

SJUTS, Johann  
Osnabrück

## **ChatGPT – beständig unzuverlässig oder unbeständig zuverlässig?**

Schon beim Lösen schulmathematischer Probleme zeigt sich die Ambivalenz der Nutzung von ChatGPT. Mal generiert der KI-Chatbot falsche, mal richtige Antworten auf die Eingaben mathematischer Aufgaben und Probleme. Dabei zeigt sich, dass ChatGPT die Richtigkeit der selbst erstellten Antwort nicht sicher zu beurteilen vermag. Das wirft Fragen zum Einsatz von ChatGPT beim Lehren und Lernen von Mathematik auf.

### **1. Einleitende Bemerkungen zum Einsatz von ChatGPT beim Lösen mathematischer Problemaufgaben**

Künstliche Intelligenz (KI) erfreut sich aktuell zunehmender Aufmerksamkeit. Stark gestiegen ist das Interesse an KI-Textrobotern (Chatbots), die ein Text-auf-Text-Werkzeug darstellen und zu eingegebenen Prompts Antworten ausgeben. Allgemein zugänglich sind ChatGPT-3.5 und die in die Suchmaschine Bing integrierte Version GPT-4. Letztere findet hier Verwendung.

Als Prompts dienten in dieser experimentellen Studie mehrere mathematische Problemaufgaben, also solche Aufgaben, die nicht mit vertrauten Routinen lösbar sind, sondern ein gewisses mathematisches Denken (Sjuts 2023) erfordern. Schulmathematisch reichen die eingesetzten Aufgaben bis zum Schuljahrgang 8. Untersuchungszeitraum waren die Monate November und Dezember 2023. Alle Prompts wurden mehrfach (in unregelmäßigen Abständen) eingegeben. Da die von ChatGPT erzeugten Antworten nicht nur kurze Ergebnisse, sondern fast immer auch erklärende Begründungen enthalten, kann die Studie Aufschlüsse über den KI-Chatbot geben.

#### **2.1 Begründung und Ergebnis nicht richtig**

Für nicht wenige Aufgaben findet ChatGPT keine Lösung. Die auf den jeweiligen Prompt gegebenen Antworten sind inkorrekt oder widersprüchlich, also insgesamt falsch. Hier ein Beispiel:

Aufgabe (Sjuts 2022, 376): Denkt man sich eine Zahl und addiert dann ihre Quersumme und ihr Querprodukt, erhält man 1. – Für welche Zahlen ist das möglich?

[Ergebnis: Das ist für unendlich viele Zahlen möglich, nämlich 10, 100, 1000, ...]

ChatGPT-Antwort: „Die Quersumme einer Zahl ist die Summe ihrer Ziffern, und das Querprodukt ist das Produkt ihrer Ziffern. Wenn die Summe der

*Quersumme und des Querprodukts einer Zahl 1 ergibt, dann muss die Zahl selbst 0 oder 1 sein.*

*Betrachten wir die beiden Möglichkeiten:*

*1. Wenn die Zahl 0 ist, dann ist sowohl ihre Quersumme als auch ihr Querprodukt 0, und die Summe von beiden ist ebenfalls 0, was nicht 1 entspricht.*

*2. Wenn die Zahl 1 ist, dann ist sowohl ihre Quersumme als auch ihr Querprodukt 1, und die Summe von beiden ist 2, was nicht 1 entspricht.*

*Daher gibt es keine Zahl, für die die Summe ihrer Quersumme und ihres Querprodukts 1 ergibt.“*

Kommentar: Durchweg ist festzustellen, dass – wie auch hier – die von ChatGPT erzeugten Antworten sich durch ihre Darstellungsqualität auszeichnen: Sie sind gegliedert und bestehen aus vollständigen Sätzen mit korrekter Zeichensetzung, Rechtschreibung und Grammatik. Bezeichnend ist auch, dass zu Beginn der Inhalt des Prompts rekapituliert und die Begrifflichkeit geklärt wird, bevor dann der Lösungsvorschlag entsteht.

Falsch im ersten Absatz ist der Schluss, dass die gedachte Zahl selbst 0 oder 1 sein muss. (Richtig wäre, dass die Quersumme 1 und das Querprodukt 0 sein muss.) Zutreffend sind hingegen die Begründungen dafür, dass die Zahlen 0 und 1 nicht in Frage kommen können, da in beiden Fällen die Summe aus Quersumme und Querprodukt nicht 1 ist.

Falsch ist am Ende, dass es keine Zahl gibt, die die gestellte Bedingung erfüllt. Denn für unendlich viele Zahlen (10, 100, 1000, ...) ist die jeweilige Summe aus Quersumme und Querprodukt 1.

Die Aufgabe ist somit ungelöst. Die Bearbeitung enthält Mängel, das Ergebnis ist falsch. Es finden sich in der Antwort auch keine für einen Lösungsweg aussichtsreichen Überlegungen.

## **2.2 Begründung und Ergebnis teilweise richtig**

Bei etlichen Aufgaben ist die von ChatGPT erstellte Bearbeitung unvollständig oder fehlerhaft, also nur teilweise korrekt und insgesamt noch unzureichend. Hier ein Beispiel:

Aufgabe (Känguru der Mathematik 2016, Klassenstufen 7 und 8, Nr. 28): An der Tafel stehen mehrere voneinander verschiedene natürliche Zahlen. Das Produkt der beiden kleinsten dieser Zahlen ist 16. Das Produkt der beiden größten dieser Zahlen ist 225. – Wie groß ist die Summe aller Zahlen an der Tafel?

[Ergebnis: Die Rechnung lautet:  $2 + 8 + 9 + 25 = 44$ . Die Summe ist 44.]

ChatGPT-Antwort: „Die beiden kleinsten natürlichen Zahlen, deren Produkt 16 ist, sind 1 und 16 oder 2 und 8 oder 4 und 4. Da die Zahlen verschieden sein müssen, kommen nur die Paare 1 und 16 oder 2 und 8 in Frage.

Die beiden größten natürlichen Zahlen, deren Produkt 225 ist, sind 1 und 225 oder 3 und 75 oder 5 und 45 oder 9 und 25 oder 15 und 15. Da die Zahlen verschieden sein müssen, kommen nur die Paare 1 und 225 oder 3 und 75 oder 5 und 45 oder 9 und 25 in Frage.

Da die kleinsten und größten Zahlen nicht dieselben sein können, muss es mindestens vier Zahlen geben. Die einzige Möglichkeit, bei der vier verschiedene Zahlen herauskommen, ist, wenn die kleinsten Zahlen 2 und 8 und die größten Zahlen 3 und 75 sind.

Die Summe dieser Zahlen ist  $2 + 8 + 3 + 75 = 88$ .“

Kommentar: Korrekt ist der Teil mit den angegebenen Faktorpaaren für die Produkte 16 und 225. In der Bildschirmfassung ist zu erkennen, dass ChatGPT auf eine Mathematiksoftware zugreift, die diese Faktorpaare ausgibt. Erwähnenswert ist, dass die Bedingung der Aufgabenstellung (die Zahlen müssen voneinander verschieden sein) berücksichtigt wird.

Richtig ist die Folgerung, dass 2 und 8 die beiden kleinsten Zahlen sind, falsch dagegen, dass 3 und 75 die beiden größten Zahlen sind. Dafür, dass es sich um vier (und nicht mehr) Zahlen handelt, die an der Tafel stehen, fehlt in der Antwort eine Begründung.

Die von ChatGPT kreierte Antwort ist teilweise korrekt, teilweise fehlerhaft, unvollständig und im Ergebnis falsch. Gleichwohl bietet sie Anhaltspunkte für Berichtigung und Vervollständigung.

### **2.3 Begründung und Ergebnis richtig**

Zu vielen Aufgaben generiert ChatGPT eine zufriedenstellende, eine nicht selten sogar einwandfreie Lösung. Hier ein Beispiel:

Aufgabe (Hesse 2009, 42): Gegeben ist eine Tafel Schokolade, die aus  $3 \cdot 4 = 12$  Stücken besteht. Sie soll entlang gerader Linien in ihre 12 Stücke zerbrochen werden. – Wie viel mal muss man brechen?

[Ergebnis: Man muss 11-mal brechen.]

ChatGPT-Antwort: „Um eine Tafel Schokolade, die aus 12 Stücken besteht, in ihre einzelnen Stücke zu zerbrechen, muss man 11 Mal brechen. Jedes Brechen teilt die Tafel in ein weiteres Stück, daher benötigt man immer ein Brechen weniger als die Anzahl der Stücke.“

Kommentar: Die Lösung ist kurz und knapp. (Vielfach sind die von ChatGPT erschaffenen Antworten recht länglich.)

Bemerkenswert ist hier, dass die in der Antwort stehende Begründung ausdrücklich nicht nur für den speziellen Fall ( $3 \cdot 4 = 12$  Stücke) gilt.

### **3. Vorläufige Schlussfolgerungen über die Nutzung von ChatGPT beim Lehren und Lernen von Mathematik**

Die Verantwortlichen der Suchmaschine Bing und des eingebundenen KI-Textroboters GPT-4 sind sich der bestehenden Unzulänglichkeiten bewusst. Deshalb erscheint auch stets der folgende Hinweis: „Bing basiert auf KI, so dass Überraschungen und Fehler möglich sind. Teilen Sie uns Ihr Feedback mit, damit wir uns verbessern können!“

Die Studie zeigte, dass die von ChatGPT gegebenen Antworten in der Lösungssicherheit schwanken und sich insgesamt über eine große Bandbreite von Lösungsqualität erstrecken. Teils sind die Ausführungen widersprüchlich und falsch, teils sind die Bearbeitungen ungenau und unvollständig, teils sind die Lösungen korrekt und umfassend.

Dass der generative KI-Textroboter nicht durchweg verlässlich richtig antwortet (*unbeständig zuverlässig* ist) und das auch nicht selbst feststellt (*beständig unzuverlässig* ist), wirft für den schulischen Einsatz Fragen auf: In welchem Maße müssen Lernende, bevor sie ChatGPT verwenden, über Fähigkeiten im Umgang damit verfügen? Was müssen Lehrkräfte im Mathematikunterricht dazu im Vorhinein vermitteln? Welche inhaltlichen und sprachlichen Voraussetzungen muss der Unterricht schaffen, damit Schülerinnen und Schüler in der Lage sind, Prompts zu formulieren, die in der Sache korrekt und in der Sprache präzise sind? Welche Textverstehenskompetenz muss vorhanden sein, um die ChatGPT-Antworten aufnehmen zu können? Und da diese grundsätzlich einer Überprüfung bedürfen: Welcher Stand mathematischen Wissens und Könnens ist dafür nötig? Gleichwohl ist angesichts der Qualität der ChatGPT-Versprechungen zu überlegen, wie gewinnbringend die Nutzung des Chatbots dennoch sein kann.

#### **Literatur**

- Hesse, Christian (2009). *Das kleine Einmaleins des klaren Denkens*. München: Beck.
- Känguru der Mathematik (2016). Klassenstufen 7 und 8.
- Sjuts, Johann (2022). Wettbewerbe und mehr – was eine Schule zur Talentförderung in Mathematik organisieren kann. G. Ambrus, J. Sjuts & É. Vásárhelyi (Hrsg.): *Mathematische Zeitschriften und Wettbewerbe für Kinder und Jugendliche. Förderung für Talentierte und Interessierte über Grenzen hinweg*. Münster: WTM, 363-384.
- Sjuts, Johann (2023). Illustrative Aufgaben zum mathematischen Denken (in den mittleren Schuljahrgängen). G. Ambrus, J. Sjuts & É. Vásárhelyi (Hrsg.): *Mathematik und mathematisches Denken – Ansprüche und Anforderungen vor, in und nach der Schule*. Münster: WTM, 225-248.