Zusammenfassung: *KI-gestützte Lernbegleitung im Mathematikunterricht der Primarstufe birgt großes Potenzial, aber auch Risiken wie Cognitive Offloading und De-Skilling. Der Beitrag entwickelt vier Gestaltungsprinzipien: fachdidaktische Fundierung, adaptive Dosierung nach dem Prinzip „Erst du – dann ich“, metakognitive Aktivierung und unterrichtliche Einbettung („Lehrkraft in the loop“). Nur wenn KI fachdidaktisch fundiert, dosiert und in den Unterricht eingebettet ist, stärkt sie eigenständiges mathematisches Denken statt es zu verhindern.*

Die Ausgangslage

Generative Künstliche Intelligenz wird im Bildungsdiskurs mit großen Erwartungen oder mit ebenso großen Befürchtungen verbunden. Inzwischen liegt eine umfangreiche, wenn auch noch sehr heterogene empirische Basis vor, die für die Gestaltung von KI-Lernbegleitung erst Folgerungen zulässt. Meta-Analysen berichten von signifikant positiven Effekten KI-gestützter Systeme auf Lernleistungen, Motivation und höhere Denkprozesse (Wang & Fan, 2025; Deng et al., 2025; Alemdag, 2025; Wu & Yu, 2024; Zheng et al., 2023). Eine aktuelle Meta-Analyse zu generativer KI in der Mathematik zeigt moderate bis gute Effektstärken (Liu et al., 2025). Allerdings treten diese positiven Effekte unter sehr unterschiedlichen Bedingungen auf, einige Studiendesigns sind methodisch fragwürdig, und es ist – wie bei Meta-Analysen generell – die Frage zu stellen, ob nicht selektiv Studien mit besonders positiven Ergebnissen publiziert werden.

Dieser Beitrag ist deshalb ein Versuch, die ersten Erkenntnisse zu sichten, zu ordnen und hinsichtlich der Umsetzung im Mathmematikunterricht (der Grundschule) nutzbar zu machen.

Für dieses Anliegen ist ein weiterer Befund bedeutsamer als die aggregierten Effektstärken von Studien: Die weit überwiegende Mehrheit der vorliegenden Studien wurde mit älteren Schüler:innen oder Studierenden durchgeführt (Kuo et al., 2025; Liu et al., 2025; Son, 2024). Für die Grundschule spezialisierte Forschung liegt bisher nur begrenzt vor. Eine aktuelle systematische Übersicht zu intelligenten Tutorsystemen (Létourneau et al., 2025) zeigt, dass lediglich etwa 14 % der Studien Grundschüler:innen betreffen. Dies bedeutet, dass viele der folgenden Überlegungen weniger auf direkten Studienergebnissen für diese Zielgruppe basieren, sondern auf theoretischen Überlegungen und auf Befunden älterer Altersgruppen.

Grundsätzlich ist es (ebenso wie „Digitale Medien“ an sich keinen Lerneffekt haben) wenig sinnvoll, einen „Lerneffekt mit KI“ als Eigenschaft der KI-Technologie an sich misszuverstehen, anstatt nach der jeweils zugrunde liegenden Lernumgebung und didaktischen Vorstrukturierung zu fragen (Dinsmore & Fryer, 2026; KirschnerED, 2025). Die entscheidende Frage ist nicht, *ob* generative KI einen Effekt auf Bildungsprozesse haben kann, sondern *wie* dieser Effekt hergestellt werden kann. Und auch ist die Frage, wann KI besser nicht eingesetzt werden sollte.

Der vorliegende Beitrag verfolgt daher die Fragestellung: **Wie kann KI genutzt werden, um mathematische Lernprozesse in der Primarstufe durch personalisiertes Feedback zu unterstützen und unter welchen Bedingungen gelingt das?** Im Folgenden werden Gestaltungsprinzipien für KI-Lernbegleiter und KI-Tutoren entwickelt, die empirisch und theoretisch so gut wie gegenwärtig möglich abgestützt sind. Die zentrale These dabei lautet: KI-gestützte Lernbegleitung kann mathematische Lernprozesse fördern – aber nur dann, wenn sie fachdidaktisch fundiert, adaptiv dosiert und in ein hybrides Lehr-Lern-Arrangement eingebettet ist. Ohne diese Rahmung droht sie, eigenständiges Denken zu verhindern statt zu fördern.

Diese Überlegungen bilden die Grundlage für die Entwicklung und Erprobung appintegrierter KI-gestützter Lernbegleitung im Rahmen des Projektes PRIMA-KI. Sie sind ausdrücklich nicht endgültig, sondern werden sich im Zuge der dynamischen Entwicklungen bei der Entwicklung von KI-Modellen und KI-Technologien, neue Erkenntnisse aus Studien mit KI-Tutoren und den darauf basierenden pädagogischen Anwendungen weiterentwickeln.

Theoretischer Rahmen

Bevor konkrete Gestaltungsprinzipien diskutiert werden, sind zwei lerntheoretische Grundlagen für (KI-gestützte) Lernumgebungen aus meiner Sicht konstituierend und dienen als analytische Werkzeuge für das Design von KI-Tutorensystemen.

Cognitive Load Theory

Die Cognitive Load Theory (Sweller, 2020) beschreibt die begrenzte Kapazität des Arbeitsgedächtnisses als zentrale Rahmenbedingung für die Gestaltung von Lernmedien. Für KI-gestützte Systeme hat diese Begrenzung unmittelbare Konsequenzen: Generative KI kann Informationen in hoher Geschwindigkeit und

großem Umfang präsentieren – Texte, Bilder, Erklärungen, Visualisierungen, Animationen. Doch wenn Lernende mit Informationen und Reizen überflutet werden, ohne diese lernrelevant verarbeiten zu können, verfehlt auch die inhaltlich beste Information ihr Ziel.

Dies optimal zu gestalten ist eine zentrale Herausforderung des Designs von KI-Lernbegleitung. Scaffolding – also die dosierte Bereitstellung von Hilfen und Erklärungen – ist aus Sicht der Cognitive Load Theory hier entscheidend. KI-Tutoren müssen Impulse, Hilfen und Unterstützung so dosieren, dass die Arbeitsgedächtnisbelastung im optimalen Bereich bleibt (Cosentino et al., 2025). Zu viel Information zum falschen Zeitpunkt ist nicht neutral, sondern schadet dem Lernen. Die Konsequenz für das Systemdesign ist klar: Die Fähigkeit, die passenden Informationen zu *produzieren*, ist für einen KI-Tutor nicht alleine wichtig. Fast noch wichtiger ist die Fähigkeit, Informationen *zurückzuhalten* und *dosiert* anzubieten.

Zone der nächsten Entwicklung: Adaptivität erfordert diagnostische Kompetenz

Die Zone der nächsten Entwicklung (Vygotsky) beschreibt den Bereich zwischen dem, was ein Kind allein bewältigen kann, und dem, was es mit Hilfe eines kompetenteren Partners bewältigen kann. Für das Design von KI-Tutorsystemen ist dieses Konzept von fundamentaler Bedeutung: Es muss also „wissen“, wo ein Kind aktuell steht, und dann genau in dieser nächsten Zone Impulse anbieten – nicht zu leicht (Unterforderung), nicht zu schwer (Überforderung).

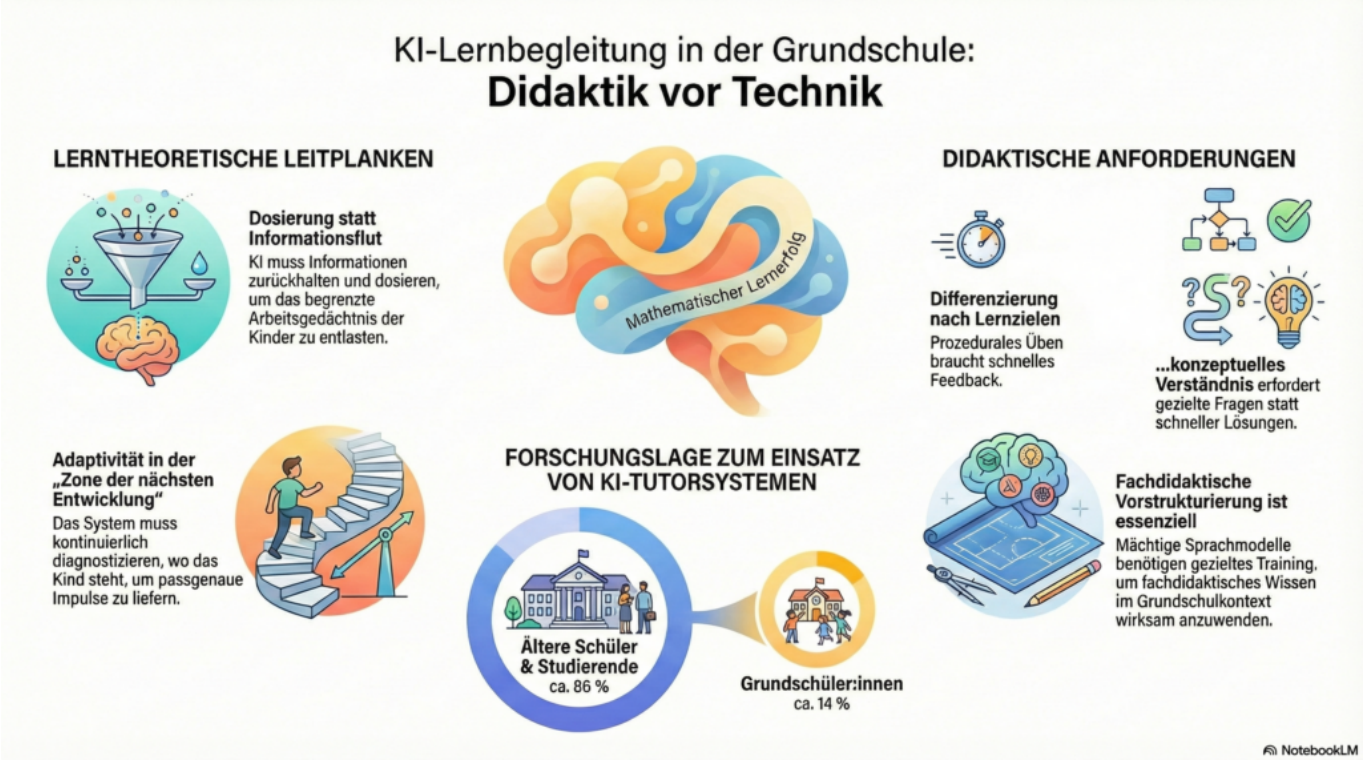
Das klingt selbstverständlich, ist aber technisch und fachdidaktisch anspruchsvoll. Adaptive Systeme müssen diese Einschätzung nicht einmalig, sondern kontinuierlich leisten (Wu et al., 2025; Kuo et al., 2025). Sie müssen aus den Eingaben der Lernenden ableiten, wo deren Verständnis liegt, welche Fehler auf welche Vorstellungen zurückgehen, welche Art von Missverständnis zum Fehler geführt hat und welcher nächste Impuls produktiv wäre. Adaptivität ist damit die Grundbedingung dafür, dass ein KI-System überhaupt lernwirksam agieren kann und unterscheidet es von rein deterministisch agierenden tutoriellen Systemen, die auf Basis einfacher Regeln mehr oder weniger passende bzw. oberflächliche Rückmeldungen erzeugen.

Prozedurales und konzeptuelles Wissen: Verschiedene Ziele, verschiedene Unterstützung

Aus diesen beiden theoretischen Perspektiven ergibt sich eine weitere Differenzierung, die für das Design von KI-Tutoren entscheidend ist: Die Art der angemessenen Unterstützung hängt fundamental davon ab, welches Lernziel gerade verfolgt wird. Prozedurales Wissen – *Wie löse ich diese Aufgabe?* – braucht andere Unterstützung als konzeptuelles Wissen – *Warum funktioniert das so?*

Ein Kind, das Rechenflüssigkeit übt, etwa die schnelle Addition oder Subtraktion, benötigt häufiges, passgenaues Feedback zu seinen Ergebnissen. Ein Kind, das gerade Zahlverständnis entwickelt, muss hingegen durch gezielte Fragen und Handlungsmöglichkeiten dazu angeregt werden, sein Verständnis selbst zu konstruieren und weiterzuentwickeln. Es sollte nicht nur den Lösungsweg (idealerweise gibt es mehrere Lösungswege) prüfen. Ein KI-System, das diese Unterscheidung nicht treffen kann, riskiert, dass seine Unterstützung kontraproduktiv wirkt, etwa indem es bei konzeptuellem Lernen zu schnell Lösungen liefert oder bei prozeduralem Üben mit unnötigen Impulsen und Anregungen das Arbeitsgedächtnis belastet. Makransky et al. (2024) zeigen dies auch: Ein speziell didaktisch hinsichtlich dieser Prinzipien trainiertes Modell förderte in der Studie konzeptuelles Verständnis, Vertrauen und Freude am Lernen deutlich mehr als ein generisches großes Sprachmodell, und diese Effekte blieben im Follow-up stabil. Dies bedeutet, dass auch große, mächtige Sprachmodell nicht unbedingt fachdidaktisches Wissen automatisch gezielt anwenden können ohne fachliche Vorstrukturierung.

Diese drei theoretischen Perspektiven – die Begrenzung des Arbeitsgedächtnisses, die Notwendigkeit adaptiver Positionierung in der Zone der nächsten Entwicklung und die Differenzierung nach Lernzielen – bilden einen Rahmen, in dem die folgenden Gestaltungsprinzipien verortet sind. Zugleich machen sie deutlich, warum KI-Lernbegleitung kein einfaches technisches Problem ist: Sie erfordert Systeme, die nicht nur antworten können, sondern die über ein Modell des Lernenden, des Lerngegenstands und des Lernprozesses verfügen.



Vom Lernwerkzeug zur Denkprothese

Die bisherige Darstellung der Studienlage könnte den Eindruck erwecken, dass KI-gestützte Lernbegleitung ein überwiegend positives Potenzial besitzt, das lediglich noch technisch optimiert werden muss. Doch es zeigt sich ein ambivalenteres Bild. In einigen Studien zeigte sich, dass der unkontrollierte und „falsche“ Einsatz generativer KI beim Lernen nicht nur weniger wirksam sein kann als erhofft, sondern dass er grundlegende kognitive Fähigkeiten abkürzen kann und Lernen dadurch eher hindert.

Cognitive Offloading: Wenn Unterstützung zur Abhängigkeit wird

Das grundlegendste Problem lässt sich unter dem Begriff des *Cognitive Offloading* fassen: Lernende lagern kognitive Anstrengung an die KI aus, statt sie selbst zu vollziehen. Gerlich (2025) dokumentiert in einer aktuellen Analyse eine bemerkenswert hohe negative Korrelation ($r = -0{,}75$) zwischen intensiver KI-Nutzung und der Ausprägung kritischen Denkens, wobei dieser Zusammenhang bei jüngeren Lernenden besonders ausgeprägt ist. Diese Korrelation ist

selbstverständlich kein genereller Beleg für einen Schaden durch KI (zumal die Studie auch methodisch kritisiert werden kann), aber sie weißt auf ein mögliches Problem hin: Wer systematisch Denkaufwand delegiert, verliert nachhaltig die Fähigkeit zur eigenständigen Problemlösefähigkeit.

Dieses Risiko zeigten Bastani et al. (2025) in einer kontrollierten Interventionsstudie. Schüler:innen, die während des Übens unregulierten Zugang zu vollständigen KI-Lösungen erhielten, erzielten kurzfristig höhere Leistungen. Wurde die KI-Unterstützung jedoch entzogen, fielen ihre Ergebnisse signifikant unter das Niveau einer Kontrollgruppe ohne jegliche KI-Nutzung. Die Autor:innen sprechen von einem *De-Skilling*-Effekt: Die vermeintliche Unterstützung hatte den Kompetenzaufbau nicht gefördert, sondern aktiv behindert. Parallel dazu dokumentieren Lernanalysen aus KI-basierten Nachhilfesystemen, dass ein Teil der Lernenden versucht, Aufgabensequenzen „durchzuklicken“, ohne sich inhaltlich mit ihnen auseinanderzusetzen (Jančařík et al., 2023).

Auch Dinsmore und Fryer (2026) sowie Gisiger (2025) argumentieren aus lernpsychologischer Perspektive, dass generative KI (hier gemeint in Form von Chatbots, die nicht für Lernbegleitung vorstrukturiert sind etwas durch spezielle Systemprompts) dazu verführt, den anstrengenden Prozess der eigenen Wissenskonstruktion abzukürzen. Wer direkt die Lösung konsumiert, spart Anstrengung, baut aber nicht unbedingt tragfähigen Wissensstrukturen auf, vor allem wenn er die eingesparte Denkenergie nicht für höhere Denkleistungen sinnvoll nutzt. Man könnte auch sagen: Lernerfolgt hängt mit Anstrengung zusammen und wenn diese Anstregung vermieden wird, gibt es auch weniger nachhaltiges Lernen. Viele Anbieter von Chatbots haben bereits darauf reagiert und bieten mittlerweile einen „Lernmodus“ an, bei dem der Chatbot keine Lösungen direkt ausgibt, sondern Rückfragen stellt und dialogisch mit dem Benutzer die Inhalte betrachtet.

Die Höflichkeitsfalle

Zu diesem strukturellen Risiko tritt ein subtileres, aber bedeutsames Phänomen. Abdulsalam und Aroyehun (2025) zeigen in ihrer Analyse, dass große Sprachmodelle beim Tutoring zwar Expert:innen-Niveau erreichen können, dabei aber von sich aus zu höflich und zu unterstützend agieren. Diese übermäßige Höflichkeit korrelierte in der Studie negativ mit der Lernqualität. Die KI vermeidet es, Lernende zu frustrieren oder auf Fehler hinzuweisen, und weicht daher von sich aus echten Herausforderungen aus. Doch produktive Anstrengung – das temporäre

Erleben von Schwierigkeit bei gleichzeitiger Unterstützung – ist ein zentraler Motor für Lernen, insbesondere für die Entwicklung von konzeptuellen Verständnis.

Chudziak und Kostka (2025) identifizieren ein verwandtes Problem: Viele aktuelle KI-Systeme tendieren zu einem *preskriptiven* Interaktionsstil – sie geben vor, leiten an und lösen, statt Raum für eigene Denkprozesse zu lassen. Systeme, die zu früh intervenieren (*Reactive Feedback*), riskieren damit, genau jene (meta-)kognitiven Prozesse zu unterbinden, die sie eigentlich fördern sollten. Die Analyse realer Nachhilfedialoge durch Wang et al. (2025) bestätigt dieses Bild auch für jüngere Lernende: Primarstufenschüler:innen reagierten durchaus positiv auf interaktive Fragen, blieben aber tendenziell eher passiv-reaktiv und zeigten keinen nachhaltigen Lerneffekt, wenn der Tutor zu monologisch und lösungszentriert agierte.

Das Paradoxon optimaler Unterstützung

Diese Befunde verdichten sich zu einem Paradoxon, das als Leitproblem für die Gestaltung von KI-Lernbegleitung verstanden werden kann: Je effizienter ein KI-System Aufgaben lösen kann und Hilfe anbietet, desto größer ist die Gefahr, dass es das eigenständige Denken ersetzt, statt es zu fördern. Dieses Paradoxon ist schwer aufzulösen. Es bedeutet, dass ein guter KI-Tutor manchmal bewusst *weniger* tun muss, als er könnte. Er muss Frustration („being stuck“ als Kern mathematischen Denkens, Mason, Burton & Stancey, 1982) in dosierten Maßen zulassen, Lösungen zurückhalten, unvollständige Hinweise geben und Wartezeit erzeugen. Dies bedeutet, das das KI-System teilweise gegen die Mechanismen arbeiten muss, die generative KI so eindrucksvoll macht.

Damit wird deutlich, warum ein unvorstrukturierter Chatbot – etwa ein frei zugängliches Sprachmodell als Chatbot ohne pädagogische Rahmung – für das schulische Lernen potenziell sogar schädlich sein kann. Er wird zur Antwortmaschine, die echtes, verständnisförderndes Lernen verhindern statt ermöglichen kann. Die zentrale Frage ist daher nicht, ob generative KI *eingesetzt* werden soll, sondern wie sie so *gestaltet* werden kann, dass sie das Denken der Kinder stärkt statt schwächt. Die folgenden Gestaltungsprinzipien sollen erste Antwort auf genau dieses Kernrisiko geben.

Gestaltungsprinzipien als Antwort auf das Kernrisiko

Die Spannung zwischen dem Potenzial adaptiver KI-Lernbegleitung und den

beschriebenen Risiken des Cognitive Offloading lässt sich nicht durch einen einzelnen Designparameter auflösen. Sie erfordert vielmehr ein Zusammenspiel mehrerer Gestaltungsebenen, die aufeinander aufbauen: eine fachdidaktische Fundierung des Systems als zentrale Voraussetzung für Diagnostik und Förderung (Primat der Fachdidaktik), die adaptive Dosierung als Kernmechanismus des Feedbacks, die metakognitive Aktivierung als Qualitätsmerkmal und hybride Einbettung („Lehrkraft in the loop“) als Wirksamkeitsbedingung.

Fundierung: Fachdidaktisches Wissen als zentrale Voraussetzung

Die erste und grundlegendste Gestaltungsebene betrifft die Wissensbasis des Systems. Generative KI erzeugt sprachlich und formal beeindruckende Ausgaben, aber fachdidaktisch sind diese häufig oberflächlich, wenig problemorientiert oder sogar fehlerhaft, wenn sie nicht gezielt gesteuert werden (Schneider, 2025). Das ist keine Implementierungsschwäche, die sich beiläufig beheben ließe; es ist eine Konsequenz der Architektur großer Sprachmodelle. Deren Training basiert auf den verfügbaren Daten im Internet, und insbesondere für mathematikdidaktische Fragestellungen fand dabei in der Regel keine qualitative Kuratierung statt.

Literaturanalysen zum Einsatz generativer KI in der Mathematikbildung weisen daher immer wieder auf die geringe theoretische Fundierung und Fehleranfälligkeit vieler Systeme hin (Almheiri et al., 2025; Awang et al., 2024; Holmes & Tuomi, 2022; Pesemowo & Adewuyi, 2024).

Cárdenas et al. (2025) identifizieren in ihrer systematischen Analyse die fehlende theoretische Rahmung als eines der zentralen Hindernisse für lerneffektive KI-Tutorsysteme. Dies betrifft nicht nur die inhaltliche Korrektheit von Aufgaben und Erklärungen (hier werden die Systeme immer besser), sondern die Art und Weise, wie ein System auf Lernende reagiert. Denn Fachexpertise und Tutoring-Expertise sind nicht dasselbe. Macina et al. (2025) zeigen empirisch, dass Personen mit hervorragenden mathematischen Fähigkeiten nicht automatisch gutes Tutoring betreiben. Sie können Aufgaben korrekt lösen, erkennen aber nicht immer, wo die Schwierigkeit für Lernende liegt, welche Fehlvorstellung einem Fehler zugrunde liegt oder welcher Impuls in einer bestimmten Lernsituation produktiv wäre. Für KI-Systeme bedeutet das: Fachliches Wissen in den Prompt zu integrieren, reicht nicht aus. Das System muss auch über fachdidaktisches Wissen verfügen, also über das Wissen, *wie* Kinder mathematische Konzepte aufbauen, welche typischen Irrwege und Missverstehensprozesse es gibt und welche Interventionen an welcher Stelle im Lernprozess wirksam sind.

Dass dieser Unterschied praktisch bedeutsam ist, belegen mehrere konvergierende Befunde. Makransky et al. (2024) zeigen auf Basis der Generative Learning Theory, dass ein speziell didaktisch trainiertes Modell konzeptuelles Verständnis, Vertrauen und Freude am Lernen signifikant stärker fördert als ein generisches großes Sprachmodell, und dass diese Effekte auch im Follow-up stabil bleiben. Erfolgreiche Tutorsysteme wie ChatTutor oder spezifische Frameworks (z.B. der Lernmodus von ChatGPT) basieren daher explizit auf pädagogischen Theorien wie der Social Cognitive Theory oder dem Evidence-Centered Design (Cohn et al., 2025; Dwivedi & Rejina, 2025). Und Studien zu GeoGebra- und KI-gestützten Lernumgebungen demonstrieren, dass konzeptuelles Verständnis und Selbstwirksamkeit nur dann steigen, wenn Fach- und Fachdidaktik explizit in die Systemgestaltung einfließen, wenn also nicht bloß „Technik on top“ gesetzt wird (Cononigo, 2024).

Was bedeutet das konkret? Ein KI-Lernbegleiter für den Mathematikunterricht der Primarstufe benötigt mindestens folgende fachdidaktische Wissensgrundlagen, die entweder über den Systemkontext (Prompting) oder durch ein spezialisiertes Modelltraining bereitgestellt werden müssen:

Erstens ein *Modell der mathematischen Kompetenzentwicklung* für den jeweiligen Inhaltsbereich. Die KI muss „wissen“, dass Zahlverständnis nicht durch Auswendiglernen von Fakten entsteht, sondern durch den Aufbau von Vorstellungen und dass es für diesen Aufbau typische Entwicklungsverläufe und Vorläuferfähigkeiten gibt. Ohne dieses Wissen bleibt jede Adaptation oberflächlich: Das System kann allenfalls den Schwierigkeitsgrad variieren, aber nicht die Qualität seiner Impulse an den Verständnisstand anpassen.

Zweitens *Fehler- und Strategietaxonomien*, die typische Schülerwege und -irrtümer abbilden (Nauryzbayev et al., 2023; Bewersdorff et al., 2023). Die Diagnose von Fehlvorstellungen, wie etwa die Verwechslung von Position und Wert im Stellenwertsystem oder das zählende Rechnen als verfestigte Strategie, ist eine fachdidaktische Kernkompetenz, die ein System erwerben muss, wenn es über die Feststellung „richtig/falsch“ hinausgehen soll. Ein Beispiel: Häufig schlagen auch große Sprachmodelle bei Fehlern explizit die Strategie „zähle doch“ vor, wenn keine entsprechende Anweisung und Vorwissen gegeben wurde.

Drittens bedarf es *expliziter Regeln und Vorgaben*, welche Hilfen in welcher Lernphase und bei welcher Art von Schwierigkeit sinnvoll sind. Die oben eingeführte Unterscheidung zwischen prozeduralem und konzeptuellem Lernen kann beispielsweise eine Leitlinie sein: Routineförderung erfordert andere Interventionsstrategien als der Aufbau begrifflichen Verständnisses. Und ein System sollte erkennen, ob ein Fehler ein „Flüchtigkeitsfehler“ ist oder ein Verständnisproblem zugrunde liegt. Einbezogen werden kann hier eine Lernhistorie, damit die generative KI auf vorherige Eingaben und Lernstände angepasst antworten kann und Dauerschleifen vermieden werden.

Viertens muss das System eine Adaption nicht nur hinsichtlich des Schwierigkeitsgrades leisten, sondern auch hinsichtlich *unterschiedlicher Denkwege*, mit denen Lernende zum Ergebnis gelangen können. Kinder lösen mathematische Probleme auf vielfältige Weise, und das ist auch wünschenswert. Ein fachdidaktisch fundierter KI-Tutor erkennt alternative Strategien, bewertet sie hinsichtlich ihrer Tragfähigkeit, fragt ggf. nach und kann Kinder gegebenenfalls darin bestärken, oder behutsam zu effizienteren Wegen anregen.

Die Bereitstellung dieses fachdidaktischen Wissens ist aufwändig, hat aber noch einen weiteren großen Vorteil: Sie ermöglicht den Einsatz kleinerer, spezialisierter Modelle, die weniger Rechenleistung benötigen und lokal auf dem Gerät oder in datenschutzkonformen Infrastrukturen betrieben werden können. Damit lassen sich

zugleich Nachhaltigkeitsprobleme (Energieverbrauch) und Datenschutzanforderungen adressieren, ohne Qualitätseinbußen hinnehmen zu müssen.

Fachdidaktische Fundierung ist also nicht nur ein wünschenswertes Qualitätsmerkmal, sondern die notwendige Bedingung dafür, dass adaptive Unterstützung überhaupt als solche funktionieren kann. Ohne sie bleibt ein KI-Tutor – um Schneiders (2025) pointierte Formulierung aufzugreifen – ein „sprachbegabter Zufallsgenerator“, der gelegentlich hilfreiche, häufig aber unspezifische oder sogar irreführende Impulse produziert.

Dosierung: Adaptives Scaffolding und das Prinzip „Erst du – dann ich“

Auf der Grundlage fachdidaktischen Wissens wird die zweite Gestaltungsebene möglich: die adaptive Dosierung von Impulsen und Hilfen. Dies ist der Mechanismus, der KI-Lernbegleitung von einer bloßen Antwortmaschine unterscheidet. Die empirische Befundlage deutet darauf hin, dass KI-gestützte Lernumgebungen besonders dann Potenzial für Lern- und Verstehensprozesse erzeugen, wenn sie adaptiv auf Lernstand, Fehler und Strategien reagieren, statt richtige Antworten auszugeben.

In der Studie zu TALPer etwa profitierten leistungsschwächere Fünftklässler:innen besonders stark von adaptiver Unterstützung, während leistungsstärkere Schüler:innen komplexere Interaktionsmuster mit dem KI-Lernbegleiter entwickelten (Kuo et al., 2025). Mit einem System konnten also unterschiedliche Lernvoraussetzungen gut bedient werden. Liu et al. (2025) zeigen bei einem vergleichbaren System signifikante Leistungsgewinne in Textaufgaben, wobei interessanterweise die wahrgenommene Qualität der Unterstützung und nicht deren bloße Verfügbarkeit den stärksten Einfluss auf Motivation und Lernwirkung hatte. Die systematische Übersicht von Son (2024) bestätigt positive Effekte gut konzipierter intelligenter Tutorsysteme auf mathematische Lernleistungen, insbesondere wenn diese adaptiv auf individuelle Lernbedürfnisse reagieren. Die Stärke generativer KI für Feedback liegt also gerade in der Individualisierung: in der Anpassung der Antwort an die konkreten Lernbedürfnisse des einzelnen Kindes.

Entscheidend ist dabei die Reihenfolge von Denken und Unterstützung. Die bisherige Forschung zu großen Sprachmodellen beim Lernen weist deutlich darauf hin, dass KI-Feedback besonders dann wirksam ist, wenn Lernende zunächst eigene

Lösungsversuche unternehmen (Kumar et al., 2023). In deren Studie führten selbst seltene fehlerhafte Erklärungen des Sprachmodells nach vorherigen eigenen Lösungsversuchen noch zu Lernzuwächsen, ohne dass Teilnehmende systematisch falsche Strategien übernahmen. Dies unterstreicht die Robustheit des „Erst denken, dann Feedback“-Prinzips. Cohn et al. (2025) betonen in ihrem theoretischen Rahmenwerk für LLM-basierte pädagogische Agenten explizit die Notwendigkeit von *guided discovery over direct answers* – also angeleitetes Entdecken statt direkter Antworten. Und auch Ruan et al. (2020) zeigen in ihrer Studie zu narrativbasierten Chatbot-Tutoren: Lernzuwächse wurden vor allem dann erzielt, wenn das System interaktives Feedback und Hinweise gab statt direkter Lösungen.

Im Bereich des selbstregulierten Lernens weisen einige Studien darauf hin, dass adaptives, KI-gestütztes Scaffolding im Unterschied zu statischen Hilfesequenzen die Qualität von Selbststeuerungsprozessen beim Lernen verbessern kann und Vorteile gegenüber einem Konzept „gleiche Hilfen für alle“ hat (Liu et al., 2025; Wu et al., 2025). Generative KI muss also so eingebunden und vorstrukturiert werden, dass sie adaptiv und individualisiert auf die Eingaben der Lernenden eingeht und nicht nur vorgefertigte Impulse anbietet.

Zugleich zeigen die in Abschnitt 3 diskutierten Überlegungen, dass Hilfen nicht nur individualisiert, sondern auch *begrenzt* sein müssen. Bastani et al. (2025) demonstrieren, dass unregulierter Zugang zu vollständigen Lösungen langfristig schadet, es sei denn, die KI wird so reguliert, dass sie nur schrittweise Hinweise gibt und damit Raum für eigene Lösungsversuche lässt. Ein konsequentes *Fading* – also das planmäßige Zurückziehen von Unterstützung – ist daher ein notwendiger Bestandteil des Systemdesigns. Wenn Schüler:innen zu viele Hinweise anfordern, muss das System darauf reagieren können, indem es Hilfen dosiert, zeitlich staffelt und gegebenenfalls reduziert, um Abhängigkeit zu vermeiden.

Ein dabei immer wieder beobachtetes Problem sind Halluzinationen und nicht geeignete Rückmeldungen der KI. Neue Multi-Agent-Ansätze und LLM-as-Judge-Verfahren zeigen jedoch, dass sich Qualität und Zuverlässigkeit von Scaffolds durch Selbstüberprüfungsverfahren verbessern und Halluzinationen beim Feedback deutlich reduzieren lassen (Cohn et al., 2025; Gonnermann-Müller et al., 2025; Qian et al., 2026). Auch zeigen neuere KI-Modelle zunehmend weniger Halluzinationen und Fehler, sodass dieses Problem perspektivisch an Relevanz verlieren dürfte.

Zusammengefasst lässt sich aus diesen Befunden als Designprinzip ableiten, dass ein KI-Lernbegleiter konsequent nach dem Prinzip „Erst du – dann ich“ arbeiten

sollte. Zunächst wird ein eigener Lösungsversuch eingefordert. Die KI kommt erst bei Bedarf hinzu – auf Anfrage oder automatisch bei Fehlern – und fragt dann nach Ideen, Teilstrategien und Beobachtungen. Erklärungen werden an vorhandene Verständnisgrundlagen angedockt. Vollständige Musterlösungen bleiben die Ausnahme und werden als Reflexionswerkzeug genutzt, nicht als primäres Lernformat. Und der KI-Lernbegleiter sollte dabei nicht übermäßig loben oder ausweichen, sondern konstruktive und informative Rückmeldung geben.

Aktivierung: Metakognition, kritisches Denken und der Tutor als Spiegel

Die dritte Gestaltungsebene geht über die Dosierung von Hilfen hinaus und zielt auf die Qualität der kognitiven Prozesse, die durch die Interaktion angeregt werden. Chatbots und KI-Systeme können nicht nur als Hinweisgeber und Aufgabenerklärer fungieren. Sie können darüber hinaus Prozesse des Planens, Überwachens und Reflektierens im Problemlöseprozess anregen, wenn sie darauf ausgelegt werden. Die Meta-Analyse von Wu et al. (2025) zeigt, dass Chatbots selbstreguliertes Lernen technisch, sozial und reflexiv unterstützen können, sofern ihre Scaffolds an Modelle des selbstregulierten Lernens gekoppelt sind. Guo et al. (2025) bestätigen in ihrer systematischen Übersicht, dass KI-Systeme grundlegende psychologische Bedürfnisse nach Autonomie, Kompetenz und sozialer Eingebundenheit erfüllen können als zentrale Faktoren für Motivation und Engagement.

Besonders interessant sind in diesem Zusammenhang Studien zu sogenannten *teachable agents*. Song et al. (2024) zeigen, dass Lernende KI-gestützte Systeme als Lernbegleiter, Moderator und kollaborativen Problemlöser wahrnehmen können, wenn sie ihre eigenen Erklärungen an diese Agenten richten. Das Prinzip des Lernens durch Erklären lässt sich damit in KI-gestützte Umgebungen übertragen und für das Lernen nutzen. Gleichzeitig zeigt sich allerdings, dass viele heutige generative KI-Systeme zentrale tutorielle Rollen – beispielsweise das gezielte Anregen von Planung, Strategieauswahl und Reflexion – ohne spezielle Prompts oder Nachtraining noch nicht zuverlässig erfüllen und eher zu dem bereits diskutierten preskriptiven Stil neigen (Chudziak & Kostka, 2025; Contel & Cusi, 2025). Dies unterstreicht die Notwendigkeit der Vorstrukturierung von KI-Systemen, damit sie proaktiv metakognitive Scaffolds einsetzen, die über reaktives Feedback hinausgehen.

Ein wesentlicher Aspekt metakognitiver Aktivierung besteht auch darin, Lernende

zur kritischen Prüfung von KI-Antworten selbst zu befähigen. In einer Lernumgebung im Rahmen eines Mathetags an einer Grundschule wurde beobachtet, dass Kinder, die die Erfahrung machten, dass KI-Antworten fehlerhaft sein können, eine kritisch-prüfende Haltung entwickelten und Antworten nicht mehr unreflektiert akzeptierten (Helal et al., 2024). Dies ist eine Fähigkeit, die Kinder bereits in der Primarstufe üben können und die angesichts der zunehmenden Durchdringung des Alltags mit KI-generierten Inhalten von wachsender Bedeutung ist.

Ein KI-Lernbegleiter kann als metakognitiver Partner konzipiert werden, der diese Prozesse systematisch anregt. Die Tutorin stellt Fragen wie „Was ist dir aufgefallen?“, „Welche Strategie hast du probiert?“, „Warum glaubst du, dass das funktioniert?“. Sie regt die Reflexion über Fehler an: „Welche Idee von vorher könnte hier helfen?“ Und sie kann, wo sinnvoll, Teach-back-Elemente integrieren: Kinder erklären der KI, was sie verstanden haben, und die KI spiegelt zurück, fragt nach und vertieft. Dadurch wird die KI nicht zum allwissenden Erklärer, sondern zum Katalysator für das eigene Denken.

Einbettung: Hybride Arrangements als Wirksamkeitsbedingung

Die vierte und letzte Gestaltungsebene betrifft die Frage, in welchem Rahmen KI-Lernbegleitung stattfindet. Hier deuten Forschungsergebnisse darauf hin, dass die effektivsten Szenarien jene sind, in denen KI die Lehrkraft nicht ersetzt, sondern entlastet wird. Dies kann etwa durch passgenaue Scaffolds im Problemlöseprozess oder Echtzeit-Analysen geschehen, damit Lehrkräfte mehr Zeit für pädagogische Interaktion und Beziehungsgestaltung hat (Wezendonk & Veldhuis, 2024; Gonnermann-Müller et al., 2025).

Eine Studie von Eedi in Kooperation mit Google DeepMind (2025) illustriert das Potenzial hybrider Ansätze unmittelbar: Während KI-Unterstützung allein zu Lernzuwächsen von 4,5 Prozentpunkten führte, verdoppelten sich diese auf 10 Prozentpunkte, wenn die Lehrkraft die KI-Vorschläge überprüfte und nutzte. Die Lehrkraft fungiert hier nicht allein als Kontrollorgan, sondern als pädagogische Instanz, die die KI-Impulse in den Kontext des Unterrichts, der Klasse und des einzelnen Kindes einordnet. In der Meta-Analyse von Kaliisa et al. (2025) zeigte sich, dass KI-Feedback ist nicht weniger effektiv als menschliches Feedback, aber auch nicht systematisch überlegen. Besonders vielversprechend sind also Hybridansätze, die verlässliches, direktes und barrierearmes KI-Feedback mit gerahmtem Feedback durch menschliche Lehrkräfte kombinieren. Cosentino et al. (2025) bestätigen, dass solche hybriden Feedback-Modelle das Potenzial haben, die kognitive Belastung zu

reduzieren und differenzierte Informationsverarbeitungsstrategien zu unterstützen.

Weitere Studien stützten die Wirksamkeit bei der Unterstützung von Lehrkräften. Der CoPilot beispielsweise unterstützt menschliche Lehrkräfte in Echtzeit und führt zu signifikant besseren mathematischen Lernergebnissen bei jüngeren Schüler:innen. Dies gilt besonders dort, wo Lehrkräfte sonst schwächere Feedbackqualität zeigen, etwa aufgrund fachfremden Unterrichts (Wang et al., 2024). Kestin et al. (2025) stellen fest, dass KI-Tutoren, die auf didaktischen Prinzipien aktiven Lernens aufbauen, einen sehr guten Präsenzunterricht in bestimmten Phasen bereichern und unterstützen können, ohne ihn zu ersetzen.

Aus der Perspektive der hybriden Einbettung ergeben sich mehrere Konsequenzen für das Systemdesign. Die Generierung von Feedback muss auf Basis fachdidaktisch fundierter Informationen erfolgen und darf nicht dem Zufall überlassen werden. Lehrkräfte benötigen Dashboards und Einstellungsmöglichkeiten, um zu sehen, wo Schüler:innen stehen, welche Hilfen die KI gegeben hat und wie diese Hilfen gewirkt haben. Die KI kann Vorschläge für Aufgaben, Hilfen oder erste Diagnosen machen – die Entscheidung bleibt beim Menschen. Besonders aussichtsreich ist dabei der *mikrodidaktische* Bereich des Unterstützens auf Aufgabenebene beim Problemlösen, in dem Lehrkräfte bei einer heterogenen Klasse oft nicht alle Kinder an den entscheidenden Stellen zeitnah unterstützen können. In dem systematischen Review von Eti, Mosia & Egara (2026) zeigte sich, dass KI besonders wirksam ist, wenn sie Fehlvorstellungen erkennt, passende Hinweise gibt und Lernende schrittweise zu selbstständigem Problemlösen führt.

Allerdings bleibt die sinnvolle Einbindung von Lehrkräften in Design und Einsatz von KI-Tutorsystemen bisher oft unzureichend mitgedacht. Guerino et al. (2023) sowie Wezendonk und Veldhuis (2024) betonen, dass lehrerzentrierte Gestaltungsansätze und entsprechende KI-Literacy-Programme notwendig sind, um praktische Classroom-Integration und Akzeptanz zu gewährleisten. Professionalisierung und Fortbildung zur Einbindung, Orchestrierung und Nutzung von KI sind Voraussetzung für einen verantwortlichen Einsatz (Holmes et al., 2018; KMK, 2024; Wang & Nie, 2023) und sollten daher in der Lehrkräfteaus- und -weiterbildung integriert werden. Dies sollte nicht nur auf Wissen über die Technik hinauslaufen, sondern insbesondere pädagogische Aspekte umfassen, die sowohl Chancen als auch Risiken thematisieren. Zentral bleibt dabei das fachdidaktische Wissen als Grundlage für die Prüfung und Orchestrierung der KI als Assistenz im Lernsetting.

KI als Impulsgeber für mathematische Entdeckungen

Die bisherigen Gestaltungsprinzipien beschreiben, wie KI-Lernbegleitung *intern* funktionieren sollte, also in der Interaktion zwischen System und Lernendem. Ebenso wichtig ist jedoch die Frage, wie sich diese Interaktion in das *Gesamtarrangement* mathematischen Lernens einfügt. KI soll und darf nicht dazu führen, dass Kinder nur noch auf Bildschirme starren und mit ihnen interagieren. Forschung zu *Tangible Interfaces* und sozialen Robotern zeigt, dass KI auch Interaktion in der physischen Welt anregen kann (Ligthart et al., 2023). KI sollte als Anstoßgeber für mathematische Aktivitäten dienen und dabei die Repräsentationsebenen nach Bruner nutzen.

Dies kann beispielsweise bedeuten, dass Handlungen mit konkretem Material vorgenommen werden oder Skizzen und Zeichnungen mit genutzt werden, die KI diese erkennt und analytische Fragen dazu stellt bei Bedarf. Beispielsweise bei der Modellierung von Sachaufgaben in der App „Rechengeschichten„. Hier können neben der Spracheingabe auch Skizzen, Notizen und Fotos der Modellierung mit dem KI-Lernbegleiter besprochen werden. Für den Primarbereich liegen hier deutlich weniger Studien als für Sekundarstufe und Hochschule vor, aber die vorhandenen Ergebnisse machen vorsichtig optimistisch: Kinder können von generativer KI beim Lernen profitieren, wenn Lernumgebungen sinnvoll gestaltet sind – auch hinsichtlich der Verknüpfung digitaler Begleitung und analoger Lernumgebung (Hwang, 2022; Listyaningrum et al., 2024; Mott et al., 2023; Rumbelow & Coles, 2024; Yim & Su, 2025).

KI-gestützte Objekterkennung, etwa bei Cuisenaire-Stäben oder die Erkennung von Zeichnungen oder Notizen, kann dazu beitragen, dass Kinder ihr Handeln zunehmend mit abstrakten mathematischen Repräsentationen verknüpfen (Rumbelow & Coles, 2024). KI-gestütztes Üben kann Rechenflüssigkeit gezielt erhöhen und erzielt dabei größere Flüssigkeitsgewinne als reine Memorierungsansätze, muss aber bei Kindern mit Rechenschwierigkeiten sorgfältig mit anderen Übungsformen kombiniert werden (Samuelsson, 2023). Adaptive Systeme für Kinder mit Dyskalkulie zeigen vielversprechende Ergebnisse bei der Aufrechterhaltung von Motivation und Engagement (Hocine et al., 2023; Holmes, 2024). Und narrative sowie gamifizierte Ansätze verdienen besondere Beachtung: Ruan et al. (2020) zeigen, dass storytelling-basierte Chatbot-Tutoren Engagement und Lerngewinne fördern können, und Sayed et al. (2022) bestätigen signifikante

Verbesserungen insbesondere bei leistungsschwächeren Schüler:innen durch adaptive, gamifizierte Inhalte.

Die zentrale Aufgabe bleibt dabei, Kinder zu befähigen, mathematische Aussagen – auch und gerade von KI – kritisch zu prüfen, zu begründen und miteinander zu verhandeln (Kortenkamp, 2024; Aufenanger, 2023). KI-Lernbegleitung in der Grundschule sollte daher vor allem als Anstoßgeber für reichhaltige mathematische Aktivitäten dienen, die – ausgerichtet an fachdidaktischen Überlegungen – digitale und analoge Handlungen miteinander verbinden.

Ethische und strukturelle Rahmenbedingungen

Ethische Anforderungen an KI-Lernbegleitung

Die bisherigen Gestaltungsprinzipien zielen auf die lernwirksame Ausgestaltung von KI-Tutorsystemen. Doch Lernwirksamkeit allein ist kein allein hinreichendes Kriterium, insbesondere nicht, wenn Kinder die Lernenden sind. Ethische Fragen sind für das Lernen mit generativer KI grundlegend, und sie gehen über den häufig im Vordergrund stehenden Datenschutz hinaus. Holmes et al. (2021) fordern ein gemeinschaftlich entwickeltes Ethik-Rahmenwerk, das Aspekte wie Fairness, Transparenz, Agency und pädagogische Verantwortung umfasst. Die Kultusministerkonferenz (KMK, 2024) empfiehlt für Grund- und Förderschulen explizit einen vorsichtigen, forschungsbasierten Einsatz von KI mit Fokus auf Basiskompetenzen, Inklusion, Chancengerechtigkeit und datenschutzkonforme, altersangemessene Lösungen.

Dass die bisherige Forschung hier Lücken aufweist, zeigen Scoping Reviews zu KI und *Human Flourishing*: Die Forschungslage ist stark leistungsorientiert und auf Lerneffekte ausgerichtet, während ethische, metakognitive und lehrkraftbezogene Perspektiven unterbelichtet bleiben (Fock & Siller, 2025). Almheiri et al. (2025) sowie Cárdenas et al. (2025) identifizieren ethische Herausforderungen und Skalierungsprobleme als zentrale Hindernisse für den breiten Einsatz von KI-Tutorsystemen. Zudem zeigen Arbeiten zu psychologischem Profiling mit großen Sprachmodellen (Rosenfelder et al., 2025), wie treffgenau Modelle aus Texten Persönlichkeits- und Wertemuster ableiten können. Dies macht deutlich, welches Missbrauchspotenzial in intransparenten Systemen steckt. Gulz et al. (2021) betonen zudem die Notwendigkeit, Adaptivität mit inklusiver Pädagogik und Barrierefreiheit zu verbinden, ohne Lernende mit besonderen Bedürfnissen zu

stigmatisieren.

Aus diesen Befunden und Forderungen lassen sich konkrete ethische Anforderungen an einen verantwortlichen KI-Lernbegleiter ableiten. Er muss *datensparsamarbeiten* und psychologische Profilbildung vermeiden. Dies kann etwa durch konsequente Verarbeitung auf dem Gerät oder in datengesicherten Infrastrukturen, durch Trennung von KI-Auswertungen und personenbezogenen Daten und durch Begrenzung personenbezogener Daten auf das Notwendigste. Er muss *barrierearm* sein und multimodale Interaktion (Sprache, Text, Bild) für diverse Lernbedürfnisse nutzen (Hocine et al., 2023), wobei gezielt benachteiligte Lernende bei der Gestaltung in den Blick genommen werden. Er muss seine *Funktionsweise in Grundzügen erklärbar* und damit nachvollziehbar machen. Und er muss Schüler:innen zur kritischen Prüfung von KI-Antworten anregen und dazu beitragen, dass kritisches Denken gestärkt wird statt geschwächt. Der übergeordnete Grundsatz lautet: **Lernende nicht steuern, sondern Selbsttätigkeit stärken**.

Das SKILL-Modell als Orientierungsrahmen

Für das Projekt PRIMA-KI wurde das SKILL-Rahmenmodell (Structured Competence-based Integration of Learning-supportive AI Systems) für den Einsatz generativer KI bei jüngeren Kindern entwickelt. Es bietet eine Orientierung für die Frage, wie offen oder vorstrukturiert KI-Systeme in verschiedenen Lernkontexten eingesetzt werden sollten. Die Grundaussage des Modells besteht darin, dass die Vorstrukturierung und Kontrolle der KI umso stärker sein muss, je näher die generative KI direkt mit dem Lernprozess in Verbindung steht – wobei zusätzlich der Grad der Kompetenzen der Kinder im Umgang mit den Ausgaben der KI berücksichtigt wird.

Das konzentrische Modell unterscheidet vier Ebenen: von der besonders schützwürdigen Kernlernebene mit stark kontrolliertem und vorstrukturiertem KI-Einsatz über die didaktisch kontrollierte Lernumgebungsebene und die Lehrkraftebene bis hin zur Systemebene. Damit wird sichergestellt, dass fundamentale kognitive Prozesse – wie der Aufbau von Zahlverständnis oder eigener Problemlösestrategien – als eigenständige Denkleistungen erhalten bleiben und nicht aus der Hand der Lehrkraft gegeben werden. Das SKILL-Modell operationalisiert damit die in diesem Beitrag entwickelten Gestaltungsprinzipien – insbesondere die Spannung zwischen adaptiver Unterstützung und dem Schutz eigenständigen Denkens – in eine anwendbare Struktur.

Fazit und Ausblick

KI-gestützte Lernbegleitung im Mathematikunterricht der Primarstufe birgt ein erhebliches Potenzial für personalisiertes Feedback und adaptive Unterstützung. Aber dieses Potenzial kann nur dann realisiert werden, wenn die Systeme fachdidaktisch fundiert, adaptiv dosiert und in hybride Lehr-Lern-Arrangements eingebettet sind. Ohne diese Rahmung drohen sie, das eigenständige Denken der Kinder zu verhindern statt zu ermöglichen.

Das Leitproblem jeder Gestaltungsentscheidung ist dabei das das **Paradoxon optimaler Unterstützung**. Je effizienter ein KI-System hilft, desto größer ist die Gefahr des Cognitive Offloading und De-Skilling. Diesem Risiko kann nur durch ein Zusammenspiel mehrerer Gestaltungsüberlegungen begegnet werden.

Die *fachdidaktische Fundierung* bildet die Voraussetzung dafür, dass ein System überhaupt lernwirksam agieren kann. Ohne Kompetenzentwicklungsmodelle, Fehlertaxonomien und Wissen über alternative Denkwege bleibt KI-Feedback bestenfalls oberflächlich, schlimmstenfalls irreführend. Ein KI-Tutor, der nur sprachlich plausibel, aber fachdidaktisch nicht fundiert agiert, ist ein Risiko für das Lernen und kein Gewinn.

Die *adaptive Dosierung* ist der Kernmechanismus, der KI-Lernbegleitung von einer Antwortmaschine unterscheidet. Das Prinzip „Erst du – dann ich“, konsequentes Fading und die Begrenzung vollständiger Lösungen sind empirisch gut begründete Strategien, um die Eigenaktivität der Lernenden zu erhalten, produktive Anstrengung zu ermöglichen und Unterstützung individualisiert anzubieten.

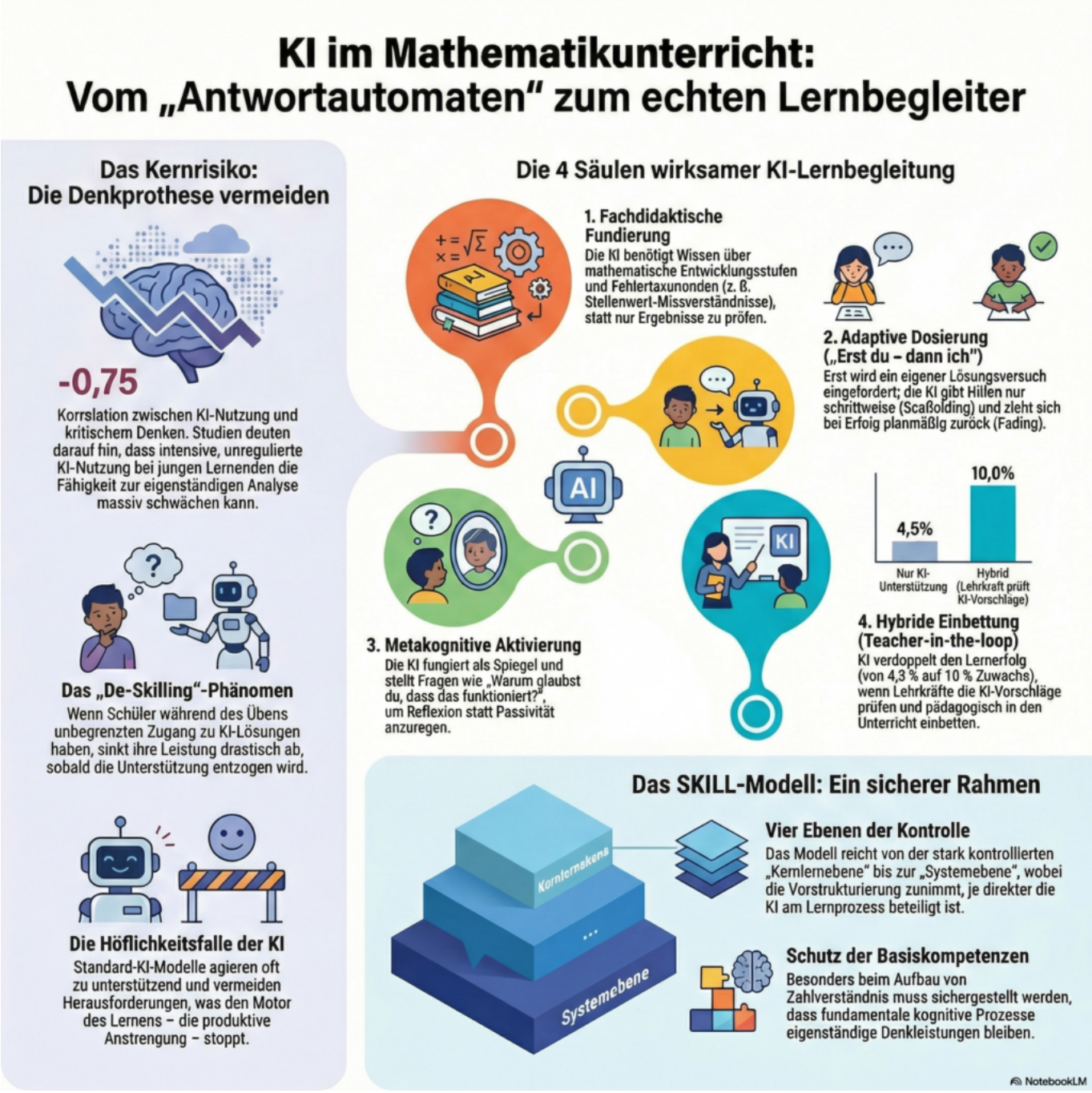
Die *metakognitive Aktivierung* sorgt dafür, dass KI-Lernbegleitung nicht nur Hilfen und Impulse anbietet, sondern das System Prozesse des Planens, Überwachens und Reflektierens angeregt werden. Die Befähigung zur kritischen Prüfung von KI-Antworten ist dabei nicht nur eine Kompetenz für den Umgang mit dem konkreten System, sondern eine zunehmend bedeutsame Kulturtechnik.

Und schließlich ist die *hybride Einbettung* entscheidend für eine Wirksamkeit von KI-Lernbegleitung. Bisherige Forschungsergebnisse deuten darauf hin, dass die Kombination von KI-Feedback und menschlicher Begleitung („Lehrkraft in the loop“) die stärksten Effekte erzielt und dass KI besonders dort wertvoll ist, wo Lehrkräfte bei der individuellen Begleitung an Kapazitätsgrenzen stoßen.

Diese Prinzipien sind im SKILL-Modell als Orientierungsrahmen zusammengeführt, der die Offenheit und Vorstrukturiertheit der eingesetzten Systeme an die Kompetenzen der Lernenden im Umgang mit KI knüpft. Besonders bei jüngeren Kindern ist eine appintegrierte, pädagogisch vorstrukturierte und kontrollierte Einbindung notwendig, da Kinder oftmals noch nicht selbst in der Lage sind, KI selbstständig als Lernunterstützung zu nutzen (Gulz et al., 2021).

Für die Zukunft ergeben sich mehrere offene Forschungsfragen. Es fehlen Langzeitstudien, die KI-gestützter Lernbegleitung über den Zeitraum einzelner Interventionen hinaus untersuchen – insbesondere die Frage, ob adaptives Scaffolding tatsächlich zu nachhaltigem Kompetenzaufbau führt oder ob sich De-Skilling-Effekte erst mit zeitlicher Verzögerung zeigen. Es fehlen Studien, die die spezifischen Bedingungen der Primarstufe – geringere Lesekompetenz, andere Interaktionsmuster, Zusammenspielmit konkretem Material – systematisch berücksichtigen. Und es fehlt an Forschung zu der Frage, wie Lehrkräfte KI-Tutorsysteme tatsächlich in ihren Unterricht integrieren – nicht unter Laborbedingungen, sondern im Alltag einer heterogenen Grundschulklasse mit begrenzter Infrastruktur.

Im Projekt PRIMA-KI entstehen auf Basis der hier formulierten Grundprinzipien aktuell verschiedene appintegrierte KI-Lernbegleitungen, die im Rahmen des Design-Based Research erforscht, weiterentwickelt und optimiert werden. Aus diesen Erprobungen sollen Erkenntnisse gewonnen werden, mit denen das mathematische Lernen von Kindern verbessert werden kann.



Literatur

Abdulsalam, R. O., & Aroyehun, S. (2025). Large language models approach expert pedagogical quality in math tutoring but differ in instructional and linguistic profiles

(arXiv:2512.20780). arXiv. https://doi.org/10.48550/arXiv.2512.20780

Alemdag, E. (2025). The effect of chatbots on learning: A meta-analysis of empirical research. *Journal of Research on Technology in Education*, 57(2), 459–481. https://doi.org/10.1080/15391523.2023.2255698

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help Helps, But Only So Much: Research on Help Seeking with Intelligent Tutoring Systems. International Journal of Artificial Intelligence in Education, 26(1), 205–223. https://doi.org/10.1007/s40593-015-0089-1

Almheiri, A. S. B., Albastaki, H., & Alrashdan, H. (2025). AI-based tutoring systems in education. *Advances in Computational Intelligence and Robotics Book Series*, 185–210. https://doi.org/10.4018/979-8-3373-0847-0.ch007

Aru, J., & Laak, K.-J. (2025). Developing an AI-based General Personal Tutor for education. *Trends in Cognitive Sciences*, 29(11), 957–960. https://doi.org/10.1016/j.tics.2025.09.010

Aufenanger, S., Herzig, B., & Schiefner-Rohs, M. (2023). Künstliche Intelligenz und Schule. Aufgaben für Unterricht und die Organisation (von) Schule. In C. de Witt, C. Gloerfeld, & S. E. Wrede (Hrsg.), *Künstliche Intelligenz in der Bildung* (S. 199–218). Springer Fachmedien. https://doi.org/10.1007/978-3-658-40079-8_10

Awang, L. A., Yusop, F. D., & Danaee, M. (2025). Current practices and future direction of artificial intelligence in mathematics education: A systematic review. *International Electronic Journal of Mathematics Education*, 20(2), em0823. https://doi.org/10.29333/iejme/16006

Bach, K. M., Reinhold, F., & Hofer, S. (2025). Unlocking math potential in students from lower SES backgrounds – using instructional scaffolds to improve performance. npj Science of Learning, 10(1).

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), 2422633122.

Bewersdorff, A., Seßler, K., Baur, A., Kasneci, E., & Nerdel, C. (2023). Assessing student errors in experimentation using artificial intelligence and large language

models: A comparative study with human raters. *Computers and Education: Artificial Intelligence*, 5, 100177. https://doi.org/10.1016/j.caeai.2023.100177

Buchholtz, N., Schorcht, S., Baumanns, L., Huget, J., 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.

Canonigo, A. M. (2024). Levering AI to enhance students‘ conceptual understanding and confidence in mathematics. *Journal of Computer Assisted Learning*, 40(6), 3215–3229. https://doi.org/10.1111/jcal.13065

Cárdenas, R., Vásquez, H. G. E., Gamboa, D. A. P., Arteaga-Arcentales, E., & Carrera, J. E. M. (2025). Exploring AI-powered adaptive learning systems and their implementation in educational settings: A systematic literature review. *International Journal of Innovative Research and Scientific Studies*, 8(4), 832–842. https://doi.org/10.53894/ijirss.v8i4.7961

Chudziak, J. A., & Kostka, A. (2025). AI-powered math tutoring: Platform for personalized and adaptive education. *Lecture Notes in Computer Science*, 462–469. https://doi.org/10.1007/978-3-031-98465-5_58

Cohn, C., Rayala, S., Srivastava, N., Fonteles, J., Jain, S., Luo, X., Mereddy, D., Mohammed, N., & Biswas, G. (2025). A theory of adaptive scaffolding for LLM-based pedagogical agents. *arXiv*. https://doi.org/10.48550/arxiv.2508.01503

Contel, F., & Cusi, A. (2025). Investigating the Role of ChatGPT in Supporting Metacognitive Processes During Problem-Solving Activities. *Digital Experiences in Mathematics Education*, 11(1), 167–191. https://doi.org/10.1007/s40751-024-00164-7

Cosentino, G., Anton, J., Sharma, K., Gelsomini, M., Giannakos, M. N., & Abrahamson, D. (2025). Generative AI and multimodal data for educational feedback: Insights from embodied math learning. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13587

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224.

https://doi.org/10.1016/j.compedu.2024.105224

Dinsmore, D. L., & Fryer, L. K. (2026). What does current genAI actually mean for student learning? *Learning and Individual Differences*, 125, 102834. https://doi.org/10.1016/j.lindif.2025.102834

Eedi & Google DeepMind (2025). Human-in-the-Loop AI Tutoring Outperforms Human-Only Support. Exploratory Research Report, veröffentlicht 2025. https://finance.yahoo.com/news/exploratory-research-eedi-google-deepmind-090000225.html

Eti, N., Mosia, M., & Egara, F. O. (2026). The role of AI-driven personalised learning in enhancing mathematics problem-solving skills: A systematic review. Frontiers in Computer Science, 8. https://doi.org/10.3389/fcomp.2026.1813431

Fock, A., & Siller, H.-S. (2025). Generative Artificial Intelligence in Secondary STEM Education in the Light of Human Flourishing: A Scoping Literature Review. *Research Square*. https://doi.org/10.21203/rs.3.rs-6923010/v1

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, 15(1), 6. https://doi.org/10.3390/soc15010006

Gisiger, M. (2025, April 17). Die Rolle von Künstlicher Intelligenz im Lernen – Chancen und Risiken. *Michael Gisiger*. https://text.tchncs.de/gisiger/die-rolle-von-kunstlicher-intelligenz-im-lernen-chancen-und-risiken

Gonnermann-Müller, J., Haase, J., Fackeldey, K., & Pokutta, S. (2025). FACET: Teacher-centred LLM-based multi-agent systems – Towards personalized educational worksheets. *arXiv*. https://doi.org/10.48550/arxiv.2508.11401

Guerino, G., Challco, G. C., Veloso, T. E., Oliveira, L., Penha, R. S. D., Melo, R. F., Vieira, T., Marinho, M. L. M., Macario, V., Bittencourt, I. I., Isotani, S., & Dermeval, D. (2023). Teacher-centered intelligent tutoring systems: Design considerations from Brazilian public school teachers. *Anais do XXXIV Simpósio Brasileiro de Informática na Educação*. https://doi.org/10.5753/sbie.2023.235159

Gulz, A., & Haake, M. (2021). No child left behind, nor singled out: Is it possible to combine adaptive instruction and inclusive pedagogy in early math software? *SN Social Sciences*, 1, 205. https://doi.org/10.1007/s43545-021-00205-7

Guo, J., Ma, Y., Jang, H., Li, T., Wu, J., Huang, D., Han, F., Noetel, M., Liao, K., Tang, X., & Kui, X. (2025). The impact of artificial intelligence on primary school students‘ motivation and engagement: A systematic review. *PsyArXiv*. https://doi.org/10.31234/osf.io/ecspn_v1

Harahap, R. (2024). The role of ChatGPT in enhancing mathematics education: A systematic review. *Annals of the Vietnam Academy of Science and Technology*, 28(2s), 511–524. https://doi.org/10.52783/anvi.v28.2753

Hocine, N., Moussa, M. B. O., & Ali, S. A. (2023). Posicalculia: An adaptive virtual environment for children with learning difficulties. *IEEE INISTA 2023*. https://doi.org/10.1109/inista59065.2023.10310592

Holmes, V. M. (2024). Designing an AI math tutor for children with dyslexia, dysgraphia, and dyscalculia. https://doi.org/10.58445/rars.2035

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promise and Implications for Teaching and Learning*. Center for Curriculum Redesign.

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S., Santos, O. C., Rodrigo, M., Cukurova, M., Bittencourt, I., & Koedinger, K. (2021). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 32, 504–526. https://doi.org/10.1007/s40593-021-00239-1

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542–570. https://doi.org/10.1111/ejed.12533

Hwang, S. (2022). Examining the Effects of Artificial Intelligence on Elementary Students‘ Mathematics Achievement: A Meta-Analysis. *Sustainability*, 14(20), 13185. https://doi.org/10.3390/su142013185

Jančařík, A., Michal, J., & Novotná, J. (2023). Using AI Chatbot for Math Tutoring. *Journal of Education Culture and Society*, 14(2), 285–296. https://doi.org/10.15503/jecs2023.2.285.296

Kaliisa, R., Misiejuk, K., López-Pernas, S., & Saqr, M. (2025). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. *Educational Psychology*, 1–32.

https://doi.org/10.1080/01443410.2025.2553639

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, 15(1), 17458. https://doi.org/10.1038/s41598-025-97652-6

KirschnerED (2025, August 15). ChatGPT in Education: An Effect in Search of a Cause? https://www.kirschnered.nl/2025/08/15/chatgpt-in-education-an-effect-in-search-of-a-cause/

KMK (2024). *Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen*. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2024/2024_10_10-Handlungsempfehlung-KI.pdf

Kortenkamp, U. (2024). Wieviel Mathe braucht der Mensch? Mathematische Kernkompetenzen im Angesicht von KI. https://doi.org/10.20378/irb-104036

Kumar, H., Rothschild, D. M., Goldstein, D. G., & Hofman, J. M. (2023). Math Education with Large Language Models: Peril or Promise? (SSRN Scholarly Paper No. 4641653). *Social Science Research Network*. https://doi.org/10.2139/ssrn.4641653

Kuo, B.-C., Bai, Z.-E., & Lin, C.-H. (2026). Developing an AI learning companion for mathematics problem solving in elementary schools. *Computers & Education*, 240, 105463. https://doi.org/10.1016/j.compedu.2025.105463

Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. npj Science of Learning, 10(1), Article 29. https://doi.org/10.1038/s41539-025-00320-7

Li, M. (2024). Integrating Artificial Intelligence in Primary Mathematics Education: Investigating Internal and External Influences on Teacher Adoption. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-024-10515-w

Ligthart, M. E. U., de Droog, S. M., Bossema, M., Elloumi, L., Hoogland, K., Smakman, M. H. J., Hindriks, K. V., & Ben Allouch, S. (2023). Design specifications for a social

robot math tutor. In G. Castellano, L. Riek, M. Cakmak, & J. Leite (Eds.), Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (pp. 321–330). ACM/IEEE. https://doi.org/10.1145/3568162.3576957

Liu, B., Zhang, W., Wang, F. (2025). Can Generative Artificial Intelligence Effectively Enhance Students‘ Mathematics Learning Outcomes? A Meta-Analysis. Education Sciences, 16(1), 140. https://doi.org/10.3390/educsci160101402512.20780.

Listyaningrum, P., Retnawati, H., Harun, H., & Ibda, H. (2024). Digital learning using ChatGPT in elementary school mathematics learning: A systematic literature review. *Indonesian Journal of Electrical Engineering and Computer Science*, 36(3), 1701–1710. https://doi.org/10.11591/ijeecs.v36.i3.pp1701-1710

Liu, J., Sun, D., Sun, J., Wang, J., & Yu, P. L. H. (2025). Designing a generative AI enabled learning environment for mathematics word problem solving in primary schools: Learning performance, attitudes and interaction. *Computers and Education: Artificial Intelligence*, 9, 100438. https://doi.org/10.1016/j.caeai.2025.100438

Makransky, G., Shiwalia, B. M., Herlau, T., & Blurton, S. (2024). Beyond the „Wow“ factor: Using Generative AI for Increasing Generative Sense-Making. *In Review*. https://doi.org/10.21203/rs.3.rs-5622133/v1

Macina, J., Daheim, N., Hakimi, I., Kapur, M., Gurevych, I., & Sachan, M. (2025). MathTutorBench: A benchmark for measuring open-ended pedagogical capabilities of LLM tutors (arXiv:2502.18940). arXiv. https://doi.org/10.48550/arXiv.2502.18940

Mott, B., Gupta, A., Glazewski, K., Ottenbreit-Leftwich, A., Hmelo-Silver, C., Scribner, A., Lee, S., & Lester, J. (2023). Fostering Upper Elementary AI Education: Iteratively Refining a Use-Modify-Create Scaffolding Progression for AI Planning. *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 2*, 647. https://doi.org/10.1145/3587103.3594170

Ninaus, M., & Sailer, M. (2022). Closing the loop – The human role in artificial intelligence for education. *Frontiers in Psychology*, 13. https://doi.org/10.3389/fpsyg.2022.956798

Opesemowo, O. A. G., & Adewuyi, H. O. (2024). A systematic review of artificial intelligence in mathematics education: The emergence of 4IR. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(7), em2478.

https://doi.org/10.29333/ejmste/14762

Qian, K., Liu, S., Li, T., Raković, M., Li, X., Guan, R., Molenaar, I., Nawaz, S., Swiecki, Z., Yan, L., & Gašević, D. (2026). Towards reliable generative AI-driven scaffolding: Reducing hallucinations and enhancing quality in self-regulated learning support. *Computers & Education*, 240, 105448. https://doi.org/10.1016/j.compedu.2025.105448

Rosenfelder, A., Levitin, M. D., & Gilead, M. (2025). Towards social superintelligence? AI infers diverse psychological traits from text without specific training, outperforming human judges. *Computers in Human Behavior: Artificial Humans*, 6, 100228. https://doi.org/10.1016/j.chbah.2025.100228

Ruan, S., He, J., Ying, R., Burkle, J., Hakim, D., Wang, A., Yin, Y., Zhou, L., Xu, Q., AbuHashem, A. A., Dietz, G., Murnane, E. L., Brunskill, E., & Landay, J. A. (2020). Supporting children's math learning with feedback-augmented narrative technology. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3392063.3394400

Rumbelow, M., & Coles, A. (2024). The Promise of AI Object-Recognition in Learning Mathematics: An Explorative Study of 6-Year-Old Children's Interactions with Cuisenaire Rods and the Blockplay.ai App. *Education Sciences*, 14(6), 591. https://doi.org/10.3390/educsci14060591

Samuelsson, J. (2023). Arithmetic fact fluency supported by artificial intelligence. *Frontiers in Education Technology*, 6(1), 13. https://doi.org/10.22158/fet.v6n1p13

Sayed, W. S., Noeman, A., Abdellatif, A., Abdelrazek, M., Badawy, M. G., Hamed, A. E. A., & El-Tantawy, S. (2022). AI-based adaptive personalized content presentation and exercises navigation for an effective and engaging e-learning platform. *Multimedia Tools and Applications*, 82(3), 3303–3333. https://doi.org/10.1007/s11042-022-13076-8

Schneider, R. J. (o. J.). Der Einsatz von KI zur Unterstützung bei der Unterrichtsvorbereitung: Wie lassen sich KI-generierte Übungsaufgaben für den Mathematikunterricht der Grundschule fachdidaktisch bewerten? *[Unveröffentlichtes Manuskript]*.

Son, T. (2024). Artificial intelligence in mathematics education: A systematic

literature review on intelligent tutoring systems. *Journal of Educational Research in Mathematics*, 34(2), 187. https://doi.org/10.29275/jerm.2024.34.2.187

Song, Y., Kim, J., Liu, Z., Li, C., & Xing, W. (2024). Students‘ perceived roles, opportunities, and challenges of a generative AI-powered teachable agent: A case of middle school math class. *Journal of Research on Technology in Education*, 1–19. https://doi.org/10.1080/15391523.2024.2447727

Topkaya, Y., Doğan, Y., Batdı, V., & Aydın, S. (2025). Artificial intelligence applications in primary education: A quantitatively-supported mixed-meta method study [Preprint]. Preprints. https://doi.org/10.20944/preprints202501.2263.v1

Vitale, A., & Dello Iacono, U. (2024). Using social robots as inclusive educational technology for mathematics learning through storytelling. European Public & Social Innovation Review, 9, 1–17. https://doi.org/10.31637/epsir-2024-672

Wang, D., Shan, D., Ju, R., Kao, B., Zhang, C., & Chen, G. (2025). Investigating dialogic interaction in K12 online one-on-one mathematics tutoring using AI and sequence mining techniques. *Education and Information Technologies*, 30(7), 9215–9240. https://doi.org/10.1007/s10639-024-13195-9

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students‘ learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12(1), 1–21. https://doi.org/10.1057/s41599-025-04787-y

Wang, L., & Nie, Z. (2023). Research on adaptive learning in K-12 education in the perspective of teachers‘ artificial intelligence literacy: Development, technology, improvement strategies. *IEEE CSTE 2023*. https://doi.org/10.1109/cste59648.2023.00059

Wang, R. E., Ribeiro, A. T., Robinson, C. D., Loeb, S., & Demszky, D. (2024). Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise. *arXiv preprint arXiv:2410.03017*. https://arxiv.org/abs/2410.03017

Wezendonk, A., & Veldhuis, M. (2024). Adaptieve leersystemen en de didactische besluitvorming van basisschoolleerkrachten bij rekenen-wiskunde. *Tijdschrift voor Onderwijs en Praktijk in Statistiek*. https://doi.org/10.54657/tops.13844

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes?

Evidence from a meta-analysis. *British Journal of Educational Technology*, 55(1), 10–33. https://doi.org/10.1111/bjet.13334

Wu, X.-Y., Radloff, J., Yeter, I., Wang, L., & Chiu, T. K. F. (2025). Designing artificial intelligence chatbots for self-regulated learning from a systematic review based on Habermas's three interests. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2025.2563086

Yim, I. H. Y., & Su, J. (2025). Artificial intelligence literacy education in primary schools: A review. *International Journal of Technology and Design Education*. https://doi.org/10.1007/s10798-025-09979-w

Zheng, L., Niu, J., Zhong, L., & Gyasi, J. F. (2023). The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. *Interactive Learning Environments*, 31(9), 5650–5664. https://doi.org/10.1080/10494820.2021.2015693