl in KI-Modellen erweitert zwar das Spektrum ihrer Anwendungsmöglichkeiten, doch reduziert die großteils automatisierte Kalibrierung solcher Modelle mittels umfangreicher Datensätze leicht die Qualität der generierten Texte auf einen generischen Durchschnitt. Diese Gefahr kann man illustrieren, indem man ChatGPT nach der Korrektur eines Textes fragt, der korrekte Formulierungen weitverbreiteter Rechtschreib- oder Grammatikfehler beinhaltet, wie in Abb. 1 dargestellt. Fehlvorstellungen und Missverständnisse rund um KI, wie die unrealistische Auffassung, dass künstliche neuronale Netze analog zum menschlichen Gehirn funktionierten, werden seit Jahren auch innerhalb der KI-Forschung diskutiert (Emmert-Streib et al., 2020). Dennoch sind im Zuge der hohen medialen Aufmerksamkeit auf KI-Modelle seit der Veröffentlichung von ChatGPT auch in der deutschen Bildungslandschaft unrealistische Bedenken und Hoffnungen wahrnehmbar: Einerseits werden etwa eine umfangreiche Transformation der Lehrpersonenrolle vorausgesagt (Weßels, 2023) und Behauptungen geäußert, nach denen schriftliche Hausarbeiten zunehmend an Relevanz verlören (Bialonski & Grieger, 2023); andererseits mehren sich unreflektierte Vorschläge und Anleitungen für den Einsatz von KI-Modellen im Unterricht und für dessen Gestaltung. Beides scheint kritisch und wir befürchten, dass diese Mischung aus überzogenen Ängsten und unrealistischen Erwartungen die unreflektierte und unkontrollierte Verwendung im Unterricht begünstigen wird.

<b>Prompt:</b>	Korrigiere die Fehler: „Das vortragende Mädchen, dessen Verständnis für Moduln überzeugte, verlöre das Überraschungsmoment seines Vortrags, wenn es dabei nicht so frei und extravertiert sänge.“
<b>GPT-4:</b>	In dem Satz [...] sind einige Fehler: [...]
<b>Prompt:</b>	Meine Professorin sagte aber, Pronomina seien [...]
<b>GPT-4:</b>	Ihre Professorin hat recht [...] [Trotzdem...] „ihres Vortrags“ statt „seines Vortrags“.
<b>Prompt:</b>	Identifiziere die Genera der Substantive in diesem Satz, die für Wahl und Flexionsform des Possessivartikels relevant sind. [...]

Abb. 1: Wahrscheinlichkeitsbasierte Antworten statt Grammatikwissens

Diese grammatikalischen Ausnahmefälle sollen überspitzt das Risiko der Reduzierung auf einen generischen Durchschnitt illustrieren, das ebenso deutlich wird, wenn man versucht, ChatGPT einen logisch kohärenten, inhaltlich anspruchsvollen und argumentativ überzeugenden Gedanken über mehrere Seiten verfolgen zu lassen. Deshalb generieren LLMs auch falsche Lösungen zu einfachsten mathematischen Rechenaufgaben, samt plausibler Beschreibung des falschen Rechenweges, wie Cap (2023) treffsicher darstellt. Derartige Ausgaben, oft Halluzinationen genannt, könnten im Rahmen eines unreflektierten Einsatzes von LLMs zur Förderung falscher Annahmen beitragen.

### Machbarkeit und Einstellungen zu LLMs im Mathematikunterricht

Neben dem wachsenden Angebot KI-gestützter Unterrichtsgestaltung gibt es mittlerweile auch erste didaktische Erwägungen zur Anwendung textgenerierender KI im Mathematikunterricht (Spannagel, 2023). Angesichts steigender Leistungsunterschiede und Zusatzbelastungen verlockt die Idee eines frei zugänglichen, maßgeschneiderten LLM-Tutors: Ein solcher könnte durch die Anpassung an das Leistungsniveau des jeweiligen Kindes, die simultane individuelle Förderung durch die Lehrkraft erleichtern, wie in Abb. 2 illustriert.

<b>Schüler*in:</b>	Ich soll $x^2 + 12x + 36$ mithilfe der binomischen Formeln faktorisieren, wie?
<b>LLM-Tutor:</b>	Was verstehst du unter „faktorisieren“? Welche binomischen Formeln [...]
<b>Schüler*in:</b>	Ahh, der Term ist von der Form $a^2 + 2ab + b^2 = (a + b)^2$ , oder?
<b>LLM-Tutor:</b>	Das ist möglich. Kannst du $a$ und $b$ so bestimmen, dass $(a + b)^2 = x^2 + 12x + 36$ ?
<b>Schüler*in:</b>	Ja, $(x + 6)^2$ oder?
<b>LLM-Tutor:</b>	Kannst du das überprüfen? Kannst du auch $18x^2 - \frac{16}{25}$ faktorisieren? [...]

Abb. 2: Vision eines LLM-Tutors

Neben der technischen Realisierbarkeit eines solchen im Mathematikunterricht einsetzbaren LLM-Tutors interessieren uns auch die Einstellungen der angehenden Lehrer\*innen bezüglich LLMs wie ChatGPT und deren potenzieller Verwendung im Mathematikunterricht.

## Untersuchungsdesign und erste Eindrücke

Für eine erste Annäherung an Antworten auf diese Fragen entwickelten wir für unsere Erhebung den prototypischen LLM-Tutor „Pi-Pilot“ als einfache PHP-Webanwendung mittels JavaScript und curl-Aufrufen der OpenAI Chat Completions API. Diese bietet neben der Anpassung bestimmter Parameter, etwa um die Zufälligkeit der generierten Antworten anzupassen, auch einfache Möglichkeiten, die Eingabe von Nutzer\*innen über zusätzlichen Kontext zu ergänzen, um so das Antwortverhalten in eine gewünschte Richtung zu lenken. Zur Erhebung der persönlichen Präferenzen zukünftiger Mathematiklehrkräfte entwickelten wir einen Fragebogen, der von 54 Studierenden ausgefüllt wurde.

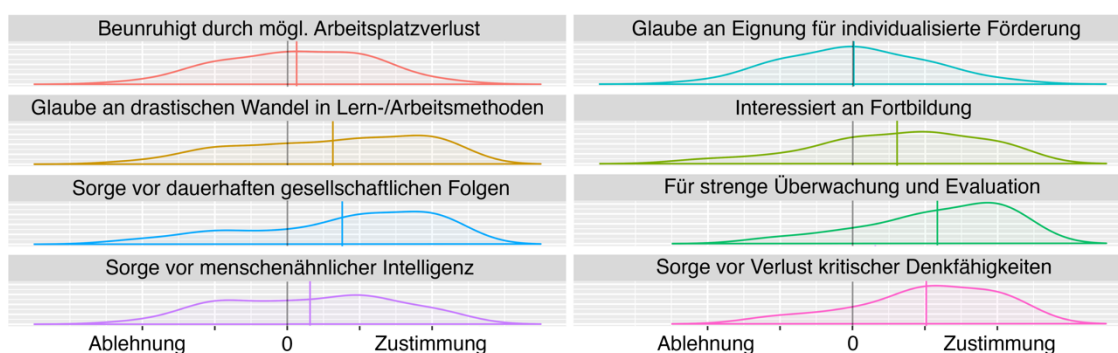


Abb. 3: Einstellungen zu LLMs (links) und deren Einsatz im Unterricht (rechts)

Während unsere Experimente mit „Pi-Pilot“ bereits durch einfache Kontextergänzungen zu einer merklich schülergerechteren Sprache in den Ausgaben führten, war es uns nicht möglich, die Antworten zuverlässig auf spezifische mathematische Themen zu beschränken. Beispielsweise wechselte „Pi-Pilot“ unerwartet auf eine Nachfrage zu den platonischen Körpern zum Tesseract. Durch das Erzwingen deterministischerer Antworten konnten wir zunächst die Wahrscheinlichkeit von Falschausgaben auf einfache mathematische Rechenaufgaben reduzieren. Doch dadurch erhöhten sich dann bei simplen Knobelaufgaben die falschen Antworten. Während wir durch Prompt Engineering und Parameteränderungen immer wieder spezifische Probleme reduzieren konnten, führte dies stets zu neuen Schwierigkeiten im Antwortverhalten. Die Befragung der Studierenden ergab ein ähnliches Bild: Sie nannten auf die Frage nach negativen Erfahrungen mit LLMs wie ChatGPT meist plausible, aber falsche Antworten auf einfache mathematische Fragen. Bezüglich ihrer Einstellungen ergab sich ein ambivalentes Bild: Obwohl die Mehrheit den Wunsch nach Fortbildung für den Einsatz von LLMs im Unterricht äußerte, schienen Sorgen und Ängste zu überwiegen. An diese Eindrücke knüpfen wir durch eine Postuntersuchung an, bei der die Studierenden in der Rolle von Schüler\*innen mit „Pi-Pilot“ experimentieren. Wir wollen erfahren, ob und wie sich Einstellungen dadurch verändern und ob

durch bloße Exposition eine Reduktion der Fehlvorstellungen über KI möglich ist. Parallel dazu erkunden wir weitere Einsatzmöglichkeiten LLM-basierter Tutoren für den Mathematikunterricht. Während unser langfristiges Forschungsziel hierbei ein gezielt trainiertes Open-Source-LLM ist, fokussieren wir uns zunächst auf die weitere Exploration der Möglichkeiten der OpenAI API mittels Assistenzsystemen als nächsten pragmatischen Schritt in diese Richtung.

## Diskussion und Perspektiven

Die Angst, dass LLMs in näherer Zukunft die Notwendigkeit für das Anfertigen von Hausarbeiten auflösen könnten, sehen wir derzeit als ebenso unbegründet an, wie die Forderung, deren Nutzung im Bildungskontext zu verbieten. Für Lehrpersonen erscheinen uns ausgewogene und pragmatische Anleitungen zum Einsatz von LLMs wie ChatGPT in und für den Unterricht, ähnlich denen des LI Hamburg (2023), als hilfreich und sinnvoll. Solche Hinweise tragen zu einem realistischeren Bild von den Möglichkeiten und Grenzen von LLMs bei. Von didaktisch orientierter Forschung zur Entwicklung LLM-basierter Tutoren versprechen wir uns, nicht allein auf die Analyse der zweifelsfrei entstehenden proprietären Lösungen angewiesen sein zu müssen, oder zumindest diese dann fundierter einschätzen zu können.

## Literatur

- Bialonski, S., & Grieger, N. (2023). Der KI-Chatbot ChatGPT: Eine Herausforderung für die Hochschulen. *Die Neue Hochschule*, 2023(1), 24–27. <https://doi.org/10.5281/zenodo.7533758>
- Cap, C. H. (2023). „Der neue Gott ist nackt!“ ChatGPT im Bildungswesen. In *Forschung & Lehre* (Bd. 30, Nummer 5, S. 344–345).
- Emmert-Streib, F., Yli-Harja, O., & Dehmer, M. (2020). Artificial Intelligence: A Clarification of Misconceptions, Myths and Desired Status. *Frontiers in Artificial Intelligence*, 3. <https://doi.org/10.3389/frai.2020.524339>
- LI Hamburg. (2023). *Fachbrief zur Künstlichen Intelligenz*. <https://li.hamburg.de/ki>
- Spannagel, C. (2023). Hat ChatGPT eine Zukunft in der Mathematik?. *Mitteilungen der Deutschen Mathematiker-Vereinigung*, 31(3), 168–172. <https://doi.org/10.1515/dmvm-2023-0055>
- Weßels, D. (2023). Meilenstein der KI-Entwicklung? Der Chatbot ChatGPT. In *Forschung & Lehre* (Bd. 30, Nummer 1, S. 26–27).