

Wie nutzen mathematisch interessierte Kinder KI beim Problemlösen?

Eine explorative Untersuchung

JULIAN KRIEGEL, DRESDEN & LUKAS BAUMANN, ESSEN

Zusammenfassung: Generative Künstliche Intelligenz (KI) wird zunehmend als Werkzeug mit Potenzial für das Mathematiklernen diskutiert. Insbesondere beim mathematischen Problemlösen, bei dem Schüler:innen auf Barrieren und inhaltliche Hürden stoßen, könnten KI-Chatbots Unterstützung bieten. Die konkreten Potenziale dieser Technologien sind bislang jedoch kaum empirisch untersucht. Die vorliegende Studie gewährt daher explorative Einblicke in die Nutzung generativer KI beim mathematischen Problemlösen. Dazu wurden 27 mathematisch interessierte Schüler:innen der Klassenstufen drei bis fünf in Paaren bei der freiwilligen Nutzung eines KI-Chatbots während des Problemlösens videografiert. Qualitative Analysen ihres Nutzungsverhaltens zeigen, dass der KI-Chatbot unterschiedliche Funktionen im Problemlöseprozess einnehmen kann. Jene KI-Nutzungsweisen sind eng mit bestimmten Phasen des Problemlöseprozesses verknüpft und wirken sich auf den Heuristikeinsatz aus.

Abstract: Generative artificial intelligence (AI) is increasingly being discussed as a tool with potential for supporting mathematics learning. In particular, AI-chatbots may offer valuable assistance when students encounter barriers and conceptual difficulties during mathematical problem solving. However, the specific potential of these technologies has scarcely been investigated empirically. This exploratory study provides insights into the use of generative AI in the context of mathematical problem solving. To do this, 27 mathematically interested students from grades three to five were video-recorded in pairs while voluntarily using an AI chatbot during problem-solving. Qualitative analyses of their usage behavior reveal that the chatbot can serve a variety of functions throughout the problem-solving process. These forms of AI use appear to be closely linked to specific phases of problem solving and influence students' use of heuristic strategies.

1. Einleitung

Generative Künstliche Intelligenzen (KI) gewinnen durch ihre breite Verfügbarkeit zunehmend an Bedeutung im schulischen Lehren und Lernen (SWK, 2024; KMK, 2024) und spielen somit auch für das Lehren und Lernen von Mathematik eine immer wichtigere Rolle. Innerhalb der Mathematikdidaktik

werden Herausforderungen und Implikationen generativer KI seit deren Verbreitung aus verschiedenen Perspektiven beleuchtet (Buchholtz et al., 2023; 2024; Spannagel, 2023; Wardat et al., 2023). Dabei werden verschiedene Schwerpunkte und Einsatzgebiete ausgelotet. Unterscheiden kann man zwischen der Nutzung durch Lehrkräfte und Dozierende (z. B. für Unterrichts- bzw. Veranstaltungsplanung; Buchholtz & Huget, 2024; Cameron & Mesiti, 2024; Helfrich-Schkarbanenko, 2023; Schorcht et al., 2024c) und der Verwendung durch Schüler:innen beim Mathematiklernen (Jauhiainen & Guerra, 2023). Gerade jener KI-Einsatz durch Schüler:innen ist bislang weniger intensiv beleuchtet.

Angesichts der bereits gut untersuchten Nutzungsweisen und Potenziale anderer digitaler Technologien im Mathematikunterricht erscheint es folgerichtig, auch für KI-Chatbots empirische Erkenntnisse zu gewinnen. So liegen Analysen zur Gestaltung und zum Einsatz von Erklärvideos (z. B. Kulgemeyer, 2020; Korntreff & Prediger, 2021), zum Taschenrechnereinsatz im Hinblick auf Kopfrechenfähigkeiten (z. B. Meißner, 2006; Spiegel, 1988) sowie zur Nutzung von Tablet-Apps und virtuellen Manipulativen (z. B. Tucker et al., 2017) vor. Auch die Kultusministerkonferenz (KMK, 2024) schlägt in ihrer Handlungsempfehlung zum Umgang mit KI in Bildungsprozessen eine verstärkte Auseinandersetzung mit dem Einsatz von KI im Primarbereich vor.

Auszuhandeln ist, in welchen Bereichen der Einsatz generativer KI im Mathematikunterricht sinnvoll zur Förderung des Lernens genutzt werden kann. Anknüpfend an die Bildungsstandards (KMK, 2022) sollten fachdidaktische Überlegungen zur zielführenden Nutzung von generativer KI vertieft werden. Eine Kompetenz, die im Mathematikunterricht gleichsam bedeutsam (Rott et al., 2023) wie unterrepräsentiert (Herold-Blasius et al., 2019) ist, ist das mathematische Problemlösen. Problemlösen stellt Schüler:innen vor besondere Herausforderungen, z. B. beim Aufgabenverständnis, der Auswahl und Umsetzung von Problemlösestrategien sowie durch fehlendes Vorwissen und mangelnde Rechenfähigkeit (Schoenfeld, 1985; Rott et al., 2023). Aktuelle Ergebnisse der TIMSS-Studie 2023 zeigen zudem, dass sich die Leistungen von Schüler:innen im Bereich des Problemlösens im Vergleich zu 2019 signifikant verschlechtert

haben (Schwippert et al., 2024), was den Handlungsbedarf in diesem Bereich unterstreicht. Aber auch (angehende) Lehrkräfte stehen vor Schwierigkeiten beim Unterrichten des Problemlösens (Herold-Blasius et al., 2019). Eine Schwierigkeit besteht darin, dass es sich beim Problemlösen um einen individuellen Prozess handelt, bei dem unterschiedliche Strategien der Schüler:innen zur Lösung beitragen können. Genau in solchen Lernprozessen kann der Einsatz generativer KI eine unterstützende Rolle spielen, etwa durch adaptive Rückmeldungen und die Berücksichtigung individueller Lernverläufe. Chiu et al. (2023) heben hervor, dass KI insbesondere durch die Förderung von Adaptivität und die Ausrichtung auf individuelle Kompetenzen Potenziale für das Lernen bietet.

Obwohl erste konzeptionelle Überlegungen und Einzelstudien zur Nutzung generativer KI im Bildungskontext vorliegen, gibt es bislang nur sehr wenige empirische Erkenntnisse darüber, wie Schüler:innen – insbesondere im Primarbereich – KI-Chatbots im Kontext des mathematischen Problemlösens nutzen. Diese Erkenntnisse sind jedoch notwendig, um produktive Nutzungsformen zu identifizieren und Unterstützungskonzepte zu entwickeln, in denen KI-Chatbots gezielt zum Erwerb von Problemlösefähigkeiten beitragen können.

In der vorliegenden Studie werden explorative Einblicke in die Nutzung von KI-Chatbots durch mathematisch interessierte Kinder beim Problemlösen gewährt. Dazu wird sich am Begriff der *Nutzungsweisen* nach Walter (2018, S. 130) orientiert, welcher darunter „Varianten der Verwendung eines (virtuellen) Arbeitsmittels [versteht], die im Rahmen spezifischer Arbeitsprozesse zur Lösung von Mathematikaufgaben von den Kindern verfolgt werden.“ Die vorliegende Studie adressiert daher die Zwecke des Einsatzes eines KI-Chatbots durch Lernende. Darüber hinaus werden folgende Aspekte untersucht: Die Zeitpunkte innerhalb des Problemlöseprozesses, zu denen Kinder einen KI-Chatbot verwenden, sowie der Zusammenhang zwischen jenen KI-Nutzungsweisen und dem Heuristikeinsatz. Ziel ist es, ein tieferes Verständnis für die Rolle von KI-Chatbots beim mathematischen Problemlösen zu entwickeln.

2. Theoretischer Hintergrund

2.1 Generative KI für das Mathematiklernen

Künstliche Intelligenz nimmt für viele Personen seit der Veröffentlichung von ChatGPT einen immer größeren Teil des Alltags ein und spielt auch im schulischen Lernen eine zunehmend wichtige Rolle. KI-

Technologien können in verschiedenen Bereichen eingesetzt werden, um das Lernen zu unterstützen. Dabei lassen sich unterschiedliche Formen von KI unterscheiden. Das Maschinelle Lernen nutzt Algorithmen, um aus großen Datenmengen Muster zu erlernen und diese auf neue, unbekannte Daten anzuwenden (Buchholtz et al., 2024). Im Rahmen von Learning Analytics werden zum Beispiel Klassifikations- und Clusterverfahren verwendet, um das Verhalten von Lernenden in digitalen Lernumgebungen zu erfassen und mittels maschineller Lernverfahren auszuwerten (Aldowah et al., 2019; Hershkovitz et al., 2024). Diese Analysen ermöglichen, individuelle Schwierigkeiten von Lernenden zu erkennen.

Eine andere Form des Maschinellen Lernens, die in diesem Artikel im Fokus steht, sind sogenannte große Sprachmodelle (Large Language Models, LLM), die in der Lage sind, lange Abhängigkeiten in Texten zu verarbeiten und aufgrund stochastischer Berechnungen kontextuell relevante Antworten zu generieren. Zu diesen Modellen gehören auch Generative Pre-trained Transformer (GPT), auf denen unter anderem ChatGPT basiert (OpenAI, 2018). Diese Modelle verstehen und generieren Output in Form von Text, Sprache oder Bildern (Schorcht et al., 2023; 2024a) und sind insbesondere darauf trainiert, menschenähnlich zu antworten.

Wie Spannagel (2023) hervorhebt, lässt sich der dialogische Austausch zwischen Lernenden und einem KI-System als ein Beispiel für verteilte Kognition (Salomon, 1993) verstehen. Diese Perspektive betrachtet kognitive Prozesse nicht ausschließlich als Eigenschaft einzelner Personen, sondern als in soziale und materielle Konstellationen ausgelagert und zwischen ihnen koordiniert. Übertragen auf die vorliegende Studie bedeutet dies, dass Schüler:innen beim Problemlösen Teile ihrer Denkprozesse – etwa das Formulieren, Überprüfen oder Präzisieren von Lösungsideen – an einen KI-Chatbot auslagern und dessen Rückmeldungen wiederum in eigene Überlegungen integrieren. Der KI-Chatbot fungiert als kognitives Werkzeug bzw. externer Partner im Problemlöseprozess, ohne ihm dabei eigenständige Intentionalität oder Verständnis zuzuschreiben.

Der Einsatz großer Sprachmodelle im Bildungskontext wird derzeit intensiv und kontrovers diskutiert (Buchholtz et al., 2023; 2024; Trust et al., 2023). Didaktische Fragestellungen werden dabei mit ethischen und rechtlichen Aspekten verbunden. Platz und Plote (2025) stellen hierzu wesentliche Kritikpunkte in Bezug auf KI dar, die von mangelnder Zugänglichkeit für alle Menschen bis hin zu Auswir-

kungen auf Arbeit, Kreativität und Umwelt reichen. Gerade die erschwerte Erklärbarkeit von KI-Technologien sowie die Generierung falscher bzw. verzerrter Ausgaben erscheinen beim Einsatz von LLM im Mathematikunterricht relevant. Sie führen zu einer Auseinandersetzung mit notwendigen „Schlüsselkompetenzen“ (Platz & Plote, 2025) von Schüler:innen und Lehrkräften im verantwortungsvollen und reflektierten Umgang mit KI-Technologien. Entsprechende Kompetenzrahmen liegen bereits vor (Miao & Cukurova, 2024; Miao et al., 2024).

Neben diesen allgemeinen Kritikpunkten wird insbesondere der Einsatz von LLM in der Primarstufe kontrovers diskutiert. Einerseits befürwortet die Ständige Wissenschaftliche Kommission (SWK) einen weitgehenden Verzicht von textgenerierenden LLM in der Grundschule und Sekundarstufe I mit Verweis auf die notwendige Stärkung basaler Lese- und Schreibkompetenzen der Schüler:innen (SWK, 2024). Andererseits hält die Konferenz der Kultusminister (KMK) den Einsatz von LLM in diesen Schulstufen für erörterungsbedürftig, indem etwa vielfältige KI-Anwendungen zur Förderung eben jener Basiskompetenzen entwickelt und eingesetzt werden könnten (KMK, 2024). Unklar ist, inwiefern generative KI fachspezifische Kompetenzen von Schüler:innen in der Grundschule und frühen Sekundarstufe fördern kann. Wir wollen daher in dieser Studie mögliche Potenziale von LLM in jenen Schulstufen forschungsbasiert ausloten, ohne mögliche Herausforderungen außer Acht zu lassen. Dafür fokussieren wir das Lösen mathematischer Probleme.

Hinsichtlich des Einsatzes von LLM beim mathematischen Problemlösen gibt es bislang wenig Forschung. Bekannt ist, dass ChatGPT aufgrund von Halluzinationen mathematische Probleme nicht zuverlässig löst (Frieder et al., 2024; Plevris et al., 2023; Schorcht et al., 2024b). Insbesondere scheinen verschiedene Prompt-Techniken, also die spezifischen Eingaben in ChatGPT, die Qualität des Outputs zu beeinflussen. So erhöhen Prompts, die explizit ein schrittweises Vorgehen des LLM evozieren sollen, die Sichtbarkeit eines strategischen Vorgehens bei LLM (Schorcht et al., 2024b). Hinsichtlich der Interaktion von Schüler:innen mit LLM beim Problemlösen gibt es jedoch – nach aktuellem Stand – keine systematische Forschung. Hingegen liegen erste Untersuchungen vor, welche die Nutzung von LLM durch Lehramtsstudierende etwa beim Problem Posing (Baumanns & Pohl, 2024) und beim mathematischen Beweisen (Dilling & Herrmann, 2024) untersuchen.

2.2 Mathematisches Problemlösen

Ein zentrales Ziel des Mathematikunterrichts über alle Schulstufen hinweg ist, dass Schüler:innen Probleme mathematisch lösen (KMK, 2022). Probleme zeichnen sich im Mathematikunterricht in Abgrenzung zu Routineaufgaben insbesondere dadurch aus, dass im Moment der Bearbeitung kein passendes Verfahren zur Lösung vorliegt (Büchter & Leuders, 2011; Schoenfeld, 1985). Lösende sind mit einer Barriere konfrontiert, wobei die Wahrnehmung als Barriere von den Fähigkeiten des Lösenden abhängt (Vollrath, 1992). Während die Addition von Brüchen für Schüler:innen zu Beginn der Sekundarstufe I noch ein Problem darstellen kann, ist sie für Schüler:innen der Sekundarstufe II eine Routineaufgabe. Ein Problemlöseprozess liegt nach Rott (2013) entsprechend vor, wenn kein „Verfahren zur Lösung der gestellten Aufgaben bekannt ist und angewendet wird“ oder „das Verfahren während der Bearbeitung verworfen [wird]“ (S. 36). Diese übergreifenden Charakterisierungen mathematischer Probleme gelten grundsätzlich schulstufenunabhängig. Für die Primarstufe wurden etwa bei Rasch (2001) schulstufengemäße Kontexte und Zugänge entwickelt, um spezifische Lösungsstrategien von Grundschulkindern beim Bearbeiten mathematischer Probleme zu untersuchen.

Etablierte Einflussfaktoren erfolgreichen Problemlösens finden sich bei Schoenfeld (1985):

- 1) *Resources*: Verfügbares mathematisches Fakten- (deklarativ) und Fertigkeitswissen (prozedural) einer Person zur Lösung von mathematischen Problemen,
- 2) *Heuristics*: Verfügbare Problemlösestrategien einer Person zur Überwindung von Barrieren bei der Bearbeitung mathematischer Probleme,
- 3) *Control*: Selbstregulierung beim Problemlösen,
- 4) *Belief systems*: Einstellungen und Perspektiven auf Mathematik und mathematische Aufgaben.

Diese Wissens- und Verhaltenskategorien rahmen die Förderung und Beforschung von Problemlösefähigkeiten. Besonders die Ausbildung heuristischer Fähigkeiten wurde intensiv untersucht (Collet, 2009; Herold-Blasius, 2021; Rott & Gawlik, 2014). In dieser Studie richten wir den Blick auf den Einsatz von Heuristiken während des Problemlösens unter Nutzung eines KI-Chatbots. Nachfolgend werden Problemlösemodelle sowie Erkenntnisse zum Heurismeneinsatz vorgestellt, welche als theoretische Grundlage der Untersuchungen dienen.

2.2.1 Phasen des Problemlösens

Die Grundlage vieler Phasenmodelle für das Problemlösen bildet die wegweisende Arbeit von Pólya (1949). Nach Pólya durchläuft der Prozess des Problemlösens vier Phasen: *Verstehen der Aufgabe*, *Ausdenken eines Plans*, *Ausführen des Plans* und *Rückschau*. Darauf implizit aufbauend hat Schoenfeld (1985) Pólyas Modell weiterentwickelt, indem er die Phasen präzierte und den Fokus auf kognitive und metakognitive Strategien legte. In Schoenfelds Modell durchläuft der Problemlöseprozess die Phasen *Analysis*, *Design*, *Exploration*, *Implementation* und *Verification*. Diese Modelle sind normativ zu verstehen, d. h. sie bilden einen idealtypischen Problemlöseprozess ab und dienen Lehrkräften zu didaktischen Zwecken, um Lernende im Problemlösen zu unterstützen und Fähigkeiten aufzubauen.

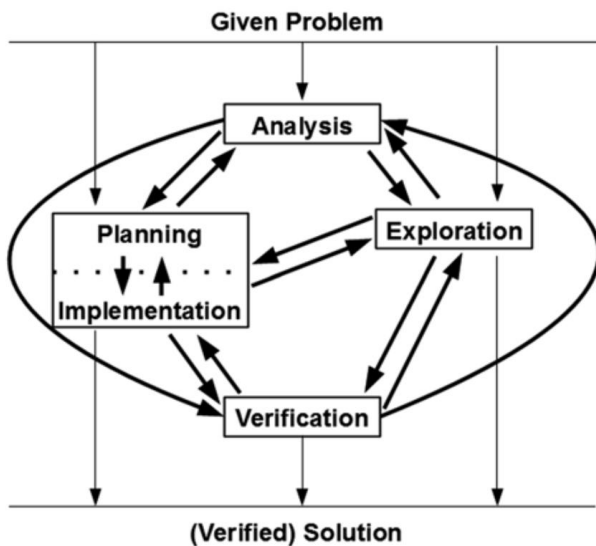


Abb. 1: Deskriptives Phasenmodell von Problembearbeitungsprozessen nach Rott (2014, S. 273)

Deskriptive Modelle zielen darauf ab, tatsächliche Problemlöseprozesse detailliert und realitätsnah abzubilden (Rott et al., 2021; Yimer & Ellerton, 2010). Sie berücksichtigen die oft unregelmäßigen und komplexen Verläufe dieser Prozesse und versuchen, die Dynamik zwischen verschiedenen Phasen und Strategien empirisch zu erfassen. Diese Studie baut auf dem deskriptiven Phasenmodell von Rott (2014; vgl. Abb. 1) auf:

- **Analysis:** Verstehen der Aufgabenstellung; Einnehmen bestimmter Perspektiven, Umformulieren des Problems,
- **Exploration:** Erkundung der Aufgabe und des Problemraums; Anwendung von Heuristiken,
- **Planning:** Explizites Formulieren eines Plans zur Bearbeitung der Aufgabe,
- **Implementation:** Ausführen des Plans; Kontrolle einzelner Schritte,

- **Verification:** Überprüfen der Lösung; Rückschau halten.

Dieses Phasenmodell bietet Einblicke in die Problemlöseprozesse von Schüler:innen, insbesondere durch die Möglichkeit, deren individuelle Merkmale zu vergleichen und zu charakterisieren. In einer Studie mit Lehramtsstudierenden des Fachs Mathematik arbeiteten Rott et al. (2021) heraus, dass bei Routineaufgaben meist lineare Prozesse ohne Exploration beobachtet werden können, während Problemlöseprozesse einen zyklischen Ablauf zeigten. Das Modell ermöglicht außerdem die Identifikation eines Problemlöseverhaltens, das von Schoenfeld (1992) als „wild goose chases“ (vergebliches Umherirren) beschrieben wurde. Rott et al. (2021) zeigten, dass dieses Verhalten statistisch mit erfolglosen Lösungsversuchen in Zusammenhang steht, was einen wichtigen Hinweis auf typische Herausforderungen im Problemlöseprozess liefert.

Aufbauend auf diesen Erkenntnissen wird in der vorliegenden Studie dieses Modell genutzt, um die KI-gestützten Problemlöseprozesse mathematisch interessierter Kinder zu untersuchen. Bislang fehlen Erkenntnisse dazu, in welchen Phasen des Problemlösens KI-Chatbots genutzt werden.

2.2.2 Heuristiken beim Problemlösen

Heuristiken sind „Denkoperationen“, die beim Lösen von Aufgaben „typischerweise von Nutzen sind“ (Pólya, 1949, S. 155). Sie stellen Handlungsmuster dar, die Schüler:innen dabei unterstützen, Barrieren beim Problemlösen zu überwinden (Bruder & Collet, 2011; Rott, 2014). Im Gegensatz zu Algorithmen zeichnen sich Heuristiken durch ihre Flexibilität und Übertragbarkeit in verschiedenen mathematischen Kontexten und Problembereichen aus (Kaitera & Harmoinen, 2022). Deswegen ist ein Ziel des Mathematikunterrichts, „in der Auseinandersetzung mit Aufgaben Problemlösefähigkeiten, die über die Mathematik hinaus gehen, (heuristische Fähigkeiten) zu erwerben“ (Winter, 1996, S. 35).

Bruder und Collet (2011) unterscheiden zwischen *heuristischen Hilfsmitteln*, *heuristischen Strategien* und *heuristischen Prinzipien*. Heuristische Hilfsmittel umfassen beispielsweise das Skizzieren einer Tabelle oder das Zeichnen einer informativen Figur, was oft mit einem Wechsel der Repräsentation einhergeht. Heuristische Strategien sind Methoden zum Lösen mathematischer Probleme wie das Vorwärts- und Rückwärtsarbeiten, das Variieren der Aufgabenstellung oder das systematische Probieren (Stender, 2021). Heuristische Prinzipien wiederum, wie das

Analogieprinzip oder das Invarianzprinzip, bieten eine übergeordnete Orientierungshilfe. Schoenfeld (1982) stellt heraus, dass der Einsatz jener Heuristiken personenspezifisch, d. h. abhängig vom Vorwissen und den Vorerfahrungen ist.

Die Forschung zeigt, dass der Einsatz von Heuristiken zum einen trainiert werden kann (Sturm, 2018) und dass er dabei hilft, Problemlösebarrieren zu überwinden (Bruder & Collet, 2011; Rott, 2018; Schoenfeld, 1985). Zudem wirkt sich die Nutzung von Heuristiken positiv auf den Problemlöseerfolg aus (Herold-Blasius, 2021; Kaitera & Harmoinen, 2022; Sturm, 2018). Allerdings bietet der Heuristikeinsatz, anders als die korrekte Anwendung von Algorithmen, keine Garantie für die Lösung eines Problems (Bruder & Collet, 2011). Interessanterweise zeigen Studien, dass mathematisch besonders fähige Schüler:innen weniger Heuristiken verwenden als ihre Altersgenossen (Rott, 2018). Dieses Phänomen erklärt Rott (2018) mit der höheren geistigen Beweglichkeit mathematisch besonders fähiger Kinder. Des Weiteren zeigte Fuchs (2006) in ihrer Untersuchung, dass mathematisch potenziell begabte Kinder unterschiedliche Problemlösestile verwenden, welche sich teilweise über einen längeren Zeitraum verfestigen. Am häufigsten tritt dabei der Problemlösestil „*Abwechselndes Überlegen und Probieren – Suchen nach Lösungsmustern*“ (Fuchs, 2006, S. 279) auf. Indem manche Schüler:innen intuitiv und systemhaft vorgehen oder hartnäckig probieren, legen die Ergebnisse zudem nahe, dass der Heuristikeinsatz mathematisch besonders fähiger Kinder differieren kann.

2.3 Enrichment-Programme für mathematisch interessierte Kinder

Mathematische Begabung ist ein domänenspezifisches und multifaktorielles Konstrukt, das sich in unterschiedlichen Fähigkeiten und Eigenschaften äußert (Käpnick, 1998). Die Fähigkeit, mathematische Probleme zu lösen, wird häufig als zentraler Indikator herangezogen (Käpnick, 1998; Kießwetter, 1985; Krutetskii, 1976). Mathematisch begabte Kinder zeichnen sich u. a. durch Merkmale wie die Fähigkeit aus, mathematische Sachverhalte durch Strukturerkennung zu memorieren, Strukturen zu konstruieren und zu nutzen, zwischen Repräsentationsmodi zu wechseln, Gedankengänge umzukehren sowie komplexe Strukturen zu erfassen und mit ihnen zu arbeiten (Aßmus, 2018; Käpnick, 1998). Ergänzend dazu spielen begabungsstützende allgemeine Persönlichkeitseigenschaften wie intellektuelle Neugier, Freude am Problemlösen und Beharrlichkeit eine

wichtige Rolle (Käpnick, 1998). Zudem führt Heinze (2005) das „Bedürfnis nach plausiblen mathematischen Erklärungen und das Streben nach Erkenntnissen“ (S. 39) als weiteres Begabungsmerkmal an, das sich im Problemlösen prinzipiell auch durch die Nutzung von KI-Chatbots adressieren lässt, sofern Prompts gezielt auf Erklärungen ausgerichtet sind. So können KI-Chatbots als kognitive Werkzeuge fungieren, die mathematisch begabte Kinder dabei unterstützen, tiefergehende Einsichten zu gewinnen.

Das Bearbeiten von Problemfeldern steht entsprechend im Zentrum von Förderkonzepten mathematisch begabter Kinder (Bardy & Bardy, 2020; Fritzlar, 2008). Beispiele hierfür sind die Enrichment-Programme „Mathe für kleine Asse“ (Universität Münster; Käpnick et al., 2021), „PriMa“ (Universität Hamburg; Nolte & Pamperien, 2024) sowie „Mathe für Cracks“ (TU Dresden), auf welches in der vorliegenden Studie zurückgegriffen wird (siehe 4.3). „Mathe für Cracks“ wird in Zusammenarbeit mit der Deutschen Gesellschaft für das hochbegabte Kind (DGhK) durchgeführt (Schorcht & Peters, 2020) und richtet sich als extracurriculares Programm an mathematisch interessierte Kinder aus den Klassenstufen 3–5, die von ihren Mathematiklehrkräften als besonders fähig eingeschätzt wurden. Ein standardisierter Intelligenztest wird nicht durchgeführt, weil aus einem hohen IQ nicht automatisch eine mathematische Begabung folgt (z. B. Nolte & Pamperien, 2024). Spezielle Tests zur Identifikation mathematischer Begabung werden ebenfalls nicht eingesetzt, jedoch weisen die gezielte Auswahl und freiwillige Teilnahme darauf hin, dass die Teilnehmenden tendenziell mathematisch befähigter sind als ihre Peers. Aufgrund fehlender Testungen sprechen wir im Folgenden von *mathematisch interessierten Kindern* (zum Mathematikinteresse siehe Stefan, 2012).

3. Forschungsfragen

Die vorliegende explorative Studie untersucht, *wie mathematisch interessierte Kinder einen KI-Chatbot beim Problemlösen nutzen*. Folgende Forschungsfragen lassen sich ableiten und werden fokussiert:

- 1) Welche Nutzungsweisen eines KI-Chatbots lassen sich bei mathematisch interessierten Kindern im Setting des Problemlösens beobachten?
- 2) Zu welchen Zeitpunkten innerhalb des Problemlöseprozesses nutzen mathematisch interessierte Kinder KI-Chatbots?
- 3) In welchem Zusammenhang stehen die Nutzungsweisen des KI-Chatbots und der Heuristikeinsatz?

4. Methodik der Studie

4.1 Problemlöseaufgaben

Um die obigen Forschungsfragen zu beantworten, wurde eine qualitative Studie durchgeführt. Es wurden drei Problemlöseaufgaben ausgewählt, wobei sich an den folgenden Kriterien von Herold-Blasius et al. (2017) orientiert wurde:

- *Vorwissen*: Die notwendigen Vorkenntnisse der Schüler:innen zur Bearbeitung und Lösung der mathematischen Probleme sollten möglichst gering sein.
- *Offenheit*: Die Problemlöseaufgaben sollten über genau eine richtige Lösung verfügen, welche jedoch auf mehreren unterschiedlichen Wegen erreicht werden kann.
- *Heurismen*: Zu den verschiedenen Lösungsweisen zählt insbesondere die Vielzahl an Heurismen, mit denen die mathematischen Probleme (potenziell) gelöst werden können.

Für die Orientierung an jenen Kriterien spricht die Ähnlichkeit des Untersuchungssettings und der Einsatz gleicher Problemlöseaufgaben (s. u.). Andere Autor:innen geben ähnliche Kriterien wie z. B. Problemhaltigkeit, Verständlichkeit und Offenheit der Vorgehensweisen an und ergänzen darüber hinaus zeitliche Aspekte und eine Vielfalt angesprochener mathematischer Inhaltsbereiche (Fuchs, 2006). Hinsichtlich der Inhaltsbereiche mussten wir Einschränkungen vornehmen. Aufgrund des Einsatzes generativer KI wurde als zusätzliches Kriterium berücksichtigt:

- *Lösbarkeit durch KI*: Die Aufgaben sollen durch den KI-Chatbot potenziell und textbasiert lösbar sein. Geometrische Problemlöseaufgaben werden daher ausgeschlossen, weil diese zum Zeitpunkt der Datenerhebung mithilfe des eingesetzten KI-Chats nicht ikonisch dargestellt werden können.

Aufgabe 1 – Sieben Tore

Ein Mann geht Äpfel pflücken. Um mit seiner Ernte in die Stadt zu kommen, muss er 7 Tore passieren. An jedem Tor steht ein Wächter und verlangt von ihm die Hälfte seiner Äpfel und einen Apfel mehr. Am Schluss bleibt dem Mann nur ein Apfel übrig. Wie viele hatte er am Anfang? (Bruder et al., 2005, S. 145)

Die Sieben-Tore-Aufgabe geht auf Leonardo de Pisa (1202) zurück und wurde bereits in vielfältigen Studien zur Untersuchung von Problemlöseprozessen herangezogen (Herold-Blasius et al., 2017; Herold-Blasius, 2021; Rott, 2013). Das *Rückwärtsarbeiten* ist eine mögliche heuristische Strategie zur Lösung dieses mathematischen Problems. Unter mehrmaliger Verwendung des Schemas „+1 und Verdoppeln“ kann ausgehend vom Endzustand (ein Apfel) schrittweise der Startzustand vor dem Passieren des ersten Tores (382 Äpfel) ermittelt werden.

Aufgabe 2 – Reihenfolgezahlen

Reihenfolgezahlen sind direkt aufeinanderfolgende Zahlen. Beispiele für Reihenfolgezahlen sind:

- 18, 19
- 5, 6, 7
- 31, 32, 33, 34

Findet alle möglichen Plusaufgaben mit Reihenfolgezahlen. Das Ergebnis darf nicht größer sein als 25. (in Anlehnung an Schwätzer & Selter, 1998)

Schwätzer & Selter (1998) führen an, dass es sich bei der Aufgabe „Reihenfolgezahlen“ um eine substantielle Aufgabenstellung handelt, da sie verschiedene Lösungsansätze und Schwierigkeitsstufen der Bearbeitung ermöglicht. Allen voran kann hier die Heuristik des *systematischen Probierens* zielführend sein. Die Autoren beobachteten hierzu eine Vielzahl an systematischen Vorgehensweisen von Schüler:innen, z. B. das Verlängern oder Verkürzen der Additionsaufgaben durch Hinzufügen oder Wegnehmen von Summanden. Durch Anwenden jener Strategien können sukzessive alle 27 Möglichkeiten bestimmt werden (vgl. Tab. 1).

Zwei Summanden	Drei Summanden	Vier Summanden	Fünf Summanden	Sechs Summanden
1+2	1+2+3	1+2+3+4	1+2+3+4+5	1+2+3+4+5+6
2+3	2+3+4	2+3+4+5	2+3+4+5+6	
3+4	3+4+5	3+4+5+6	3+4+5+6+7	
4+5	4+5+6	4+5+6+7		
5+6	5+6+7			
6+7	6+7+8			
7+8	7+8+9			
8+9				
9+10				
10+11				
11+12				
12+13				

Tab. 1: Lösungsmöglichkeiten der Aufgabe Reihenfolgezahlen (Schwätzer & Selter, 1998)

Aufgabe 3 – Kleingeld

Wie könnt ihr einen Geldbetrag von genau 31 Cent hinlegen, wenn ihr nur 10-Cent-, 5-Cent- und 2-Cent-Münzen zur Verfügung habt? Gebt alle Möglichkeiten an! (PISA-Konsortium Deutschland, 2006, S. 177)

Neben dem Anfertigen einer *Tabelle* bietet sich bei dieser Problemlöseaufgabe das *systematische Probieren* in Kombination mit dem *Zerlegungs- und Ergänzungsprinzip* als Lösungsstrategie an. Die 10-Cent-Münzen können schrittweise in 5-Cent- und 2-Cent-Münzen zerlegt und somit neue Möglichkeiten gefunden werden (Herold-Blasius et al., 2017; Rott, 2013). Darüber hinaus kann das *Extremalprinzip* genutzt werden, indem zunächst mit dem größten Münzwert begonnen und dann schrittweise „aufgefüllt“ wird (Bruder & Collet, 2011). Insgesamt ergeben sich somit sechs Möglichkeiten (vgl. Tab. 2).

Lösung	10-Cent	5-Cent	2-Cent
1	2x	1x	3x
2	1x	1x	8x
3	0x	1x	13x
4	1x	3x	3x
5	0x	3x	8x
6	0x	5x	3x

Tab. 2: Lösungsmöglichkeiten der Kleingeld-Aufgabe (Herold-Blasius et al., 2017, S. 105)

4.2 Generative KI

Eingesetzt wurde der DSGVO-konforme KI-Chatbot von fobizz, der den Lernenden optional zur Verfügung gestellt wurde (<https://tools.fobizz.com/>). Fobizz ermöglicht die Nutzung des KI-Chatbots, ohne dass die Schüler:innen einen persönlichen Account benötigen. Dadurch sowie durch die Nutzung des KI-Chatbots unter Aufsicht der Forschenden konnte vom Mindestalter einer eigenständigen KI-Nutzung von 13 Jahren (OpenAI, 2024) abgewichen werden.

Bei der Verwendung des KI-Chatbots von fobizz können verschiedene Sprachmodelle ausgewählt werden (GPT-3.5, GPT-4, GPT-4o, GPT-4o-mini, Claude 3, Claude Sonnet 3.5, Mistral AI und Luminous; Stand: November 2024). Während der Datenerhebung wurden die LLM GPT-4 Turbo und GPT-4o genutzt.

Im Vergleich zum Vorgängermodell GPT-3.5 wurde das Kontextfenster von GPT-4o auf 128.000 Tokens erhöht, wodurch auch längere, komplexere Prompts sprachlich erfasst und entsprechende Antworten generiert werden können (OpenAI, 2023). Neu ist außerdem, dass neben Text- auch Bild- und Spracheingaben möglich sind. In fobizz ist die Sprachfunktion allerdings nicht verfügbar. Jedoch können mithilfe

der Sprache-zu-Text-Funktion der Tastatur eingesprochene Prompts transkribiert und abgeschickt werden.

4.3 Stichprobe

Für die hier vorgestellte Untersuchung wurden im Enrichment-Kurs „Mathe für Cracks“ (Schorcht & Peters, 2020) im Sommersemester 2024 Daten von insgesamt 35 mathematisch interessierten Kindern aus dem Raum Dresden erhoben. Für die Erhebung arbeiteten die Kinder in Tandems an Problemlöseaufgaben, um lautes Denken („talk/think aloud“) anzuregen (Goos & Galbraith, 1996) sowie deren Problemlöseprozesse unter Nutzung des KI-Chatbots zu rekonstruieren. Die Arbeit in Tandems diente außerdem dazu, eine natürliche Kommunikationssituation zu schaffen und den Druck in Einzelsituationen zu verringern, für die forschende Person ein mathematisches Produkt „liefern“ zu müssen (Schoenfeld, 1985, S. 178). Darüber hinaus betonen Johnson und Johnson (1999), dass kooperative Lernformen wie Tandemarbeit Einblicke in das Denken von Schüler:innen ermöglichen und damit besondere Beobachtungschancen bieten. Das Matching der Lernenden erfolgte zufällig durch die Versuchsleiter:innen. Es wurde darauf geachtet, dass jede Problemlöseaufgabe durch möglichst gleich viele Tandems bearbeitet wird. Eine freie Aufgabenwahl bestand daher nicht. Aufgrund des Zeitrahmens des Enrichment-Programms war es nicht möglich, dass alle Kinder jedes mathematische Problem bearbeiten.

Insgesamt wurden 32 Problemlöseprozesse von 20 Tandems erhoben, wovon ...

- zwölf Tandems ohne Partnerwechsel an zwei Problemlöseaufgaben arbeiteten (24 Kinder),
- ein Tandem ohne Partnerwechsel an einer Aufgabe arbeitete (zwei Kinder),
- sieben Tandems mit Partnerwechsel und unterschiedlichen Zusammensetzungen an einer Aufgabe arbeiteten (neun Kinder) (vgl. Tab. 3).

Die Partnerwechsel ergaben sich aus der inhaltlichen Ausgestaltung des Enrichment-Programms. Von den 32 Problemlöseprozessen wurden elf für die Analyse ausgeschlossen, weil entweder keine KI-Nutzung (fünf Prozesse) oder wenig bis keine Kommunikation der Schüler:innen stattfand (sechs Prozesse). Diese sind in Tab. 3 in runden Klammern dargestellt. Nach Ausschluss dieser Fälle lagen 21 Problemlöseprozesse von 15 unterschiedlichen Tandems zur Analyse vor. Die Tandems setzten sich aus 27 Kindern

zusammen, darunter zehn Mädchen und 17 Jungen. Von ihnen besuchten 13 die 3. Klasse, zwölf die 4. Klasse, ein Kind die 5. Klasse und ein Kind nahm am jahrgangsübergreifenden Unterricht teil. In den 21 Problemlöseprozessen haben sieben Paare die Aufgabe *Sieben Tore*, sechs Paare die Aufgabe *Reihenfolgezahlen* und acht Paare die Aufgabe *Kleingeld* bearbeitet (vgl. Tab. 3; siehe 4.1).

Tandem	Sieben Tore	Reihenfolgezahlen	Kleingeld
1	(x)	(x)	
2	(x)	(x)	
3	(x)	x	
4	x	(x)	
5	x	x	
6	x		(x)
7	(x)		
8	x		x
9	x		x
10	x		x
11	x		
12			x
13			(x)
14		x	x
15		x	(x)
16		x	
17			x
18			x
19		(x)	
20		x	x
Gesamt	7 (+4)	6 (+4)	8 (+3)

Tab. 3: Erhobene Problemlöseprozesse (eigene Darstellung)

4.4 Ablauf der Datenerhebung

Zu Beginn wurde allen Tandems ein 1,5-minütiges Instruktionsvideo gezeigt, welches das Aufgaben-Setting erläutert. Darunter zählt die Erklärung der zur Verfügung stehenden Materialien (Aufgabenkarte, Schreibmaterial) und technischen Geräte (Tablet, Tastatur) sowie basaler Bedienmöglichkeiten des KI-Chatbots (Starten eines neuen Chats, Eingabe eines Prompts, Lesen des LLM-Outputs im Ausgabefeld, Möglichkeit des Stellens einer Nachfrage). Weitere Aspekte (Wechsel des Sprachmodells, Einsatz von KI-Agenten, Prompt-Techniken zur Verbesserung der Ausgabequalität) wurden nicht thematisiert. Durch das Instruktionsvideo wurde eine vergleichbare Arbeitsgrundlage für alle Tandems gewährleistet. Insbesondere konnte nicht vorausgesetzt werden, dass die Kinder bereits über Vorerfahrungen mit KI-Chatbots verfügen. Anschließend hatten die Tandems 15 Minuten Zeit, gemeinsam und unter optionaler Verwendung des KI-Chatbots die jeweilige Problemlöseaufgabe zu bearbeiten. Als Abschluss jedes Problemlöseprozesses diente ein leitfadengestütztes Interview, um zur Reflexion des KIEinsatzes anzuregen. Die Schüler:innen wurden

gefragt, zu welchem Zweck sie den KI-Chatbot bei der Aufgabebearbeitung verwendeten. Fand keine KI-Nutzung statt, so wurden die Gründe gegen die Verwendung erfragt.

Um ein umfassendes Bild über die Nutzung eines KI-Chatbots zu erhalten, wurden neben Notizen der Kinder und KI-Chatverläufen (*Produkte*) auch die Problemlöseprozesse videografisch erfasst (*Prozess*). Erhoben wurden zwei Video-Perspektiven (frontal, seitlich) sowie Screenrecordings der Tablets, um alle Eingabeversuche im KI-Chatbot zu erfassen.

4.5 Datenanalyse

Die videografierten Daten (Kamera- und Bildschirm-aufnahmen) wurden in einem ersten Schritt gesichtet, einzelne Problemlöseprozesse wurden im Zuge dessen ausgeschlossen (siehe 4.3). Zu allen übrigen Prozessen wurde eine Prozessbeschreibung in Anlehnung an Rott (2013) angefertigt. Jene Prozessbeschreibung ermöglicht zugleich die Kodierung der Episodeneinteilung, der eingesetzten Heuristiken und der KI-Nutzung.

Die Datenkodierung folgt der qualitativen Inhaltsanalyse nach Mayring (2022) als bewährte Methode zur Auswertung qualitativer Daten. Dazu wurden Techniken der strukturierenden und zusammenfassenden Inhaltsanalyse genutzt. Insbesondere wurden sowohl vorab formulierte Kategorien *deduktiv* auf das Datenmaterial angewandt (siehe 4.5.1 und 4.5.2) als auch neue Kategorien *induktiv* aus dem Datenmaterial heraus gebildet (siehe 4.5.3). Als Grundlage zur Erstellung des Kodierleitfadens dienten das Kodiermanual von Rott (2013) und die Erweiterungen von Herold-Blasius (2021). Mithilfe des induktiv erweiterten und endgültigen Kategoriensystems wurden alle Daten final kodiert. Die Auswertungen erfolgten durch den Erstautor. Die Analysen und auftretenden Zweifelsfälle wurden gemeinsam diskutiert; anschließend wurde die gesamte Datenanalyse vom Zweitautor validiert, um Transparenz und intersubjektive Nachvollziehbarkeit zu gewährleisten.

4.5.1 Episodeneinteilung

Anhand der Prozessbeschreibung, d. h. der inhaltlichen Zusammenfassung der Problemlöseprozesse, wurde eine Einteilung in Episoden vorgenommen. Diese sind aus der Problemlöseforschung bekannt und wurden schrittweise erweitert (siehe 2.2.1; Herold-Blasius, 2021; Rott, 2013, 2014; Schoenfeld, 1985). Eine Übersicht zu den kodierten Episoden mit Kategoriendefinitionen ist im Anhang A zu finden. Es

handelt sich bei dieser Art der Kodierung um ein *event sampling*, weil die Episodengrenzen anhand festgelegter Kriterien, etwa bestimmte Handlungen und Aktivitätswechsel, festgelegt werden (Rott, 2013). Durch die KI-Nutzung ergeben sich Besonderheiten im Hinblick auf die Problemlöseepisoden, etwa wenn die KI-Nutzung allein als Plan festgelegt und implementiert wird.

4.5.2 Heurismeneinsatz

Die Heurismenkodierung folgt der Dreiteilung in heuristische Hilfsmittel, Strategien und Prinzipien (Bruder & Collet, 2011) sowie den zugehörigen Unterkategorien (Herold-Blasius, 2021; Rott, 2013; siehe Anhang B). Da das Ende eines Heurismeneinsatzes schwer bestimmbar ist, werden Heurismen nach Rott (2013) als *Impulse* verstanden und deren Auftreten per *event sampling* anhand beobachtbarer Handlungen bzw. Versprachlichungen kodiert. Daraus folgt eine wesentliche Unterscheidung bei der Kodierung sich wiederholender Heurismen: Strukturgleiche Wiederholungen desselben Heurismus werden nur einmal kodiert. Heurismenwechsel sowie gleichnamige Heurismen in unterschiedlichen Strukturzusammenhängen (Schemata) werden hingegen mehrfach kodiert. Gleichnamige Heurismen werden in diesem Fall nummeriert. Deutlich wird diese Unterscheidung am Beispiel der Sieben-Tore-Aufgabe. Der Heurismus des *Rückwärtsarbeitens* kann durch verschiedene Schemata, etwa „+1 und Verdoppeln“ oder „Verdoppeln und +1“, angewandt werden. Die Heurismen sind zwar gleichnamig, stellen aber verschiedene mathematische Strukturen zur Ergebnisbestimmung dar.

Besonderheiten ergeben sich weiterhin durch das parallele Auftreten mehrerer Heurismen und sogenannter *Strategiekeime* (Stein, 1995). Zu letzteren zählen Heurismen, welche nur in Ansätzen erkennbar sind und nicht gefestigt verwendet werden.

4.5.3 KI-Nutzung

Im Gegensatz zur Kodierung der Episoden und des Heurismeneinsatzes liegen noch keine deduktiven Kategorien zur KI-Nutzung beim mathematischen Problemlösen vor. Stattdessen werden diese aus den Prompts der Schüler:innen induktiv durch Zusammenfassen und Strukturieren gebildet. Des Weiteren wird die Häufigkeit der KI-Nutzung erfasst. Analog zu den Heurismen liegt ein *event sampling* vor. Die KI-Nutzung wird als *Impuls* betrachtet, d. h. der Startpunkt einer konkreten Handlung am KI-Chatbot wird kodiert. Die genaue Bestimmung des

Endpunktes ist hingegen wenig zielführend, gerade vor dem Hintergrund von Eingabepausen und anderen Tätigkeiten, welche die Schüler:innen zwischenzeitlich von der Fortsetzung eines Prompts abhalten.

4.5.4 Zusammenlegung der Kodierungen in einer Grafik

Um die drei beschriebenen Kodierebenen gemeinsam zu betrachten, werden sie in einer Grafik zusammengelegt. Diese orientiert sich an Herold-Blasius (2021) und besteht aus drei Zeilen, welche den Problemlöseprozess in seinem zeitlichen Verlauf darstellen (vgl. Abb. 3–5). Die gewählte Darstellung ermöglicht das Erkennen von Mustern und Zusammenhängen zwischen den drei Kodierebenen. Insbesondere wird ermöglicht, die KI-Nutzung den jeweiligen Episoden zuzuordnen sowie Zusammenhänge zum Heurismeneinsatz erkennbar zu machen.

Während die Problemlöse-Episoden mit genauem Start- und Endpunkt dargestellt werden, dienen für die Heurismen und KI-Nutzung vertikale Striche als Kennzeichnung des Impulses. Diese werden durch Codes beschriftet. Arbeiten Partner:innen getrennt voneinander, wird dies in der Grafik durch umrandete, separate Zeilen visualisiert (vgl. Abb. 4).

5. Ergebnisse

Nachfolgend werden die induktiv ermittelten Kategorien zu den Nutzungsweisen eines KI-Chatbots beim mathematischen Problemlösen systematisch vorgestellt (vgl. Abb. 2). Zu den einzelnen KI-Nutzungsweisen werden ausgewählte Beispielprozesse analysiert. Die hier beschriebenen Prozesse sind insofern komplementär zueinander, dass sie ein möglichst breites Spektrum der KI-Nutzung mathematisch interessierter Kinder beim Problemlösen aufzeigen. Außerdem stellen sie typische Zeitpunkte der KI-Nutzung im Problemlöseprozess sowie den Zusammenhang zwischen der Nutzung eines KI-Chatbots und dem Heurismeneinsatz dar. Im Anschluss wird jeweils ein allgemeiner Blick auf die erhobenen Problemlöseprozesse geworfen.

5.1 Rechenhilfe

5.1.1 Rechenhilfe zur Bestimmung einer Teillösung

Die drei gewählten Problemlöseaufgaben können durch Berechnungen systematisch gelöst werden, wenn ihre dahinter liegende mathematische Struktur erkannt wurde. Dies wird im Problemlöseprozess von Pia und Lola beispielhaft deutlich (vgl. Abb. 3).

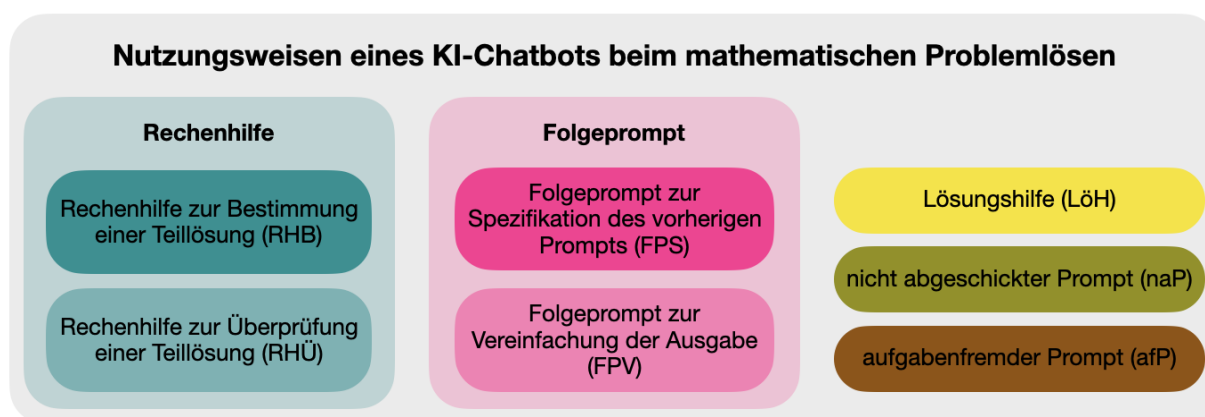


Abb. 2: Nutzungsweisen eines KI-Chatbots beim mathematischen Problemlösen (eigene Darstellung)

Die Schülerinnen arbeiten an der Sieben-Tore-Aufgabe und gehen dabei systematisch vor. Nach einer *Explorations*-Episode, in denen sie verschiedene Lösungsansätze des *Rückwärts- und Vorwärtsarbeitens* (*RwA 1*, *RwA 2* und *VwA 1*) diskutieren, gehen sie in eine *Planung-Implementation* über (05:41–10:31).

Pia gibt an, dass der Mann vor dem letzten Tor vier Äpfel habe. Sie erklärt: „Vier durch zwei ist zwei (..) plus einen Apfel gibt er drei weg hat er noch einen übrig“ (06:12). Lola nimmt sich die Tastatur und schickt einen Prompt („Was ist 4.“) ab, worauf die beiden das Geteilt-durch-Zeichen suchen. Ein weiterer Prompt („was ist 4:“) wird nicht abgeschickt. Sie stellen schließlich das Tablet zur Seite und lösen die Aufgabe gemeinsam. Pia notiert $4 : 2 = 2 + 1$ und $4 - 3 = 1$. Lola stellt fest: „Das geht auch aber wir können das doch auch anders machen“ (07:53). Nach kurzer Stille ergänzt sie: „Egal wir machen das halt so wie du gesagt hast“ (08:06).

In der eben beschriebenen Episode wird *vorwärts gearbeitet* (*VwA 1*), um die Anzahl vier Äpfel nach dem sechsten Tor zu begründen (Anwendung des Schemas „Halbieren und -1“). Lola versucht zunächst allein und dann gemeinsam mit Pia, die von Pia genannte Rechnung in den KI-Chatbot einzugeben. Der KI-Chatbot wurde demnach als *Rechenhilfe*, wenngleich auch erfolglos, eingesetzt. Bei Pia ist die strukturierte *Implementation des Plans* bereits beim Erklären und Notieren erkennbar. Lola verspricht die Absicht, Pias Plan nun umzusetzen.

Daraufhin notiert Lola beim siebten Tor eine vier, beim sechsten Tor eine neun und beim fünften Tor eine 19. Danach nimmt sie sich die Tastatur. Pia unterbricht Lola: „Warte das sind neununddreißig“ (08:29) und schreibt eine 39 beim vierten Tor auf. Die beiden bestimmen mithilfe des KI-Chatbots die Anzahl der Äpfel für die noch übrigen Tore und notieren diese (79, 157, 315). Die 315 Äpfel des ersten

Tores notieren sie zugleich als Anzahl der Äpfel, die der Mann am Anfang hatte.

Lola und Pia *rechnen rückwärts* (*RwA 2*) und wenden dabei das Schema „Verdoppeln und +1“ an. Addition und Multiplikation werden somit in ihrer Reihenfolge vertauscht. Der KI-Chatbot dient den Kindern als *Rechenhilfe zur Bestimmung von Teillösungen* (*RHB*), indem sie die einzelnen Additionsaufgaben nacheinander als Prompt eingeben (z. B. „was ist $39+39$ “) und dann noch im Kopf eins addieren. Für das zweite Tor wird allerdings $78+78$ anstatt $79+79$ eingegeben, wodurch sie nicht das richtige Endergebnis von 382 Äpfeln erhalten.

Die Analyse aller Problemlöseprozesse zeigt, dass Schüler:innen den KI-Chatbot häufig während der Exploration und der (Planung-)Implementation nutzen, um Zwischenschritte oder Teilergebnisse zu *berechnen*. Genauer nutzen die Tandems in 13 der 21 ausgewerteten Problemlöseprozesse den KI-Chatbot mindestens einmal als *Rechenhilfe zur Bestimmung einer Teillösung* (*RHB*). In diesen Fällen haben die Kinder bereits durch den Einsatz aufgabenbezogener Heuristiken eine Lösungsidee entwickelt, welche im Anschluss durch Nutzung des KI-Chatbots umgesetzt wird. Der Heuristikeinsatz stellt somit den *Auslöser* der KI-Nutzung dar.

5.1.2 Rechenhilfe zur Überprüfung einer Teillösung

Die Nutzung des KI-Chatbots für Berechnungen konnte nicht nur während Episoden der Exploration und Implementation beobachtet werden, sondern auch zur *Überprüfung von Teillösungen* (*RHÜ*) im Rahmen einer *Verifikations*-Episode. Nach Rott (2013) liegt eine Verifikation vor, wenn Schüler:innen bereits Zwischen- bzw. Endergebnisse ermittelt haben, die auf ihre Korrektheit überprüft werden sollen.

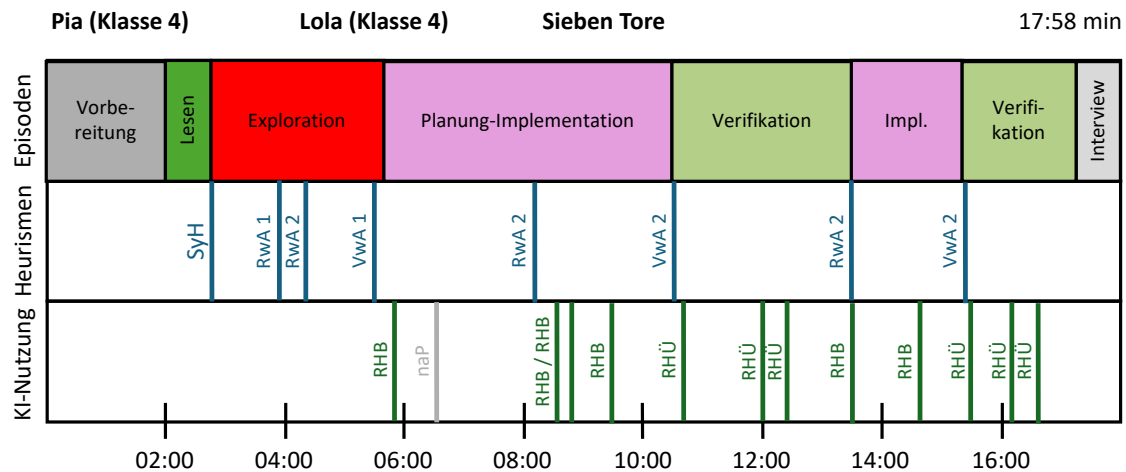


Abb. 3: Pia und Lola – Sieben Tore (eigene Darstellung)

Der Problemlöseprozess von Pia und Lola eignet sich, diese KI-Nutzungsweise einzuordnen (vgl. Abb. 3). Nachdem alle Teilergebnisse notiert wurden (siehe 5.1.1), diktiert Pia die Aufgabe: „Was ist (...) drei eins fünf [...] minus eins [...] durch zwei“ (10:31). Lola schlägt schließlich die Aufgabe $314 : 2$ vor, welche sie in den KI-Chatbot eingeben. Danach tippt Pia für das nächste Tor den Prompt „was ist $157:2$ “ ein. Nach der Ausgabe „156,5“ ändert sie ihren Prompt zu „Was ist $156:2$ “. Nun erhalten sie das Ergebnis „78“ und stellen einen Unterschied zu ihrer notierten Lösung auf dem Papier fest: „Wir haben da neunundsiebzig“ (12:53). Pia schlussfolgert: „Das heißt die Zahl und die Zahl ist falsch“ (13:24) und tippt dabei mit dem Stift auf die 157 und 315. Lola widerspricht ihr mehrmals.

Wie aus der Beschreibung hervorgeht, leitet Pia eine *Verifikation* der bereits ermittelten Ergebnisse ein, indem sie die Aufgabe $(315 - 1) : 2$ diktiert (10:31–13:32). Dazu wird das *Vorwärtsarbeiten* (VwA 2) nach dem Schema „-1 und Halbieren“ angewandt. Die Umsetzung des Heurismus wird durch den KI-Chatbot im Sinne einer *Rechenhilfe zur Überprüfung einer Teillösung* (RHÜ) unterstützt. Durch die Nutzung des KI-Chatbots als Hilfestellung konnten die Schülerinnen einen Rechenfehler im Lösungsweg identifizieren.

In der darauffolgenden *Implementations*-Episode (13:32–15:21) gelingt es Pia und Lola mithilfe des *Rückwärtsarbeitens* (RwA 2) und unter KI-Nutzung, jenen Rechenfehler zu korrigieren. Sie erhalten als neues Endergebnis 319 Äpfel, welche der Mann zu Beginn hatte. Auch dieses Ergebnis *verifizieren* sie erneut durch *Vorwärtsarbeiten* (VwA 2) und die Eingabe entsprechender Prompts (15:21–17:13).

Das Fallbeispiel von Pia und Lola verdeutlicht, dass der KI-Chatbot die beiden Schülerinnen in verschiedenen Episoden des Problemlöseprozesses und für unterschiedliche Zwecke unterstützen kann. Sie nutzen den KI-Chatbot zielgenau zur Implementation und Verifikation ihres Plans und wenden mithilfe des LLM aufgabenspezifische Heuristiken an.

Insgesamt sind Pia und Lola das einzige von 15 Tandems, die den KI-Chatbot während der *Verifikation* als *Rechenhilfe zur Überprüfung einer Teillösung (RHÜ)* nutzen. Auch bei dieser Nutzungsweise des KI-Chatbots konnten wir beobachten, dass die Schülerinnen von einem gewählten Heurismus ausgehen und diesen mithilfe des LLM umsetzen. Demnach nimmt der Heurismeneinsatz auch hier die Funktion des *Auslösers* der KI-Nutzung ein.

5.2 Lösungshilfe

Der Problemlöseprozess von Jonas und Kilian (vgl. Abb. 4) ist ein typisches Beispiel für die KI-Nutzungsweise *Lösungshilfe* (*LÖH*), welche in vielen anderen Prozessen strukturell ähnlich beobachtet wurde. Unter der Kategorie *Lösungshilfe* (*LÖH*) wurden alle Versuche zusammengefasst, die gesamte Aufgabenstellung oder einen Teil dieser als Prompt einzugeben. Insbesondere wird mit der KI-Nutzung als Lösungshilfe das Ziel verfolgt, das Endergebnis bzw. alle Möglichkeiten zu erhalten. Dabei wird nicht unterschieden, ob der Aufgabentext 1:1 übernommen oder umformuliert wird.

Jonas und Kilian arbeiten gemeinsam an der Problemlöseaufgabe „Reihenfolgezahlen“. Nachdem sie die Aufgabenstellung gelesen und danach die Problemstellung analysiert haben, gehen sie sofort in eine *Planungs- und Implementations*-Episode über (02:42–06:05).

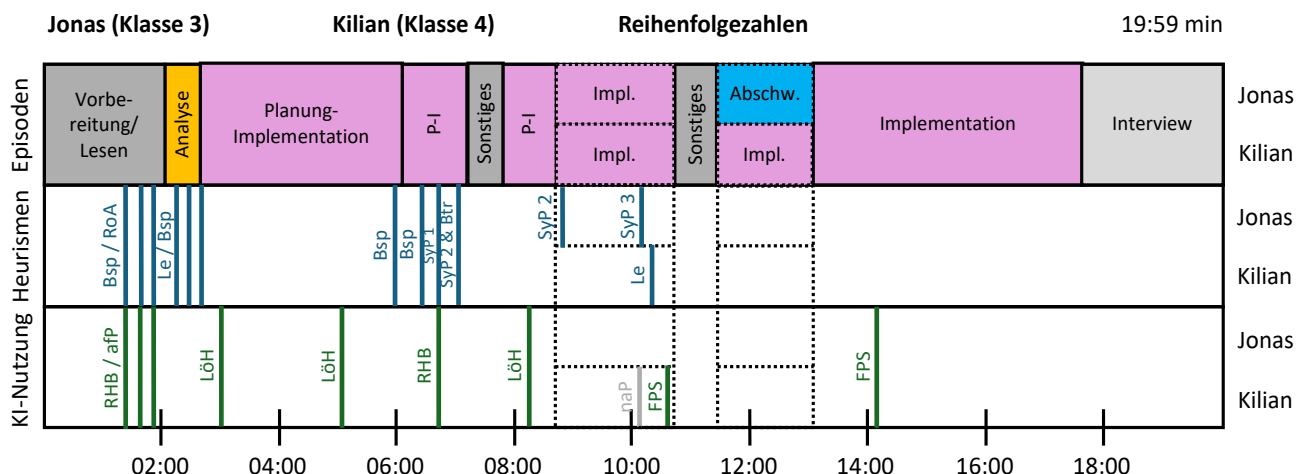


Abb. 4: Jonas und Kilian – Reihenfolgezahlen (eigene Darstellung)

Kilian notiert eine erste Additionsaufgabe auf dem Papier ($12+13=25$). Danach gibt er gemeinsam mit Jonas die Aufgabenstellung in eigenen Worten in den KI-Chatbot ein („nenne mir alle Reihenfolgen zahlen die addiert Maximal 25 ergeben [sic]“). Als Output erhalten sie 26 Additionsaufgaben bestehend aus je zwei Summanden und der Summe 25 ($0+25=25$, $1+24=25$, $2+23=25$, ..., $25+0=25$). Kilian schreibt die ersten beiden Aufgaben ab ($0+25$, $1+24$). Danach stellt er fest: „Das sind keine Reihenfolgezahlen“ (04:51), woraufhin Jonas ihm zustimmt und als Beispiel die Zahlen von fünf bis zehn angibt. Kilian gibt einen neuen Prompt in den KI-Chatbot ein („Nenne mir alle Reihenfolgezahlen [sic] bis 13“). Analog zum vorherigen Prompt gibt der KI-Chatbot Additionsaufgaben mit der Summe 13 aus ($0+13=13$, $1+12=13$, $2+11=13$, ..., $13+0=13$). Kilian fügt an: „Das wollte ich gar nicht“ (05:46) und streicht die notierten Aufgaben durch. Er beginnt, eigene Beispiele aufzuschreiben ($1+2$ und $0+1$).

Indem der Aufgabentext in verkürzter Form als Prompt eingegeben wird, fungiert der KI-Chatbot als *Lösungshilfe* (LÖH). Der Plan von Jonas und Kilian, das mathematische Problem zu lösen, liegt in der Nutzung des KI-Chats. Der LLM-Output enthält jedoch, mit Ausnahme der Additionsaufgaben $12+13$ und $13+12$ bzw. $6+7$ und $7+6$, keine Additionsaufgaben mit Reihenfolgezahlen im Sinne der Aufgabenstellung. Dies stellt die Lernenden vor neue inhaltliche Herausforderungen. Kilian bemerkt zwar nach beiden Prompts Differenzen zwischen der Aufgabenanforderung und dem LLM-Output, kann daraus allerdings (noch) keine Schlussfolgerungen für die Spezifikation seines Prompts ziehen.

Aus der Analyse aller Videodaten geht hervor, dass viele Tandems den KI-Chatbot als *Lösungshilfe* (LÖH)

verwenden. Dies ist in insgesamt 14 der 21 ausgewerteten Problemlöseprozesse mindestens einmal der Fall. Besonders an der Lösungshilfe ist, dass der Plan zur Lösung des mathematischen Problems allein in der KI-Nutzung begründet ist. In der anschließenden *Implementation* des Plans, also der Eingabe des Aufgabentexts in den KI-Chatbot, sind keine eigenen Heurismen der Lernenden erkennbar – es findet *KI-Nutzung ohne Heurismeneinsatz* statt. Die Lösungshilfe unterscheidet sich damit klar von der Rechenhilfe (siehe 5.1).

5.3 Folgeprompt

Unter dem Begriff Folgeprompt fassen wir alle zusätzlichen neuen Eingaben in den KI-Chatbot, welche direkt auf einen Prompt folgen und sich auf diesen beziehen. Im Rahmen der Studie konnten zwei Arten von Folgeprompts während des Problemlösens zu Kategorien zusammengefasst werden: *Folgeprompts zur Spezifikation des vorherigen Prompts* (FPS) (siehe 5.3.1) sowie *Folgeprompts zur Vereinfachung der Ausgabe* (FPV) (siehe 5.3.2).

5.3.1 Folgeprompt zur Spezifikation des vorherigen Prompts

Im Problemlöseprozess von Jonas und Kilian werden *Folgeprompts* eingesetzt (vgl. Abb. 4). Im Anschluss an das Erkennen der fehlerhaften LLM-Outputs (siehe 5.2) verfolgen Jonas und Kilian zunächst einen anderen Plan, um weitere Additionsaufgaben mit Reihenfolgezahlen *ohne KI-Nutzung* zu bestimmen. Ab 08:44 arbeiten die beiden getrennt voneinander. Jonas ermittelt durch das Entfernen von Summanden („Vorne verkürzen“) und das Erhöhen aller Summanden („Alle erhöhen“) schrittweise weitere Möglichkeiten (Schwätzer & Selter, 1998, S. 127). Währenddessen gibt Kilian ab 10:34 einen neuen Prompt

ein und schickt ihn bei 13:03 ab („reihenfolgenzahlen sind direkt aufeinanderfolgende zahlen z.b. 1,2 finde alle möglichen plusaufgaben mit reihenfolgenzahlen. Das Ergebnis darf nicht größer sein als 25. [sic]“). Der LLM-Output enthält nun 26 Additionsaufgaben mit zwei oder mehr Summanden unter Verwendung von Reihenfolgezahlen. Von diesen 26 Aufgaben werden im LLM-Output zehn als „nicht gültig“ erklärt, da ihre Summe über 25 liegt.

Das getrennte Arbeiten von Jonas und Kilian macht die *Implementation* zweier verschiedener *Pläne* deutlich. Während Jonas durch *Systematisches Probieren* (SyP 2 und SyP 3) weitere Möglichkeiten bestimmt, setzt Kilian seinen Plan, den KI-Chatbot zur Lösung des mathematischen Problems zu nutzen, fort. Durch seinen Prompt *spezifiziert* er seine vorherige Eingabe, indem er zusätzliche relevante Informationen zur Lösung der Aufgabe bereitstellt (FPS). Durch die Erklärung des Begriffs „Reihenfolgezahlen“ und die Ergänzung eines Beispiels kann der KI-Chatbot den Prompt im Sinne der Aufgabenstellung offenbar besser verarbeiten. Der LLM-Output enthält nun 16 von 27 richtigen Möglichkeiten, welche systematisch – etwa durch das Entfernen von Summanden oder das Erhöhen aller Summanden – dargestellt werden. Das Stellen eines Folgeprompts führt in diesem Fall zu einer Steigerung der mathematischen Qualität und Korrektheit der Ausgabe, wodurch er gleichzeitig als *Lösungshilfe* (LöH) dient (siehe 5.2). Analog zur Lösungshilfe setzt Kilian zur Umsetzung seines Plans keine eigenen Heuristiken ein, stattdessen findet auch hier *KI-Nutzung ohne Heurismeneinsatz* statt.

Mit Absenden des Prompts arbeiten Jonas und Kilian wieder zusammen. Beide lesen sich die LLM-Ausgabe durch. Jonas stellt fest: „Es gibt über viele“ (13:22). Ab 14:09 gibt Kilian seinen Prompt nochmals, aber leicht umformuliert, ein

(„Reihenfolgenzahlen sind direkt aufeinanderfolgende zahlen z.b. 1,2,3 u.s.w finde alle möglichen plusaufgaben die als Ergebnis nicht mehr als 25 haben [sic]“). Bei 15:45 schickt Kilian den Prompt ab. Erneut lesen sich die beiden den LLM-Output durch. Dieser enthält nun 39 Additionsaufgaben mit Reihenfolgezahlen, von denen zwölf als „nicht gültig“ gekennzeichnet sind. Jonas merkt in Bezug auf die Additionsaufgaben an: „Die müssen wir nur noch alle aufschreiben“ (16:13). Kilian ergänzt: „Und ein paar davon wegstreichen“ (16:18). Daraufhin stimmt Jonas ihm zu: „Ja alle wo nicht gültig steht“ (16:21).

Ein weiteres Mal nutzt Kilian einen *Folgeprompt zur Spezifikation des vorherigen Prompts* (FPS), indem er das Beispiel für Reihenfolgezahlen auf drei Zahlen ausweitet. Der KI-Chatbot geht nun noch systematischer vor, indem mit der kleinsten Aufgabe begonnen und diese so lange um weitere Summanden verlängert wird, bis die Summe über 25 liegt ($1+2=3$, ..., $1+2+3+4+5+6=21$). Danach wird mit den nächstgrößeren Additionsaufgaben aus zwei Summanden ($2+3=5$, $3+4=7$, ...) analog verfahren. Der KI-Chatbot listet nun alle 27 Möglichkeiten auf, wodurch der Prompt seinen Zweck als *Lösungshilfe* (LöH) erfüllt. Ob Jonas und Kilian die Systematik der KI-generierten Lösung verstanden haben, kann nicht abschließend beurteilt werden. Die beiden haben keine Zeit mehr, die einzelnen Möglichkeiten auf ihr Blatt zu übertragen. Allerdings können sie als ungültig gekennzeichnete Additionsaufgaben aufgrund ihrer zu hohen Summe ausschließen.

5.3.2 Folgeprompt zur Vereinfachung der Ausgabe

Als weiteres Beispiel für die Nutzung von Folgeprompts beim mathematischen Problemlösen dient der Prozess von Bruno und Max. Beide bearbeiten die Kleingeld-Aufgabe (vgl. Abb. 5).

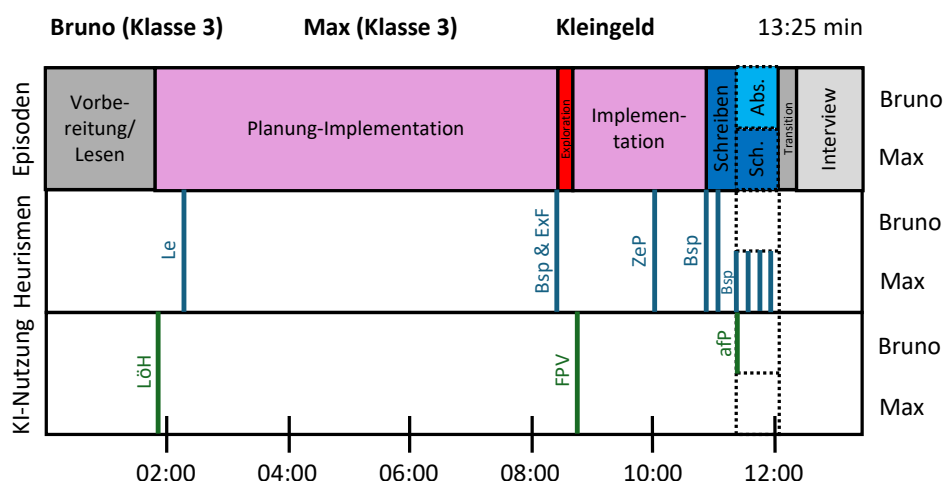


Abb. 5: Bruno und Max – Kleingeld (eigene Darstellung)

Wie bereits im Prozess von Jonas und Kilian beobachtet, legen auch Bruno und Max die KI-Nutzung im Sinne einer *Lösungshilfe* (LöH) als *Plan* fest und führen diesen anschließend in der *Implementation* aus (01:47–08:25). Bruno tippt zu Beginn des Problemlöseprozesses die Aufgabenstellung als Prompt ein („Wie viele Möglichkeiten gibt es 31 Cent nur mit 10-Cent -, 5-Cent- und 2-Cent zu legen? [sic]“). Bei 06:49 schickt er den Prompt ab. Max sagt: „So viel kann ich nicht lesen“ (06:55). Bruno liest einen Teil des LLM-Outputs vor: „Zusammenfassend es gibt die folgenden Kombinationen ja wir haben es“ (07:37). Daraufhin zeigt er auf den Bildschirm und schlussfolgert, dass es drei Möglichkeiten gäbe. Dann liest Bruno weiter vor und korrigiert, dass es insgesamt sechs Möglichkeiten gäbe.

Die besondere Schwierigkeit der KI-Nutzung in diesem Problemlöseprozess liegt in der *Komplexität des LLM-Outputs*, welcher Gleichungen und Variablen enthält. Der KI-Chatbot stellt zur Lösung der Aufgabe die Gleichung $10x + 5y + 2z = 31$ auf, wobei x die Anzahl der 10-Cent-, y die Anzahl der 5-Cent- und z die Anzahl der 2-Cent-Münzen repräsentieren. Bruno bezieht sich mit seinem Kommentar zur Anzahl der Möglichkeiten womöglich auf folgenden Ausschnitt des LLM-Outputs:

Zusammenfassend gibt es die folgenden Kombinationen:

- Bei $x = 0$: $(y, z) = (1, 13), (3, 8), (5, 3)$
- Bei $x = 1$: $(y, z) = (1, 8), (3, 3)$
- Bei $x = 2$: $(y, z) = (1, 3)$

Aus diesen drei Stichpunkten schließt Bruno, dass es drei Möglichkeiten gäbe. Erst durch weiteres Lesen erfährt er, dass es sechs Möglichkeiten gibt. Der KI-Chatbot ermittelt zwar das mathematisch korrekte Ergebnis der Kleingeld-Aufgabe und dient somit als *Lösungshilfe* (LöH), Bruno und Max können das symbolisch-algebraische Darstellungsregister der Lösung jedoch nicht für die weitere Problembearbeitung nutzen (Prediger & Wessel, 2011).

Ab 08:40 beginnt Bruno mit der Eingabe eines neuen Prompts („Die sechs Lösungen einfach“), welchen er bei 10:17 absendet. Max hat unterdessen zwei eigene Beispiele (2x 10ct, 1x 5ct, 3x 2ct sowie 1x 10ct, 3x 5ct und 3x 2ct) notiert. Im LLM-Output können Max und Bruno lesen:

Hier sind die sechs möglichen Lösungen, um 31 Cent mit 10-Cent-, 5-Cent- und 2-Cent-Münzen zu bilden:

1. 0 10-Cent-Münzen, 1 5-Cent-Münze, 13 2-Cent-Münzen

2. 0 10-Cent-Münzen, 3 5-Cent-Münzen, 8 2-Cent-Münzen
3. 0 10-Cent-Münzen, 5 5-Cent-Münzen, 3 2-Cent-Münzen
4. 1 10-Cent-Münze, 1 5-Cent-Münze, 8 2-Cent-Münzen
5. 1 10-Cent-Münze, 3 5-Cent-Münzen, 3 2-Cent-Münzen
6. 2 10-Cent-Münzen, 1 5-Cent-Münze, 3 2-Cent-Münzen

Jede dieser Kombinationen summiert sich zu genau 31 Cent.

Max sagt daraufhin: „Man müsste jetzt einfach nur noch“ (10:35). Bruno ergänzt: „Jetzt schreiben wir einfach ab“ (10:37). Bruno schlägt vor, Max solle seine bereits notierten Beispiele durchstreichen, da diese falsch seien. Max widerspricht zunächst. Von 10:56 bis 12:08 schreibt er die sechs Möglichkeiten, die im Output aufgelistet sind, nacheinander ab.

Durch seinen Prompt leitet Bruno eine neue *Implementations*-Episode ein (08:40–10:56). Sein *Plan*, die Kleingeld-Aufgabe allein durch KI-Nutzung zu lösen, wird somit fortgeführt. Die Generierung und Anwendung von Heurismen spielt hingegen keine Rolle, wodurch auch in diesem Problemlöseprozess eine *KI-Nutzung ohne Heurismeneinsatz* stattfindet. Stattdessen gelingt es Bruno durch die Eingabe eines Folgeprompts, die inhaltliche Hürde des vorherigen Outputs zu überwinden. Im Vergleich zur Spezifikation des vorherigen Prompts (siehe 5.3.1) werden keine zusätzlichen Informationen zur Lösung der Aufgabe genannt. Vielmehr wird ein Anstoß zur *Vereinfachung der Ausgabe* (FPV) gegeben, indem die Möglichkeiten „einfach“ ausgegeben werden sollen. Die gewünschte Vereinfachung könnte sich hierbei auf das Entfallen von Erklärungen, Gleichungen und Variablen beziehen. Der KI-Chatbot berücksichtigt diesen Folgeprompt und reduziert die Komplexität der Ergebnisdarstellung. Anstatt von Wertepaaren (y, z) in Abhängigkeit von x werden nun alle sechs Möglichkeiten, um mit 10-Cent-, 5-Cent- und 2-Cent-Münzen genau 31 Cent zu legen, in einer Liste dargestellt. Somit findet durch den Einsatz des Folgeprompts ein *intermodaler Transfer* vom symbolisch-algebraischen zum symbolisch-numerischen Darstellungsregister statt (Bruner, 1966; Prediger & Wessel, 2011). Max und Bruno nutzen diese angepasste Darstellung des LLM-Outputs zur anschließenden Notation aller Möglichkeiten. In der *Schreiben*-Episode (10:56–12:08) kommt die inhaltliche Arbeit zum Erliegen. Insbesondere wird durch das sofortige Abschreiben der KI-generierten Lösung nicht ersichtlich, inwiefern Bruno und Max die

Systematik hinter dem Finden aller sechs Möglichkeiten verstanden haben. Stattdessen wird die KI-generierte Lösung scheinbar nicht hinterfragt. Bruno schließt sogar die Richtigkeit der Beispiele von Max aus, obwohl diese der 5. und 6. Möglichkeit des LLM-Outputs entsprechen.

Wie aus der Analyse aller 21 Problemlöseprozesse hervorgeht, nutzen die Tandems in fünf Prozessen *Folgeprompts zur Spezifikation des vorherigen Prompts (FPS)* und in zwei Prozessen *Folgeprompts zur Vereinfachung der Ausgabe (FPV)*. Sie treten in Episoden der *Implementation* sowie stets *ohne eigenen Heurismeneinsatz* auf. Oftmals stehen Folgeprompts in Zusammenhang mit der KI-Nutzung als Lösungshilfe.

5.4 Weitere KI-Nutzungsweisen

Neben der Verwendung als Rechen- und Lösungshilfe sowie dem Stellen von Folgeprompts konnten noch zwei weitere KI-Nutzungsweisen mathematisch interessierter Kinder beim Problemlösen induktiv ermittelt werden. Diese sind zwar zur Unterstützung der Lernenden im Problemlöseprozess nicht förderlich, stellen aber dennoch nicht zu vernachlässigende KI-Nutzungsweisen der Kinder dar.

Das Fallbeispiel von Bruno und Max verdeutlicht beide KI-Nutzungsweisen (vgl. Abb. 5). Nachdem Max mit dem Abschreiben der sechs Lösungsmöglichkeiten für die Kleingeld-Aufgabe beginnt, schaut Bruno ihm zunächst zu. Ab 11:24 gibt Bruno einen neuen Prompt ein, sendet diesen jedoch nicht ab („hvfzftvnfhvjfhnfhgfvfhtvujhghjzngnnfhjctzrc“).

Es handelt sich somit um einen *nicht abgeschickten Prompt (naP)*. Diese Prompts können sowohl aufgabenbezogener (Rechenhilfe, Lösungshilfe, Folgeprompt) als auch aufgabenfremder Natur sein. In 16 der 21 ausgewerteten Problemlöseprozesse werden Prompts eingegeben, aber nicht abgeschickt.

Der von Bruno eingetippte Prompt ist zugleich ein *aufgabenfremder Prompt (afP)*, da dieser keinen inhaltlichen Bezug zur mathematischen Aufgabe erkennen lässt. Bruno befindet sich daher in einer *Abschweifungs-Episode* (11:24–12:08). Zur Kategorie *aufgabenfremder Prompt (afP)* zählen all diejenigen Eingaben, die entweder nicht den inhaltlichen Kern des mathematischen Problems ansprechen oder sich mit außermathematischen Themen beschäftigen („Spaßeingaben“). Letztere treten in Episoden der *Abschweifung* auf, da keine inhaltliche Arbeit an der Problemlöseaufgabe stattfindet. Insbesondere wenden die Lernenden beim Abschweifen keine Heurismen an, sodass auch aufgabenfremde

Prompts eine *KI-Nutzung ohne Heurismeneinsatz* implizieren. Insgesamt wird in 12 Problemlöseprozessen mindestens ein aufgabenfremder Prompt eingesetzt.

6. Diskussion

6.1 Zusammenfassung der Ergebnisse

Mithilfe unserer explorativen Studie konnten wir erste Erkenntnisse zum Nutzungsverhalten mathematisch interessierter Kinder im Umgang mit einer generativen KI gewinnen. Wir fokussierten dabei den Einsatz eines KI-Chatbots beim mathematischen Problemlösen und untersuchten KI-Nutzungsweisen, deren Verortung im Problemlöseprozess sowie deren Zusammenhänge zum Heurismeneinsatz.

Aus den erhobenen Daten konnten verschiedene *Nutzungsweisen eines KI-Chatbots bei mathematisch interessierten Kindern im Setting des Problemlösens* induktiv abgeleitet werden. Oftmals wird der KI-Chatbot als *Rechenhilfe* genutzt, um (Zwischen-)Ergebnisse im Problemlöseprozess durch Eingabe entsprechender Terme oder Gleichungen zu bestimmen bzw. zu überprüfen. Als *Lösungshilfe* dient der KI-Chatbot dann, wenn Schüler:innen die Aufgabenstellung vollständig oder in Teilen als Prompt eingeben. In Zusammenhang mit der Lösungshilfe stehen sogenannte *Folgeprompts*, d. h. zusätzliche Eingaben in den KI-Chatbot, welche direkt auf einen Prompt folgen und sich auf diesen beziehen. Hierbei können wir Folgeprompts unterscheiden, die einer Spezifikation der vorherigen Eingabe oder der Vereinfachung des LLM-Outputs dienen sollen. Schließlich beobachteten wir als weitere KI-Nutzungsweisen *nicht abgeschickte Prompts* sowie *aufgabenfremde Prompts*. Letztere beschäftigen sich mit außermathematischen Themen.

KI-Nutzungsweisen zeigen sich zu unterschiedlichen *Zeitpunkten innerhalb des Problemlöseprozesses* und stehen mit ihrer *Funktion* in Beziehung. Während der *Exploration* dient der KI-Chatbot als Rechenhilfe, ohne dass ein Lösungsplan vorliegt. Lösungsideen werden im Tandem diskutiert und mithilfe des KI-Chatbots ausprobiert. Während der (*Planung-*)*Implementation* kommt die Rechenhilfe zur Bestimmung von Teillösungen bei der Umsetzung eines bestehenden Plans zum Einsatz. Einen Sonderfall während der Implementation stellt die Lösungshilfe dar: Indem die Aufgabenstellung als Prompt dient, besteht der Plan zur Lösung des mathematischen Problems einzig in der KI-Nutzung. Folgeprompts treten ebenfalls während Implementations-Episoden auf, da sie sich auf vorhergehende Lösungshilfen

beziehen. In *Verifikations*-Episoden prüfen Schüler:innen Teillösungen durch die Nutzung des KI-Chatbots als Rechenhilfe. Eine aufgabenfremde KI-Nutzung trat ausschließlich während einer *Ab-schweifung* auf. Insgesamt zeigte sich die Nutzung des KI-Chatbots in nahezu allen Episoden des Problemlösens – mit Ausnahme der *Analyse*.

Die Darstellungsform der Problemlöseprozesse erlaubt, *Zusammenhänge zwischen den KI-Nutzungsweisen und des Heurismeneinsatzes* zu beschreiben. Es zeigen sich drei Auffälligkeiten:

- 1) *Heurismus als Auslöser für die KI-Nutzung*: Die Schüler:innen einigen sich zunächst auf einen bekannten Heurismus oder entwickeln gemeinsam einen neuen. Der Lösungsansatz steht demnach bereits fest und wird anschließend unter Verwendung des KI-Chatbots umgesetzt.
- 2) *KI-Nutzung ohne Heurismeneinsatz*: Den KI-Nutzungsweisen Lösungshilfe, Folgeprompts und aufgabenfremde Prompts ist gemein, dass keine eigenen Heurismen der Schüler:innen erkennbar sind.
- 3) *Heurismen ohne KI-Nutzung*: Tandems entwickeln und wählen in unterschiedlichen Episoden aufgabenbezogene Heurismen aus und wenden diese ohne den KI-Chatbot an.

Ein vierter denkbarer Zusammenhang, die *KI-Nutzung als Auslöser für einen Heurismeneinsatz*, konnte in der Studie nicht beobachtet werden.

6.2 Interpretation der Ergebnisse

Die ermittelten KI-Nutzungsweisen mathematisch interessierter Kinder beim Problemlösen bieten neue Erkenntnisse für die Problemlöseforschung. Offenbar sind die Nutzungsweisen eines KI-Chatbots mit entsprechenden Problemlöse-Episoden verknüpft. Man kann von *episodenspezifischen KI-Nutzungsweisen* sprechen. Die Analysen in dieser Studie ermöglichen es, die Rolle von KI-Chatbots als Werkzeug beim Problemlösen (Dörfler, 1993) einzuordnen und zu diskutieren.

Aus Studien zum Heurismeneinsatz beim Problemlösen ist bekannt, dass dieser zur Überwindung inhaltlicher Hürden führen kann (Bruder & Collet, 2011; Rott, 2018; Schoenfeld, 1985). Auch in unserer Untersuchung setzten Kinder unterschiedliche aufgabenbezogene Heurismen ein. In Zusammenhang mit der Nutzung eines KI-Chatbots beim Problemlösen lässt dies verschiedene Perspektiven zu. Besonders die *KI-Nutzung ohne Heurismeneinsatz* sollte kritisch

betrachtet werden. Werden bei der Lösungshilfe, den Folgeprompts und den aufgabenfremden Prompts keine eigenen Heurismen entwickelt, so liegt die Vermutung nahe, dass der KI-Einsatz zu einer Verhinderung der Heurismenentwicklung und -anwendung durch Problemlösende führt. Es stellt sich insbesondere die Frage, ob die KI-Nutzung als Lösungshilfe den Schüler:innen Denkprozesse ab- und vorwegnimmt. Unsere Fallbeispiele legen nahe, dass Kinder die KI-generierte Lösung beim Abschreiben nicht reflektieren und sich vermutlich unzureichend mit der Systematik hinter den Problemlöseaufgaben beschäftigen. Dies widerspricht dem zentralen Ziel des Problemlösens, geeignete Heurismen zu entwickeln und anzuwenden (KMK, 2022). Platz und Plote (2025, S. 8) sprechen von einer möglichen Abnahme menschlicher Kreativität und kritischen Denkens, wenn Nutzende „zunehmend passiv auf KI-Lösungen zurückgreifen, anstatt aktiv Probleme zu lösen.“

Gleichwohl ermöglicht die Rechenhilfe, dass ein von Schüler:innen zuvor ausgewählter *Heurismus* auch *als Auslöser für die KI-Nutzung* dienen kann. Viele Kinder setzen generative KI als Rechenhilfe beim Problemlösen ein (siehe 5.1.1). Indem der KI-Chatbot dabei die Funktion eines Taschenrechners einnimmt, könnte von einer Entlastung des Problemlösens durch eine Verlagerung von Tätigkeiten auf das LLM ausgegangen werden (Ruthven & Chaplin, 1997). Kognitive Prozesse werden verteilt und Routineaufgaben wie das Berechnen von Teillösungen auf den KI-Chatbot ausgelagert (Salomon, 1993). Die Lernenden lassen demnach den KI-Chatbot Berechnungen durchführen, um sich selbst womöglich stärker auf die mathematische Struktur und das Entwickeln von Heurismen fokussieren zu können. Im Prozess von Pia und Lola wird dies insofern deutlich, dass beide zunächst eigene Lösungsideen diskutieren und für die ersten Teilschritte (Tore) eigenständig umsetzen, bevor sie den KI-Chatbot als Rechenhilfe nutzen. Sie wenden zur Bearbeitung der Sieben-Tore-Aufgabe eine Kombination aus Vorwärts- und Rückwärtsarbeiten an und müssen demnach Gedankengänge mehrmals umkehren, um zu einer Lösung zu gelangen. Diese *Reversibilität* stellt so gleich ein mathematikspezifisches Begabungsmerkmal dar (Fuchs, 2006; Kämpnick, 1998; Krutetskii, 1976).

Allerdings können KI-Chatbots keine zuverlässige Entlastung des Kopfrechnens bieten, da der LLM-Output aufgrund stochastischer Berechnungen generiert wird und infolgedessen durch Halluzinationen geprägt ist (Frieder et al., 2024; Plevris et al.,

2023; Schorcht et al., 2024b). Insofern unterscheiden sich KI-Chatbots maßgeblich von anderen digitalen Technologien wie Taschenrechnern und Apps, welche ebenfalls zur Lösungsfindung eingesetzt werden (Meißner, 2006; Spiegel, 1988; Tucker et al., 2017). Jene Halluzinationen können nicht nur bei der Rechenhilfe, sondern auch bei den anderen KI-Nutzungsweisen zu fehlerhaften Ausgaben des LLM führen. Unsere Studienergebnisse zeigen, dass mathematisch interessierte Kinder meist die fachliche Korrektheit des LLM-Outputs nicht hinterfragen. Stattdessen vertrauen sie oft den Antworten des KI-Chatbots. Zudem haben die Kinder den KI-Chatbot bereits für einfache Aufgaben zur Hilfe genommen (z. B. „was ist $39+39$ “; siehe 5.1.1), welche als Teil der arithmetischen Basiskompetenzen i. d. R. im Kopf lösbar sein sollten (KMK, 2022).

Offen bleibt, welche Prompt-Techniken den Problemlöseprozess gezielt unterstützen können. Unsere Untersuchungsergebnisse deuten darauf hin, dass nur wenige Kinder Ansätze für die Nutzung von Prompt-Techniken beim Problemlösen, durch das Formulieren von *Folgeprompts*, entwickeln. Ein Grund dafür könnte im wahrscheinlich mangelnden Vorwissen der Schüler:innen zum Prompt Engineering liegen. Dies spiegelt sich auch in der häufigen Verwendung einfacher Eingaben (Zero Shot) wider. Weiterhin könnte die hohe Komplexität des LLM-Outputs für Lernende eine Herausforderung darstellen (siehe 5.3.2). Diese Komplexität erschwert es Kindern, den Wechsel des Darstellungsregisters – etwa von der symbolisch-numerischen zur grafisch-visuellen Darstellung (Prediger & Wessel, 2011) – durch gezielte Prompts und unter Verwendung von Fachsprache herbeizuführen. Zudem ist zu berücksichtigen, dass LLM Schwierigkeiten haben, Darstellungen in unterschiedlichen Kontexten flexibel zu vernetzen, weil sie v. a. auf sprachlichen bzw. linguistischen Trainingsdaten basieren (Hankeln, 2024). Jedoch wird die Fähigkeit, selbstständig zwischen Repräsentationsebenen zu wechseln, als zentraler Bestandteil mathematischer Begabung hervorgehoben (Fuchs, 2006; Käpnick, 1998; Krutetskii, 1976). Es stellt sich daher die Frage, inwiefern KI-Technologien diese Fähigkeit künftig im Mathematikunterricht fördern können.

Gleichzeitig verdeutlicht die Beobachtung, dass einige Schüler:innen Folgeprompts nutzen, um den LLM-Output besser zu verstehen, das ausgeprägte Bedürfnis mathematisch interessierter Kinder nach plausiblen mathematischen Erklärungen (Heinze, 2005). Dies zeigt, dass generative KI durchaus das Potenzial hat, dieses Lernbedürfnis aufzugreifen.

Um jenes Potenzial künftig besser auszuschöpfen, erscheint es sinnvoll, Kindern unterstützende Strategien im Umgang mit Prompt-Techniken an die Hand zu geben, die ihnen ermöglichen, spezifische Anforderungen – wie etwa Darstellungswechsel – zielgerichtet zu formulieren. Allerdings legen die Analysen auch nahe, dass die bloße Verwendung von Folgeprompts nicht zwangsläufig mit einem tiefen Verständnis der zugrunde liegenden mathematischen Strukturen einhergeht. So bleibt beispielsweise offen, ob Bruno und Max die mathematische Struktur zur Lösung der Kleingeld-Aufgabe (siehe 5.3.2) tatsächlich verstanden oder sich primär an den Ausgaben des KI-Chatbots orientiert haben, ohne diese inhaltlich zu reflektieren. Zukünftige Studien könnten daher untersuchen, inwieweit die KI-Nutzung tatsächlich zu einer kognitiven Aktivierung der Lernenden beiträgt und welche Rahmenbedingungen dafür förderlich sind.

6.3 Grenzen der Studie

Unsere qualitative Studie untersuchte gezielt das KI-Nutzungsverhalten *mathematisch interessierter* Kinder. Die zugrunde liegende Stichprobe ist daher auf eine Positivauswahl beschränkt, auch wenn zuvor keine Testung auf eine mathematische Begabung stattfand. Zukünftige Studien sollten den Einsatz von KI-Chatbots in heterogeneren Lerngruppen analysieren, um ggf. weitere Nutzungsweisen und mögliche Unterschiede zwischen Schüler:innen verschiedener Leistungsniveaus zu identifizieren. Gleichzeitig konnten auch innerhalb der vorliegenden Stichprobe Unterschiede im Strategie- und Lösungsverhalten sowie in der Lösungsqualität der mathematischen Probleme festgestellt werden, die auf die Vielfalt individueller Ansätze hindeuten.

Ebenfalls kann von unterschiedlichen Vorerfahrungen der teilnehmenden Schüler:innen im Umgang mit KI-Chatbots ausgegangen werden. In der vorliegenden Untersuchung wurden diesbezügliche Vorkenntnisse und Erfahrungen nicht systematisch erfasst. Es ist jedoch anzunehmen, dass sich die dynamische Entwicklung generativer KI-Modelle auf die Dynamik der Nutzung durch Schüler:innen auswirkt. Zukünftige Forschung ist erforderlich, um besser zu verstehen, wie unterschiedliche Vorerfahrungen und technologische Fortschritte die Interaktion mit KI-Chatbots beim Problemlösen beeinflussen. Um die Ausgangsbedingungen in dieser Studie zu vereinheitlichen, wurde zu Beginn jedes Problemlöseprozesses ein Instruktionsvideo gezeigt, das grundlegende Funktionen des KI-Chatbots erklärte. In diesem Video wurde ein Beispieldialog im Sinne einer

Rechenhilfe gezeigt („Was ist $53 + 49$?“), an dem sich einige Kinder womöglich orientiert haben. Weiterhin wurde im Instruktionsvideo auf die Möglichkeit des Stellens einer Nachfrage hingewiesen. Inwiefern die spezifische Gestaltung des Videos Einfluss auf die Nutzungsweisen der Kinder nahm, könnte in zukünftigen Studien weiter untersucht werden.

Die Ergebnisse der Studie sind außerdem durch die Auswahl der Aufgaben limitiert. Es standen drei arithmetische Aufgaben im Fokus, die durch Nutzung des KI-Chatbots schriftlich und (potenziell) unterschiedlich gelöst werden können. Problemlöseaufgaben, die eine Arbeit an grafischen Darstellungen erfordern, wurden nicht eingesetzt.

Nicht zuletzt kam in der Untersuchung kein eigens für das Problemlösen in der Grundschule entwickelter KI-Chatbot bzw. KI-Agent zum Einsatz (Herold-Blasius et al., 2025), woraus mathematische Fehler im LLM-Output sowie inhaltlich-sprachliche Hürden für die Schüler:innen beim Verstehen des Outputs verstärkt entstehen können.

6.4 Implikationen für den Mathematikunterricht

Bei der KI-Nutzung als Rechenhilfe und als Lösungshilfe ist zu bedenken, dass generative KI mathematische Probleme nicht zuverlässig löst (Frieder et al., 2024; Plevris et al., 2023; Schorcht et al., 2024b). Lernende sollten KI-Chatbots daher nicht als Taschenrechner-Ersatz verstehen. Mathematikunterricht sollte Halluzinationen nicht nur problematisieren, sondern aktiv zur Reflexion über die Plausibilität von KI-generierten Ausgaben anregen, um kritisches Denken sowie die *Verifikation* der Ergebnisse zu fördern. Ein möglicher Ansatz könnte darin bestehen, verschiedene LLM-Ausgaben zu einem identischen Prompt gemeinsam mit der Klasse zu analysieren, um auf Unterschiede aufmerksam zu machen. Damit steht die Stärkung der Medienkompetenz der Lernenden im Sinne eines *Lernens über KI* (KMK, 2024) in Zusammenhang.

Aus der Nutzung von Folgeprompts können ebenfalls erste Folgerungen für die Unterrichtspraxis gezogen werden. Um einen didaktisch wünschenswerten Output zu erhalten, muss präzise gepromptet werden. Nur wenige Kinder nutzten beim Problemlösen die Möglichkeit, einen vorherigen Prompt zu spezifizieren. Das zielgerichtete Prompten sollte demnach als übergreifende Kompetenz in den Unterricht integriert werden, damit die Schüler:innen *mit KI* (KMK, 2024) gewinnbringend lernen können.

Weiterhin folgt aus der seltenen Nutzung von Folgeprompts zur Vereinfachung der Ausgabe die Notwendigkeit, Darstellungsvernetzungen beim Mathematiklernen verstärkt zu fördern. Die flexible Vernetzung von Repräsentationsmodi mathematischer Lerninhalte ist eine zentrale Aufgabe des Mathematikunterrichts (Prediger & Wessel, 2011), welche durch den Umgang mit KI-Chatbots zusätzliche Möglichkeiten erhält. Die Lernenden sollten beim Lesen und Verstehen des LLM-Outputs jene Darstellungsvernetzungen eigenständig vornehmen oder durch gezieltes Prompten herbeiführen können. Zukunftsweisend könnten KI-Agenten sein, die durch spezielle Systemprompts und Hintergrundinformationen darauf „trainiert“ sind, passende Repräsentationen und Sprachniveaus zu wählen.

6.5 Ausblick

Für die Mathematikdidaktik ergeben sich zahlreiche Anknüpfungspunkte für weiterführende Forschung beim Problemlösen mit KI-Chatbots. Unklar bleibt etwa, inwiefern sich die KI-Nutzungsweisen beim Problemlösen verändern, wenn Kinder Hilfestellungen im Sinne eines Scaffoldings erhalten. Eigens für das Problemlösen in der Grundschule entwickelte KI-Agenten könnten dabei eine entscheidende Rolle spielen (Herold-Blasius et al., 2025).

Fachdidaktische Forschung sollte daran anknüpfend untersuchen, welche KI-Nutzungsweisen sich unterstützend oder hinderlich auf Problemlöseprozesse auswirken. Besonders relevant ist die Frage, wie Halluzinationen vermieden oder die mathematische Qualität des Outputs verbessert werden können. Wie auch Schorcht et al. (2024b) herausstellen, könnte der Einsatz spezieller Prompt-Techniken vielversprechend sein.

Künftige Weiterentwicklungen von LLM sollten zudem in Untersuchungen einbezogen werden. Darunter zählt u. a. die Bearbeitung von Problemlöseaufgaben unterschiedlicher Leitideen der Bildungsstandards (KMK, 2022), wenn LLM auch grafische (Hilfs-) Elemente zuverlässig und aufgabenbezogen generieren können. Zudem könnten datenschutzkonforme LLM, welche eine Spracheingabe und -ausgabe sowie eine menschenähnliche Interaktion ermöglichen, neue Potenziale für ein KI-gestütztes Problemlösen eröffnen.

Literatur

- Aldowah, H., Al-Samarraie, H. & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13–49. <https://doi.org/10.1016/j.tele.2019.01.007>
- Aßmus, D. (2018). Characteristics of mathematical giftedness in early primary school age. In F. M. Singer (Hrsg.), *Mathematical creativity and mathematical giftedness: Enhancing creative capacities in mathematically promising students* (S. 145–167). Springer.
- Bardy, T. & Bardy, P. (2020). *Mathematisch begabte Kinder und Jugendliche. Theorie und (Förder-)Praxis*. Springer.
- Baumanns, L. & Pohl, M. (2024). Leveraging ChatGPT for problem posing: an exploratory study of pre-service teachers' professional use of AI. In E. Faggiano, A. Clark-Wilson, M. Tabach & H.-G. Weigand (Hrsg.), *Mathematics Education in the Digital Age 4 (MEDA4). Proceedings of the 17th ERME Topic Conference* (S. 57–64). University of Bari Aldo Moro and ERME.
- Bruder, R., Büchter, A. & Leuders, T. (2005). Die „gute“ Mathematikaufgabe – ein Thema für die Aus- und Weiterbildung von Lehrerinnen und Lehrern. In G. Graumann (Hrsg.), *Beiträge zum Mathematikunterricht 2005* (S. 139–146). WTM.
- Bruder, R. & Collet, C. (2011). *Problemlösen lernen im Mathematikunterricht*. Cornelsen.
- Bruner, J. S. (1966). *Towards a theory of instruction*. Norton.
- 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.
- Buchholtz, N. & Huget, J. (2024). ChatGPT as a reflection tool to promote the lesson planning competencies of pre-service teachers. In E. Faggiano, A. Clark-Wilson, M. Tabach & H.-G. Weigand (Hrsg.), *Mathematics Education in the Digital Age 4 (MEDA4). Proceedings of the 17th ERME Topic Conference* (S. 129–136). University of Bari Aldo Moro and ERME.
- Buchholtz, N., Schorcht, S., Baumanns, L., Huget, J., Peters, F., Noster, N., Rott, B., Siller, H.-S. & Sommerhoff, D. (2024). Damit rechnet niemand! Sechs Leitgedanken zu Implikationen und Forschungsbedarfen zu KI-Technologien im Mathematikunterricht. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 117, 15–29.
- Büchter, A. & Leuders, T. (2011). *Mathematikaufgaben selbst entwickeln. Lernen fördern – Leistung überprüfen*. 5. Auflage. Cornelsen.
- Cameron, S. & Mesiti, C. (2024). What kind of mathematics teacher is ChatGPT? Identifying the pedagogical practices preferred by generative AI tools when preparing lesson plans. In J. Visnovska, E. Ross & S. Getenet (Hrsg.), *Proceedings of the 46th annual conference of the Mathematics Education Research Group of Australasia*. MERGA.
- Chiu, T. K., Xia, Q., Zhou, X., Chai, C. S. & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, 100118. <https://doi.org/10.1016/j.caeai.2022.100118>
- Collet, C. (2009). *Problemlösekompetenzen in Verbindung mit Selbstregulation fördern. Wirkungsanalysen von Lehrerfortbildungen*. Waxmann.
- de Pisa, L. (1202). *Liber Abaci*.
- Dilling, F. & Herrmann, M. (2024). Using large language models to support pre-service teachers mathematical reasoning – an exploratory study on ChatGPT as an instrument for creating mathematical proofs in geometry. *Frontiers in Artificial Intelligence*. <https://doi.org/10.3389/frai.2024.1460337>
- Dörfler, W. (1993). Computer use and views of the mind. In C. Keitel & K. Ruthven (Hrsg.), *Learning from Computers: Mathematics Education and Technology* (S. 159–186). Springer.
- Frieder, S., Pinchetti, L., Griffiths, R. R., Salvatori, T., Lukasiewicz, T., Petersen, P. & Berner, J. (2024). Mathematical capabilities of ChatGPT. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt & S. Levine (Hrsg.), *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*.
- Fritzlar, T. (2008). Förderung mathematisch begabter Kinder im mittleren Schulalter. In C. Fischer, F. J. Mönks & U. Westphal (Hrsg.), *Individuelle Förderung: Begabungen entfalten – Persönlichkeit entwickeln. Fachbezogene Förder- und Förderkonzepte* (S. 61–77). LIT.
- Fuchs, M. (2006). *Vorgehensweisen mathematisch potentiell begabter Dritt- und Viertklässler beim Problemlösen*. LIT.
- Goos, M. & Galbraith, P. (1996). Do it this way! Metacognitive strategies in collaborative mathematical problem solving. *Educational Studies in Mathematics*, 30(3), 229–260. <https://doi.org/10.1007/BF00304567>
- Hankeln, C. (2024). Challenges in using ChatGPT for assessing conceptual understanding in mathematics education. *Journal of Mathematics Education*, 17(1), 1–15. <https://doi.org/10.26711/007577152790171>
- Heinze, A. (2005). *Lösungsverhalten mathematisch begabter Grundschulkinder – aufgezeigt an ausgewählten Problemstellungen*. LIT.
- Helfrich-Schkarbanenko, A. (2023). *Mathematik und ChatGPT: Ein Rendezvous am Fuße der technologischen Singularität*. Springer.
- Herold-Blasius, R., Rott, B. & Leuders, T. (2017). Problemlösen lernen mit Strategieschlüsseln. Zum Einfluss von flexiblen heuristischen Prompts bei Problemlöseprozessen von Dritt- und Viertklässlern. *mathematica didactica*, 40(2), 99–122. <https://doi.org/10.18716/ojs/md/2017.1235>
- Herold-Blasius, R., Holzäpfel, L. & Rott, B. (2019). Problemlösestrategien lehren lernen – Wo die Praxis Probleme beim Problemlösen sieht. In A. Büchter, M. Glade, R. Herold-Blasius, M. Klinger, F. Schacht & P. Scherer (Hrsg.), *Vielfältige Zugänge zum Mathematikunterricht. Konzepte und Beispiele aus Forschung und Praxis* (S. 295–310). Springer.
- Herold-Blasius, R. (2021). *Problemlösen mit Strategieschlüsseln. Eine explorative Studie zur Unterstützung von Problembearbeitungsprozessen bei Dritt- und Viertklässlern*. Springer.
- Herold-Blasius, R., Baumanns, L., Gruhn, K. & Kleinschmidt, V. (2025). Adaptive Hilfen beim Problemlösen. Individuelle Rückmeldungen durch KI-Chatbots. *Mathematik Differenziert*, 17(4), 40–45.
- Hershkovitz, A., Noster, N., Siller, H. S. & Tabach, M. (2024). Learning analytics in mathematics education: the case of feedback use in a digital classification task on reflective symmetry. *ZDM – Mathematics Education*, 56, 727–739. <https://doi.org/10.1007/s11858-024-01551-5>
- Jauhainen, J. S. & Guerra, A. G. (2023). Generative AI and ChatGPT in School Children's Education: Evidence from a School Lesson. *Sustainability*, 15(18), 14025. <https://doi.org/10.3390/su151814025>

- Johnson, D. & Johnson, R. (1999). *Learning together and alone: cooperative, competitive and individualistic learning*. Prentice Hall.
- Käpnick, F. (1998). *Mathematisch begabte Kinder: Modelle, empirische Studien und Förderprojekte für das Grundschulalter*. Peter Lang.
- Käpnick, F., Kaiser, J., Strübbe, F. & Witte, A. (2021). Ein Erfahrungsbericht zur Entwicklung digitaler Förderformate im Lehr-Lern-Labor Mathe für Kleine Asse. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 111, 12–19.
- Kaitera, S. & Harmoinen, S. (2022). Developing Mathematical Problem-Solving Skills in Primary School by Using Visual Representations on Heuristics. *Lumat Special Issue: Mathematical Thinking and Understanding in Learning of Mathematics*, 10(2), 111–146. <https://doi.org/10.31129/LU-MAT.10.2.1696>
- Kießwetter, K. (1985). Die Förderung von mathematisch besonders begabten und interessierten Schülern – ein bislang vernachlässigtes sonderpädagogisches Problem. *MNU*, 5(39), 300–306.
- [KMK] Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2022). *Bildungsstandards für das Fach Mathematik Primarbereich*.
- [KMK] Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2024). *Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen*.
- Korntreff, S. & Prediger, S. (2021). Verstehensangebote von YouTube-Erklärvideos – Konzeptualisierung und Analyse am Beispiel algebraischer Konzepte. *Journal für Mathematik-Didaktik*, 43(2), 281–310. <https://doi.org/10.1007/s13138-021-00190-7>
- Krutetskii, V. A. (1976). *The psychology of mathematical abilities in schoolchildren*. University of Chicago Press.
- Kulgemeyer, C. (2020). A framework of effective science explanation videos informed by criteria for instructional explanations. *Research in Science Education*, 50(6), 2441–2462. <https://doi.org/10.1007/s11165-018-9787-7>
- Mayring, P. (2022). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 13., überarbeitete Auflage. Beltz.
- Meißner, H. (2006). Taschenrechner im Mathematikunterricht der Grundschule. *mathematica didactica*, 29(1), 5–25. <https://doi.org/10.18716/ojs/md/2006.1059>
- Miao, F. & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000391104>
- Miao, F., Shiohira, K. & Lao, N. (2024). *AI competency framework for students*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000391105>
- Nolte, M. & Pamperien, K. (2024). Mathematik ist PriMa – Begabtenförderung in der Grundschule und darüber hinaus. In S. Schiemann (Hrsg.), *Interesse für Mathematik wecken – Talente fördern* (S. 57–71). Springer.
- OpenAI (2018). *Improving language understanding with unsupervised learning*. <https://openai.com/index/language-unsupervised/>
- OpenAI (2023). *New models and developer products announced at DevDay*. <https://openai.com/index/new-models-and-developer-products-announced-at-devday/>
- OpenAI (2024). *Nutzungsbedingungen für Europa*. <https://openai.com/de-DE/policies/eu-terms-of-use/>
- PISA-Konsortium Deutschland (Hrsg.) (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Waxmann.
- Platz, M. & Plote, C. (2025). KI und Ethik im Klassenzimmer. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 118, 6–13.
- Plevris, V., Papazafeiropoulos, G. & Jiménez Rios, A. (2023). Chatbots put to the test in math and logic problems: a comparison and assessment of ChatGPT-3.5, ChatGPT-4, and Google bard. *AI*, 4, 949–969. <https://doi.org/10.3390/ai4040048>
- Pólya, G. (1949). *Schule des Denkens. Vom Lösen mathematischer Probleme*. Narr Francke Attempo.
- Prediger, S. & Wessel, L. (2011). Relating registers for fractions: Multilingual learners on their way to conceptual understanding. In M. Setati, T. Nkambule & L. Goosen (Hrsg.), *Proceedings of the ICMI Study 21 Conference: Mathematics Education and Language Diversity* (S. 324–333). Sao Paulo, Brazil.
- 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*. Franzbecker.
- Rott, B. (2013). *Mathematisches Problemlösen. Ergebnisse einer empirischen Studie*. WTM.
- Rott, B. & Gawlick, T. (2014). Explizites oder implizites Heuristentraining – was ist besser? *mathematica didactica*, 37, 191–212. <https://doi.org/10.18716/ojs/md/2014.1126>
- Rott, B. (2014). Mathematische Problembearbeitungsprozesse von Fünftklässlern – Entwicklung eines deskriptiven Phasenmodells. *Journal für Mathematik-Didaktik*, 35, 251–282. <https://doi.org/10.1007/s13138-014-0069-2>
- Rott, B. (2018). Empirische Zugänge zu Heuristiken und geistiger Beweglichkeit in den Problemlöseprozessen von Fünft- und Sechstklässlern. *mathematica didactica*, 41(1), 47–76. <https://doi.org/10.18716/ojs/md/2018.1135>
- Rott, B., Specht, B. & Knipping, C. (2021). A descriptive phase model of problem-solving processes. *ZDM – Mathematics Education*, 53(4), 737–752. <https://doi.org/10.1007/s11858-021-01244-3>
- Rott, B., Bruder, R., Heinrich, F. & Bauer, C. (2023). Problemlösen lernen. In R. Bruder, A. Büchter, H. Gasteiger, B. Schmidt-Thieme & H.-G. Weigand (Hrsg.), *Handbuch der Mathematikdidaktik* (S. 313–339). Springer.
- Ruthven, K. & Chaplin, D. (1997). The Calculator as a Cognitive Tool: Upper-Primary Pupils Tackling a Realistic Number Problem. *International Journal of Computers for Mathematical Learning*, 2, 93–124. <https://doi.org/10.1023/A:1009747024970>
- Salomon, G. (1993). No distribution without individual's cognition. In G. Salomon (Hrsg.), *Distributed cognitions: Psychological & educational considerations* (S. 111–138). Cambridge Univ. Press.
- Schoenfeld, A. H. (1982). Some thoughts on problem-solving research and mathematics education. In F. K. Lester & J. Garofalo (Hrsg.), *Mathematical problem solving: Issues in research* (S. 27–37). Franklin Institute Press.

- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.
- Schoenfeld, A. H. (1992). On paradigms and methods: What do you do when the ones you know don't do what you want them to? Issues in the analysis of data in the form of videotapes. *The Journal of the Learning Sciences*, 2(2), 179–214.
- Schorcht, S. & Peters, F. (2020). Mathe für Cracks: Problemlösen für mathematisch interessierte Kinder der 3. bis 5. Klasse. *Grundschulunterricht Mathematik*, 2020(3), 14–18.
- 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, 12–23.
- Schorcht, S., Baumanns, L., Buchholtz, N., Huget, J., Peters, F. & Pohl, M. (2024a). Lernt die KI nun Sehen und Zeichnen? Chancen und Herausforderungen der Bildgenerierung und Bildinterpretation mit ChatGPT für die mathematikdidaktische Forschung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 116, 22–29.
- Schorcht, S., Buchholtz, N. & Baumanns, L. (2024b). Prompt the Problem. Investigating the Mathematics Educational Quality of AI-Supported Problem Solving by Comparing Prompt Techniques. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1386075>
- Schorcht, S., Peters, F. & Kriegel, J. (2024c). Communicative AI Agents in Mathematical Task Design: A Qualitative Study of GPT Agents Network acting as a Multi-Professional Team. *Digital Experience in Mathematics Education*. <https://doi.org/10.1007/s40751-024-00161-w>
- Schwätzer, U. & Selter, C. (1998). Summen von Reihenfolgezahlen — Vorgehensweisen von Viertkläßlern bei einer arithmetisch substantiellen Aufgabenstellung. *Journal für Mathematikdidaktik*, 19(2–3), 123–148. <https://doi.org/10.1007/BF03338865>
- Schwippert, K., Kasper, D., Eickelmann, B., Goldhammer, F., Köller, O., Selter, C. & Steffensky, M. (Hrsg.). (2024). *TIMSS 2023. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Waxmann. <https://doi.org/10.31244/9783830999591>
- Spannagel, C. (2023). Hat ChatGPT eine Zukunft in der Mathematik? *Mitteilungen der Deutschen Mathematiker-Vereinigung*, 31(3), 168–172.
- Spiegel, H. (1988). Vom Nutzen des Taschenrechners im Arithmetikunterricht der Grundschule. In P. Bender (Hrsg.), *Mathematikdidaktik. Theorie und Praxis* (S. 177–189). Cornelsen.
- Stefan, G. (2012). *Motivation und Interesse im Mathematikunterricht der Grundschule: Genese – Indizierung – Förderung. Evaluation und Reflexion des Unterrichtsprojekts Sprech- und Schreibanlässe im Mathematikunterricht der dritten und vierten Jahrgangsstufe*. Dr. Kovač.
- Stein, M. (1995). Elementare Bausteine von Problemlöseprozessen. Gestaltorientierte Verhaltensweisen. *mathematica didactica*, 18, 59–84. <https://doi.org/10.1007/BF03338881>
- Stender, P. (2021). *Heuristische Strategien in der Schulmathematik: Eine Methodendidaktik*. Springer.
- Sturm, N. (2018). *Problemhaltige Textaufgaben lösen. Einfluss eines Repräsentationstrainings auf den Lösungsprozess von Drittklässlern*. Springer.
- [SWK] Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (2023). *Large Language Models und ihre Potenziale im Bildungssystem. Impulspapier der Ständigen Wissenschaftlichen Kommission (SWK) der Kultusministerkonferenz*.
- Trust, T., Whalen, J. & Mouza, C. (2023). Editorial: ChatGPT: Challenges, Opportunities, and Implications for Teacher Education. *Contemporary Issues in Technology and Teacher Education*, 23(1), 1–23.
- Tucker, S. I., Lommatsch, C. W., Moyer-Packenham, P. S., Anderson-Pence, K. L. & Symanzik, J. (2017). Kindergarten children's interactions with touchscreen mathematics virtual manipulatives: An innovative mixed methods analysis. *International Journal of Research in Education and Science*, 3(2), 646–665.
- Vollrath, H.-J. (1992). Zur Rolle des Begriffs im Problemlöseprozeß des Beweisen. *Mathematische Semesterberichte*, 39, 127–136. <https://doi.org/10.1007/BF03186465>
- Walter, D. (2018). *Nutzungsweisen bei der Verwendung von Tablet-Apps. Eine Untersuchung bei zählend rechnenden Lernenden zu Beginn des zweiten Schuljahres*. Springer.
- Wardat, Y., Tashtoush, M. A., AlAli, R. & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(7), em2286. <https://doi.org/10.29333/ejmste/13272>
- Winter, H. (1996). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Deutschen Mathematiker-Vereinigung*, 4, 35–41.
- Yimer, A. & Ellerton, N. F. (2010). A five-phase model for mathematical problem solving: Identifying synergies in pre-service-teachers' metacognitive and cognitive actions. *ZDM – Mathematics Education*, 42, 245–261. <https://doi.org/10.1007/s11858-009-0223-3>

Anschrift der Verfasser:innen

Julian Kriegel
Technische Universität Dresden
Fakultät Erziehungswissenschaften
Professur für Grundschulpädagogik/Mathematik
Münchner Straße 1
01062 Dresden
julian.kriegel@tu-dresden.de

Lukas Baumanns
Universität Duisburg-Essen
Fakultät für Mathematik
Thea-Leymann-Str. 9
45127 Essen
lukas.baumanns@uni-due.de

Anhang

Anhang A: Episoden-Kodierung

Episode	Definition
Inhaltliche bzw. inhaltstragende Episodentypen	
Analysis (Analyse)	Verstehen und Durchdringen des Problems, Auswahl von Perspektiven, Umformulierung des Problems
Exploration (Erkundung)	Generieren von Ideen, Erarbeitung von Lösungsansätzen
Planning (Planung)	Aufstellen eines Plans zur Aufgabenlösung
Implementation (Ausführung)	Abarbeiten des Plans, Kontrollen und Evaluationen des Rechenwegs
Verification (Verifikation)	Überprüfen des Ansatzes und der Lösung, Rückschau halten
Nicht-inhaltliche bzw. nicht-inhaltstragende Episodentypen	
Vorbereitung	Begrüßung und Kennenlernen, Präsentation des Instruktionsvideos
Reading (Lesen)	Lesen der Aufgabenstellung, Lesen der Aufgabenbedingungen, Zeit der Stille nach dem Lesen
Schreiben	Aufschreiben von Antwortsätzen, schriftliches Zusammenfassen der (bisherigen) Erkenntnisse
Organisation	organisatorische Tätigkeiten, Phasen, in denen aufgabenbezogen, aber nicht inhaltlich gearbeitet wird (z. B. Anlegen von Tabellen)
Abschweifung	aufgabenferne Tätigkeiten (Arbeit an einer anderen Aufgabe, Rumaßern)
Interview	Phase nach Beendigung der Aufgabe, Gespräch zwischen Interviewer:in und Schülerin bzw. Schüler
Sonstiges	Phasen, die eigene Episoden bilden, aber keinem anderen Typ zugeordnet werden können
Transition	Übergang zwischen zwei Episoden, Phase, die keiner der beiden Episoden zugeordnet werden kann oder nicht im Transkript erfassbar ist

Die Definitionen sind dem Kodiermanual von Rott (2013) sowie den Erweiterungen von Herold-Blasius (2021) entnommen.

Anhang B: Heurismen-Kodierung

Heurismus	Definition
Heuristische Hilfsmittel	
(Glg) Gleichung	Das Aufstellen von Gleichungen, in denen Zusammenhänge dargestellt werden (dies hängt häufig mit dem Transformationsprinzip zusammen).
(SyH) Systematisierungshilfe	Das Einführen ordnender Elemente, die bei der Ausführung und Überwachung einer Tätigkeit/ eines Planes helfen.
(Tab) Tabelle oder Liste	Anlegen einer Tabelle, um gegebene Werte zu ordnen und um evtl. vorhandene (funktionale) Zusammenhänge sichtbar zu machen. Oft verwendet als Hilfsmittel bei der Suche nach Mustern oder beim Systematischen Probieren.
Heuristische Strategien	
(Btr) Backtracking	Arbeiten nach dem "Versuch-Irrtum-Prinzip": Es wird solange vorwärts gearbeitet bzw. ausprobiert, bis es nicht weiter geht. Dann werden solange Schritte zurückgenommen, bis alternative Wege ausprobiert werden können. Generieren, Nennen oder Fixieren eines Beispiels.
(Bsp) Beispiel	
(Mus) Suche nach Mustern	Die Suche nach und das (vermeintliche) Erkennen von mathematischen Regelmäßigkeiten (geometrische Figuren, Zahlenfolgen) im Aufgabenkontext.
(SyP) Systematisches Probieren	Testen von mehreren Elementen (Einsetzen von Werten/ Betrachten von Fällen) mit dem Ziel, sich der Lösung anzunähern. Systematisch wird es durch Techniken, die helfen, möglichst alle Elemente zu berücksichtigen (Auflistung aller möglichen Werte bzw. Fallunterscheidung).
(RwA) Rückwärtsarbeiten	Betrachtung des Zielzustands/ des Gesuchten; davon ausgehend (evtl. durch Zerlegung des Problems in Teilprobleme) zum Anfangszustand zu gelangen. Teilweise müssen dazu vorgegebene Operationen umgekehrt werden.
(VwA) Vorwärtsarbeiten	"Drauf los"-Arbeiten vom Anfangszustand; das Gegebene wird verwendet, um zum Zielzustand zu gelangen.
Heuristische Prinzipien	
(ExF) Extremalprinzip / Extremfall	Betrachten von extremen Fällen oder Werten, die nicht unbedingt angenommen werden können müssen. Z. B. Verlegen von Punkten ins Unendliche oder Betrachten entarteter Figuren bei geometrischen Aufgaben; Einsetzen von besonders großen Werten oder Grenzwerten bei algebraischen Aufgaben.
(Tfo) Transformationsprinzip	Das Übersetzen der Aufgabe in (eine andere) "mathematische Sprache"; das Finden einer (anderen) Modellierung; oder das Erweitern der Aufgabenstruktur.
(ZeP) Zerlegungs- und Ergänzungsprinzip	Zerlegung der Aufgabe in (leichter zu lösende) Teilaufgaben. Dies können einzelne Aufgabenschritte sein oder tatsächliche (geometrische) Teilfiguren. Die Suche nach bekannten Elementen durch Zergliedern und ggf. akzentuiertes Zusammenfügen von Informationen.
Weitere Kategorien	
(Le) Erneut lesen	Nochmaliges Lesen der Aufgabe.
(RoA) Routineaufgabe	Vereinfachung des Problems auf eine Rechenaufgabe; (ggf. sinnfreier) Einsatz möglichst vieler gegebener Werte.

Die Definitionen sind dem Kodiermanual von Rott (2013) sowie den Erweiterungen von Herold-Blasius (2021) entnommen. Es wurden nur diejenigen Kategorien aufgelistet, welche bei der Datenauswertung mindestens einmal kodiert wurden.