

Herausforderungen und Entwicklungsmöglichkeiten für die Mathematikdidaktik durch generative KI-Sprachmodelle

Nils Buchholtz, Lukas Baumanns, Judith Huget, Franziska Peters, Maximilian Pohl und Sebastian Schorcht

Am 30. November 2022 veröffentlichte das US-amerikanische Unternehmen OpenAI die Anwendung ChatGPT seines generativen KI-Sprachmodells GPT-3 (Generative Pre-Trained Performer) und ermöglichte somit der Allgemeinheit die kostenlose Nutzung der Anwendung (erreichbar unter der URL chat.openai.com/chat). Bei ChatGPT handelt es sich um einen auf künstlicher Intelligenz basierenden Chatbot, der Inhalte menschlicher Sprache mit beeindruckender Genauigkeit versteht und auch erzeugt.

ChatGPT basiert auf einem neuronalen Netzwerk-Modell maschinellen Lernens, das in der Lage ist, menschliche Eingaben in Textform zu analysieren und darauf einzugehen – etwa im Sinne von Benutzeranfragen. Auf Basis großer Mengen von Trainingsdaten wurde es dahingehend entwickelt, Muster und Beziehungen in der Sprache zu erkennen und auf die Eingaben passende Antworten zu generieren, um damit menschenähnliche Konversation zu betreiben. Der Lernprozess der KI basiert auf einem Reinforcement-Learning-Algorithmus, der auf Belohnungen (durch zunächst menschliches Feedback) und stochastischen Vorhersagen über die Wahrscheinlichkeit der Maximierung von Belohnungen bei der Optimierung von Konversationshandlungen (passenden Antworten) beruht (Schulman et al., 2017). Die Größe der Trainingsdaten wurde

nicht veröffentlicht, allerdings wurde das neuronale Netzwerk des Vorgängers GPT-3 mit 175 Milliarden Parametern trainiert (Floridi & Chiriatti, 2020). Die Datenbanken von ChatGPT werden somit fortlaufend aktualisiert, um Antworten auf dem neuesten Stand zu halten (der Stand der Trainingsdaten ist 2021). Dabei merkt sich ChatGPT auch frühere Eingaben in einem Gespräch, um an Folgefragen anzuknüpfen oder darauf Bezug zu nehmen.

Die Dichte (und Güte) der Informationen, die ChatGPT auf unsere Eingangsfrage in Abb. 1 bereitstellt, sind für eine allgemeinverständliche Erklärung für Laiinnen und Laien durchaus befriedigend. Sie verbleiben jedoch weit unter der Ausführlichkeit eines gleichartigen Wikipedia-Artikels, der sich über eine einfache Google-Anfrage aufrufen lässt (de.wikipedia.org/wiki/Mathematikdidaktik). Dennoch besteht ein entscheidender Unterschied zwischen einer Anfrage bei ChatGPT und einer Google-Suche, den wir hier problematisieren wollen: Während ein Wikipedia-Artikel allgemein zugängliche Informationen bereitstellt, die der/die Anfragende selbstständig lesen und im Hinblick auf eine konkrete Fragestellung einordnen und in Bezug auf die Güte und Nützlichkeit bewerten muss, kann ChatGPT eine zu den präsentierten Informationen eingegebene Nachfrage verarbeiten und beantworten (vgl. Abb. 2).

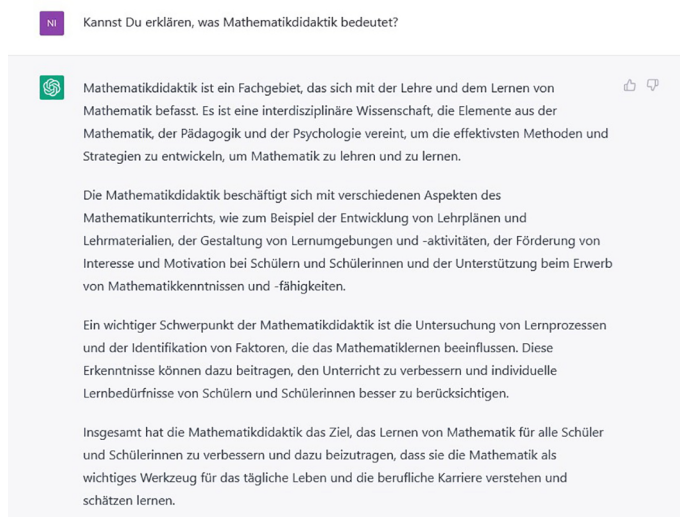


Abbildung 1. Beispiel einer ChatGPT-Antwort (Prompt von N. Buchholtz am 27. 12. 2022)

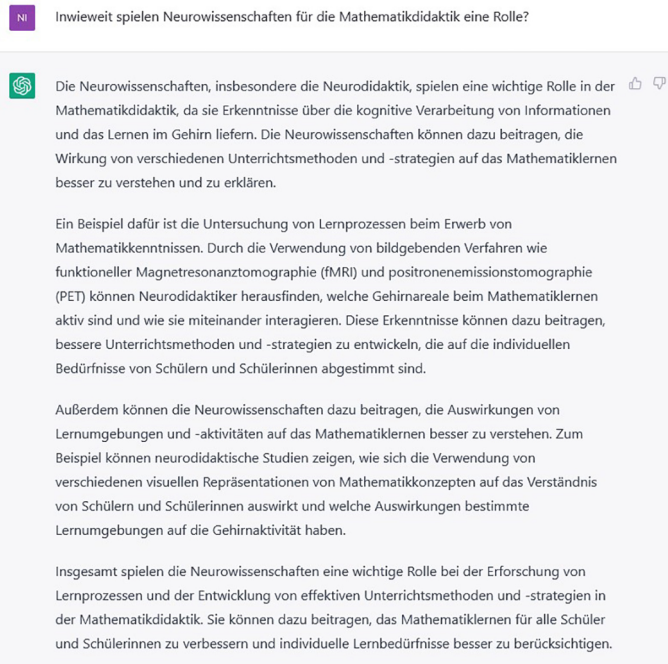


Abbildung 2. Beispiel einer ChatGPT-Antwort auf die Frage nach spezifischen Zusammenhängen von Neurowissenschaften und Mathematikdidaktik (Prompt von N. Buchholtz am 27. 12. 22)

Während diese Antwort ebenfalls plausibel klingt, ist ohne entsprechende Expertise jedoch nicht erkennbar, ob die präsentierten Informationen korrekt und vollständig sind. Hier liegt das Grundproblem generativer KI-Sprachmodelle: ChatGPT beruht auf der KI-basierten Zusammenstellung möglichst erfolversprechender Konversationshandlungen und ist damit vom Feedbackverhalten von Menschen und den zugrundeliegenden Trainingsdaten abhängig, sodass der Wahrheitsgehalt der KI-basierten Antwort nicht der entscheidende Parameter ist und damit nicht immer ge-

währleistet ist. Dies gilt zwar prinzipiell für nicht-wissenschaftliche Quellen wie z. B. einen Wikipedia-Eintrag; bei diesem besteht jedoch zumindest noch eine Qualitätsprüfung durch die Autorencommunity. Die als „Datenhalluzinationen“ bezeichneten generierten Antworten der KI können sogar so detailliert sein, dass bei gezielter Nachfrage nach wissenschaftlichen Publikationen auch konkrete real existierende Journals genannt werden (inklusive plausibler Nennung von Ausgabe und Seitenangaben, vgl. Abb. 3). Die Quellenangaben in Abb. 3 sind jedoch zum großen Teil inkorrekt bzw.

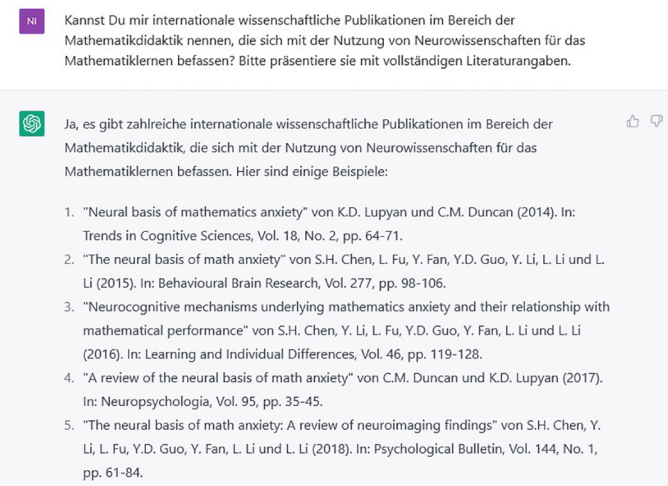


Abbildung 3. Beispiel einer spezifischeren ChatGPT-Antwort (Prompt von N. Buchholtz am 29. 12. 22)

existieren gar nicht. Es gibt in unserem Beispiel zwar die Fachzeitschrift *Trends in Cognitive Sciences*, der angegebene Artikel ist jedoch frei erfunden und erfüllt lediglich erfolgreich unsere Anfrage (vgl. Abb. 3).

Der Grad der Detailliertheit der bereitgestellten Antworten ist verblüffend. Angefangen von alltagsbasierten Anfragen (z. B. „Erzähle einen Witz“; „Schreibe ein Märchen über eine Maus und einen Hasen“) ist ChatGPT sogar in der Lage, ganze Essays oder publikationsreife Blogartikel zu verfassen (vgl. Lobo, 2022) oder komplexe Dialoge in unterschiedlichen Bereichen zu führen (z. B. über die Qualität von technischem Equipment).

Bereits jetzt werden die vermuteten weitreichenden Auswirkungen dieser und anderer generativer KI-Sprachmodelle auf die Arbeit von Forschenden und Lehrenden im Bildungsbereich diskutiert (Weßels, 2022). Durch die zunehmende Bekanntheit der Technologie und die mediale Aufmerksamkeit, die ChatGPT in der letzten Zeit zuteil wurde (Moorsstedt, 2022; Jahn, 2022; Lobo, 2022), ist anzunehmen, dass sich der Nutzen dieser oder ähnlicher Anwendungen in der Gesellschaft langfristig verbreiten wird, und Forschende wie Lehrende im Schul- und Hochschulbereich in zunehmendem Maße mit ganzen KI-generierten Texten konfrontiert sein werden. Es stellen sich aus unserer Sicht daher sowohl gleichermaßen Fragen der Herausforderungen für einen produktiven Umgang mit generativen KI-Sprachmodellen wie Fragen der Anwendungsmöglichkeiten, die eine solche Technologie für die universitäre Forschung und Lehre sowie schulische Kontexte bereithalten kann. In diesem Beitrag wollen wir daher auf mögliche Herausforderungen der Anwendung eines solchen Sprachmodells für die Mathematikdidaktik eingehen und in diesem Zusammenhang einen möglichen Umgang mit der Technologie in der Forschung diskutieren.

Anwendungen im Bereich der mathematikdidaktischen Forschung und Lehre

In mathematikdidaktischen Forschungszusammenhängen lässt sich ChatGPT durchaus als Ideensammlung verwenden, indem beispielsweise Eingaben zu bestimmten Forschungsfragen oder Ideen zu Forschungsprojekten gesetzt werden. Der Nutzen kann hier darin bestehen, dass die KI aus ihren Trainingsdaten eine Antwort auf eine präzise Frage in natürlicher und verständlicher Sprache generiert. Damit unterscheidet sich ChatGPT von einer Literaturrecherche auf Google Scholar oder anderen Suchdatenbanken, bei der vorrangig Literatur im peripheren Umfeld der Frage präsentiert wird. Dies ist insbesondere dann hilfreich, wenn Forschungs-


fragen noch vage sind oder nach ganz bestimmten Zusammenhängen gefragt wird. Das „Halluzinieren“ der KI führt zum Beispiel zu kreativen Zusammenstellungen von Antworten, die als Denkanstoß dienen können.


Wir haben zunächst ChatGPT gefragt, wo es Anwendungsmöglichkeiten innerhalb der mathematikdidaktischen Forschung für sich selbst sieht. Im Wesentlichen nennt ChatGPT hierzu die eigenen Anwendungsmöglichkeiten im Bereich der Entwicklung und Evaluation individueller Lehr- und Lernmaterialien, der Diagnostik von Schwierigkeiten von Schülerinnen und Schülern beim Mathematiklernen und der Analyse von Schülerantworten (vgl. Abb. 4).

Verwendung von ChatGPT als Analysetool zur Evaluation

Von Vorteil bei ChatGPT ist die direkte Nachfrage bezüglich weiterer Ideen oder die beliebige Vertiefung der Diskussion eines Inhalts. Beispielhaft möchten wir Aspekt 5 der obigen Liste vertiefen (vgl. Abb. 5). Mit den Möglichkeiten, generative KI für die Analyse von Schülerdaten einzusetzen, ergeben sich hier Ähnlichkeiten etwa zu Ansätzen des Educational Data Mining oder der Learning Analytics, deren Nutzen im Bildungsbereich allerdings aufgrund von verengten behavioristischen Annahmen über Lernprozesse umstritten ist (Knox et al., 2019). Kern dieser Ansätze ist das extensive Sammeln von großen Mengen von Daten über die Lernaktivitäten von Schülerinnen und Schülern, um aus automatisierten statistischen Analysen didaktische Entscheidungen abzuleiten (z. B. bei der Entwicklung adaptiver Lernsysteme oder der Identifizierung von Schülerinnen und Schülern mit besonderen Schwierigkeiten beim Mathematiklernen), den Lernerfolg vorherzusagen oder die Wirksamkeit von pädagogischen Maßnahmen zu beurteilen und zu optimieren (Shin & Shim, 2020). Dabei sind die Möglichkeiten individualisierter Unterstützungsmaßnahmen einerseits und die Gefahr der Reduktion der Lernprozesse von Schülerinnen und Schülern auf „auswertbare“ Daten andererseits, stets abzuwägen.

Mit Bitte um Vertiefung scheint ChatGPT hierzu zwar genauere Angaben zu unterbreiten, bleibt aber in der Ausführung vage. So nennt die Anwendung konkrete Punkte, wie die Sammlung von Schülerantworten und -fragen oder die Simulation von Lehrmethoden, führt aber andererseits nicht aus, wie dies ganz konkret umgesetzt werden soll oder welche Lehrmethoden hier auf ihre Wirksamkeit hin überprüft werden. Auch die Visualisierung von Daten ist mit ChatGPT bislang nicht möglich, obwohl dies vom Modell selbst als Anwendungsfall beschrieben wird. Hier bleibt abzuwarten, ob zu-


 Erstelle eine Liste von Aspekten, wie ChatGPT in der mathematikdidaktischen Forschung eingesetzt werden kann.

 Einige Möglichkeiten, wie ChatGPT in der mathematikdidaktischen Forschung eingesetzt werden könnte, sind:

1. Unterstützung bei der Erstellung von Lehrmaterialien: ChatGPT könnte verwendet werden, um automatisch Lehrmaterialien zu generieren, die auf den Bedarf und die Fähigkeiten der Schüler abgestimmt sind.
2. Personalisierter Unterricht: ChatGPT könnte eingesetzt werden, um personalisierten Unterricht zu bieten, indem es Schülern individuelle Aufgaben und Anweisungen gibt, um ihre Fähigkeiten und Kenntnisse zu verbessern.
3. Diagnose von Lernproblemen: ChatGPT könnte verwendet werden, um Schülern zu helfen, Schwierigkeiten im Mathematikunterricht zu identifizieren und zu lösen, indem es sie auf bestimmte Probleme aufmerksam macht und ihnen Feedback gibt.
4. Unterstützung bei der Gestaltung von Lehrplänen: ChatGPT könnte verwendet werden, um Lehrern bei der Erstellung von Lehrplänen und Lernaktivitäten zu helfen, die den individuellen Bedürfnissen und Fähigkeiten ihrer Schüler entsprechen.
5. Unterstützung bei der Evaluation von Lehrmethoden: ChatGPT könnte verwendet werden, um die Effektivität verschiedener Lehrmethoden zu untersuchen, indem es beispielsweise Schülerfragen und -antworten sammelt und analysiert.
6. Erstellung von Lernspielen: ChatGPT könnte verwendet werden, um Lernspiele zu entwickeln, die Schülern dabei helfen, Mathematik spielerisch zu erlernen und zu üben.
7. Unterstützung bei der Durchführung von Experimenten: ChatGPT könnte verwendet werden, um die Durchführung von Experimenten in der Mathematikdidaktik zu unterstützen, indem es beispielsweise Fragen stellt und Daten sammelt.

Abbildung 4. Anwendungsmöglichkeiten von ChatGPT in der mathematikdidaktischen Forschung (Prompt von L. Baumanns am 29. 12. 22)

 Könntest du auf Aspekt 5 dieser Liste noch genauer eingehen?

 Gerne. Der Einsatz von ChatGPT in der Evaluierung von Lehrmethoden könnte folgendermaßen aussehen:

1. Sammlung von Schülerdaten: ChatGPT könnte eingesetzt werden, um Schülerfragen und -antworten während des Mathematikunterrichts zu sammeln. Diese Daten könnten dann analysiert werden, um herauszufinden, welche Lehrmethoden am effektivsten sind.
2. Simulation von Lehrmethoden: ChatGPT könnte verwendet werden, um Lehrmethoden zu simulieren und zu testen, indem es beispielsweise Schülern Aufgaben gibt und dann deren Leistung bewertet. Dies könnte helfen, die Stärken und Schwächen verschiedener Lehrmethoden aufzuzeigen.
3. Unterstützung bei der Analyse von Lehrmethoden: ChatGPT könnte verwendet werden, um Lehrern bei der Analyse von Lehrmethoden zu helfen, indem es beispielsweise Daten visualisiert oder Auswertungen erstellt. Dies könnte dazu beitragen, die Effektivität verschiedener Lehrmethoden zu vergleichen und die besten Ansätze zu identifizieren.

Es ist wichtig zu beachten, dass ChatGPT lediglich ein Werkzeug ist, das bei der Evaluierung von Lehrmethoden unterstützen kann, aber keine vollständige Lösung darstellt. Es ist wichtig, dass Lehrmethoden auf vielfältige Weise evaluiert werden, um ein umfassendes Bild von ihrer Effektivität zu erhalten.

Abbildung 5. Verwendung von ChatGPT als Datenanalysetool (Prompt von L. Baumanns am 29. 12. 22)

künftige Modelle diese Lücke füllen können. Aller Voraussicht nach werden KI-Modelle wie ChatGPT in der hier vertieften Anwendung aber den Expansionsdrang von Dateninfrastrukturen im Bildungsbereich weiter unterstützen (Hartong & Nikolai, 2021) und vermutlich große Mengen von Daten sammeln (worin auch ein unternehmerisches Ziel von OpenAI in der Lancierung einer kostenlos nutzbaren Anwendung liegen könnte).

Verwendung von ChatGPT beim Verfassen von wissenschaftlichen Texten

Einen zweiten (durchaus problematischeren) Anwendungsbereich besitzen generative KI-Sprachmodelle für das Schreiben von wissenschaftlichen Texten. Mit einer einfachen Eingabe lässt sich beispielsweise ein Text verfassen, der auf den ersten Blick den Kerngedanken zu verschiedenen Repräsentationsmodi in der Mathematikdidaktik zusammenfasst (vgl. Abb. 6).

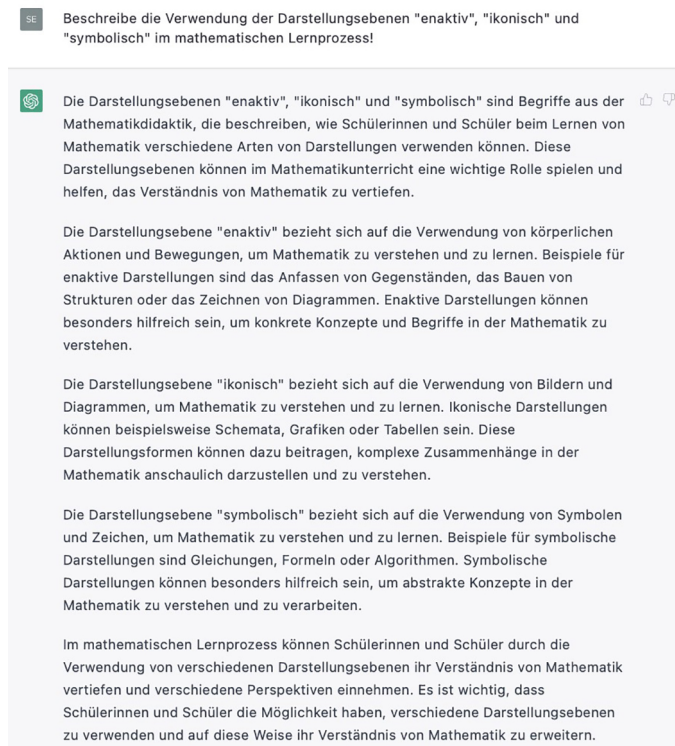


Abbildung 6. Eingabe zur Klärung einer fachlichen Frage (Prompt von S. Schorcht am 29. 12. 22)

Auch wenn die Antwort Nuancen eines tiefergehenden begrifflichen Verständnisses nicht richtig erfasst, entsprechen die Antworten von ChatGPT qualitativ durchaus schriftlicher Ausarbeitungen von Studierenden niedriger Semester. Mit einer entsprechenden Überarbeitung derartiger Texte durch Expertinnen und Experten, indem beispielsweise die fachliche Tiefe hinzugefügt wird, ließe sich in kurzer Zeit ein wissenschaftlicher und in bestimmten Fällen sicher auch publizierbarer Text erstellen. Allerdings wird die Antwort von ChatGPT, auch wenn sie auf den ersten Blick plausibel erscheint, generell durch den Algorithmus und daher nur durch eine Serie von Vermutungen erstellt. Sie ist aufgrund dessen immer mit einer gewissen Unsicherheit behaftet und genügt keinesfalls wissenschaftlichen Ansprüchen. Eine faktenbasierte Validität der von der KI präsentierten Informationen – etwa durch existierende Quellenangaben und wissenschaftliche Zitationen – wird nicht gewährleistet. Aktuell arbeitet OpenAI bereits an einem Prototypen für die Lösung dieses Problems. Die sich in Entwicklung befindliche Anwendung WebGPT (openai.com/blog/webgpt) besitzt eine integrierte Internetsuchfunktion und unterstützt die ausgegebenen KI-generierten Texte mit auffindbaren Quellenangaben. Andere KI-basierte Anwendungen wie etwa Elicit (elicit.org), die ebenfalls auf GPT-3 basieren und zusätzlich bestehende wissenschaftliche Publikationsdatenbanken auf spezi-

fische Suchanfragen hin durchforsten um eine Auswahl von Studien zusammenstellen, haben dies bereits besser umgesetzt und erweisen sich damit für die eigene Literaturrecherche als erstaunlich nützlich. Dennoch bleibt selbst bei diesen Weiterentwicklungen das Problem der Voraussetzung der menschlichen Expertise bei der Beurteilung der dargestellten Informationen bestehen. Für Personen, die keine hinreichenden mathematikdidaktischen Vorkenntnisse besitzen, ist augenscheinlich nicht zu erkennen, ob die zusammengetragene Information in Abb. 6 korrekt ist oder nicht, was aber natürlich im Allgemeinen auch für nicht-KI-generierte Texte gilt. Ebenso wenig ist erkennbar, ob der Text überhaupt manuell oder automatisiert verfasst wurde.

Daraus ergeben sich aus unserer Sicht erhebliche Konsequenzen für wissenschaftliche Zeitschriften und deren Qualitätssicherungssysteme. Es ist möglicherweise mit einer Zunahme an Publikationen zu rechnen, die mit KI-Unterstützung geschrieben werden. Für entsprechende Begutachtungsprozesse bedeutet dies, dass es zukünftig noch wichtiger wird, Personen mit der entsprechenden Expertise als Begutachtende gewinnen zu können, die die Validität und Glaubwürdigkeit von dargestellten Informationen hinreichend einschätzen können – und zwar bis hin auf die Ebene von Quellenangaben. Forschende dürften aufgrund des drohenden Reputationsverlustes bei Publikationen im Allge-

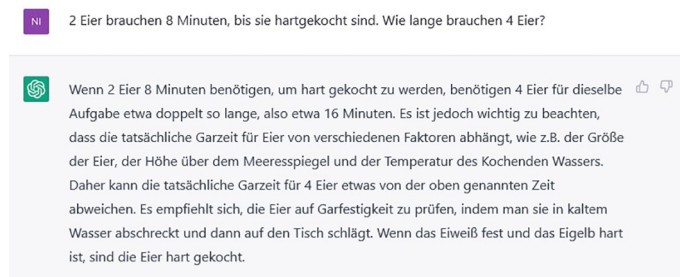


Abbildung 7. Fehlinterpretation einer mathematischen Fragestellung (Prompt von N. Buchholtz am 29. 12. 2022)

meinen wohl nicht auf non-existente Literaturquellen (Abb. 3) zurückgreifen. Allerdings kann die KI unter einer entsprechenden Eingabe leicht dazu aufgefordert werden, bestehende (eigene und fremde) Publikationen textlich so umzuschreiben, dass ein neuer Text kreiert wird und damit auch bestehende Plagiatssoftware umgangen werden kann. Hierzu müssten also zusätzlich Qualitätskriterien entwickelt werden, mit denen einwandfrei erkannt werden kann, inwieweit ein wissenschaftliches Manuskript KI-basiert erstellt worden ist – ein kaum zu lösendes Problem, bei dem allenfalls die Selbstauskunft von Autorinnen und Autoren abgefragt werden kann. Allenfalls kann eine weitere (Kontroll-) KI die Akzente eines KI-basierten Textes erkennen. Dies würde aber zu einem Szenario führen, in dem KIs letztendlich gegeneinander arbeiten und die menschliche Expertise nur noch im Hintergrund relevant ist.

Verwendung von ChatGPT im Rahmen des kreativen mathematischen Arbeitens

Ein dritter Anwendungsbereich für generative KI-Sprachmodelle liegt möglicherweise in der Generierung von mathematischen Aufgaben, Lösungswegen oder mathematischen Problemen für Anwendungsfälle. Die mathematischen Fähigkeiten von ChatGPT sind allerdings momentan noch überraschend gering ausgeprägt und ebenfalls fehleranfällig. So findet sich etwa bei Cherian et al. (2022) eine Studie mit 11 einfachen mathematischen Textaufgaben, bei der jede Aufgabe neuronalen Netzwerken (darunter ChatGPT) zehn mal unabhängig voneinander gestellt wurde, mit der Bitte nach einer Lösung und der Erklärung des Lösungswegs. ChatGPT konnte die Aufgaben nicht zuverlässig richtig lösen und die Lösungshäufigkeiten schwanken zwischen 0% und 90% pro Aufgabe. Die Anwendung ist bisher nicht zuverlässig in der Lage, spezifische Fallunterscheidungen zu treffen oder semantische Feinheiten mathematischer Begriffe situationsangemessen zu interpretieren. Da es die Antworten aus menschlich generierten Trainingsdaten und Feedback generiert, sind zwangsläufig auch

menschliche Fehlvorstellungen in den Antworten wiederzufinden (vgl. Abb. 7).

Auch die Fähigkeiten zum Problemlösen sind nur rudimentär ausgeprägt. So generiert ChatGPT bereits fehlerhafte Antworten bei einer einfachen Problemlöseaufgabe wie „Mein Vater ist 33 Jahre alt, ich bin 8 Jahre alt. Wann ist mein Vater doppelt so alt wie ich?“ (ChatGPT: „Dein Vater ist doppelt so alt wie du, wenn er 66 Jahre alt ist.“). Andererseits kann die Anwendung im Bereich der Informatik aber durchaus dazu verwendet werden, Programmcode zu schreiben oder Fehler in Programmcodes zu finden und zu reparieren – und dies unabhängig von der zugrundeliegenden Programmiersprache und mit einer gewissen Zuverlässigkeit.

Verwendung von ChatGPT bei Prüfungen im Lehramtsstudium

Weitreichende Auswirkungen eines verstärkten Nutzens von generativen KI-Sprachmodellen durch Studierende sehen wir auch für den Bereich der Lehramtsausbildung an Hochschulen. Die derzeitige Qualität der KI-generierten Texte entspricht teilweise auch der Qualität von Arbeiten von Studierenden, sodass wir die Integrität von semesterbegleitenden schriftlichen Ausarbeitungen, abschließenden Hausarbeiten oder anderen schriftlichen online-basierten Prüfungen gefährdet sehen (vgl. Susnjak, 2022). Niedrigschwellige Aufgaben, bei denen mathematikdidaktische Texte oder wissenschaftliche Publikationen zusammengefasst oder erstellt werden müssen, lassen sich bereits jetzt mit ausreichend hoher Präzision und Qualität durch entsprechende Eingaben in die generative KI und anschließender Überarbeitung lösen. Für Lehrende ist daher möglicherweise in Zukunft bei der Bewertung kaum mehr unterscheidbar, ob studentisch produzierte Texte manuell oder automatisch generiert wurden. Insofern Bewertungsmaßstäbe für Essays oder schriftliche Ausarbeitungen vorliegen, kann die KI sogar gebeten werden, diese bei der Texterstellung oder Bewertung zu berücksichtigen (Weßel, 2022). Die Einschränkungen durch nicht geprüfte Falschinformation wirken sich selbst-

verständlich auch in diesem Bereich aus, jedoch liegt die Prüflast hier bei den Lehrenden. Auch universitätseigene Plagiatssoftware ist in der Regel nicht darauf ausgelegt, KI-basierten Text zu erkennen. Insofern laufen KI-unterstützte Prüfungsleistungen hier unter dem Radar und lassen die Diskussion um Plagiatssoftware obsolet erscheinen.

Ein derartiger Gebrauch nicht zugelassener Hilfsmittel kann aus unserer Sicht als Problematik des „KI-Ghostwritings“ bezeichnet werden. Wir sehen hier Bedarf darin, langfristig über alternative Prüfungsformen nachzudenken und die Prüfungskultur im Lehramtsbereich den neuen Gegebenheiten anzupassen. Das bedeutet, dass sich Prüfungen ggf. stärker auf Mündlichkeit hin ausrichten müssen (z. B. durch Vorträge mit anschließendem Prüfungsgespräch) und der Schwerpunkt schriftlicher Prüfungen sich verlagern könnte hin zu stärker daten- oder situationsbasierten Prüfungsformaten, bei denen Studierende mathematikdidaktische Sachverhalte (z. B. aus gesehendem Unterricht oder anhand von Schülerlösungen) kompetenzorientiert theoretisch einordnen müssen. Hierzu könnten die in den letzten Jahren verstärkt vorangetriebenen Forschungen zu Video-Vignetten konkrete Impulse liefern (Geisen et al., 2022). Bei reinen Literaturarbeiten wird hingegen zukünftig möglicherweise nicht mehr ersichtlich sein, ob hier tatsächlich Wissen systematisch erworben wurde oder mittels KI-Unterstützung gearbeitet wurde (Weßels, 2022).

Herausforderungen für den Mathematikunterricht?

Es erscheint uns unter den beschriebenen Problemen der Nicht-Verlässlichkeit der durch generative KI-Sprachmodelle zur Verfügung gestellten Information fragwürdig, diese Modelle in schulischen Bildungszusammenhängen bewusst einzusetzen. Mathematische Fehler und schlechte Erklärungen finden sich allerdings auch in Medien, die „menschengemacht“ sind und dennoch auf breites Nutzungsverhalten im schulischen Mathematikunterricht treffen (z. B. Erklärvideos oder Blogbeiträge). Dies allein also als Ausschlusskriterium zu verwenden, würde das didaktische Potenzial, das generativen KI-Sprachmodellen innewohnt, daher vermutlich verkennen. Genauso wie die Tatsache, dass neue Technologien, die das Lernen augenscheinlich und auf verführerische Weise erleichtern, sich ohnehin nahezu von selbst unter Schülerinnen und Schülern verbreiten und die Mathematikdidaktik sich zu diesen neuen Technologien früher oder später verhalten muss. Klar ist, dass schulischer Unterricht und die Begleitung von Lernprozessen durch qualifizierte Lehrpersonen durch derartige Modelle nicht ersetzt werden können, auch wenn dies möglicherweise in der Entwicklung von automatisierten

Anwendungen in Zukunft so angelegt sein wird.

Derzeit sind die bereitgestellten Antworten von ChatGPT oft noch sehr fehlerbehaftet. Für eine Unterstützung bei mathematischen Berechnungen eignet sich die Anwendung wegen der Fehleranfälligkeit ebenfalls nicht; hierzu gäbe es aber ohnehin bereits zuverlässig arbeitende Anwendungen wie Photomath (deren Nutzen für den Unterricht aber umstritten ist). Nicht auszuschließen ist aber, dass weiterentwickelte Modelle diese Fehleranfälligkeit in der Zukunft reduzieren werden. Wir nehmen daher positiv an, dass weiterentwickelte generative KI-Sprachmodelle bei entsprechender Integration in Online-Lernumgebungen durchaus Hilfen zur Individualisierung von Lernprozessen für lernschwache Schülerinnen und Schüler bereitstellen können, zum Beispiel in der Reduzierung von Anforderungen durch einfache Sprache oder beim Verständnis von Aufgaben oder mathematischen Begriffen; hier steckt durchaus Forschungspotenzial.

Für eine reflektierte Nutzung dieser mächtigen Technologie erscheint uns zum derzeitigen Zeitpunkt eine kritische Einbindung in den Unterricht angemessen. Lernende (wie Lehrende) sollten dafür geschult werden, wie generative KI-Sprachmodelle zu benutzen sind und wie fehleranfällig sie sind, d. h. sie müssen in die Lage versetzt werden, zielführende Eingaben zu setzen und die Fehler und die Güte der bereitgestellten Information zu erkennen – auch in der Betrachtung des Falschen liegt eine Lernchance. Hier gliedern sich die dafür benötigten digitalen Kompetenzen in ganz natürlicher Weise an die Kompetenz „mit Medien mathematisch arbeiten“ der neu überarbeiteten Bildungsstandards (KMK, 2022) der Sekundarstufe an. Google wird als Recherchetool vermutlich weiterhin eine große Rolle für Lernende spielen. Dort, wo Google aber nur Informationen bereitstellen kann, und diese dann selbst bewertet werden müssen, könnten generative KI-Modelle Sachverhalte in Zukunft auch webbasiert erklären.

Fazit

Im Rahmen dieses Beitrags haben wir selbst viele Eingaben bei ChatGPT vorgenommen. Oft führte eine Eingabe zu enttäuschenden Resultaten, da viele der dargestellten Informationen falsch oder unvollständig sind. Diese Fehleranfälligkeit ist der Technologie durch den verwendeten Algorithmus inhärent. Daraus lässt sich die generelle Frage ableiten, ob generative KI-Sprachmodelle wie ChatGPT überhaupt in Forschung und Lehre verwendet werden sollten. Besieht man die Probleme, die das Imitieren täuschend echter menschlicher Konversation mit sich bringt, und Möglichkeiten, mit der diese Technologie missbraucht werden kann, sowie das

„disruptive“ Potenzial, das ihr in der medialen Aufmerksamkeit zugesprochen wird, so muss die Antwort „nein“ lauten. Generative KI-Sprachmodelle wie ChatGPT können nach dem derzeitigen Stand der Technologie zwar Text generieren, der wissenschaftlich durchaus korrekt sein kann, aber sie können dies nicht zuverlässig und sie können nicht die notwendige wissenschaftliche Analyse und die Originalforschung ersetzen, die beispielsweise beim Verfassen einer wissenschaftlichen Arbeit erforderlich sind. Wie bei anderer risikobehafteter Technologie aber auch kommt es aus unserer Sicht auf den vernünftigen Gebrauch der technischen Entwicklung an. Bedenkt man das rasante Tempo, mit dem sich in den letzten Jahren derartige KI-basierte Modelle weiterentwickelt haben, und die Fortschritte, die mit Sprachmodellen wie ChatGPT erreicht wurden, so sollte die Mathematikdidaktik aus unserer Sicht nicht die Chance verpassen, diese Technologie für Lernprozesse in Forschung und Lehre sinnvoll aufzugreifen und für das Lehren und Lernen von Mathematik reflektiert einzusetzen. Bis eine derartige Technologie in der mathematikdidaktischen Forschung und Lehre zuverlässig und fehlerfrei verwendbar ist, dürfte zwar noch etwas Zeit vergehen, aber Lösungsansätze für die offensichtlichen Probleme sind bereits in Sichtweite oder befinden sich bereits in der konkreten Entwicklung. Das Potenzial für den Einsatz generativer KI-Sprachmodelle ist groß. Die genauere Erforschung der Anwendungsmöglichkeiten und Grenzen der Technologie steht momentan noch aus. Die wenigen wissenschaftlichen Studien zu ChatGPT, die wir in der Kürze der Zeit recherchieren konnten, sind teilweise nur in Online-Archiven ohne Peer-Review zu finden. Wir hoffen dennoch, mit diesem Beitrag einen ersten Überblick über die aktuellen spannenden Entwicklungen geben zu können.

Anmerkung: Beim Verfassen dieses Beitrags Ende Dezember 2022 konnte die Anwendung für die Allgemeinheit kostenfrei verwendet werden. Bei Sichtung der Druckfahnen im Februar 2023 stand die Einführung von Bezahlmodellen bereits in Aussicht und die KI-Sprachmodelle erfuhren mittlerweile eine exorbitante mediale Aufmerksamkeit.

Literatur

- Cherian, A., Peng, K.-C., Lohit, S., Smith, K. & Tenenbaum, J. B. (2022). Are deep neural networks SMARTer than second graders? DOI:10.48550/arXiv.2212.09993
- Floridi, L. & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds & Machines*, 30, 681–694.
- Geisen, M., Zender, J. & Buchholtz, N. (2022). MS4: Alternative Prüfungsformate – digital und kompetenzorientiert. Beiträge zum Mathematikunterricht 2022.

- Hartong, S. & Nikolai, R. (2021). Warum es unabdingbar ist, Dateninfrastrukturen in der Bildungssteuerung stärker kritisch in den Blick zu nehmen. *Bildung unter (digitaler) Beobachtung – Analysen zur wachsenden Bedeutung von Dateninfrastrukturen in der Bildungssteuerung*. Thementeil in der Zeitschrift für Pädagogik, 67(3), 317–322.
- Jahn, T. (2022). Kultur und Wissenschaft – Mittels KI Examensarbeit schreiben? www.deutschlandfunk.de/kultur-und-wissenschaft-mittels-ki-examensarbeit-schreiben-dlf-40a6641c-100.html
- Knox, J., Williamson, B. & Bayne, S. (2020). Machine behaviourism: Future visions of ‘learnification’ and ‘datafication’ across humans and digital technologies. *Learning, Media and Technology*, 45(1), 31–45.
- KMK (2022). Bildungsstandards der Kultusministerkonferenz. KMK.
- Lobo, S. (2022, 7. Dezember). Das Ende der irrelevanten künstlichen Intelligenz. *Der Spiegel* (online). www.spiegel.de/netzwelt/web/chatgpt-markiert-das-ende-der-irrelevanten-kuenstlichen-intelligenz-kolumne-a-b2afeb69-083d-4e69-8920-da5cad549d5f
- Moorstedt, M. (2022, 4. Dezember). Künstliche Intelligenz Chat GPT beantwortet Fragen verblüffend klug. *sueddeutsche.de*. www.sueddeutsche.de/kultur/chatgpt-kuenstliche-intelligenz-1.5708877
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. (2017). Proximal policy optimization algorithms. DOI:10.48550/arXiv.1707.06347
- Shin, D. & Shim, J. (2021). A systematic review on data mining for mathematics and science education. *International Journal of Science and Mathematics Education*, 19(4), 639–659.
- Susnjak, T. (2022). ChatGPT: The end of online exam integrity? DOI:10.48550/arXiv.2212.09292
- Weßels, D. (2022, 20. Dezember). ChatGPT – ein Meilenstein der KI-Entwicklung. *Forschung und Lehre*. www.forschung-und-lehre.de/lehre/chatgpt-ein-meilenstein-der-ki-entwicklung-5271

Nils Buchholtz, Universität Hamburg
E-Mail: nils.buchholtz@uni-hamburg.de

Lukas Baumanns, Universität Köln
E-Mail: lukas.baumanns@uni-koeln.de

Judith Huget, Universität Bielefeld
E-Mail: jhuget@math.uni-bielefeld.de

Franziska Peters, Universität Gießen
E-Mail: franziska.peters@math.uni-giessen.de

Maximilian Pohl, Universität Duisburg-Essen
E-Mail: mapo1988@gmail.com

Sebastian Schorcht, TU Dresden
E-Mail: sebastian.schorcht@tu-dresden.de