

LEIFHEIT, Luzia; LENZ, Katja; SCHREITER, Saskia & LADEL, Silke  
Schwäbisch Gmünd

## **Die sprachbasiert-probabilistische Funktionsweise generativer KI und die formale Logik als Fundament der Mathematik – Ein Widerspruch?**

Mathematik und Informatik beruhen auf formal-logischen Regeln, klaren Definitionen abstrakter Konzepte und Zusammenhänge, sowie deterministischen Prozessen. In der Aussagenlogik werden Aussagen als Behauptungen definiert, die entweder wahr oder falsch sind (vgl. Albertini & Huber, 2003). Durch die logische Verknüpfung von Aussagen (z. B. Konjunktion, Implikation oder Äquivalenz) können neue Aussagen gebildet werden, deren Wahrheitsgehalt z. B. mit Wahrheitstabellen bestimmt werden muss. Die formale Logik zeichnet sich aus durch Determinismus, Konsistenz und Eindeutigkeit in der Anwendung von Regeln. Werden bei gleichen Ausgangswerten die gleichen Regeln für Berechnungen oder Schlussfolgerungen im Sinne von Inferenzen angewandt, so führen diese immer zu denselben Ergebnissen. Damit schafft die formale Logik die Grundlage dafür, dass Argumentationen, Theorien und Berechnungen verlässlich, nachvollziehbar und überprüfbar sind, was das Fundament wissenschaftlicher Exaktheit und Sicherheit in Mathematik und Informatik bildet. Über die Verknüpfung von Aussagen können neue Aussagen gebildet werden (z. B. Implikation oder Äquivalenz), deren Wahrheitsgehalt wiederum aus dem Wahrheitsgehalt der ursprünglichen Aussagen und der Art der Verknüpfung ableitbar ist (vgl. Albertini & Huber, 2003). Im Kontrast dazu arbeiten generative KI-Modelle nicht mit formaler Logik, sondern mit Wahrscheinlichkeiten und Mustererkennung, die Verbindungen zwischen Wörtern und Sätzen erfasst, was eine ganz andere Funktionsweise und Ergebnisstruktur hervorbringt.

### **LLM-basierte KI-Tools und logisches Schlussfolgern**

KI-Tools, die auf großen Sprachmodellen (Large Language Models, kurz: LLMs) basieren, funktionieren durch die Mustererkennung in riesigen Textmengen (Myers et al., 2024). Anstatt auf der Grundlage der Anwendung logischer Schlussregeln (z. B. Modus ponens, Kettenschluss, Reductio ad absurdum) Schlussfolgerungen zu ziehen, generieren sie Antworten, indem sie statistische Wahrscheinlichkeiten für Wortsequenzen vorhersagen (vgl. Leifheit et al., 2024). Der Prozess der Erstellung einer automatisiert erzeugten Textausgabe kann augenscheinlich wie logisches Denken wirken, erfolgt jedoch ohne tatsächliches Wissen oder inhaltliches Verständnis der Ein- und Ausgabe (vgl. Trott et al., 2023). Die Funktionsweise generativer KI ist dadurch charakterisiert, dass bestehende Muster identifiziert und mittels

In: L. Schick, M. Platz & A. Lambert (Hrsg.),  
Beiträge zum Mathematikunterricht 2025.

stochastischer Algorithmen zu neuen Ausgaben synthetisiert werden (vgl. McCoy et al., 2023). Logische Regeln, wie sie in der formalen Logik zum Tragen kommen, finden dabei keine Anwendung. Folglich können in der Regel formal korrekte und logisch erscheinende Formulierungen erzeugt werden, ohne dabei die Konzepte der Logik auf kognitiver Ebene zu erfassen oder auf formaler Ebene anzuwenden.

### **Probabilistische Generierung von Output durch sprachbasierte KI**

Die Berechnung des Outputs bei LLMs ist eine Art statistische Musteranpassung, die auf Wahrscheinlichkeiten von Wortfolgen basiert (vgl. Leifheit et al., 2024). Dadurch erscheinen ihre Antworten oft kohärent und logisch, da sie typische Argumentationsmuster und sprachliche Strukturen replizieren können. Sie imitieren Argumente, indem sie bekannte Zusammenhänge nachbilden, die oft in Texten vorkommen. Jedoch fehlt ihnen konzeptuelles Wissen zum Aufbau sachlogischer Argumente, da sie kein Bewusstsein für die Bedeutung der Wörter und Konzepte haben, die sie verwenden (vgl. Trott et al., 2023). Während ein Mensch bei einer logischen Schlussfolgerung aktiv Wissen, Erfahrung und ein (intuitives) Verständnis der Anwendung logischer Regeln einsetzt, erkennt ein LLM keine inhaltlichen Bedeutungen oder logischen Zusammenhänge in diesem Sinne. Es interpretiert keine Konzepte und kann keine logischen Inferenzregeln anwenden oder abstrakte Prinzipien verstehen – es weiß nur, dass bestimmte Wortfolgen in der Vergangenheit oft zusammen aufgetreten sind.

Anders als eine klare, deterministische Berechnung folgt die Antwort eines LLMs zudem keinem systematisch-logischen, sondern einem probabilistischen Aufbau mit pseudozufälligen Abweichungen. Pseudozufällige Abweichung bedeutet, dass kein „echter“ Zufall zum Tragen kommt, sondern Zufall im Rahmen des Machine-Learning-Modells simuliert wird (vgl. Vadhan, 2012), das zufällige Entscheidungen im Rahmen vordefinierter Muster zulässt. Konkret bedeutet dies, dass Wahrscheinlichkeiten für jedes mögliche nächste Wort berechnet werden, sodass es nicht nur eine „richtige“ Antwort gibt, sondern eine Vielzahl wahrscheinlicher Fortsetzungen. Diese Fortsetzungen werden mit beabsichtigter Variabilität aus mehreren möglichen Wörtern mit hoher Wahrscheinlichkeit ausgewählt, um nicht immer dieselben Antworten zu liefern und somit kreativer und damit menschlicher zu wirken. Dies erzeugt eine Varianz im Output, d. h. auf dieselbe Eingabe erfolgt in der Regel nicht jedes Mal derselbe Output (vgl. McCoy et al., 2023).

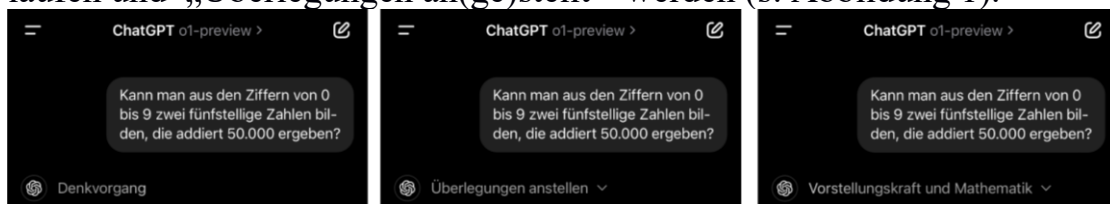
An dieser Stelle kommt es zu einem Konflikt: Der formal-logische Ansatz der Mathematik verlangt nach festen, deterministischen Antworten, während die probabilistische Funktionsweise der KI-Modelle darauf ausgelegt ist,

sprachliche Flexibilität und Variationen zu erzeugen. Diese probabilistische Struktur scheint aus mathematischer Sicht kontraintuitiv, da das Resultat auf einem Training in Sprachmustern basiert und nicht auf festgelegten mathematischen Axiomen und Kalkülen. Der Output wird durch ein Netzwerk von Parametern berechnet, die Zusammenhänge in der Sprache darstellen, ohne dass ein logischer Argumentations- oder Schlussfolgerungsprozess befolgt wird. Aus mathematischer Sicht wirkt es daher wenig nachvollziehbar, dass diese Art von Berechnung sinnvoll und logisch erscheinende – unter anderem, da meist sprachlich und grammatikalisch korrekt und einwandfrei formulierte –, aber nicht immer tatsächlich logische oder mathematisch korrekte Antworten ausgeben kann (vgl. Ma et al., 2024; Li et al., 2024).

### Wichtige Aspekte für Lehrende der Mathematik und Informatik

Lehrende der Mathematik und Informatik sollten verstehen, dass LLMs nicht auf formaler Logik und regelgeleiteten Schlussfolgerungen basieren, sondern auf Mustererkennung in großen Sprachkorpora. Der reflektierte Einsatz dieser Tools erfordert das Bewusstsein, dass die Modelle keine logischen Fähigkeiten aufweisen, sondern Aussagen anhand von wahrscheinlichen Wortfolgen innerhalb sprachlicher Kontexte generieren. Gerade für Lehrende im Bereich Mathematik und Informatik ist es daher entscheidend zu erkennen, dass probabilistische, primär sprachbasierte KI-Modelle keine zuverlässig logisch korrekte Wiedergabe mathematischer Zusammenhänge gewährleisten können.

Besonders irreführend ist dabei, dass textgenerierende KI-Anwendungen wie z. B. ChatGPT o1 in der Wartezeit während der Verarbeitung des Prompts und der Textgenerierung, suggerieren, dass „Denkvorgänge“ ablaufen und „Überlegungen an(ge)stellt“ werden (s. Abbildung 1).



**Abb. 1:** ChatGPT suggeriert Denken und die Anwendung mathematischer Regeln

Dies birgt die Gefahr, dass die Nutzung von KI im Bereich der Mathematik und Informatik überschätzt wird und fehlerhafte Ausgaben nicht hinterfragt und erkannt werden (vgl. Antonenko & Abramowitz, 2023; Bewersdorff et al., 2023). Lehrende sollten sich daher beim Einsatz von KI bewusst sein, dass die Antworten rein sprachabhängig-statistisch erzeugt und daher mit Vorsicht zu interpretieren sind. Derartiges Hintergrundwissen hilft dabei, generative KI-Tools effektiv und kritisch für Aufgaben im Kontext von Mathematik und Informatik einzusetzen, ohne ihren augenscheinlich logisch

wirkenden Ausgaben unhinterfragt zu vertrauen. Ein besseres Verständnis der Funktionsweise von KI-Anwendungen kann dazu führen, dass geeignete Einsatzmöglichkeiten von (angehenden) Lehrkräften besser erkannt werden und passgenauere Anwendungen der KI-Tools möglich werden. Im Hinblick auf die Weiterentwicklung von KI-Systemen scheint es für die Anwendung in der Mathematik und Informatik interessant, formale Logik in KI-Systeme zu integrieren, um Modelle zu schaffen, die sowohl probabilistische Musterverarbeitung beherrschen als auch logisch konsistent arbeiten.

## Literatur

- Albertini, C., & Huber, M. (2023). *Grundbegriffe der Mathematik: Logik-Mengen-Relationen und Funktionen-Zahlbegriff*. Carl Hanser Verlag GmbH Co KG.
- Antonenko, P., & Abramowitz, B. (2023). In-service teachers'(mis) conceptions of artificial intelligence in K-12 science education. *Journal of Research on Technology in Education*, 55(1), 64-78.
- Bewersdorff, A., Zhai, X., Roberts, J., & Nerdel, C. (2023). Myths, mis- and preconceptions of artificial intelligence: A review of the literature. *Computers and Education: Artificial Intelligence*, 4, 100143.
- Leifheit, L., Loefflad, D., Belschner, S., Beuttler, B., Winkelmann, J., Meurers, W. D., & Holz, H. (2024). KI im Unterricht: Entwicklung von Lehrveranstaltungen für Lehramtsstudierende der Sprach- und MINT-Fächer. *Ludwigsburger Beiträge zur Medienpädagogik*, 24, 1-19.
- Ma, J., Dai, D., Sha, L., & Sui, Z. (2024). Large language models are unconscious of unreasonability in math problems. *arXiv preprint, arXiv: 2403.19346*.
- McCoy, R. T., Smolensky, P., Linzen, T., Gao, J., & Celikyilmaz, A. (2023). How much do language models copy from their training data? Evaluating linguistic novelty in text generation using raven. *Transactions of the Association for Computational Linguistics*, 11, 652-670.
- Myers, D., Mohawesh, R., Chellaboina, V. I., Sathvik, A. L., Venkatesh, P., Ho, Y. H., Henshaw, H., Alhawawreh, M., Berdik, D. & Jararweh, Y. (2024). Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Computing*, 27(1), 1-26.
- Trott, S., Jones, C., Chang, T., Michaelov, J., & Bergen, B. (2023). Do large language models know what humans know? *Cognitive Science*, 47(7), e13309.
- Vadhan, S. P. (2012). Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3), 4-5.
- Li, X., Wang, W., Li, M., Guo, J., Zhang, Y., & Feng, F. (2024). Evaluating Mathematical Reasoning of Large Language Models: A Focus on Error Identification and Correction. *arXiv preprint, arXiv: 2406.00755*.