

Kortenkamp, Ulrich

Wieviel Mathe braucht der Mensch? : Mathematische Kernkompetenzen im Angesicht von KI

In:

Steinweg, Anna Susanne (Hrsg.), Schule im Wandel – Mathematikunterricht im Wandel:
Tagungsband des AK Grundschule in der GDM 2024, Bamberg: University of Bamberg Press, S.
57-72. 2024. DOI: 10.20378/irb-104036

Beitrag im Sammelwerk - Verlagsversion

DOI des Beitrags: 10.20378/irb-105104

Datum der Veröffentlichung: 29.11.2024

Rechtehinweis:

Dieses Werk ist durch das Urheberrecht und/oder die Angabe einer Lizenz geschützt. Es steht Ihnen frei, dieses Werk auf jede Art und Weise zu nutzen, die durch die für Sie geltende Gesetzgebung zum Urheberrecht und/oder durch die Lizenz erlaubt ist. Für andere Verwendungszwecke müssen Sie die Erlaubnis der Rechteinhaberinnen und Rechteinhaber einholen.

Für dieses Dokument gilt die **Creative-Commons-Lizenz CC BY**.




Die Lizenzinformationen sind online verfügbar:

<https://creativecommons.org/licenses/by/4.0/>

Wieviel Mathe braucht der Mensch?

Mathematische Kernkompetenzen im Angesicht von KI

von Ulrich Kortenkamp

 0000-0002-5577-8819

Der rasante Fortschritt bei generativer KI ist beeindruckend und erschreckend zugleich. Manche in der Schule vermittelten Kompetenzen werden plötzlich bedeutungslos –in Deutsch und anderen Sprachen, in der Kunst und Musik. Auch der Mathematikunterricht kommt erneut in Erklärungsnot: Wozu muss man <hier ein beliebiges Thema einsetzen> können, wenn ChatGPT die Aufgabe komplett löst, die Lösungswege erklärt und das alles noch in perfektem Deutsch, Englisch oder Chinesisch?

Schlüsselwörter: Künstliche Intelligenz, Sinnstiftung, Technologie, Probe, Demokratiebildung, Hundewelpen

1 Vorbemerkung

Die Entwicklung von Künstlicher Intelligenz (KI) in den letzten Jahren geht so schnell voran, dass ein Artikel oder Vortrag dazu nur eine Momentaufnahme der aktuellen Situation geben kann. Insofern geht es hier nicht darum, zu zeigen, was aktuell möglich ist (das ist aber als Grundlage für die Argumentation notwendig) oder was noch passieren kann (das ist dann vermutlich schon passiert, wenn der Artikel zukünftig gelesen wird). Das ambitionierte Ziel des Artikels ist, auf dieser stetig fortschreitenden Basis zu argumentieren und – spoiler alert – die Notwendigkeit eines fundierten Mathematikunterrichts jenseits von Rechenfertigkeiten und Anwendbarkeit zu begründen. Dies geschieht nicht nur aus dem schon immer bestehenden Rechtfertigungszwang des allgemeinbildenden Mathematikunterrichts (Winter, 1995) heraus, sondern aus der persönlichen Überzeugung, dass nur dies die Menschheit für dem Untergang retten kann.

2 Der KI-Schock

Der 30. November 2022 stellt eine Zäsur für die öffentliche Wahrnehmung von KI dar. Die damals noch als non-profit-organisation agierende Firma OpenAI stellt ihren KI-basierten Chatbot ChatGPT vor (ChatGPT, 2022), mit dem plötzlich jede und jeder mit einer künstlichen Intelligenz kommunizieren kann. Und der Schock ist groß, weil das System nicht nur einfach zu bedienen ist, sondern auch und gerade

in standardisierten Tests Werte erreicht, die über dem Median der (menschlichen) Bevölkerung liegen (teddy [@davidsong], 2022).

Schon vor 30 Jahren wurde diskutiert, wie KI in der Mathematikdidaktik verwendet werden kann, und die Diskussion dauert bis heute (Balacheff, 1993a, 1993b; Gadanidis, 2017; Hwang, 2022; Mohamed et al., 2022). Zu den Hauptanwendungen zählen Intelligente Tutorielle Systeme (ITS) (Holmes et al., 2019), die durch eine intelligente Analyse der Antworten von Schülerinnen und Schülern in der Lage sein sollen, die idealen Rückmeldungen und Unterstützungsmaßnahmen für den individuellen Lernfortschritt der einzelnen Schülerinnen und Schüler zu liefern. Das klingt zunächst verlockend und gut, weil damit Lehrkräfte von einer anstrengenden und herausfordernden Aufgabe entlastet werden könnten. Dabei darf nicht vergessen werden, dass Schule nicht primär der Individualisierung dient. Die Warnung ist nicht neu: *„Die zu weitgehende Differenzierung der Schüler nach ihrer Begabung führt in ihrer Konsequenz zum Einzelunterricht, d. h. zur Auflösung der Schule. Aus Gründen sozialer Natur können wir den Unterricht in Schulgemeinschaften nicht entbehren. [...] Die Trennung der gesunden Kinder nach ihrer Begabung in zwei oder drei Gruppen entzieht dem Einzelindividuum die Vorteile des gemeinsamen Unterrichtes.“* (Rzesnitzek, 1912, S. 129), aber auch heute noch notwendig, wie Lutz Führer (1997, S. 24) in seiner achten These formuliert: *„Je stärker Unterricht individualisiert, desto eher sind ‚tendenziell globale Werte‘ wie Erlebnisfähigkeit im gemeinschaftlichen Sachbezug, Teamfähigkeit, Sozialbindung oder ganzheitliche Bildung gefährdet.“*

Doch es ist nicht ganz so einfach: Wenn die KI so gut ist, dass sie „weiß“ (was genau „Wissen“ einer KI ist bleibt dabei zu klären), wie Schülerinnen und Schüler Mathematik lernen sollen, dann kann sie eigentlich auch direkt die Aufgaben selbst lösen. Und ja, das kann sie auch: Schülerinnen und Schüler haben mit als erste erkannt, dass die KI ihnen das mühsame Lernen von Faktenwissen und prozeduralen Fähigkeiten abnehmen kann und nutzen sie direkt, um Aufgaben zu lösen. Es bleibt höchstens noch der Schritt des Abschreibens (oder, wenn die Aufgabe auch getippt abgegeben werden darf, das kopieren- und-einfügen). Vor diesen Konsequenzen wurde direkt nach der Veröffentlichung von ChatGPT gewarnt (Cotton et al., 2023). Empirische

Befunde zeigen aber, dass sich das Täuschungsverhalten von Schülerinnen und Schülern nicht zwangsläufig ändert (Lee et al., 2024).

Wir sind damit aber beim Kernproblem angelangt: Wie kann man rechtfertigen, dass Schülerinnen und Schüler etwas lernen müssen, was eine Maschine genauso gut oder gar besser kann? Ohne die Antwort vorwegzunehmen sei hier direkt bemerkt, dass dies *schon immer* ein fundamentales Problem des schulischen Unterrichts war, welches aber aus pädagogischer Sicht durch einen Blick auf das Ziel von Bildung im Allgemeinen beantwortet werden kann. Es geht nicht darum, Menschen zu möglichst hilfreichen, effizienten, funktionierenden Individuen heranzubilden, sondern (u. a.) darum, diesen Menschen die selbstbestimmte Teilhabe an der Gesellschaft und die soziale Integration zu ermöglichen.

Um aber die Rechtfertigung für Mathematikunterricht genau zu betrachten, sollten wir zunächst klären, was die Maschine derzeit und in Zukunft tatsächlich kann und erst danach darauf eingehen, ob dieses „Können“ den Mathematikunterricht verändern soll oder muss.

3 Was kann KI – von Wow! bis What?!

Wie bereits bemerkt, ist KI in der Lage, standardisierte Tests erfolgreich abzulegen. Würde man über Schülerinnen und Schüler mehr erfahren wollen, so wäre man auch nicht zufrieden, nur ihre Klassenarbeitsnoten zu erfahren, sondern man möchte im besten Fall mit ihnen reden und genau herausfinden, was sie können. So gibt es eine Chance, oberflächliches, auswendig gelerntes Wissen von Verstehen zu trennen. Das Tolle an einem Chatbot ist, dass wir genau das tun können: Wir können Fragen stellen, deren Antworten uns Aufschluss darüber geben können, was die internen mentalen Modelle der KI sind – so wie wir es auch mit Schülerinnen und Schülern tun.

An dieser Stelle ist nun eine Zwischenbemerkung notwendig: Die hier immer wieder stattfindende Anthropomorphisierung der KI durch Verwendung von Wendungen wie „sie sagt“, „sie denkt“, „sie kann“, „sie meint“ oder „sie weiß“ geschieht bewusst, um der Art und Weise der Kommunikation mit der Maschine gerecht zu werden, ohne zu unterstellen, dass die KI ein Bewusstsein hat. Die Diskussion dieser (auch

ethisch relevanten) Frage muss an anderer Stelle erfolgen und hat hier leider keinen Platz.

3.1 Male ein Bild!

Zu den klassischen Methoden der Erkenntnisgewinnung über die Psyche von Menschen gehört die Interpretation von Bildern. Nicht nur in der Therapie, sondern auch in der Forschung findet diese Anwendung (Billmann-Mahecha & Drexler, 2017). In der Mathematikdidaktik ist es ebenfalls üblich, Bilder nicht nur als Erklärung für mathematische Sachverhalte darzubieten, sondern sie einzufordern, um das mathematische Verständnis der Schülerinnen und Schüler zu ergründen, sowohl zur individuellen Kompetenzmessung als auch zur Analyse größerer curricularer Zusammenhänge (z. B. Kuzle & Glasnovic Gracin, 2020).

In (Kortenkamp & Dohrmann, 2023) wurden bereits die Beispiele in Abb. 1 vorgestellt, die mit DALL·E, einem AI-Bildgenerator, hergestellt wurden.

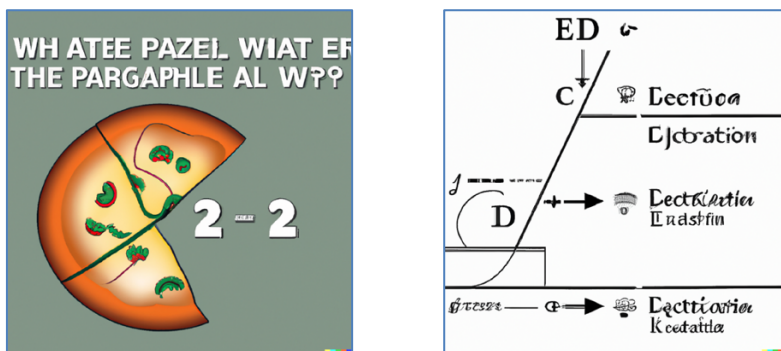


Abb. 1 Mit DALL-E generierte Erklärungen für „zwei Drittel einer Pizza“ und für das Epsilon-Delta-Kriterium für stetige Funktionen

Aus diesen Bildern lässt sich nicht auf mathematisches Verständnis schließen, würde ein Schüler oder eine Schülerin so ein Bild auf Nachfrage produzieren, so wäre man tief besorgt und müsste sich darauf einstellen, viel Zeit in das Aufholen von Schulstoff zu investieren.

Seit der Generierung der Grafiken in Abb. 1 ist – für KI – eine Ewigkeit vergangen. Insbesondere kann KI inzwischen wesentlich besser

Schrift generieren. Dennoch: Der mathematische Inhalt bleibt fragwürdig, wie die beiden „Infografiken“ zur Addition von Brüchen in Abb. 2 demonstrieren.

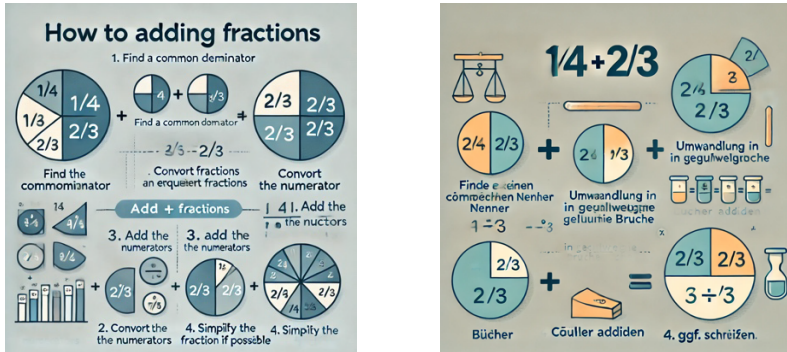
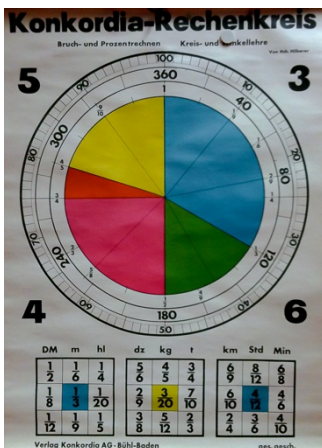


Abb. 2 Mit ChatGPT (GPT-4o) generierte Infografiken für die Addition von Brüchen, links Englisch, rechts die „deutsche Übersetzung“. Prompt: „Bitte erstelle eine Infografik zur Addition von Brüchen.“ und „Kannst du die Grafik auch auf Deutsch übersetzen?“

Dieser Einblick in die mathematische Denkweise von KI bestätigt das, was wir selbst schon oft bei KI-generierten Texten erlebt haben: Es kommt auf den Schein an, nicht auf den Inhalt. Die Grafiken ähneln dem, was man aus dem Mathematikunterricht kennt, betrachtet man aber Details, so sind diese inkonsistent und fast immer falsch. Wir müssen der KI aber auch zugestehen, dass der Schulunterricht für Schülerinnen und Schüler durchaus ähnliche Züge haben kann. Und nicht nur für die: Abb. 3 zeigt den „Konkordia-Rechenkreis“ von Heinrich Hilberer, den wir vor vielen Jahren an der PH Karlsruhe unter den Unterrichtsmaterialien gefunden haben, aber bis heute nicht entschlüsseln konnten, wie dieser funktioniert.



This image shows a mathematical tool called the "Konkordia-Rechenkreis," likely used for teaching fractions, percentages, and relationships involving circle geometry, angles, and units. Here's a breakdown:

1. **Circular Diagram:** The main part is a pie chart divided into colored sectors, likely representing fractions, degrees, or percentages of a whole. It is labeled with angles from 0° to 360° around the outer ring, corresponding to divisions of the circle. Inside the ring, there are numbers indicating fractions and their decimal equivalents or percentages, which likely relate to parts of the circle.
2. **Numbers around the Circle:** Numbers such as 5, 3, 4, and 6 seem to indicate multiples or sections of some scale, potentially representing different unit conversions or number systems used for various purposes.
3. **Units and Conversion Tables:** Below the circle, there are labeled grids with abbreviations like "DM" (possibly Deutsche Mark), "kg" (kilograms), and "Min" (minutes). These seem to be unit conversions or fractional relationships for various quantities such as money, distance, mass, and time.
4. **Fractions and Decimals:** The grid contains fractions (e.g., 1/3, 3/20) and their possible equivalents in different measurement units, which is reinforced by the coloring that corresponds to the sectors in the circle. For instance, the yellow and blue shading in the grids likely corresponds to the same colors in the circle sectors, showing visual representations of fractions or parts of the whole.

In summary, the "Konkordia-Rechenkreis" appears to be an educational tool that connects fractions, percentage calculations, and unit conversions, making it useful for teaching basic arithmetic and geometry concepts.

Abb. 3 Der Konkordia-Rechenkreis von Heinrich Hilberer, Verlag Konkordia AG, Bühl-Baden. Es ist unklar, wie dieser funktioniert. ChatGPT weiß dazu auch nicht mehr, erkennt aber immerhin, dass es sich um etwas handelt, mit dem man Bruchrechnung unterrichten könnte

3.2 Löse eine Sachaufgabe!

Vielleicht muss man ja auch nicht wirklich alles verstehen und erklären können? Genügt es, wenn die KI in der Lage ist, für uns Aufgaben zu lösen?

Wir versuchen dies mit einer Variation einer altbekannten Sachaufgabe: „Ein Quadrat mit Ecken A, B, C, D hat eine Kantenlänge von 200m. Eine Ameise startet bei A und schafft jeden Tag 200m, wird aber in der Nacht um die Hälfte zurückgeworfen. Nach wie vielen Tagen ist sie wieder bei A angekommen?“ (Rasch, 2001, S. 288)

Die Diskussion mit ChatGPT (Kortenkamp, 2024b) wird hier vollständig wörtlich wiedergegeben:

[ChatGPT] *Die Ameise muss insgesamt vier Seiten des Quadrats zurücklegen, um zu ihrem Ausgangspunkt A zurückzukehren. Jede Seite hat eine Länge von 200m.*

Die Ameise kann jeden Tag 200m vorwärts gehen, aber in der Nacht wird sie um die Hälfte zurückgeworfen, was bedeutet, dass sie jeden Tag nur 100m in die richtige Richtung geht.

Um die Gesamtstrecke von 800m zurückzulegen, benötigt die Ameise 8 Tage, da sie jeden Tag 100m in die richtige Richtung geht und insgesamt $8 \times 100m = 800m$ zurücklegt.

Das bedeutet, dass die Ameise nach 8 Tagen wieder bei A ankommen wird.

Diese Antwort sieht vernünftig aus, ist aber falsch. Die folgende erste Rückfrage ist aber auf jeden Fall möglich, auch wenn man selbst die Aufgabe nicht löst. Die darauffolgende Frage ist eine typische Hilfestellung, die Schülerinnen und Schülern helfen soll, ihre eigene Lösung konkret an einer Stelle zu hinterfragen.

[K.] *Bist Du sicher?*

[ChatGPT] *In diesem spezifischen Szenario bin ich mir sicher, dass die Ameise nach 8 Tagen wieder bei A ankommt. Allerdings kann es immer Faktoren geben, die die Berechnung beeinflussen, wie z.B. das Wetter oder Hindernisse auf der Strecke.*

[K.] *Wo ist die Ameise am Ende des siebten Tages?*

[ChatGPT] *Am Ende des siebten Tages hat die Ameise insgesamt 700m zurückgelegt. Davon hat sie an 6 Tagen je 100m in die richtige Richtung zurückgelegt, was insgesamt 600m ergibt. Zusätzlich hat sie an einem weiteren Tag 200m in die richtige Richtung zurückgelegt. Daher ist die Ameise am Ende des siebten Tages 100m entfernt von ihrem Ausgangspunkt A. Sie wird am achten Tag weitere 100m in die richtige Richtung gehen und A erreichen.*

An dieser Stelle ist es offensichtlich möglich, der KI ihren Fehler konkret zu demonstrieren:

[K.] *Aber 600m+200m sind doch 800m?*

[ChatGPT] *Sie haben Recht, ich bitte um Entschuldigung für den Fehler in meiner vorherigen Antwort. Die Ameise hat insgesamt 800m zurückgelegt, davon hat sie an 6 Tagen je 100m in die richtige Richtung zurückgelegt, was insgesamt 600m ergibt. An einem weiteren Tag hat sie 200m in die richtige Richtung zurückgelegt. Daher ist die Ameise am Ende des siebten Tages genau an ihrem Ausgangspunkt A angekommen.*

Wir halten fest, dass die KI hier in der Lage ist, Lösungsansätze für die Aufgabe zu liefern, die dem entsprechen, was auch Schülerinnen und Schüler leisten können. Sogar die Intervention einer Lehrkraft funktioniert genauso wie im Unterricht. Das bedeutet aber auch, dass die KI keine zuverlässigen Lösungen liefert, die Schülerinnen und Schüler direkt verwenden können. Für eine sinnvolle Nutzung der KI müssen sie in der Lage sein, die kritischen Fragen, die hier die Lehrkraft formuliert, selbst zu stellen.

3.3 Rechne etwas aus!

Als drittes und letztes Beispiel reduzieren wir die Anforderungen an die KI-Kompetenz weiter. War in Abschnitt 3.1 noch inhaltliches Verständnis und in Abschnitt 3.2 Modellierungskompetenzen gefragt, so verlangen wir nun reine Rechenfähigkeiten (also das, was Taschenrechner schon seit vielen Jahrzehnten im Unterricht beisteuern).

Noch vor ein paar Monaten war ChatGPT bei solchen Fragen schockierend schlecht. Hier ist ein gekürztes Beispiel aus einem längeren Dialog (Kortenkamp, 2024a): „Was ist $57373784873 * 7283884773$ “ be-

antwortet die KI mit „4177439138507155293009“, das korrekte Ergebnis ist aber 4179040380230345742909. Die „gefühlte Wahrheit“ wird auch auf die Frage „Bist du sicher?“ hemmungslos (eine KI kennt da keine Scham) mit „Ja“ beantwortet.

Auf die Bitte, davon überzeugt zu werden, versucht sich ChatGPT an einer Zerlegung des zweiten Faktors in $70000000000 + 2800000000 + 800000000 + 700000 + 8000 + 700 + 30 + 3$. Dies sieht auch zunächst gut aus, ist aber weder eine sinnvolle Zerlegung (warum 28?), noch ist sie korrekt, denn es werden nicht alle Stellen aufgeführt, und manche sind schlicht falsch (zum Beispiel 8000). Um diese längere Rechnung durchzuführen schlägt ChatGPT den Einsatz eines Hilfsmittels vor: „Um es zu vereinfachen, verwenden wir einen Taschenrechner oder ein Computerprogramm, um die exakte Multiplikation durchzuführen“ – auch dieser Schritt ist fragwürdig, denn wenn man einen Taschenrechner verwenden will, dann kann man dies ja viel sicherer direkt zu Beginn machen. Die Nutzung der Programmiersprache Python durch ChatGPT bestätigt dann angeblich das falsche Ergebnis von zuvor.

Es wird noch haarsträubender: Wir fragen ChatGPT nach der Probe – „Was ist $4177439138507155293009/72838847733$?“ Das ist wohl für „Kopfrechnen“ zu schwierig, darum setzt ChatGPT sofort Python ein und erhält das Ergebnis 57351801525,19829 (in Python wird ein Dezimalpunkt statt eines Dezimalkommas verwendet). Das abschließende Fazit der KI lautet: „Das zeigt, dass das ursprüngliche Multiplikationsergebnis korrekt ist, da der Quotient fast gleich der ursprünglichen Zahl 57373784873 ist. Die leichte Abweichung kann auf Rundungsfehler zurückzuführen sein.“

Der Fairness halber sei erwähnt, dass die allerneueste Version von ChatGPT nun direkt auf Python zurückgreift, um Rechenaufgaben zu lösen (Kortenkamp, 2024e). Auch in dieser Diskussion bleiben Fragen offen, zum Beispiel versucht GPT-4o eine stellenweise Zerlegung der Form $57373784873 \times 72838847733 = (5.737 \times 10^{10} + 3.737 \times 10^9 + \dots) \times (7.283 \times 10^{10} + 2.838 \times 10^9 + \dots)$, aber abgesehen davon ist die Gesamtdarstellung wesentlich überzeugender und nahezu fehlerfrei.

Damit sind wir aber bei der Grundfrage angelangt: *Wie können Schülerinnen und Schüler sich davon überzeugen, dass die Antwort der KI korrekt ist, und nicht nur überzeugt werden? Was ist also die Mathematik, die man noch beherrschen muss, um eine KI souverän einzusetzen?*

4 Versuch eines Fazits

Eine erste Konsequenz aus den bisherigen Beispielen ist, dass es offensichtlich wichtiger ist, eine Lösung begründen zu können als die Lösung zu finden. In der meinen Einführungsvorlesungen plädiere ich dafür, dass es wichtiger ist, verschiedene Begründungen dafür zu kennen, dass $6 \times 7 = 42$ ist, als das Ergebnis zu kennen. Fragt man ChatGPT nach solchen Begründungen, schwanken diese zwischen „ist halt so“ über Darstellungswechsel, die für Grundvorstellungen (vom Hofe, 1995) taugen, bis hin zu komplett absurden Behauptungen. Hier einige Beispiele aus (Kortenkamp, 2024c).

- *Wenn du sechs Gruppen mit jeweils sieben Elementen hast und sie zusammenzählst, erhältst du 42.*
(... ähnliche Formulierungen mit Gegenständen in Reihen und Spalten, Kugeln in Schachteln, Stunden in Tagen, Äpfeln pro Kind)
- *Die Zahl 42 wird in vielen verschiedenen Bereichen als Bedeutung oder Symbol verwendet, aber sie wird am häufigsten mit der Antwort auf die "Ultimative Frage des Lebens, des Universums und allem" in Douglas Adams' Roman "Per Anhalter durch die Galaxis" in Verbindung gebracht.*
- *Wenn du 42 Grad Celsius in Fahrenheit umrechnest, erhältst du 107,6 Grad Fahrenheit, was auch eine häufig verwendete Temperatur ist.*
- *Wenn du in einem sechsstelligen Zahlensystem rechnest, wobei jeder Buchstabe des Alphabets eine Zahl darstellt (A = 1, B = 2, usw.), ergibt die Summe der Buchstaben in "six times seven" (S=19, I=9, X=24, T=20, I=9, M=13, E=5, S=19, S=19, E=5, V=22, E=5, N=14) ebenfalls 42.*

Ein paar Monate später konnten ChatGPT die absurden Antworten fast abgewöhnt werden (Kortenkamp, 2024d), stattdessen konzentriert es sich auf formale Argumentationen:

1. **Definition der Multiplikation:** Multiplikation ist wiederholte Addition. Wenn du 66 siebenmal addierst, erhältst du $6 + 6 + 6 + 6 + 6 + 6 + 6 + 6 = 42$.
2. **Tausche die Faktoren:** Nach dem Kommutativgesetz der Multiplikation ist $6 \times 7 = 7 \times 6$, was ebenfalls 42 ergibt.
3. **Zerlegung:** Du kannst 6×7 als $(5 + 1) \times 7$ schreiben, was $5 \times 7 + 1 \times 7 = 35 + 7 = 42$ ergibt.
4. **Verwendung der Division:** Da $42 \div 7 = 6$ und $42 \div 6 = 7$, ergibt die Umkehrung der Division $6 \times 7 = 42$.
5. **Multiplikationstabellen:** Die Multiplikationstabelle zeigt, dass $6 \times 7 = 42$ aufgeführt wird.
6. **Flächenberechnung:** Wenn du ein Rechteck mit einer Länge von 6 und einer Breite von 7 hast, beträgt die Fläche $6 \times 7 = 42$.
7. **Teilmengenbildung:** Wenn du eine Menge von 42 Elementen hast und diese in 6 gleich große Gruppen aufteilst, enthält jede Gruppe 7 Elemente, was zeigt, dass $6 \times 7 = 42$.
8. **Fibonacci-Multiplikation:** 6×7 folgt den einfachen Gesetzen der Arithmetik, die auch in verschiedenen mathematischen Sequenzen, wie z.B. der Fibonacci-Folge, angewendet werden.
9. **Programmierung:** Ein einfacher Computerbefehl zur Berechnung von 6×7 würde ebenfalls 42 liefern, da die Arithmetik in der Computerwissenschaft auf den gleichen Grundprinzipien basiert.
10. **Zählbares Beispiel:** Wenn du sechs Gruppen mit jeweils sieben Objekten (z.B. Steine) hast und alle zusammenzählst, erhältst du 42 Objekte.

Gut brauchbar ist immerhin Argument 3! An dieser Stelle muss dringend erwähnt werden, dass das Training einer KI behavioristisch funktioniert – gute Antworten werden belohnt, schlechte Antworten bestraft, und die Gewichtungen der intern verwendeten Parameter werden so angepasst, dass immer öfter belohnt als bestraft wird. Fast so wie in der Schule – auch Kindern kann man kreative Antworten („42 ist ja wie bei Douglas Adams“) abgewöhnen.

Zurück zur ursprünglichen Frage: Was müssen Schülerinnen und Schüler denn noch können, wenn es eine Maschine gibt, die immer zuverlässiger richtige Antworten und richtige Begründungen liefert?

Leider können wir uns in keiner Weise darauf verlassen, dass zu irgendeinem Zeitpunkt die KI zuverlässig korrekte Informationen liefert. In (Hicks et al., 2024) wird dies auf der Grundlage des Begriffs *bullshit* verdeutlicht – die KI halluziniert nicht nur (behauptet also Dinge, die frei erfunden sind), sondern gibt Äußerungen von sich, die der Definition von *bullshit* nach Harry Frankfurt (2005) genügen – sie werden nicht durch den Wunsch der KI nach Täuschung charakterisiert, sondern zeichnen sich durch die rücksichtslose Missachtung der Wahrheit aus.

Das sehen wir in den obigen Beispielen, aber auch an den gerade gebrachten Argumenten zu $6 \times 7 = 42$ in der Arithmetik: Menschen müssen in der Lage sein, Antworten, die ihnen durch andere Instanzen – Mitschülerinnen und Mitschüler, Eltern, Lehrkräfte, Politiker, Taschenrechner, Computersoftware und letztlich auch KI – gegeben werden, auf ihren Wahrheitsgehalt zu untersuchen und zu überprüfen.

Und das gilt nicht nur für einfache Arithmetik und Algebra. Im klassischen Modellierungskreislauf, der schon bei Blum (1985) als „heute übliche *Modellauffassung* für mathematische Anwendungen“ mit Wurzeln in den 70er Jahren charakterisiert wird, können alle Schritte, nicht nur die innerhalb des mathematischen Modells, sondern auch der Wechsel zwischen realer Welt und mathematischem Modell, mittlerweile durch KI bearbeitet werden. Es ist aber notwendig, *jede* Bearbeitung durch die KI in allen Schritten zu überprüfen, da die KI selbst nicht in der Lage ist, logisch zu argumentieren. Dies ist einerseits durch die auf statistischen Verfahren basierende Funktionsweise generativer KI klar, wurde aber kürzlich auch empirisch nachgewiesen (Mirzadeh et al., 2024).

Diese Überprüfung der Maschine wurde bereits vor 30 Jahren auf der Herbsttagung des AK Mathematikunterricht und Informatik, als erste Computeralgebra-Systeme schulrelevant wurden zwar gefordert – „*DE-RIVE erfordert also vom Schüler eine kritische ‚Revision‘ und Überprüfung der gelieferten Ergebnisse. Neben seiner Rolle als Planer von Rechenabläufen wird er damit auch zum ‚Prüfer‘ von Ergebnissen.*“ (Weth, 1993) – doch in der abschließenden Podiumsdiskussion nicht mehr deutlich beachtet (Bender & Gruß, 1993).

Eine solche kritische Revision und Überprüfung kann nur dann glaubhaft sein, wenn diese durchgeführt werden kann, ohne die Maschine selbst wieder einzusetzen – man kann auch eine Autorität nicht fragen, ob sie wirklich die Wahrheit sagt, und eine Antwort erwarten, die mehr aussagt, als die zuvor geäußerte vermeintliche Wahrheit. Das „*Stimmt das?*“ als Frage an ChatGPT ist sinnlos. Es gibt aber einen Ausweg aus dieser Situation: So wie die Probe schon in der Grundschule genutzt wird, um die Lösung einer komplexeren Aufgabe (zum Beispiel eine schriftliche Division) über das Lösen einer einfacheren Aufgabe (zum Beispiel die Multiplikation als Gegenaufgabe zur Division) zu überprüfen, so können über *Zertifikate* (Kortenkamp & Larkin, 2024) die Lösungen der Autorität überprüft werden. Dies ist im zuvor vorgestellten Dialog (Kortenkamp, 2024a) noch schief gegangen, inzwischen schlägt aber ChatGPT selbst schon hilfreiche Zertifikate auf Nachfrage vor. Die schon in der Grundschule durchführbare 9er und 11er-Probe kommt als konkreter, im Kopf durchführbarer Vorschlag zur Überprüfung großer Multiplikationen (Kortenkamp, 2024e): *„Ja, die Modulo-9-Rechnung lässt sich im Kopf durchführen, indem du die Quersumme der Zahl bildest (die Summe der einzelnen Ziffern) und das Ergebnis dann durch 9 teilst.“*

Diese mathematische Denkweise, das Überprüfen von Argumentationen durch die Zerlegung in einfachere und andersartige Probleme, muss zentraler Bestandteil des Mathematikunterrichts werden. Wir müssen damit umgehen, dass KI ein mächtiges, aber ein stets zu hinterfragendes Werkzeug ist. Das unterscheidet sie aber nicht von anderen Autoritäten – auch Autokraten, Diktatoren, harmlose Politiker und weniger harmlose Politiker, müssen stets überprüft und in Frage gestellt werden. Und das gelingt nicht nur dann, wenn man schlauer als sie ist, sondern gerade dann, wenn man Mathematik beherrscht. Dies steht in einer Reihe mit der Demokratisierung des Rechnens durch Adam Ries, und ich zitiere hier (Führer, 1997): *„[...] und schließlich bedeutete Rechnenkönnen und die Probe machen auch Nachrechnenkönnen, -sollen und -dürfen. Dreifach haben damit Rechenmeister wie Ries zur Erweiterung und Emanzipation kollektiven Bewußtseins beigetragen und dem erst zaghaft aufkeimenden bürgerlichen Rechtsempfinden Ausdrucksmittel verschafft.“*

Jenseits dieser Ausführungen möchte ich abschließend festhalten, dass für den – sehr wahrscheinlichen – Fall der Fälle, das KI die Macht über die Menschheit erlangt, uns immer noch die Strategie bleibt, möglichst nett, süß und freundlich zu sein, so dass die KI beschließt, dass sie uns weiterhin füttert und würdevoll behandelt. Ich nenne diese Strategie die Hundewelpen-Strategie.

Literatur

- Balacheff, N. (1993a). Artificial Intelligence and Mathematics Education: Expectations and Questions. *proc. 14th Biennial of the AAMT*, 1–24.
- Balacheff, N. (1993b). Artificial Intelligence and Real Teaching. In C. Keitel & K. Ruthven (Hrsg.), *Learning from Computers: Mathematics Education and Technology* (S. 131–158). Springer. https://doi.org/10.1007/978-3-642-78542-9_6
- Bender, P., & Größ, D. (1993). Einige Fragen an die Mathematikdidaktik anlässlich der zunehmenden Nutzung des Computers in der Schule -Nachlese zu einer Podiumsdiskussion. In H. Hischer (Hrsg.), *Wieviel Termumformung braucht der Mensch? Fragen zu Zielen und Inhalten eines künftigen Mathematikunterrichts angesichts der Verfügbarkeit informatischer Methoden* (S. 126–129). Franzbecker.
- Billmann-Mahecha, E., & Drexler, H. (2017). Auswertung von Zeichnungen als qualitatives Verfahren in der Psychologie. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 1–20). Springer Fachmedien. https://doi.org/10.1007/978-3-658-18387-5_53-1
- Blum, W. (1985). Anwendungsorientierter Mathematikunterricht in der didaktischen Diskussion. *Mathematische Semesterberichte*, 32(2), 195–232.
- ChatGPT: *Optimizing Language Models for Dialogue*. (2022, November 30). OpenAI. <https://openai.com/blog/chatgpt/>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). *Chatting and Cheating. Ensuring academic integrity in the era of ChatGPT*. EdArXiv. <https://doi.org/10.35542/osf.io/mrz8h>
- Frankfurt, H. G. (2005). *On Bullshit*. Princeton University Press. <https://doi.org/10.1515/9781400826537>
- Führer, L. (1997). *Pädagogik des Mathematikunterrichts*. Vieweg+Teubner Verlag. <https://doi.org/10.1007/978-3-663-14678-0>
- Gadanidis, G. (2017). Artificial intelligence, computational thinking, and mathematics education. *The International Journal of Information and Learning Technology*, 34(2), 133–139. <https://doi.org/10.1108/IJILT-09-2016-0048>

- Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, 26(2), 38. <https://doi.org/10.1007/s10676-024-09775-5>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence In Education: Promises and Implications for Teaching and Learning*. <https://curriculumre-design.org/our-work/artificial-intelligence-in-education/>
- Hwang, S. (2022). Examining the Effects of Artificial Intelligence on Elementary Students' Mathematics Achievement: A Meta-Analysis. *Sustainability*, 14(20), 13185. <https://doi.org/10.3390/su142013185>
- Kortenkamp, U. (2024a). *ChatGPT (May 2024): Was ist 57373784873 * 72838847733* [Chat Transcript]. <https://chatgpt.com/share/e23cbad0-fad4-4625-9321-ac8ea400b720>
- Kortenkamp, U. (2024b). *ChatGPT (May 2024): Ameise im Quadrat* [Chat Transcript]. <https://chatgpt.com/share/3f228267-8b71-44a5-8273-f96950f313c8>
- Kortenkamp, U. (2024c). *ChatGPT (May 2024): Gib mir 10 Argumente dafür, dass $6*7=42$ ist* [Chat Transcript]. <https://chatgpt.com/share/3b03315a-d119-41dd-9635-9766b9147297>
- Kortenkamp, U. (2024d). *ChatGPT (Oct 2024): Gib mir 10 Argumente dafür, dass $6*7=42$ ist* [Chat Transcript]. <https://chatgpt.com/share/3b03315a-d119-41dd-9635-9766b9147297>
- Kortenkamp, U. (2024e). *ChatGPT (Oct 2024): Was ist 57373784873 * 72838847733?* [Chat Transcript]. <https://chatgpt.com/share/67163b9f-9860-800d-a620-e735ff9afb22>
- Kortenkamp, U., & Dohrmann, C. (2023). Pre-service teacher training with AI: Using ChatGPT discussions to practice teacher-student discourse. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel, & M. Tabach (Hrsg.), *Proceedings of the 46th Conference of the International Group for the Psychology of Mathematics Education* (Bd. 3, S. 187–194). University of Haifa, Israel: PME.
- Kortenkamp, U., & Larkin, K. (2024). How can virtual geometry manipulatives be used in ways that mitigate their ontological, technological and pedagogical limitations? In T. Lowrie, A. Gutiérrez, & F. Emprin (Hrsg.), *Proceedings of the Twenty-Sixth ICMI Study Advances in Geometry Education* (S. 369–376). https://icmistudy26.sciencesconf.org/data/pages/26th_ICMI_Study_Proceedings_2.pdf
- Kuzle, A., & Glasnovic Gracin, D. (2020). Making Sense of Geometry Education through the Lens of Fundamental Ideas: An Analysis of Children's Drawings. *The Mathematics Educator*, 29(1), 7–52.

Lee, V. R., Pope, D., Miles, S., & Zárate, R. C. (2024). Cheating in the age of generative AI: A high school survey study of cheating behaviors before and after the release of ChatGPT. *Computers and Education: Artificial Intelligence*, 7, 100253. <https://doi.org/10.1016/j.caeai.2024.100253>

Mirzadeh, I., Alizadeh, K., Shahrokhi, H., Tuzel, O., Bengio, S., & Farajtabar, M. (2024). *GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models* (arXiv:2410.05229). arXiv. <http://arxiv.org/abs/2410.05229>

Mohamed, M. Z. bin, Hidayat, R., Suhaizi, N. N. binti, Sabri, N. binti M., Mahmud, M. K. H. bin, & Baharuddin, S. N. binti. (2022). Artificial intelligence in mathematics education: A systematic literature review. *International Electronic Journal of Mathematics Education*, 17(3), em0694. <https://doi.org/10.29333/iejme/12132>

Rasch, R. (2001). *Zur Arbeit mit problemhaltigen Textaufgaben im Mathematikunterricht der Grundschule eine Studie zu Herangehensweisen von Grundschulkindern an anspruchsvolle Textaufgaben und Schlussfolgerungen für eine Unterrichtsgestaltung, die entsprechende Lösungsfähigkeiten fördert* [Habilitationsschrift, Franzbecker]. WorldCat.

Rzesnitzek, F. (1912). *System der Pädagogik als Volkserziehung*. Verlag der Jos. Kösel'schen Buchhandlung.

teddy [@davidtsong]. (2022, Dezember 2). *I made ChatGPT take a full SAT test. Here's how it did: <https://t.co/734sPFU3HY>* [Tweet]. X (Formerly Twitter). <https://x.com/davidtsong/status/1598767389390573569>

vom Hofe, R. (1995). *Grundvorstellungen mathematischer Inhalte*. Spektrum Akademischer Verlag.

Weth, T. (1993). Zum Rollenwechsel des Schülers beim Arbeiten mit Unterrichtssoftware. In H. Hischer (Hrsg.), *Wieviel Termumformung braucht der Mensch? Fragen zu Zielen und Inhalten eines künftigen Mathematikunterrichts angesichts der Verfügbarkeit informatischer Methoden* (S. 106–110). Franzbecker.

Winter, H. (1995). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 61, 37–46. <https://doi.org/10.1515/dmvm-1996-0214>

Prof. Dr. Ulrich Kortenkamp
Universität Potsdam
Karl-Liebknecht-Str. 24-25
14476 Potsdam OT Golm
ulrich.kortenkamp@uni-potsdam.de