

KI und Ethik im Klassenzimmer

Melanie Platz und Christine Plote

Künstliche Intelligenz (KI) hat in den letzten Jahren tiefgreifende Veränderungen in allen Bereichen der Gesellschaft ausgelöst. Besonders die Einführung von ChatGPT Ende 2022 markierte einen Wendepunkt: Erstmals sind KI-Technologien für eine breite Öffentlichkeit zugänglich. Die Möglichkeit, KI intuitiv zu nutzen – mit einem Computer fast wie mit einem Menschen zu ‚sprechen‘ – weckt Begeisterung und bietet enorme Chancen, wirft jedoch ethische und gesellschaftliche Fragen auf.

Dieser Beitrag untersucht die historische Entwicklung, den aktuellen ‚Hype‘ und ethische Kriterien für KI, insbesondere generativer KI im Schulkontext. Wir laden Lesende dazu ein, die Potenziale und Herausforderungen von KI kritisch zu hinterfragen und konstruktiv über deren sinnvollen Einsatz zu diskutieren.

Künstliche Intelligenz – keine Erfindung der Neuzeit

Obwohl KI meistens als Innovation der Neuzeit gilt, gibt es schon seit der Antike Ideen, künstliches Leben zu erschaffen (z. B. Spitzer, 2023). So beschrieb Homer in der Ilias, wie der Gott Hephaistos goldene Dienerinnen erschuf, die eigenständig denken und handeln konnten. Im 17., 18. und 19. Jahrhundert entwickelten Pioniere wie Leibniz, Babbage, Lovelace und Boole mathematische, logische und technische Grundlagen für die Idee intelligenter Maschinen. Turing führte diese Ansätze im 20. Jahrhundert weiter, indem er die Berechenbarkeit formalisierte und die Frage stellte, ob Maschinen denken können.

Als Geburtsstunde der Bezeichnung ‚Künstliche Intelligenz‘ gilt das ‚Summer Research Project on Artificial Intelligence‘, das 1956 am Dartmouth College in New Hampshire (USA) stattfand (z. B. Buxmann & Schmidt, 2021). Die Bezeichnung ist jedoch umstritten, denn, obwohl das Thema vielfach diskutiert wurde, existieren bisher nur individuelle Konzepte, statt einer kohärenten Definition (Vieweg, 2023).

Ein Ziel von KI ist es, Maschinen mit Fähigkeiten zu versehen, die intelligentem (menschlichen) Verhalten ähneln (Kaplan & Haenlein, 2019), und Intelligenz außerhalb des menschlichen Gehirns zu schaffen (Buxmann & Schmidt, 2021). Dazu scheint die zeitlose Definition von Elaine Rich (1983) bis heute passend, da sie philosophische Debatten um die Bedeutung von ‚Künstlich‘ und ‚Intelligenz‘ vermeidet,

wie sie häufig im Vordergrund stehen: „Artificial Intelligence (A.I.) is the study of how to make computers do things at which, at the moment, people are better.“ (S. 1).

Mit der Prägung der Bezeichnung ‚Künstliche Intelligenz‘ waren noch keine Anwendungen geschaffen. Erst die drastisch wachsende Rechenleistung moderner Computer, Fortschritte bei künstlichen neuronalen Netzen und Deep Learning (u. a. Hinton's Backpropagation-Algorithmus 1986) und das Konzept des ‚Trainierens‘ von Algorithmen an großen Datenmengen (z. B. der Datensatz ImageNet, der in den 2010er Jahren schnelle Fortschritte im Machine Learning ermöglichte), machten generative KI möglich – wobei die nötigen ‚Big Data‘ erst durch das Internet verfügbar wurden (Ertel, 2021).

Ende 2022 brachte OpenAI mit ChatGPT – scheinbar aus dem Nichts – eine für alle frei verfügbare KI-Anwendung auf den Markt. „Seit dem 30. November 2022 kann jeder mit KI herumspielen und selbst erleben und herausfinden, was KI kann. Die meisten Leute waren völlig geplättet – um es vorsichtig auszudrücken“ (Spitzer, 2023, S. 16). Bereits zwei Monate nach der Veröffentlichung zählte die Preview-Version 100 Millionen aktive Nutzende: „Wahrscheinlich ist keine moderne Technologie gleichzeitig so verführerisch und verstörend wie generative KI. Die Emotionen kochten hoch beziehungsweise über.“ (Baker, 2024, S. 31). ‚KI‘ und ‚ChatGPT‘ wurden fortan häufig synonym verwendet.

Die Grundidee von ChatGPT ist nicht neu. Schon Jahre früher hatte es Chatbots gegeben, die allerdings auf verhältnismäßig wenig Interesse stießen. Wenn man außerdem bedenkt, dass ChatGPT „noch nicht einmal der beste Bot seiner Art, zumindest nicht in den ersten Versionen“, war (Baker, 2024, S. 43), drängt sich die Frage auf: Was hat ChatGPT so schnell so erfolgreich gemacht?

Niedrige Einstiegshürden, intuitive Bedienung, die freie Verfügbarkeit und die vielseitigen Einsatzmöglichkeiten spielten sicherlich eine große Rolle. Hinzu kam, dass, nach Jahren ohne bahnbrechende Technik-Revolutionen (der ‚iPhone-Moment‘ lag bereits 15 Jahre zurück) ChatGPT auf ein Publikum traf, das neue Technologien mit Begeisterung aufnahm (Spitzer, 2023).

Ein Schlüssel zum Erfolg von ChatGPT liegt in unserer Neigung, sprachliche Zeichenketten als sinnvolle

menschliche Äußerungen wahrzunehmen und ihnen Absicht sowie eine verantwortliche Person zuzuschreiben (Bender et al., 2021). Diese Vermenschlichung zeigte sich auch 1966, als Weizenbaum den Chatbot ELIZA entwickelte, der eine Gesprächstherapeutin simulierte. Weizenbaum sah ELIZA als ‚witziges Programm‘ und war entsetzt, als Psychiater auf einem Kongress ernsthaft erwogen, den Bot in ihrer psychotherapeutischen Arbeit einzusetzen (Krauthausen, 2020). „Ich war wirklich fassungslos und mir drängte sich die Frage auf, welches Selbstverständnis ein Psychiater haben musste, um die Idee zu entwickeln, einen wesentlichen Teil seiner Arbeit einer Maschine zu übergeben?“ (Weizenbaum & Wendt, 2006, S. 95, zit. n. Krauthausen, 2020, S. 47).

Was macht Künstliche Intelligenz (nicht) ethisch?

Unmittelbar nach dem Launch von ChatGPT wurde neben Begeisterung für die neue Technologie auch erste Kritik laut, und Stellungnahmen erschienen.

Im März 2023 forderten Expertinnen und Experten in einem offenen Brief des US-Instituts ‚Future of Life‘ eine sechsmonatige Entwicklungspause für KI (Future of Life, 2023). Im Mai warnte das ‚Center for AI Safety‘ vor existenziellen KI-Risiken (CAIS, 2023). Unter den Unterzeichnern der Briefe waren Tesla-Gründer Elon Musk und OpenAI-CEO Sam Altman. Kritiker wie Emily Bender sehen diese Warnungen als Ablenkung von realen, akuten Problemen generativer KI und vermuten in den offenen Briefen den Versuch, kleinteilige Regulierung zu verhindern (Süddeutsche Zeitung, 2023).

Generative KI stärkt die Macht von Konzernen wie Google, Amazon, Facebook, Apple und Microsoft (kurz: GAFAM), so ein Vorwurf. ‚Big Tech‘ verfügt über die nötigen Ressourcen, um KI-Anwendungen im großen Stil zu entwickeln. Dafür nutzen sie enorme Datenmengen aus dem Netz und ihrer Nutzer. OpenAI kann bspw. laut AGB jede Information, die in einen ChatGPT-Dialog eingegeben wird, analysieren und für das Training von KI-Modellen nutzen (Baker, 2004).

Hier gilt: Je größer die verfügbaren Corpora sind, desto besser wird die Leistung der zugehörigen KI werden. Erneut bietet der Größenvorteil der Plattformen und der nahezu unbegrenzte Zugang zu Daten eine unschlagbare Ausgangsposition für Big Tech. (Andree, 2023, S. 100)

Stochastisches Geplapper

2021 erschien ein Paper, das die Risiken und Verantwortlichkeiten großer Sprachmodelle und der umfang-

reichen Datensätze, auf denen sie basieren, untersuchte: ‚On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?‘ (Bender et al., 2021). Die Autorinnen prägten die Bezeichnung ‚stochastische Pageien‘, um zu illustrieren, dass Sprachmodelle Textsequenzen auf Basis ihrer Trainingsdaten wahrscheinlichkeitsbasiert kombinieren, ohne echtes Verständnis für die Bedeutung ihrer Ergebnisse (S. 616 f.). Bekannt wurden zwei der Autorinnen, Dr. Timnit Gebru und Dr. Margaret Mitchell, Gründerinnen und Leiterinnen von Googles KI-Ethikabteilung, auch, weil sie aufgrund des kritischen Artikels von ihrem Arbeitgeber Google entlassen wurden (Ryan et al., 2024). ‚On the Dangers of Stochastic Parrots‘ wurde zu einem Referenzpunkt in der Diskussion über Risiken und Verantwortlichkeiten im Umgang mit generativer KI und gleichzeitig ein Symbol für die Spannungen zwischen Ethik und wirtschaftlichem Druck in der KI-Industrie.

Einige Kritikpunkte an großen KI-Modellen sind nach Bender et al. (2021) und der UNESCO (Miao & Holmes, 2023):

- *Ungerechtigkeit*: Viele KI-Modelle werden für die Bedürfnisse derjenigen entwickelt, die bereits die meisten Privilegien in der Gesellschaft haben. Diesen Punkt greift auch die SWK (2024) in Bezug auf ein schulisches Umfeld auf: „Problematisch ist, dass LLM nicht für alle Lernenden gleichermaßen zugänglich sind. [...] Für Lernende mit niedrigem sozio-ökonomischem Hintergrund stellen finanzielle Hürden durch private Anbieter und mangelnde Unterstützungsmöglichkeiten durch die Eltern eine Zugangsbarriere dar.“ (S. 17).
- *‚Verunreinigte‘ Trainingsdaten, Bias (Verzerrung) und Diskriminierung* – „Size Doesn’t Guarantee Diversity“ (Bender et al., 2021, S. 613): Vorurteile und algorithmische Biases können diskriminierende Inhalte und soziale Ungleichheiten reproduzieren und verstärken, mit problematischen Folgen für benachteiligte Gruppen. Beispiel: Werden für das KI-Training Daten von Reddit verwendet, einer Plattform mit überwiegend männlicher Nutzerbasis in den USA, kann dies geschlechtsspezifische Verzerrungen fördern, die sich in Folgeversionen fortsetzen. Die Intransparenz von Trainingsdaten wird zunehmend kritisch, da KI-generierte Inhalte das Internet ‚verunreinigen‘ und realistischere Deep Fakes entstehen. „Im Bildungskontext ist die Frage nach der Datengrundlage und deren transparenter Veröffentlichung besonders relevant“ (SWK, 2024, S. 9).
- *Mangelnde Erklärbarkeit und Transparenz*: Die Funktionsweise von KI-Modellen ist oft intransparent, ihre Algorithmen schwer erklärbar. Diese ‚Black-Boxes‘ erschweren die Bewertung von Quellen und

Qualität KI-generierter Inhalte – ein kritisches Problem angesichts der Gefahr von Fehlinformationen, Desinformation, Halluzinationen und Biases.

- **Fehlinformationen, Halluzinationen und Manipulationsgefahr:** KI-Modelle bergen das Risiko, falsche Inhalte zu generieren, die von Nutzenden ungeprüft übernommen werden. Dies kann gezielt für Desinformation oder Manipulation missbraucht werden und stellt eine potenzielle Bedrohung für persönliche Sicherheit und demokratische Prozesse dar.
- **Abhängigkeit, Verlust menschlicher Kreativität:** KI-Systeme können eine unkritische Verwendung fördern, wodurch eigenständiges Denken und kreative Prozesse in den Hintergrund geraten. Insbesondere im Bildungsbereich besteht die Gefahr, dass Lernende und Lehrpersonen zunehmend passiv auf KI-Lösungen zurückgreifen, anstatt aktiv Probleme zu lösen. Dies könnte langfristig die Entwicklung von Kreativität und kritischem Denken beeinträchtigen.
- **Schutz geistigen Eigentums:** Generative KI-Modelle wie ChatGPT werden oft ohne Zustimmung der Urheber/innen auf großen Datenmengen aus dem Internet trainiert, darunter Texte, Bilder, Videos, Code und persönliche Daten. Mechanismen zum Schutz von Urheberrechten oder zur Entschädigung fehlen weitgehend. Nutzende können Urheberrechte nur geltend machen, wenn sie wesentliche menschlich-gestalterische Beiträge nachweisen.
- **Mangelnde Nachhaltigkeit:** KI-Modelle sind äußerst ressourcenintensiv. Allein das Training eines BERT-Basismodells verbraucht so viel Energie wie ein Transatlantikflug und verursacht erhebliche CO₂-Emissionen. Zusätzlich erfordert die Kühlung der Rechenzentren enorme Wassermengen, oft in Regionen mit knappen Wasserressourcen.
- **Arbeitsbedingungen:** Betrieb und Wartung von KI-Modellen basieren oft auf menschlicher Arbeit, etwa für Datenannotationen oder Inhaltsmoderation. Diese Aufgaben werden häufig in Billiglohnländern unter schlechten Bedingungen ausgeführt. Gleichzeitig konzentrieren sich die wirtschaftlichen Vorteile der KI-Entwicklung auf wenige große Tech-Konzerne, während benachteiligte Regionen und Arbeitskräfte kaum davon profitieren.
- **Kulturelle Perspektiven:** Es besteht die Gefahr einer einseitigen kulturellen Prägung, da die Inhalte von den Ländern oder Konzernen dominiert werden, die die KI entwickeln und trainieren.
- **Auch in der Geschwindigkeit der Entwicklung von KI** wird eine Gefahr gesehen, da diese schneller als die Anpassung der nationalen Rechtsvorschriften geschieht.

Es lässt sich folgern:

In no other field is the ethical compass more relevant than in artificial intelligence. These general-purpose technologies are re-shaping the way we work, interact, and live. [...] AI technology brings major benefits in many areas, but without the ethical guardrails, it risks reproducing real world biases and discrimination, fueling divisions and threatening fundamental human rights and freedoms. (Ramos, Assistant Director-General for Social and Human Sciences of UNESCO, o. J.)

Die Diskussion um generative KI wirft grundlegende ethische Fragen auf, die eng mit den Werten unserer Gesellschaft verbunden sind. Um diese besser zu verstehen, lohnt es sich zunächst ein paar ethische Grundlagen zu betrachten.

Ethischer Rahmen und ethische KI

Nach Vieweg (2023, S. 10) ist Ethik „die philosophische Disziplin, die sich mit der Frage nach den Kriterien für gute und schlechte Handlungen beschäftigt.“ Diese Frage ist kontextabhängig und in einer vielfältigen Gesellschaft schwer universell zu beantworten. Dennoch bietet Ethik als Theorie der gelebten Moral Orientierung für konkretes Handeln (Vieweg, 2023).

Ein strukturiertes Set aus Prinzipien, Werten und Leitlinien unterstützt dabei, komplexe ethische Fragestellungen systematisch zu analysieren und Handlungsoptionen zu bewerten. Ethische Rahmen als Satz von Leitlinien, die sicherstellen, dass Entscheidungen und Handlungen von KI-Systemen in Übereinstimmung mit sozialen, rechtlichen und moralischen Normen stehen, werden zu diesem Zweck entwickelt (Dignum, 2019).

Für den Deutschen Ethikrat (2023) steht bei der Beurteilung von KI eine zentrale Frage im Fokus: Erweitert oder vermindert der Einsatz von KI die menschliche Autorenschaft und die Bedingungen für verantwortliches Handeln? Diese Perspektive unterstreicht die Notwendigkeit, KI nicht nur technisch, sondern auch ethisch fundiert zu bewerten.

Dafür bietet die Definition der Europäischen Kommission (2022, S. 11) Orientierung: „Ethische KI bezeichnet die Entwicklung, Einführung und Nutzung von KI-Systemen, die ethischen Normen, Ethikgrundsätzen und damit verbundenen Grundwerten gerecht werden.“ Der EU AI Act (2024) ergänzt diese Perspektive, indem er ‚KI-Systeme‘ in Art. 3 als maschinengestützte Systeme definiert, die autonom aus Eingaben Ergebnisse ableiten und Entscheidungen treffen, die virtuelle oder physische Umgebungen beeinflussen. Der AI Act basiert auf dem Konzept der ‚vertrauenswürdigen KI‘, das auf vier Grundsätzen beruht:

1. Transparenz und Nachvollziehbarkeit;
2. Fairness: keine Diskriminierung aufgrund von Geschlecht, Alter, ethnischen oder sozialen Merkmalen, religiösen oder politischen Überzeugungen oder anderen Eigenschaften;
3. Robustheit: resilient gegenüber Fehlern und Angriffen;
4. Sicherheit: keine physischen oder virtuellen Schäden verursachen.

Die UNESCO (2021) führt Werte und ethische Prinzipien wie Privatsphäre, Transparenz, Erklärbarkeit und Nicht-Diskriminierung ein, die bei der Entwicklung und Anwendung von KI respektiert werden müssen. Dies soll dazu beitragen, digitale Transformation so zu gestalten, dass sie die Menschenrechte fördert und zu den nachhaltigen Entwicklungszielen der UN beiträgt.

Von ethischen Grundsätzen zur Regulierung

ChatGPT traf 2022 auf einen weitgehend unregulierten Markt. Dies ermöglichte zwar die rasante Verbreitung der Anwendung. Gleichzeitig führte der Hype um ChatGPT die Notwendigkeit einer angemessenen Regulierung von KI-Technologien vor Augen und brachte Diskussionen darüber in die breitere Öffentlichkeit.

Bereits 2017 hatte das Europäische Parlament Regeln für Robotik und KI gefordert, um wirtschaftliches Potenzial auszuschöpfen und Sicherheitsstandards zu gewährleisten (EU, 2017). Eine Chance für Europa (Delvaux, 2017), da Normen Rechtssicherheit geben, was wirtschaftliche Risiken senkt und Investitionen in KI-Technologien ermöglicht (Vieweg, 2023). Auch der KI Bundesverband (2023) sieht in einem klaren Rechtsrahmen Chancen, ‚KI Made in Europe‘ zu stärken, warnt jedoch, dass das Gesetz Europas Wettbewerbsfähigkeit gegenüber den USA und China schwächen könnte.

Aus ethischer Sicht ist es zumeist keine Frage, ob eine Technologie wie generative KI feste Korridore braucht, innerhalb derer sie sicher agieren kann. Mittelstadt et al. (2024) argumentieren, dass LLM menschenrechtliche Prinzipien wie Würde, Gleichheit und Meinungsfreiheit beeinträchtigen können, wenn sie ohne angemessene Kontrollen eingesetzt werden, und plädieren für gesetzliche Regelungen, die menschenrechtliche Prinzipien in den Mittelpunkt stellen.

Der EU AI Act ist geprägt von den Werten der EU-Grundrechtecharta – Nichtdiskriminierung, Transparenz, Verantwortlichkeit und Datenschutz. Am 1. August 2024 in Kraft getreten, ist er die weltweit erste umfassende Regulierung für KI. Er regelt den Einsatz von KI-Systemen in der EU durch einen risikobasierten Ansatz, der je nach Risikoklasse unterschiedliche Anforderungen stellt. Im Bildungssektor stuft der AI Act

nahezu alle KI-Systeme als Hochrisiko-Anwendungen ein. Bspw. dürfen Lehrkräfte die Bewertung von Schülerleistungen weder vollständig noch teilweise an KI-Systeme abgeben.

Wie streng die Vorschriften durchgesetzt werden, und ob Unternehmen wie OpenAI Urheberrechtsklagen oder milliardenschwere Geldstrafen drohen, ist noch unklar. Die EU hat Unternehmen, Wissenschaftler/innen und andere Akteure aufgefordert, an der Ausarbeitung des Regelwerks mitzuwirken. Auch Vertreter großer Tech-Konzerne sollen sich beworben haben. Kritiker fürchten, dass diese die Vorgaben verwässern könnten, um strenge Regulierungen zu verhindern (Coulter, 2024).

KI in Schule und Mathematikdidaktik

Die Potenziale von ChatGPT & Co. für die Bildung wurden schnell untersucht und offene Fragen zu den Auswirkungen diskutiert, etwa ob es ‚Schummeln‘ ist, wenn Schüler/innen ChatGPT zum Verfassen von Arbeiten nutzen. Einige Bundesländer reagierten früh mit Handreichungen und Fortbildungen.

Im März 2023 veröffentlichte der Deutsche Ethikrat die Stellungnahme ‚Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz‘. Der umfangreiche Bericht untersucht vier zentrale Anwendungsbereiche, darunter auch die schulische Bildung (Deutscher Ethikrat, 2023).

2024 veröffentlichte die Ständige Wissenschaftliche Kommission der Kultusministerkonferenz ein Impulspapier zu ‚Large Language Models und ihre Potenziale im Bildungssystem‘. Sie gibt Anregungen für Unterricht und Forschungsaufgaben und diskutiert Erprobungsphasen, domänenspezifische Tools, den altersgemäßen Einsatz, Prüfungskultur, Rahmenbedingungen, Integration in Lernplattformen und Entwicklung von Commons-Lösungen (SWK, 2024). Die ‚Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen‘, die die KMK im Oktober 2024 beschloss, baut auf diesen Impulsen der SWK auf und konkretisiert Handlungsmaßnahmen. Ziel ist eine gemeinsame Position und abgestimmte Schritte für den Einsatz von KI in Schulen (KMK, 2024). Beide Papiere betonen die Wichtigkeit, die ethischen und rechtlichen Fragen des Einsatzes von KI zu klären und das damit verbundene technologische Wissen bereits in der Primarstufe zu vermitteln (ICT-Literacy bzw. Lernen über KI).

Es überrascht nicht, dass das Thema KI auch in der Mathematikdidaktik-Community zunehmend auf Interesse stößt. In mehreren Veranstaltungen wird das Thema aufgegriffen (u. a. beim KI-Symposium in Würzburg 2024 oder bei GDM Jahrestagungen), in Special

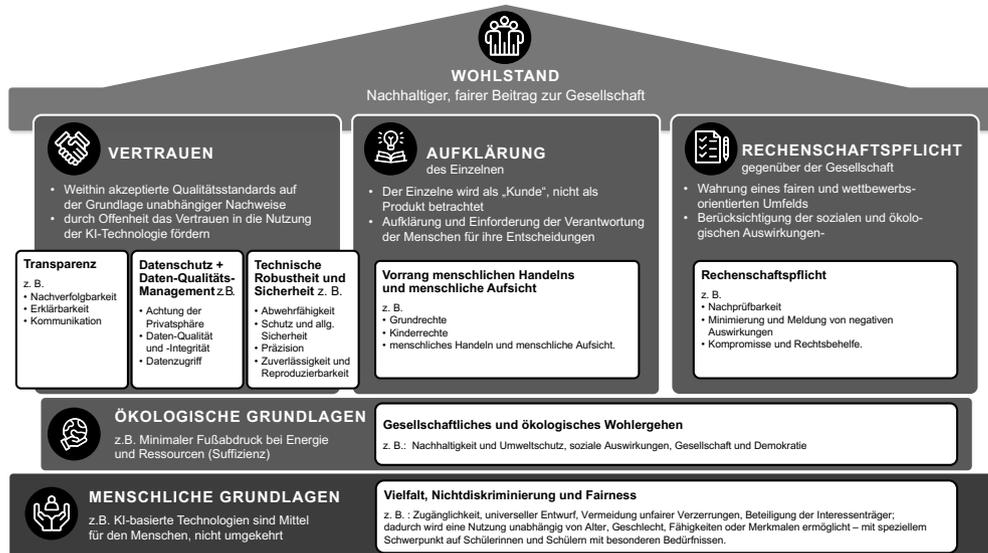


Abbildung 1. Haus der ethischen KI (grau hinterlegt; Vieweg, 2023, S. 252) mit den Ethischen Leitlinien für Lehrkräfte (weiß hinterlegt; Europäische Kommission, 2022, S. 19 f.) (eigene Darstellung)

Issues mathematikdidaktischer Journals und in den *Mitteilungen der GDM* (z. B. Hischer, 2023; Buchholtz et al., 2023, Schorch et al., 2023 oder Dilling et al., 2024). Im Bereich der Primarstufe griff u. a. die Herbsttagung des AK Grundschule das Thema auf (Kortenkamp, 2024). Auch Aspekte der KI-Ethik wurden in den *Mitteilungen der GDM* erwähnt (z. B. Buchholtz et al., 2024; Bescherer et al., 2024).

Jedes Fach, auch Mathematik, soll zur Entwicklung von Kompetenzen für ein Leben in der digitalen Welt beitragen (KMK, 2016). Mathematik eignet sich besonders, da viele KI-Grundlagen auf elementarisierbaren mathematischen Konzepten basieren (z. B. Platz, im Druck). Auch Bescherer et al. (2024) greifen dies auf:

Der Mathematikunterricht sollte zum Mündigwerden durch und gegenüber Mathematik auch unter Beachtung digitaler Perspektiven beitragen. [...] Hierzu gehört insbesondere der im Rahmen gesellschaftlicher Transformationen zunehmende Einsatz von KI in vielen Lebensbereichen, für den Lernende und Lehrkräfte sensibilisiert werden sollten und dem sie nur durch mathematische Bildung kompetent begegnen können. Dies betrifft weitreichende und für die Teilhabe an unserer Gesellschaft relevante Aspekte, z. B. Profiling oder Vergabesysteme, die im Sinne einer Algorithmen-Ethik kritisch diskutiert werden sollen [...]. (Bescherer et al., 2024, S. 10)

Ethische KI?

Vieweg (2023) beschreibt die Dimensionen ethischer KI mithilfe der Metapher des ‚Hauses der ethischen KI‘

(s. Abb. 1). Ziel ist die Schaffung nachhaltigen Wohlstands für die Gesellschaft. KI soll eine Vernetzung ermöglichen, die Würde und Aufstiegschancen für alle sichert. Dafür müssen Begünstigte mit Zugang zu Kapital und Wissen ihre treuhänderische Verantwortung gegenüber Schwächeren wahrnehmen.

Die ‚Ethischen Leitlinien für Lehrkräfte über die Nutzung von KI und Daten für Lehr- und Lernzwecke‘ der Europäischen Kommission definieren Anforderungen, die KI-Systeme im Bildungsbereich erfüllen müssen. Zu den sieben Anforderungsbereichen bieten Leitfragen Lehrkräften Orientierung und regen zur Reflexion ihrer beruflichen Praxis an (Europäische Kommission, 2022). Abb. 1 verknüpft diese Anforderungen mit dem Konzept des ‚Hauses der ethischen KI‘.

An dieser Stelle lohnt sich eine Reflexion der bisherigen Ausführungen. Mithilfe der besprochenen Werte und Rahmenbedingungen kann die Leserin/der Leser ein eigenes kleines Ethik-Audit durchführen und beurteilen, ob ChatGPT oder ein anderes KI-System ihrer/seiner Ansicht nach in die Gruppe der ethischen KI-Anwendungen eingeordnet werden kann. Die Bewertung ist nicht immer einfach, da nach außen oft alles geordnet erscheint – Vorsicht vor ‚Ethics Washing‘: Wie beim ‚Green Washing‘ wird Ethik hier genutzt, um den Anschein einer ethischen KI zu erwecken, ohne substanzielle Theorie oder echte Umsetzung (Schultz et al., 2024).

Schlüsselkompetenzen für die nächste Generation

Fengchun Miao, Leiter der Unit for Technology and AI in Education, UNESCO, formulierte folgende treffende Antwort auf die Frage, welche Rolle Kinder in

der Mensch-KI-Interaktion einnehmen sollen (Miao, 2024): Wir wollen für unsere Kinder, dass sie

- kritische KI-Bürger, keine KI-Süchtigen,
- verantwortungsbewusste Nutzer von KI, nicht Deepfake-Verbreiter,
- Mitgestalter von KI-Werkzeugen, nicht nur passive Nutzer,
- Anführer der nächsten KI-Generationen, nicht nur Konsumenten

werden.

Genau dieses Ziel sollten wir verfolgen. Viele Organisationen haben entsprechende Empfehlungen und Rahmen definiert. Noch vor der Veröffentlichung von ChatGPT wurde von der UNESCO 2021 die ‚Recommendation on the Ethics of Artificial Intelligence‘ verabschiedet. Dort wird die Vermittlung von KI-Literalität an alle Gesellschaftsschichten betont:

Member States should work with international organizations, educational institutions and private and non-governmental entities to provide adequate AI literacy education to the public on all levels in all countries in order to empower people and reduce the digital divides and digital access inequalities resulting from the wide adoption of AI systems. (UNESCO, 2021, S. 33)

Im schulischen Bereich soll dies auch in Deutschland umgesetzt werden:

Für das „Lernen über KI“ sind neben einer grundlegenden informatischen Bildung – insbesondere über KI sowie über ihre Wirkungsweisen – auch die Klärung ihrer ethischen und rechtlichen Rahmenbedingungen bei Lehrkräften, Schülerinnen und Schülern sowie in der Bildungsadministration selbst erforderlich. [...] Die Länder werden bei künftigen Anpassungen ihrer Bildungspläne und curricula- ren Vorgaben aller Fächer und Schulstufen in allen

Schularten informatische Kompetenzen sowie Kompetenzen für die Bildung in der digitalen Welt auch im Hinblick auf die Herausforderungen durch KI integrieren. (KMK, 2024, S. 5)

Zur Frage, welche Kompetenzen zu einer KI-Literalität beitragen, hat die UNESCO 2024 einen KI-Kompetenzrahmen für Schüler/innen (Miao & Shiohira, 2024; siehe Tab. 1) und einen für Lehrpersonen (Miao & Cukurova, 2024; siehe Tab. 2) veröffentlicht.

Die vier Aspekte definieren zentrale KI-Kompetenzen, die Lernende zu verantwortungsvollen Nutzenden und Mitgestaltenden machen und sie auf ihre Rollen in der nächsten KI-Generation vorbereiten sollen. Die drei Entwicklungsstufen fördern zunehmende Beherrschung und ethisches Bewusstsein im Umgang mit KI, wobei Ethik global und lokal verstanden und angewendet werden soll. Praktische Anregungen hierzu bieten z. B. Payne (2019) mit einem Lehrplan zur KI-Ethik für die Mittelstufe, Kersting et al. (2019) mit verständlichen Erklärungen zur KI, Noack und Sanner (2023) mit interaktiven Projekten sowie das Sprachmodell SoekiaGPT (www.soekia.ch), das die Funktionsweise von Textgeneratoren wie ChatGPT verdeutlicht.

Mit zunehmender Integration von KI im Bildungsbereich muss auch die Professionalisierung von Lehrkräften gestärkt werden, um Schüler/innen angemessen zu unterrichten. Die KMK empfiehlt, entsprechende Kompetenzen systematisch in Fachwissenschaften, Fachdidaktiken und Bildungswissenschaften zu integrieren (KMK, 2024). Ergänzend fordert sie ethische Leitlinien, basierend auf denen der Europäische Kommission (2022) sowie den Maßstäben des Deutschen Ethikrats (Deutscher Ethikrat, 2023).

Die Europäische Kommission (2022) nennt in ihren ‚Ethischen Leitlinien für Lehrkräfte über die Nutzung von KI und Daten für Lehr- und Lernzwecke‘ potenzielle Indikatoren, die im Rahmen des Europäischen

Tabelle 1. KI-Kompetenzrahmen für Schüler/innen (übersetzt aus dem Englischen aus Miao & Shiohira, 2024, S. 19)

Kompetenzaspekte	Stufen der Entwicklung		
	Verstehen	Anwenden	Erstellen
Menschenzentrierte Denkweise	Menschliches Handeln	Menschliche Verantwortung	Gesellschaft im Zeitalter von KI
Ethik der KI	Gelebte Ethik	Sichere und verantwortungsbewusste Nutzung	Ethik durch Design
KI-Techniken und -Anwendungen	KI-Grundlagen	Anwendungskennntnisse	Erstellung von KI-Werkzeugen
KI-System-Design	Problemuntersuchung	Architekturdesign	Iterations- und Rückkopplungsschleifen

Tabelle 2. KI-Kompetenzrahmen für Lehrpersonen (übersetzt aus dem Englischen aus Miao & Cukurova, 2024, S. 22)

Aspekte	Entwicklung		
	Erwerben	Vertiefen	Erstellen
Menschenzentrierte Denkweise	Menschliches Handeln	Menschliche Verantwortung	Soziale Verantwortung
Ethik der KI	Ethische Prinzipien	Sichere und verantwortungsbewusste Nutzung	Ko-kreation ethischer Regeln
KI-Grundlagen und -Anwendungen	Grundlegende KI-Techniken und -Anwendungen	Anwendungskennntnisse	Erstellen mit KI
KI-Pädagogik	KI-unterstütztes Lehren	Integration von KI-Pädagogik	KI-gestützte pädagogische Transformation
KI für die berufliche Weiterentwicklung	KI für lebenslanges berufliches Lernen	AI zur Verbesserung des organisatorischen Lernens	KI zur Unterstützung des beruflichen Wandels

Kompetenzrahmens ‚DigCompEdu‘ berücksichtigt werden sollten. Die UNESCO ergänzt dies mit einem KI-Kompetenzrahmen für Lehrpersonen (Miao & Cukurova, 2024; siehe Tabelle 2), der Ethik als zentrale Komponente hervorhebt, und definiert zudem detaillierte curriculare Ziele und Lernziele für jeden der fünfzehn Kompetenzblöcke (Miao & Cukurova, 2024).

Fazit

Generative KI in ihrer heutigen Form ist nicht nur ein aktueller Trend, sondern eine Technologie, die unsere Gesellschaft langfristig prägen wird. Für die Gesellschaft im Ganzen wie auch für den schulischen Bereich bietet sie enorme Chancen, birgt aber auch große ethische Herausforderungen. Ein verantwortungsvoller Umgang mit KI erfordert nicht nur das Lernen *mit* KI, sondern auch ein fundiertes Lernen *über* KI, um Kindern einen dauerhaft souveränen und kritischen Umgang mit dieser sich ständig weiterentwickelnden Technologie zu ermöglichen.

Lehrpersonen spielen dabei eine Schlüsselrolle: Sie müssen nicht nur technologische Kompetenzen vermitteln, sondern auch die ethische Brille aufsetzen, um den Einsatz von KI zu reflektieren und Kinder auf ihre Rolle als mündige Erwachsene vorzubereiten. Dies ist eine anspruchsvolle Aufgabe, da sich die Technologien schnell weiterentwickeln und ständige Anpassungen erforderlich sind. Es ist jedoch eine Investition, von der sowohl die Entscheidungsträger von morgen als auch die Gesellschaft als Ganzes profitieren werden.

Literatur

Andree, M. (2023). *Big Tech muss weg!* Campus Verlag.
 Baker, P. (2024). *ChatGPT für Dummies*. John Wiley & Sons.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 610–623.
 Beschere, C., Büscher, C., Lengnink, K., Pinkernell, G., Reinhold, F., Schacht, F., Schreiber, C. & Walter, D. (2024). Mathematische Bildung und Digitalisierung. *Mitteilungen der GDM*, 117, 6–14.
 Buchholtz, N., Baumanns, L., Huget, J., Peters, F., Pohl, M. & Schorcht, S. (2023). Herausforderungen und Entwicklungsmöglichkeiten für die Mathematikdidaktik durch generative KI-Sprachmodelle. *Mitteilungen der GDM*, 114, 19–26.
 Buchholtz, N., Schorcht, S., Baumanns, L., Huget, J., Noster, N., Rott, B., Siller, H.-S. & Sommerhoff, D. (2024). Damit rechnet niemand! *Mitteilungen der GDM*, 117, 15–24.
 Buxmann, P., & Schmidt, H. (2021). *Künstliche Intelligenz*. Springer.
 CAIS (2023). *Statement on AI Risk*. www.safe.ai/work/statement-on-ai-risk
 Coulter, M. (2024). *Tech giants push to dilute Europe’s AI Act*. Reuters. www.reuters.com/technology/artificial-intelligence/tech-giants-push-dilute-europes-ai-act-2024-09-20/
 Delvaux, M. (2017). *Europäisches Parlament 2014–2019*. Bericht mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik (2015/2103(INL)) A8-0005/2017, S. 9.
 Deutscher Ethikrat (2023). *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz*. www.ethikrat.org/publikationen/stellungnahmen/mensch-und-maschine/
 Dilling, F., Holten, K., Pielsticker, F. & Witzke, I. (2024). Aushandlungs- und Argumentationsprozesse fördern durch den Einsatz generativer KI-Sprachmodelle beim schulischen Mathematiklernen? *Mitteilungen der GDM*, 116, 14–22.
 Dignum, V. (2019). *Responsible artificial intelligence*. Springer.

- Ertel, W. (2021). *Grundkurs Künstliche Intelligenz*. Springer.
- EU AI Act (2024). *Gesetz über künstliche Intelligenz (Verordnung (EU) 2024/1689), Fassung des Amtsblatts vom 13. Juni 2024*. Interinstitutionelle Akte: 2021/0106(COD)
- Europäische Kommission (2022). *Ethische Leitlinien für Lehrkräfte über die Nutzung von KI und Daten für Lehr- und Lernzwecke*. Amt für Veröffentlichungen der Europäischen Union. data.europa.eu/doi/10.2766/494
- EU (2017). Robotic and AI. www.europarl.europa.eu/news/de/press-room/20170210IPR61808/robotik-und-kuenstliche-intelligenz-abgeordnete-fur-eu-weite-haftungsregelungen
- Future of Life (2023). *Pause Giant AI Experiments: An Open Letter*. futureoflife.org/open-letter/pause-giant-ai-experiments/
- Hischer, H. (2023). ChatGPT und Mathematikunterricht – eine didaktische Herausforderung? *Mitteilungen der GDM*, 115, 6–11.
- Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25.
- Kersting, K., Lampert, C., & Rothkopf, C. (2019). *Wie Maschinen lernen*. Springer.
- KI Bundesverband (2023). *Position Paper on the EU AI Act*. ki-verband.de/wp-content/uploads/2023/04/PositionPaper_AI-Act_German-AI-Association_April2023.pdf
- KMK (2016). *Strategie der Kultusministerkonferenz „Bildung in der digitalen Welt“* (Beschluss der Kultusministerkonferenz vom 8.12.2016). www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2016/Bildung_digitale_Welt_Webversion.pdf
- KMK (2024). *Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen*. (Beschluss der Kultusministerkonferenz vom 10.10.2024). 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*. In Steinweg, A. S. (2024) (Hrsg.). *Schule im Wandel – Mathematikunterricht im Wandel: Tagungsband des AK Grundschule in der GDM 2024* (S. 57–72). University of Bamberg Press.
- Krauthausen, G. (2020). Tablets ante portas. In B. Brandt, L. Bröll, & H. Dausend (Hrsg.), *Digitales Lernen in der Grundschule II* (S. 40–50). Waxmann.
- Miao, F. (2024). *AI competency frameworks for school students (under development)*. Vortragsfolien.
- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. unesdoc.unesco.org/ark:/48223/pf0000391104
- Miao, F. & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. unesdoc.unesco.org/ark:/48223/pf0000386693
- Miao, F., & Shiohira, K. (2024). *AI competency framework for students*. UNESCO. unesdoc.unesco.org/ark:/48223/pf0000391105
- Mittelstadt, B., Wachter, S., & Russell, C. (2024). The Unfairness of Fair Machine Learning: Leveling Down and Strict Egalitarianism by Default. *Michigan Technology Law Review*, 30(1), 3.
- Noack, P. & Sanner, S. (2023). *Künstliche Intelligenz verstehen: Eine spielerische Einführung*. Rheinwerk Verlag.
- Payne, B. H. (2019). *An ethics of artificial intelligence curriculum for middle school students*. MIT Media Lab. https://ec.europa.eu/futurium/en/system/files/ged/mit_ai_ethics_education_curriculum.pdf
- Platz, M. (im Druck). Welche Auswirkungen hat KI auf die Internetsuche? Eine Lernumgebung für den Mathematikunterricht der Primarstufe. In M. Platz & A. Steffen-Delplanque (Hrsg.), *Digitale Medien im Mathematikunterricht der Primarstufe 2024. Beiträge zur 7. PriMaMedien-Sommertagung 2024 in Saarbrücken*. Band 12 der Reihe „Lernen, Lehren und Forschen mit digitalen Medien in der Primarstufe“. WTM-Verlag.
- Ramos, G. (o. J.). *Ethics of Artificial Intelligence*. www.unesco.org/en/artificial-intelligence/recommendation-ethics
- Rich, E. (1983). *Artificial Intelligence*. McGraw-Hill.
- Ryan, M., Christodoulou, E., Antoniou, J., & Iordanou, K. (2024). An AI ethics ‘David and Goliath’: Value conflicts between large tech companies and their employees. *AI & SOCIETY*, 39(2), 557–572.
- 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 GDM*, 115, 12–23.
- Schultz, M. D., Conti, L. G., & Seele, P. (2024). Digital ethics-washing: a systematic review and a process-perception-outcome framework. *AI and Ethics*, 1–14.
- Spitzer, M. (2023). *Künstliche Intelligenz: Dem Menschen überlegen*. Droemer.
- Süddeutsche Zeitung (2023). *Was hinter der lauten Warnung vor der KI-Apokalypse steckt*. www.sueddeutsche.de/wirtschaft/ki-kuenstliche-intelligenz-ausloeschung-menschen-open-ai-1.5892228
- SWK (2024). *Large Language Models und ihre Potenziale im Bildungssystem*. Impulspapier der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz. DOI:10.25656/01:28303
- UNESCO (2021). *Recommendation on the Ethics of Artificial Intelligence*. unesdoc.unesco.org/ark:/48223/pf0000381137
- Vieweg, S. H. (2023). *KI für das Gute*. Springer.

Melanie Platz, Universität des Saarlandes
melanie.platz@uni-saarland.de

Christine Plote, Open Search Foundation e. V.
cp@opensearchfoundation.org