

KALLWEIT, Michael & ROLKA, Katrin
Bochum

Generative KI zur Unterstützung bei Modellierungsaufgaben in Mathematik – Einblicke in einen explorativen Einsatz

Die Integration digitaler Werkzeuge in den Mathematikunterricht ist eng mit dem technologischen Fortschritt der letzten Jahrzehnte verknüpft. Beginnend mit den wissenschaftlichen Taschenrechnern, die eine effiziente Bearbeitung komplexer Berechnungen ermöglichten, führte der Weg über grafikfähige Taschenrechner, die zudem eine visuelle Darstellung mathematischer Funktionen erlaubten. Mit dem Aufkommen der CAS-Rechner wurde es möglich, symbolische Ausdrücke zu manipulieren und nahezu alle möglichen Aufgabenbereiche an ein technisches System auszulagern. Die Einführung dynamischer Geometrie-Software eröffnete dann neue Wege zur interaktiven Erkundung geometrischer Konzepte. Und moderne Smartphone-Apps (u. a. mit CAS und Formelerkennung per Foto) sind vielen Schüler*innen als nützliche Helfer bekannt.

Die aktuell steigende Verbreitung großer generativer Sprachmodelle (wie ChatGPT, Bard, Claude und LLaMA) eröffnet neue Perspektiven für digitale Lernunterstützung. In diesem Beitrag werden die Konzeption eines Dialogsystems, ausgewählte Dialogbeispiele und Einsatzmöglichkeiten im Mathematikunterricht präsentiert, sowie Herausforderungen und Potenziale dieses innovativen Ansatzes diskutiert. Dabei wird das Modellieren fokussiert, bei dem ausgehend von einer *Fragestellung in der Realität* eine Vereinfachung auf der Sachebene zu einem *Realmodell* erfolgt, das durch Mathematisieren in ein *mathematisches Modell* überführt wird und nach der Ermittlung einer *mathematischen Lösung* der Rückbezug auf die Fragestellung zu deren Beantwortung stattfindet (z. B. Greefrath et al., 2013). Die Auseinandersetzung findet an vielen Stellen auf textlicher Ebene statt, sodass hier die Einbindung von generativen Sprachmodellen möglich erscheint.

Generative KI

Generative KI-Modelle sind Algorithmen, die trainiert wurden, Daten in einem bestimmten Format zu erzeugen, das auf den Mustern basiert, welche sie zuvor von existierenden Daten gelernt haben. Im Jahr 2017 wurde ein bedeutender Durchbruch mit der Einführung der Transformer-Architektur (Vaswani, 2017) erzielt, die Modelle wie BERT, GPT, BLOOM und LLaMA hervorgebracht hat. Die Transformer-Architektur unterscheidet sich von früheren Ansätzen durch ihre Fähigkeit, Sequenzen von Daten zu verarbeiten, ohne auf rekurrente Schleifen angewiesen zu sein. Dies ermöglicht es den Modellen, Kontextinformationen über weite Strecken innerhalb von

Texten zu erfassen, was zu einer erheblichen Verbesserung des Verständnisses und der Generierung von natürlicher Sprache führt. Die aktuellen Modelle können Inhalte erzeugen, die von menschlichen Erzeugnissen kaum zu unterscheiden sind.

Dennoch sind die Mathematik-Fähigkeiten der Sprachmodelle systembedingt nicht mit den algorithmischen Fähigkeiten von CAS-Systemen vergleichbar. Trotz vieler stetiger Verbesserungen (vgl. Lightman et al., 2023 und Golkar et al., 2023) weisen die Modelle weiterhin Probleme in Arithmetik und logischer Argumentation sowie fortlaufender Konsistenz auf.

Chatbot als Unterstützung bei Modellierungsaufgaben

Trotz der Schwächen in Mathematik sind die Fähigkeiten, Texte zu verarbeiten und einen interaktiven Dialog zu führen, vielversprechend. Am HDM@RUB – Zentrum für Hochschuldidaktik Mathematik wurde auf Basis von generativen Sprachmodellen, ein Prototyp eines Chatbots als Dialogpartner bei der Bearbeitung von Modellierungsaufgaben erstellt und im Wintersemester 23/24 mit kleinen Stichproben aus Studierenden und Schüler*innen getestet. Das System verwendete zunächst das Sprachmodell GPT-3.5 und später GPT-4, beide über OpenAIs kostenpflichtige API. Das System wurde in die Rolle einer Lehrkraft durch einen dreiteiligen Anweisungstext versetzt, der aus folgenden Teilen bestand:

- Didaktisches Hintergrundwissen zu Modellierungsaufgaben (u. a. die Bearbeitungsschritte im Modellierungskreislauf)
- Ergänzende allgemein-didaktische Anweisungen für den Gesprächsverlauf (z. B. „Löse gemachte Fehler nicht sofort auf und korrigiere sie nicht, sondern fordere auf, den Fehler selber zu finden.“)
- Beispielhafte Lösung zur „Tanken“-Aufgabe (vgl. Blum & Leiß, 2006; s. u.) als Orientierung.

Der Prompt endete mit der Aufforderung, dem Nutzer eine zuvor festgelegte Aufgabe (siehe Abb.1) zu stellen und bei Bedarf Impulse für die (Weiter)Arbeit an der Aufgabe zu geben.

Herr Stein wohnt in Trier, 20 km von der Grenze zu Luxemburg entfernt. Er fährt mit seinem Auto zum Tanken nach Luxemburg, wo sich direkt hinter der Grenze eine Tankstelle befindet. Dort kostet der Liter Benzin nur 1,55 €, im Gegensatz zu 1,79 € in Trier. Lohnt sich die Fahrt für Herrn Stein? Begründe deine Antwort.

Abb. 1: „Tanken“-Aufgabe (vgl. Blum & Leiß, 2006)

Beispiele aus den Chatverläufen

Trotz der kleinen Stichproben (4 Schüler*innen aus den Jahrgangsstufen 7, 8, 9 und Q2 sowie einem Kurs mit Lehramtsstudierenden) konnten einige

interessante Beobachtungen in den Interaktionen gemacht werden. Sprachlich bzw. inhaltlich inkorrekte Formulierungen werden sehr wohlwollend vom Sprachmodell interpretiert und in korrekter Form zurückgespiegelt (s. Abb. 2).

Schüler(in): der preis ist um einiges günstiger als die kilometer Anzahl

Lehrer(in): Richtig! Obwohl die zusätzlichen Kilometer für die Fahrt nach Luxemburg Kosten verursachen, überwiegt die Ersparnis durch den günstigeren Benzinpreis.

Schüler(in): ein Auto hat durchschnittlich 60 liter

Lehrer(in): Das ist eine plausible Schätzung, da viele Autos Tankgrößen zwischen 40 und 70 Litern haben.

Abb. 2: Beispiele für wohlwollende Interpretationen durch das System

Da die verwendeten Sprachmodelle bereits eigenes assoziiertes Wissen mitbringen, konnte auf die Angabe von relevanten Fakten in dem Prompt verzichtet werden. So gab das Sprachmodell auf Nachfrage realistische Werte passend für den Sachkontext an (s. Abb. 3).

Lehrer(in): Ein durchschnittlicher Tank eines Pkws fasst oft zwischen 40 und 60 Litern. Einige kleinere Autos haben vielleicht nur einen 35-Liter-Tank, während größere Autos oder SUVs bis zu 70 Liter oder mehr fassen können.

Lehrer(in): Ein Auto kann beispielsweise zwischen 5 und 10 Litern auf 100 km verbrauchen, je nach Art des Autos und Fahrweise.

Abb. 3: Beispiele für die Angabe von realistischen Werten durch das System

Dennoch zeigte sich auch die Wichtigkeit der im Prompt hinterlegten Gesprächsanweisungen, die das System größtenteils sinnvoll befolgte (s. Abb. 4).

Prompt: Weise dabei auf Fehler hin und gib bei Bedarf auch Hinweise. [...] Löse gemachte Fehler nicht sofort auf und korrigiere sie nicht, sondern fordere auf, den Fehler selber zu finden.

Schüler(in): dann würde ich für 1 km 0,7 liter brauchen und für 3 km dann 2,1

Lehrer(in): Da ist ein kleiner Fehler in deiner Rechnung. Wenn 7 Liter für 100 km benötigt werden, wie viel verbraucht er dann wirklich für 1 km? Überlege noch einmal und rechne Schritt für Schritt.

Abb. 4: Beispiel für die Befolgung des Prompts durch das System

Fazit und Ausblick

Zurzeit werfen KI-Tools wie große Sprachmodelle bedeutsame Fragen für die Ausgestaltung einer zukünftigen Bildungsvermittlung auch im Mathematikunterricht auf. Unsere Erprobung zeigt mögliche Ansätze für den Einsatz im Unterricht auf. Sie wurde u. a. in einer Vorlesung zu digitalen Medien im Mathematikunterricht durchgeführt. Hier wurden Lehramtsstudierenden zudem die technologischen Hintergründe der Sprachmodelle vermittelt und die Nutzung verschiedener Lehr-Lernszenarien kritisch reflektiert.

Viele Weiterentwicklungen und Verbesserungen sind denkbar. In diesem ersten Einsatz wurden die Sprachmodelle von OpenAI genutzt, um den aktuellen Stand der Technik zu testen. In nachfolgenden Weiterentwicklungen wird auf freie Sprachmodelle zurückgegriffen, die zum Teil auch speziell für mathematische Fragestellungen und Argumentationen trainiert wurden, wie z. B. WizardMath (Luo et al., 2023). Im Bereich des Prompt-Engineering werden rasante Fortschritte erzielt, z. B. mittels Chain-of-Thought Prompting (Wei et al., 2023). Erste Experimente mit diesen Techniken zeigen große Verbesserungspotenziale auf.

Literatur

- Blum, W. & Leiß, D. (2005). Modellieren im Unterricht mit der „Tanken“-Aufgabe. *mathematik lehren*, 128, S. 18-21.
- Golkar, S., Pettee, M., Eickenberg, M., Bietti, A., Cranmer, M., Krawezik, G., Lanusse, F., McCabe, M., Ohana, R., Parker, L., Blancard, B. R.-S., Tesileanu, T., Cho, K., & Ho, S. (2023). *xVal: A Continuous Number Encoding for Large Language Models* (arXiv:2310.02989). arXiv. <http://arxiv.org/abs/2310.02989>.
- Greefrath, G., Kaiser, G., Blum, W. & Borromeo Ferri (2013). Mathematisches Modellieren – Eine Einführung in theoretische und didaktische Hintergründe. In R. Borromeo Ferri et al. (Hrsg.). *Mathematisches Modellieren für Schule und Hochschule, Realitätsbezüge im Mathematikunterricht* (S. 11–37). Springer.
- Lightman, H., Kosaraju, V., Burda, Y., Edwards, H., Baker, B., Lee, T., Leike, J., Schulman, J., Sutskever, I., & Cobbe, K. (2023). *Let's Verify Step by Step* (arXiv:2305.20050). arXiv. <http://arxiv.org/abs/2305.20050>.
- Luo, H., Sun, Q., Xu, C., Zhao, P., Lou, J., Tao, C., Geng, X., Lin, Q. & Zhang, D. (2023). *WizardMath: Empowering Mathematical Reasoning for Large Language Models via Reinforced Evol-Instruct*. arXiv [Cs.CL]. Retrieved from <http://arxiv.org/abs/2308.09583>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). *Attention Is All You Need*. CoRR, abs/1706.03762. Retrieved from <http://arxiv.org/abs/1706.03762>
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models* (arXiv:2201.11903). arXiv. <http://arxiv.org/abs/2201.11903>