

- OpenAI. (2023c). Educator considerations for ChatGPT. Zuletzt geprüft am 9. 8. 2023. platform.openai.com/docs/chatgpt-education
- OpenAI. (2023d). Our approach to AI safety. Zuletzt geprüft am 6. 8. 2023. openai.com/blog/our-approach-to-ai-safety
- Richter, C. & Schmitz, M. (2014). Übersicht zur Geometrie in der Schule – Sekundarstufe I. www.mi-didaktik.uni-jena.de/fmididaktikmedia/dateien-ressourcen/jenaer-schriften/richter-und-schmitz-2014-uebersicht-zur-geometrie-in-der-schule-sekundarstufe-i.pdf
- Santhosh, R., Abinaya, M., Anusuya, V. & Gowthami, D. (2023). ChatGPT: Opportunities, Features and Future Prospects. In *7th International Conference on Trends 35 in Electronics and Informatics (ICOEI)* (S. 1614–1622).
- Schulministerium NRW (2023). *Umgang mit textgenerierenden KI-Systemen: Ein Handlungsleitfaden*. MSB NRW. www.schulministerium.nrw/system/files/media/document/file/handlungsleitfaden_ki_msb_nrw_230223.pdf
- Voigt, J. (1994). Entwicklung mathematischer Themen und Normen im Unterricht. In H. Maier, & J. Voigt (Hrsg.), *Verstehen und Verständigung: Arbeiten zur interpretativen Unterrichtsforschung* (S. 77–111). Aulis.
- Weßels, D. (2022, 20. Dezember). ChatGPT – ein Meilenstein der KI-Entwicklung. www.forschung-und-lehre.de/lehre/chatgpt-ein-meilenstein-der-ki-entwicklung-5271 (9. 8. 2023).
- Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L. & Tang, Y. (2023). A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development. *IEEE/CAA Journal of Automatica Sinica*, 10(5), 1122–1136.
- Frederik Dilling, Universität Siegen
E-Mail: dilling@mathematik.uni-siegen.de
- Kathrin Holten, Pädagogischen Hochschule Kärnten
E-Mail: kathrin.holten@ph-kaernten.ac.at
- Felicitas Pielsticker, Universität Siegen
E-Mail: pielsticker@mathematik.uni-siegen.de
- Ingo Witzke, Universität Siegen
E-Mail: witzke@mathematik.uni-siegen.de

Lernt die KI nun Sehen und Zeichnen? Chancen und Herausforderungen der Bildgenerierung und Bildinterpretation mit ChatGPT für die mathematikdidaktische Forschung

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

Zusammenfassung. Der vorliegende Beitrag führt die Diskussion über Herausforderungen der KI-Integration im Mathematikunterricht weiter und diskutiert aktuelle Einsatzmöglichkeiten von ChatGPT im Licht der jüngsten technischen Weiterentwicklungen. Insbesondere wird das Text-zu-Bild Modell DALL·E 3 und das Plugin Wolfram beleuchtet, welche das Large Language Modell GPT-4 befähigen, Bilder auf Basis eines Textinputs (Prompt) zu generieren. Seit September 2023 ist GPT-4 zudem in der Lage, visuelle Inhalte zu analysieren und zu deuten. In diesem Zusammenhang können Bilddateien in ChatGPT als Input dienen und mittels eines Prompts interpretiert werden. Welche Bedeutung diese neuen Funktionen hinsichtlich der Verwendung in der mathematikdidaktischen Forschung haben können, wollen wir im Anschluss diskutieren.

Einleitung

Erhebliche Fortschritte im Bereich der Künstlichen Intelligenz konnte OpenAI durch das generative KI-Sprachmodell GPT-4 schon im Sommer 2023 verzeichnen, wie wir in unserem letzten Beitrag in den *Mitteilungen der GDM 115* darlegen konnten (Schorcht et al., 2023). Die Diskussion um Künstliche Intelligenz im Bildungskontext und im Speziellen im Mathematikunterricht ist seitdem spürbar intensiver geworden (Kasneci et al., 2023; Lutz-Westphal, 2023; Schleiß et al., 2023; Spannagel, 2023). Es existieren beispielsweise erste empirische Untersuchungen zum Umgang mit und der Einstellung zu Large Language Modellen im Bildungssektor (von Garrel et al., 2023; Chan & Hu, 2023; Preiß et al., 2023; Smolansky et al., 2023). Ein Jahr nach der Veröffentlichung von GPT präsentiert sich das

Large Language Modell von OpenAI kontinuierlich mit neuen Funktionen wie beispielsweise der Einführung von Plugins wie Wolfram (Alpha) oder der Möglichkeit der Programmierung von individuell maßgeschneiderten GPTs (wie etwa der Generator für Kopfübungen von Brüggemann, 2023). Erst Anfang November brachte OpenAI das Modell GPT-4 zudem in die Lage, aus integrierten Technologien wie der bildgenerierenden KI DALL·E 3, der Datenanalyse Advanced Data Analysis, dem Browser Bing und weiteren Plugins für die Erfüllung eines Prompts eines dieser Werkzeuge selbstständig auszuwählen. Zusätzlich ist insbesondere durch die Integration des Large Language Modells in die Microsoft Office Umgebung durch M365 Copilot, welcher auf den GPT-Modellen basiert, in der Bezahlversion von Office die KI-Umgebung seit November 2023 für einen noch größeren Teil der Bevölkerung unmittelbar zugänglich und beginnt seine Integration in die Arbeitswelt. Die neuen Möglichkeiten der Bildgenerierung und Bildinterpretation erweitern daher die Nutzungsmöglichkeiten der KI auch für das mathematische Lehren und Lernen in der beruflichen Praxis von Lehrkräften und auch die Möglichkeiten der Nutzung in der Lehrkräfteausbildung noch einmal deutlich weiter. Wir stellen in diesem Beitrag daher erneut Chancen und Herausforderungen in diesem speziellen Bereich vor.

Unterschiedliche Algorithmen der Bildgenerierung

Um zu verstehen, wie die Text-zu-Bild-Modelle funktionieren, bedarf es einiger technologischer Details. Die Generierung von Bildern mit Algorithmen beruht auf der Übersetzung einer sprachlichen Eingabe in eine grafische Darstellung. Im Wesentlichen werden dabei verschiedene Algorithmen verwendet. Die bekanntesten Algorithmen sind Generative Adversarial Networks (GAN), Autoencoder und Diffusionsmodelle (Linde, 2023). Bei Generative Adversarial Networks (GAN) – erstmals beschrieben bei Goodfellow et al. (2014) – stehen zwei neuronale Netzwerke (Generator und Diskriminator) in einem Wettbewerb zueinander. Während der Generator Bilder aus zufälligem Rauschen generiert, beurteilt der Diskriminator die erzeugten Bilder hinsichtlich ihres authentischen Aussehens. In diesem iterativen Prozess verbessert der Generator kontinuierlich seine Fähigkeit, realitätsnahe Bilder zu generieren, während der Diskriminator gleichzeitig seine Kompetenz im Identifizieren von gefälschten Bildern verfeinert. Autoencoder sind generative Modelle, die aus einem Encoder zur Umwandlung von Eingaben (z. B. Bildern) in einen kompakten Vektor und einem Decoder zur Rekonstruktion oder Neuerstellung von Bildern bestehen. Im Trainingsprozess ler-

nen Autoencoder, wichtige Merkmale aus den Bildern effizient zu extrahieren (z. B. zur Bildrauschen-Reduktion), und daher können sie auch zur Generierung neuer Bilder eingesetzt werden. Diffusionsmodelle sind neuere generative Modelle, die auf stochastischen Prozessen beruhen (Linde, 2023). Sie transformieren Bilder schrittweise in Rauschen und lernen, diesen Vorgang umzukehren, um realistische Bilder zu erzeugen. In der Umkehrung wird, ausgehend von einem Bild, das ausschließlich Zufallsrauschen enthält, ein ursprüngliches Bild hineininterpretiert, das typischerweise noch sehr grob und verschwommen ist. Diese erste Version wird dann erneut mit Rauschen überlagert, woraus eine zweite Version berechnet wird. Dieser Prozess wird mehrfach wiederholt, wobei das künstlich hinzugefügte Rauschen bei jedem Durchgang schwächer wird. Auf diese Weise können schrittweise immer mehr Details aus der anfänglichen Grobskizze herausgearbeitet werden. Diese Methode kann in einigen Fällen bessere Ergebnisse als GANs oder Autoencoder erzielen und wird in bildgenerierenden KI-Modellen wie z. B. DALL·E 3 eingesetzt. Für eine Text-zu-Bild-Generierung, wie sie in DALL·E 3 Verwendung findet, sind dann drei wesentliche Schritte notwendig: Im ersten Schritt wird die Spracheingabe in einen Vektor überführt, der die Bedeutung des Textes repräsentiert – das sogenannte Embedding. Anschließend wird dieses Text-Embedding in ein Grafik-Embedding transformiert, das den Inhalt und Stil eines Bildes kodiert, ohne jedoch bereits alle Details festzulegen. Im dritten Schritt konvertiert ein Diffusionsmodell das Grafik-Embedding in ein tatsächliches Bild (Linde, 2023). Ein gemeinsames Merkmal aller dieser Algorithmen ist, dass sie aufgrund einer gewissen Unschärfe arbeiten müssen. Dies führt dazu, dass die direkte Umwandlung von Text in Bild nicht immer perfekt gelingt, da Texte üblicherweise weniger Informationen als Bilder enthalten.

Bildgenerierung in GPT-4 mit DALL·E 3

Die neueste Variante des diffusionsbasierten Text-zu-Bild-Modells DALL·E 3 arbeitet zusätzlich mit einem sog. „Image Captioner“, der jedem Bild im ursprünglichen Trainingsdatensatz eine neue Beschreibung zuordnet und somit Text-zu-Bild-Passung im Prozess des Embedding verbessert (Bettler et al., 2023). Je präziser die Prompt-Eingabe also ist, desto besser kann das KI-Modell die Bildausgabe an den Prompt anpassen. Dennoch ist das System von den notwendigen Diffusionsmodellen abhängig, weswegen die Ausgabe immer mit kreativen Elementen einhergeht – d. h. auch hier „halluziniert“ die KI. Wir haben exemplarisch Bilder zu Standardsituationen im Arithmetikunterricht in



Abbildung 1. Mit DALL·E 3 unter GPT-4 erstellte Abbildung zum Prompt Rechenrahmen (links) und Plättchen (rechts)

der Grundschule generieren lassen (Abb. 1). Hierzu haben wir in ChatGPT mit der Integration von DALL·E 3 folgende Prompts eingegeben:

Generiere mir eine Abbildung aus dem Mathematikunterricht: Zwei Kinder sitzen vor einem Rechenrahmen mit 100 Kugeln. In jeder Reihe sind 10 Kugeln. In jeder Reihe sind erst fünf Kugeln weiß, dann fünf Kugeln rot. Nach fünf Reihen wechselt die Farbe und es sind erst fünf Kugeln rot, dann fünf Kugeln weiß in jeder Reihe. Die Kinder auf dem Bild berechnen die Aufgabe $13 - 4$.

und

Generiere mir ein Bild mit einem Kind, das fünf rote und drei blaue runde Plättchen vor sich liegen hat. Daneben liegt ein Blatt Papier auf dem steht $3 + 5 = 8$.

Das linke Bild in Abbildung 1 zeigt zwei Kinder vor einem „Rechenrahmen“, der dem im Prompt beschriebenen Rechenrahmen nur teilweise genügt. Es finden sich weiße und rote Kugeln auf den Stangen, die jedoch keine erkennbare Struktur aufweisen. Die Anzahl der Kugeln pro Reihe ist nicht einheitlich und meistens nicht zehn. Statt zehn sind neun Reihen zu erkennen. Auch die Aufgabe „ $13 - 4$ “ ist nicht erkennbar. Das System halluziniert also andere Elemente in das Bild hinein, die zu einer mathematikdidaktisch unüblichen Ausgabe führen. Insgesamt erfüllt die Ausgabe also höchstens teilweise die Eingabe. Bestimmte Elemente der Ausgabe sind enthalten, weil sie im Prompt nicht explizit negiert wurden. In unseren Versuchen hat sich allerdings gezeigt, dass selbst eine Negierung vom Modell nicht immer erkannt wird. Die geforderte

Aufgabe ist zudem nicht sichtbar. Wir haben daher mit Anführungszeichen im zweiten Prompt versucht die Aufgabe sichtbar in der Darstellung zu erhalten (Abb. 1, rechts). Das zweite Bild zeigt ein Kind, das rote und blaue Plättchen vor sich liegen hat. Drei blaue Plättchen entsprechen der Eingabe, die Anzahl der roten Plättchen (fünf) wurde vom Modell ignoriert. Dieses Phänomen ist tatsächlich als „Prompt following“ bekannt, ein Problem, bei dem diffusionsbasierte Text-zu-Bild-Modelle Eingaben ignorieren oder im Verlauf des Prozesses nicht mehr korrekt verarbeiten (Rassin, Ravfogel & Goldberg, 2022; Betker et al., 2023). Die Plättchen sind allerdings entsprechend der Vorgabe rund. Zudem hat das Modell korrekt ein Blatt Papier generiert und die Aufgabe „ $3 + 5 = 8$ “ in einer Sprechblase so angeordnet, dass sie dem Blatt zugeordnet werden könnte. Interessant sind die Farbvarianten in den blauen Plättchen, die vom Modell halluziniert wurden. Der mathematikdidaktische Nutzen der auf diese Weise generierten Bilder ist allerdings noch weit von konkreten praktischen Anwendungen entfernt. Inwiefern genauere Eingaben mit Techniken des Prompt-Engineering zu besseren Ausgaben führen, bleibt noch offen. Der Abgleich der generierten Bilder mit im Internet vorhandenen Abbildungen, um beispielsweise die Darstellung des Rechenrahmens anzupassen, ist bisher noch nicht möglich. Lehrende können allerdings die Bildgenerierung schon aktuell in der Bezahlversion von ChatGPT nutzen, um gewünschte visuelle Darstellungen zu generieren. Inwiefern diese Bildgenerierung allerdings auch gewissen mathematischen und mathematikdidaktischen Qualitätsstandards entspricht, die für Visualisierungen wünschenswert sind, ist wohl individuell im Einzelfall zu entscheiden. Detaildarstellungen mit mathematischen Inhalt

ten konnten wir in unseren Versuchen bisher nicht zufriedenstellend generieren. Für eine allgemeine Darstellung einer Gruppenarbeit oder des Spiels auf dem Pausenhof kann DALL·E 3 aber schon passable Abbildungen erstellen.

Eine sich anschließende Frage, die sich durch die Verwendung des KI-Modells stellt, bezieht sich auf die Bildrechte an den generierten Bildern. OpenAI (2023) schreibt hierzu in seinen Terms of Use unter Punkt 3(a) folgendes:

(a) **Your Content.** You may provide input to the Services (“Input”), and receive output generated and returned by the Services based on the Input (“Output”). Input and Output are collectively “Content.” As between the parties and to the extent permitted by applicable law, you own all Input. Subject to your compliance with these Terms, OpenAI hereby assigns to you all its right, title and interest in and to Output. This means you can use Content for any purpose, including commercial purposes such as sale or publication, if you comply with these Terms. OpenAI may use Content to provide and maintain the Services, comply with applicable law, and enforce our policies. You are responsible for Content, including for ensuring that it does not violate any applicable law or these Terms.

Alle durch OpenAI generierten Outputs sind also frei verfügbar und selbst für kommerzielle Zwecke verwendbar.

Bildgenerierung in GPT-4 mit Plugin Wolfram

Eine Bildgenerierung, die zu mathematischen Zwecken eingesetzt werden kann, ist unter anderem auch mit dem Plugin Wolfram verfügbar. Das Plugin basiert auf dem Zugang Wolfram Alpha und der Systemsprache Wolfram Language (Wolfram, 2023). Durch die Auswahl des Plugins unter GPT-4 kann direkt auf die Webversion Wolfram Alpha zugegriffen werden. ChatGPT übersetzt dafür den Prompt in Wolfram Language und stellt eine Anfrage. Wolfram kalkuliert die benötigte Ausgabe und sendet es in Form einer URL an ChatGPT zurück. Das KI-Modell stellt die Ausgabe dann im Chat dar. Die Übersetzung kann über den Button „Wolfram benutzt“ eingesehen werden (Spannagel, 2023). Wir haben mit einigen Prompts versucht, Bilder zu erstellen, die für den Mathematikunterricht gewinnbringend genutzt werden könnten. Das Plugin Wolfram liefert uns deutlich bessere Ergebnisse als DALL·E 3, die in vielen Fällen schon den Bedürfnissen der genauen Ausgabe des mathematischen Kerns entsprechen. In Abbildung 2 sind zwei solche unterschiedlichen Ausgaben erkennbar. Im linken

Teil der Abbildung 2 soll ein Funktionsgraph zur gebrochen-rationalen Funktion $f: \mathbb{R} \setminus \{1\} \rightarrow \mathbb{R}$ mit $f(x) = \frac{x^2-1}{x^3-1}$ (mit stetig hebbarer Definitionslücke bei $x = 1$) geplottet werden. Der Aufforderung kommt ChatGPT mit Wolfram zwar nach, es lässt sich hier aber kritisieren, dass der sprachliche Rahmen des Bild-Outputs nicht ganz angemessen mit der Definitionslücke umgeht, denn es wird einerseits nicht korrekt auf das Verhalten der Funktion an der Definitionslücke verwiesen, andererseits wird die Definitionslücke im abgebildeten Graphen bereits stetig behoben, ohne dass eine stetige Fortsetzung angegeben wird (ein ähnliches Ergebnis erhält man aber übrigens auch mit GeoGebra). Im rechten Teil der Abbildung 2 wird eine quadratische Pyramide mit farbigen Flächen und Kanten perspektivisch dargestellt, die gut brauchbar für das Arbeiten mit Visualisierungen geometrischer Körper im Unterricht erscheint.

Das Plugin Wolfram stellt aus unserer Sicht eine bessere Möglichkeit als DALL·E 3 zur Generierung von Visualisierungen und Darstellungen für den Mathematikunterricht dar – aber die Nützlichkeit lässt sich natürlich immer nur im Einzelfall und vor dem Hintergrund des Zwecks der Erstellung von Abbildungen entscheiden. Die Halluzinationen im Wolfram Plugin erscheinen weniger gravierend, möglicherweise auch, weil das Modell anhand großer naturwissenschaftlicher Daten trainiert wurde und auf einen anderen bildgenerierenden Algorithmus zurückgreift. Deutlich wird aus unseren Versuchen aber, dass selbst hier eine kritische Prüfung des KI-Outputs nicht unterbleiben kann.

Bildinterpretation mit GPT-4

Als zweite wichtige Anwendung der neuen Funktionen im mathematikdidaktischen Forschungszusammenhang thematisieren wir das Interpretieren von Abbildungen. Für eine Bildinterpretation können nun Bilder und PDF-Dokumente in ChatGPT hochgeladen werden. ChatGPT ist mit Bilderkennungs- und Verarbeitungsalgorithmen ausgestattet, die eine Interpretation der hochgeladenen Dateien ermöglichen. Das Sprachmodell kann die erkannten Informationen mit dem Kontext der vorherigen Konversationen verbinden, um kontextbezogene Antworten zu generieren. Diese Möglichkeit erweitert damit die Bandbreite der Interaktionen mit der KI deutlich. Angeregt durch diese Neuerungen haben wir dem Modell diagnostische Aufgaben gestellt, indem wir Lösungen von Lernenden bereitstellten und es zur Interpretation aufforderten. Dazu nutzten wir einmal eine Termumformungsaufgabe mit einem algebraischen Fehler (Abb. 3, links) und eine Lösung einer Femi-Aufgabe, in der die Höhe einer menschlichen Pyramide, bestehend

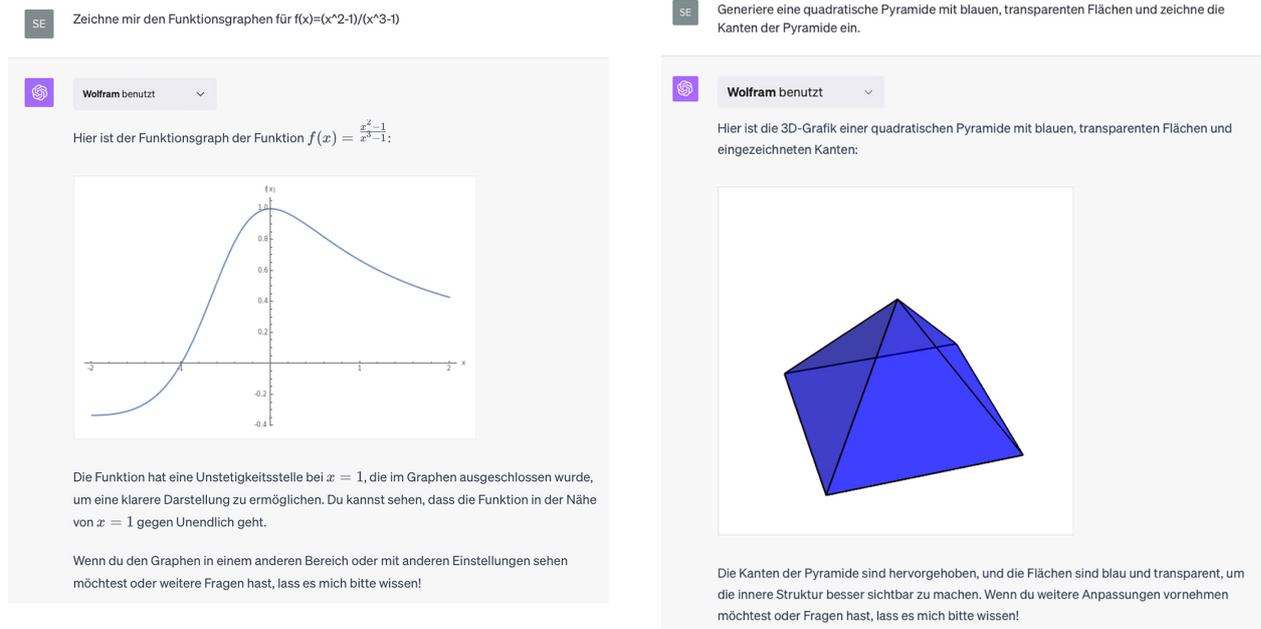


Abbildung 2. Mit Plugin „Wolfram“ unter ChatGPT-4 erstellte Darstellung eines Funktionsgraphen (links) und eines geometrischen Körpers (rechts)

aus den Kindern der eigenen Grundschule, modelliert und berechnet werden sollte (Abb. 3, rechts). Die Eingabe von echten individuellen Schülerlösungen aus dem Mathematikunterricht in ChatGPT ist angesichts datenschutzrechtlicher Vorgaben gerade bei vulnerablen Personengruppen natürlich bedenklich. Dazu weist OpenAI in den Terms of Use unter Punkt 2(c) auch auf diese Problematik hin: „*Restriction*. You may not [...] (vii), send us any personal information of children under 13 or the applicable age of digital consent.“ Um dennoch einen Eindruck von den Möglichkeiten der Bildinterpretation des KI-Modells zu erhalten, haben wir die Inhalte anonymisiert und mit einer neuen Handschrift versehen. Personengebundene Informationen von Lernenden (und wenn es nur Handschriften sind) sind somit nicht mehr enthalten.

Im ersten eher offenen Prompt zur Fehleranalyse in Abbildung 3 (links) erfasst das KI-Modell die Handschrift korrekt und analysiert die Lösung der Lernenden anhand der Rechenkompetenz, Problemlösekompetenz, Kommunikation und Argumentation. Diese Kompetenzen stimmen allerdings nur teilweise mit denen aus den Bildungsstandards überein. Das KI-Sprachmodell scheint also nicht automatisch über das Internet auf die KMK-Standards zuzugreifen. Die Quelle der angegebenen Kompetenzen wird hier auch nicht ersichtlich. Eine „Halluzination“ des KI-Sprachmodells ist in Bezug auf die Kompetenzen daher nicht auszuschließen. Der Schülerfehler hingegen wird erstaunlich gut und

korrekt erkannt, und auf die richtige Lösung hingewiesen. Positiv erwähnenswert ist auch der Abschnitt zur Argumentation, in der ChatGPT aufgrund der mangelnden diagnostischen Informationen gerade keine Aussage über die Begründungsfähigkeit der Lernenden macht. Das Potential der diagnostischen Fähigkeiten des KI-Sprachmodells zeigt sich unserer Meinung deutlich, eine Referenz auf die KMK-Bildungsstandards wäre hier aber natürlich wünschenswert gewesen.

Um die Beliebigkeit der Interpretation mathematischer Kompetenzen einzugrenzen, ergänzten wir den zweiten Prompt daher mit einem PDF-Dokument. Der zweite, eher geschlossene Prompt (Abb. 3, rechts), gibt somit die Kompetenzen der Bildungsstandards im Bereich der Primarstufe (KMK, 2022) als Maßstab mathematischer Kompetenzen vor. Die gestellte Fermi-Aufgabe für die Kinder lautete: „Wie viele Kinder besuchen die Schule? Wie hoch ist eine menschliche Pyramide aus den Kindern unserer Schule?“ In der Schülerlösung wurde zuerst die Gesamtzahl der Lernenden der eigenen Grundschule über eine Multiplikation der beiden Faktoren „Durchschnittliche Anzahl der Kinder je Klasse“ und „Anzahl der Klassen in der Schule“ mathematisiert. Danach wurde von dieser Gesamtzahl über fortgesetzte Subtraktion die jeweilige Anzahl benötigter Kinder je Pyramidenstufe abgezogen ($128 - 1 - 2 - 3 - 4 - 5 - \dots$) bis die errechnete Gesamtzahl ausgeschöpft war. In der dargebotenen Lösung fehlt allerdings der Schritt zur Verbindung mit dem Größenbereich Längen.

You

Aufgabe: Vereinfach den Term so weit wie möglich.

a.) $3x - (x - y)$

$$3x - x + y$$

$$x(3 + y)$$

Analiere folgende Schülerlösung hinsichtlich der mathematischen Kompetenzen, wie sie in den Bildungsstandards zu finden sind.

ChatGPT

Die Schülerlösung auf dem Bild zeigt eine mathematische Aufgabe, bei der ein Term vereinfacht werden soll. Die Aufgabe lautet: Vereinfache den Term $3x - (x - y)$ so weit wie möglich.

Der Schüler hat den Term wie folgt vereinfacht:

- $3x - x + y$
- $x(3 + y)$

Bei der Analyse der Schülerlösung im Hinblick auf mathematische Kompetenzen, wie sie in Bildungsstandards vorkommen, können verschiedene Aspekte betrachtet werden:

- **Rechenkompetenz:** Der Schüler zeigt ein Verständnis der Grundrechenarten, indem er die Klammer auflöst und die Terme korrekt zusammenfasst. Das Ausklammern im zweiten Schritt ist jedoch nicht korrekt, da $3x - x$ zu $2x$ vereinfacht werden sollte, nicht zu $x(3 + y)$. Dies deutet auf ein Missverständnis in Bezug auf die Distributivgesetze hin.
- **Problemlösekompetenz:** Der Schüler hat einen Ansatz zur Problemlösung verfolgt, der aber nicht zum korrekten Ergebnis führt. Dies zeigt, dass der Schüler ein grundlegendes Verständnis für die Struktur von Termen hat, aber möglicherweise Schwierigkeiten mit komplexeren algebraischen Manipulationen hat.
- **Kommunikation:** Der Schüler hat seine Gedankengänge schriftlich festgehalten, was gut ist. Allerdings wäre eine Überprüfung und Korrektur seiner Arbeit ein wichtiger Schritt, um die Richtigkeit zu gewährleisten.
- **Argumentation:** Es ist nicht klar, ob der Schüler die Schritte seiner Vereinfachung begründen kann, da die Lösung fehlerhaft ist. Ein Verständnis darüber, warum bestimmte Schritte genommen werden, ist ein wichtiger Teil der mathematischen Argumentationskompetenz.

Die Lösung zeigt, dass der Schüler eine Vorstellung davon hat, wie Terme vereinfacht werden, aber es fehlt an Genauigkeit und einem tieferen Verständnis der algebraischen Gesetze. Es wäre wichtig, die Fehler zu korrigieren und dem Schüler zu zeigen, wie die korrekte Vereinfachung aussehen sollte:

1. $3x - x + y$ sollte zu $2x + y$ vereinfacht werden, da $3x - x = 2x$.

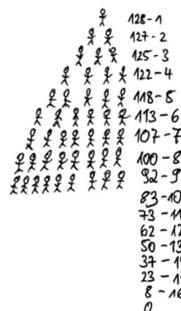
Durch eine solche Korrektur könnte der Schüler seine Fähigkeiten in den genannten Kompetenzbereichen verbessern.

You



Wie viele Kinder besuchen die Schule? Wie hoch ist eine menschliche Pyramide aus den Kindern unserer Schule?

Antwort: Es sind 128 Kinder in der Schule
Rechnung: $16 \cdot 8 = 128$ Es sind 8 Klassen



Analiere folgende Schülerlösung (sichtbar im angehängten Bild) hinsichtlich der prozessbezogenen Kompetenzen, wie sie in den Bildungsstandards zu finden sind. Nutze dafür die angehängten Bildungsstandards.

ChatGPT

Um die Schülerlösung im Bild hinsichtlich der prozessbezogenen Kompetenzen zu analysieren, werden die Kompetenzen, wie sie in den Bildungsstandards definiert sind, herangezogen:

1. **Mathematisch argumentieren:** Die Schülerlösung zeigt, dass der Schüler oder die Schülerin in der Lage war, eine mathematische Struktur zu erkennen und die Idee einer menschlichen Pyramide als ein mathematisches Problem zu repräsentieren. Sie haben eine Abnahme-Sequenz aufgestellt, um die Gesamtzahl der Schüler zu berechnen.
2. **Mathematisch kommunizieren:** Der Schüler oder die Schülerin hat seine/ihre Gedanken in schriftlicher Form dargelegt und mathematische Notationen verwendet, um die Lösung zu kommunizieren. Das zeigt eine Fähigkeit, mathematische Sachverhalte zu beschreiben und Lösungswege zu dokumentieren.
3. **Probleme mathematisch lösen:** Es wurde eine Strategie entwickelt, um das Problem schrittweise zu lösen. Der Schüler oder die Schülerin hat eine Reihe aufgestellt, die zur Lösung des Problems führt, was auf die Kompetenz hinweist, Lösungsstrategien zu entwickeln.
4. **Mathematisch modellieren:** Der Schüler oder die Schülerin hat das reale Phänomen der menschlichen Pyramide in ein mathematisches Modell übersetzt, indem eine Struktur für die Abnahme der Schülerzahl pro Reihe erstellt wurde.
5. **Mathematisch darstellen:** Die Lösung beinhaltet eine tabellarische Darstellung, die die Abnahme der Anzahl der Schüler pro Stufe der Pyramide zeigt. Diese Darstellung unterstützt das Verständnis des zugrundeliegenden Musters der Abnahme.
6. **Mit mathematischen Objekten und Werkzeugen arbeiten:** Der Schüler oder die Schülerin hat mit Zahlen und mathematischen Symbolen gearbeitet, um das Problem zu lösen. Dabei wurden mathematische Operationen wie Subtraktion und Multiplikation angewendet.

Die Schülerlösung zeigt, dass der Schüler oder die Schülerin in der Lage war, mehrere prozessbezogene Kompetenzen zu demonstrieren. Sie haben das Problem verstanden, eine Strategie zur Lösung entwickelt, ihre Gedanken klar kommuniziert und ein mathematisches Modell zur Darstellung des Problems verwendet. Das weist auf ein gutes Verständnis und die Fähigkeit hin, Mathematik zur Lösung realer Probleme anzuwenden.

Abbildung 3. Interpretation einer Schülerlösung (Reproduzierung zwecks Anonymisierung der Handschrift) in zwei Prompt-Varianten durch ChatGPT-4 (Links ein offener Prompt zur Analyse eines algebraischen Fehlers; Rechts ein geschlossener Prompt zur Analyse einer Femi-Aufgabe der Grundschule anhand der KMK-Bildungsstandards des Primarbereich)

ChatGPT differenziert in der Ausgabe (Abb. 3, rechts) prozessbezogene Kompetenzen anhand des vorgegebenen PDFs. Die Lösung des Kindes scheint jedoch bei jeder Beurteilung hinsichtlich einer Kompetenz neu interpretiert zu werden. Das KI-Modell berücksichtigt beispielsweise unter der Kompetenz „mathematisch argumentieren“ teilweise nur die erste Fragestellung der Aufgabe. Dies wird daran

deutlich, dass ChatGPT die verschiedenen Terme als den Versuch zur Bestimmung der Gesamtzahl der Lernenden in der Schule interpretiert („Sie haben eine Abnahme-Sequenz aufgestellt, um die Gesamtzahl der Schüler zu berechnen“). Andererseits wird unter der Beurteilung der Kompetenz „mathematisch modellieren“ von ChatGPT die Abnahme-Sequenz in Bezug zur Pyramide gesetzt, was auf ei-

ne teilweise korrekte Interpretation hindeutet („Der Schüler oder die Schülerin hat das reale Phänomen der menschlichen Pyramide in ein mathematisches Modell übersetzt, indem eine Struktur für die Abnahme der Schülerzahl pro Reihe erstellt wurde“). Unter der prozessbezogenen Kompetenz „mathematisch darstellen“ deutet ChatGPT die Lösung sogar als eine tabellarische Darstellung und zeigt damit, dass die Bildinterpretation auch die Anordnung der Terme berücksichtigt. GPT-4 scheint somit in der Lage zu sein, Terme zu identifizieren, diese von der gesamten Anordnung getrennt zu interpretieren und in Beziehung zu setzen. Dennoch ist die Interpretationsleistung im Kontext der Aufgabe aus unserer Perspektive noch ausbaufähig. Es bleibt abzuwarten, ob diese Kombination von PDF und Interpretation durch den automatisierten Internetzugriff in Zukunft noch verbessert werden kann, und die Ausgabe als ein erster Interpretationsvorschlag dienen kann, den Lehrkräfte durch ihre Expertise anreichern und vervollständigen bzw. verbessern.

Insgesamt zeigen unsere Versuche in der Bildinterpretation, dass ChatGPT durchaus in der Lage ist, Schülerfehler zu erkennen, wenn die zugrunde liegenden Aufgaben nicht allzu komplex sind. Die Fehleranalysen und auch die Kompetenzanalysen funktionieren allerdings noch nicht immer zuverlässig und bedürfen der individuellen Nachprüfung. Wir sehen trotzdem das Potenzial für die Nutzung im Bereich der Diagnose von Schülerleistungen, etwa, wenn Lehrkräfte in Bezug auf klassische Fehlermuster in bestimmten Bereichen (z. B. Termumformungen, Modellierungs- oder arithmetische Basiskompetenzen) Unterstützung suchen. Inwiefern diese Bildinterpretationen für diagnostische Zwecke auch verwendbar sind, und ob Lehrkräfte darauf in ihrer Praxis zurückgreifen würden, ist noch weitestgehend unerforscht. Die Möglichkeiten dazu sind jedoch offensichtlich vorhanden.

Fazit

Bildgenerierung und Bildinterpretation in ChatGPT sind mit den aktuellen Entwicklungen möglich geworden, wenn auch zum Teil noch nicht hinreichend auf die Bedürfnisse von Mathematiklehrkräften angepasst. Die Richtung der KI-Entwicklung hin zu einem stärker professionsorientierten Einsatz zeichnet sich für uns jedoch ab und es scheint nur noch eine Frage der Zeit zu sein, bis die bildgenerierenden Modelle und Bildinterpretationen die benötigte Qualität für den Anwendungsbereich Schule erreichen. Die personalisierte Bildgenerierung könnte jedenfalls bereits jetzt zu einer deutlichen Zunahme der Verwendung hochauflösender

Abbildungen im Unterricht führen. Fortwährend bleiben aber die Probleme des Umgangs mit dem Datenschutz bestehen, außerdem ist nicht klar, ob die bildgenerierenden und -interpretierenden Algorithmen verstecktem Bias unterliegen und somit Schülergruppen systematisch bevor- oder benachteiligen könnten, insofern zukünftig diagnostische Entscheidungen an die KI übertragen werden. Es ergeben sich weiterhin auch viele Fragen, deren Beantwortung momentan noch aussteht: Sind qualitativ angemessene und mathematikdidaktisch fundierte Visualisierungen mit Diffusionsmodellen überhaupt herstellbar? Bedarf es spezieller Prompt-Techniken, um die Qualität der Ausgaben zu verbessern? Inwiefern nutzen Lehrkräfte im Unterrichtsallday schon KI-Modelle zur Bildgenerierung und -interpretation und welche unterschiedlichen Arten der Nutzung können beobachtet werden? Ist die Interpretationsleistung der KI-Modelle domänenspezifisch (z. B. in Bezug auf mathematische Inhaltsbereiche wie Algebra, Stochastik oder Geometrie) und wie kann die Interpretationsleistung gesteigert werden? Die Anwendungsmöglichkeiten aber auch die Grenzen der Technologie in der mathematikdidaktischen Forschung bleiben damit ein wichtiges neues Forschungsfeld. Um damit auf den Titel unseres Beitrags zurückzukommen: wir können festhalten, dass ChatGPT mit den neuen Entwicklungen und Funktionen das „Sehen“ und „Zeichnen“ nun gelernt hat, aber aufgrund der eigenen Halluzinationen bleibt die Faktentreue auch in der Bildgenerierung und -interpretation als Problem bis auf Weiteres bestehen.

Literatur

- Betker, J., Goh, G., Jing, L., Brooks, T., Wang, J., Li, L., Zhuang, J., Lee, J., Guo, Y., Manassra, W., Dhariwal, P., Chu, C., Jiao, Y. & Ramesh, A. (2023). *Improving Image Generation with Better Captions*. cdn.openai.com/papers/dall-e-3.pdf
- Buchholtz, N., Baumanns, L., Huget, J., Peters, F., Schorcht, S. & Pohl, M. (2023). Herausforderungen und Entwicklungsmöglichkeiten für die Mathematikdidaktik durch generative KI-Sprachmodelle. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 114, 19–26.
- Brüggemann, J. (2023). *Kopfübungen automatisch erstellen mit dem Mathemia Generator für Kopfübungen*. mathemia.de/blog/2023-11-13-gpt-kopfuebungs-generator/
- Chan, C. K. Y. & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. DOI:10.1186/s41239-023-00411-8
- von Garrel, J., Mayer, J. & Mühlfeld, M. (2023). *Künstliche Intelligenz im Studium: Eine quantitative Befragung von Studierenden zur Nutzung von ChatGPT*

- & Co. opus4.kobv.de/opus4-h-da/frontdoor/index/index/docId/395
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). *Generative Adversarial Networks*. DOI:10.48550/arXiv.1406.2661
- Heaven, W. D. (2023). *Wie KI-Bild- und Textgeneratoren die Kreativ-Branche umkrempeln*. Heise Medien. www.heise.de/hintergrund/Wie-KI-Bild-und-Textgeneratoren-die-Kreativ-Branche-umkrempeln-7489270.html
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J. & Kasneci, G. (2023). *ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education*. DOI:10.35542/osf.io/5er8f
- Linde, H. (2023). *So funktionieren KI-Bildgeneratoren*. www.golem.de/news/kuenstliche-intelligenz-so-funktionieren-ki-bildgeneratoren-2305-174436.html
- Lutz-Westphal, B. (2023). ChatGPT und der „Faktor Mensch“ im schulischen Mathematikunterricht. *Mitteilungen der Deutschen Mathematiker-Vereinigung*, 31(1), 19–21.
- Preiß, J., Bartels, M., Niemann-Lenz, J., Pawlowski, J. & Schnapp, K.-U. (2023). *ChatGPT and Me. Erste Ergebnisse der quantitativen Auswertung einer Umfrage über die Lebensrealität mit generativer KI an der Universität Hamburg*. Digital and Data Literacy in Teaching Lab. Universität Hamburg. 26. 9. 2023.
- OpenAI (2023). *Terms of Use: Updated March 14, 2023*. openai.com/policies/terms-of-use
- Rassin, R., Ravfogel, S. & Goldberg, Y. (2022). *DALLE-2 is Seeing Double: Flaws in Word-to-Concept Mapping in Text2Image Models*. DOI:10.48550/arXiv.2210.10606
- Schleiss, J., Mah, D.-K., Böhme, K., Fischer, D., Meisenholler, J., Paaßen, B., Schork, S. & Schrupf, J. (2023). *Künstliche Intelligenz in der Bildung. Drei Zukunftsszenarien und fünf Handlungsfelder*. KI-Campus. DOI:10.5281/zenodo.7702620
- Schorcht, S., Baumanns, L., Buchholtz, N., Huget, J., Peters, F. & Pohl, M. (2023). Ask Smart to Get Smart: Mathematische Ausgaben generativer KI-Sprachmodelle verbessern durch gezieltes Prompt Engineering. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 115, 12–24.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2022). *Bildungsstandards im Fach Mathematik für den Primarbereich*.
- Smolansky, A., Cram, A., Radulescu, C., Zeivots, S., Huber, E. & Kizilcec, R. F. (2023). Educator and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education. In *Proceedings of the Tenth ACM Conference on Learning @ Scale (L@S '23)*. Association for Computing Machinery, New York, USA, (pp. 378–382). DOI:10.1145/3573051.3596191
- Spannagel, C. (2023). Hat ChatGPT eine Zukunft in der Mathematik? *Mitteilungen der Deutschen Mathematiker-Vereinigung*, 31(3), 168–172. DOI:10.1515/dmvm-2023-0055
- Wolfram, S. (2023). *Instant Plugins for ChatGPT: Introducing the Wolfram ChatGPT Plugin Kit: Stephen Wolfram Writings*. writings.stephenwolfram.com/2023/04/instant-plugins-for-chatgpt-introducing-the-wolfram-chatgpt-plugin-kit/
- Sebastian Schorcht, Technische Universität Dresden
E-Mail: sebastian.schorcht@tu-dresden.de
- Lukas Baumanns, Technische Universität Dortmund
E-Mail: lukas.baumanns@tu-dortmund.de
- Nils Buchholtz, Universität Hamburg
E-Mail: nils.buchholtz@uni-hamburg.de
- Judith Huget, Universität Hamburg
E-Mail: judith.huget@uni-hamburg.de
- Franziska Peters, Technische Universität Dresden
E-Mail: franziska.peters@tu-dresden.de
- Maximilian Pohl, Universität Duisburg-Essen
E-Mail: maximilian.pohl@uni-due.de