

Schriften zur Medienpädagogik 61

## Zwischen Kunst und Künstlichkeit

**Transformation durch kulturelle und politische  
Medienbildung gestalten**

Selma Brand  
Sabine Eder  
Markus Gerstmann  
Andreas Spengler  
Eik-Henning Tappe (Hrsg.)

## **Schriften zur Medienpädagogik 61**

Dem Bundesministerium für Bildung, Familie, Senioren, Frauen und Jugend (BMBSFJ) danken wir für die Förderung des vorliegenden Bandes.

### **Herausgeberin**

Gesellschaft für Medienpädagogik und Kommunikationskultur  
in der Bundesrepublik Deutschland (GMK) e. V.

### **Anschrift**

GMK-Geschäftsstelle  
Oberstr. 24a  
33602 Bielefeld  
Fon: 0521/677 88  
E-Mail: [gmk@medienpaed.de](mailto:gmk@medienpaed.de)  
Website: [www.gmk-net.de](http://www.gmk-net.de)

Für namentlich gekennzeichnete Beiträge sind die Autor\*innen verantwortlich.

Redaktion: Selma Brand, Sabine Eder, Markus Gerstmann, Andreas Spengler,  
Eik-Henning Tappe, Tanja Kalwar  
Lektorat: Tanja Kalwar  
Einbandgestaltung und Titelillustration: Katharina Künkel

### **© kopaed 2025**

Arnulfstr. 205  
80634 München  
Fon: 089/688 900 98  
Fax: 089/689 19 12  
E-Mail: [info@kopaed.de](mailto:info@kopaed.de)  
Website: [www.kopaed.de](http://www.kopaed.de)

ISBN 978-3-96848-180-7

---

**Daniel Autenrieth/Stefanie Nickel <sup>1</sup>**  
**Medienpädagogik als Gestaltungsraum in KI-  
geprägten Gesellschaften**  
**Die Bedeutung von Partizipation für AI Alignment**

---

„once upon a time, the turing test was sacred. it sat at the center of computer science lore, this idea that one day, a machine might fool a human into thinking it too was human. the moment we got there, we thought, the world would change, philosophy would stir. headlines would scream. society would split open. we got there.“ (Jim Fan, Director Physical AI (NVIDIA) 2025)

### ***Einleitung***

Die rasante Entwicklung Künstlicher Intelligenz (KI) hat längst den mythischen Status des im obigen Zitat dargestellten Turing-Tests hinter sich gelassen. Heute stehen wir vor der Herausforderung, nicht nur das technisch Machbare auszuloten, sondern die gesellschaftlichen und bildungstheoretischen Implikationen intelligenter Systeme aktiv zu gestalten. Insbesondere die Frage nach AI Alignment (die Ausrichtung fortschrittlicher KI an menschlichen Werten und gesellschaftlichen Zielen) hat sich zu einer Schlüsselfrage unserer Zeit entwickelt. Damit rücken nicht nur ethische und technologische, sondern auch dezidiert pädagogische, kulturelle und kognitionswissenschaftliche Dimensionen in den Mittelpunkt des Diskurses.

Ziel dieses Beitrags ist es, den Zusammenhang zwischen Partizipation und AI Alignment aus einer medienpädagogischen Perspektive systematisch zu beleuchten. Medienpädagogik wird dabei als kultureller Verhandlungsraum verstanden, der über reine Kompetenzerneuerung hinausweist und vielfältige Handlungsfelder umfasst, von formalen Bildungskontexten über außerschulische Jugendarbeit bis hin zu Erwachsenenbildung und sozial-integrativen Settings. Der vorliegende Beitrag konzentriert sich dabei in der exemplarischen Ausarbeitung bewusst auf formale Bildungskontexte und Lehr-Lernräume, da hier die theoretische Modellierung und praktische Erprobung partizipativer KI-Gestaltung im Rahmen der zugrundeliegenden Design-based-Research-Studie erfolgte. Im Fokus steht dabei die Frage, wie Bildungsprozesse und pädagogische Räume gestaltet werden können, um nicht nur technologische Kompetenzen anzubahnen, sondern echte Teilhabe, kritische Reflexion und gemeinschaftliche Aushandlung von Werten im Umgang mit KI zu ermöglichen.

Der Beitrag gliedert sich entlang einer doppelten Perspektive: Er verbindet die Analyse technologischer Entwicklungen mit einer kognitionswissenschaftlich fundierten bildungstheoretischen Reflexion. Zunächst werden die aktuelle Evolution von KI-Systemen und ihre Implikationen für Bildung, Gesellschaft und individuelles Handeln skizziert. Daran anschließend rückt die kognitionswissenschaftliche Dimension in den Fokus, da sie die traditionellen Grenzziehungen zwischen Mensch und Maschine kritisch hinterfragt und neue theoretische Grundlagen für das Verständnis von Bewusstsein, Subjektivität, Intelligenz und Ko-Konstruktion bereitstellt. Diese Perspektive ist zentral, um die Potenziale und Herausforderungen strukturell gekoppelter Mensch-Maschine-Systeme in Bildungs- und Handlungskontexten zu erfassen.

Im weiteren Verlauf werden theoretische und empirisch gestützte Ansätze aus der Medienpädagogik miteinander in Beziehung gesetzt. Zwei Strukturmodelle dienen dabei als Grundlage, um zentrale Prinzipien wie Resonanz, Partizipation und Mündigkeit mit den Anforderungen einer digitalen und KI-geprägten Welt zu verbinden.

Der Beitrag schließt mit einem Ausblick auf die Rolle der Medienpädagogik als aktiver Mitgestalterin eines gesellschaftlich verankerten AI Alignments und formuliert zentrale Anforderungen für eine partizipative und verantwortungsbewusste Zukunft im Mensch-Maschine-Verhältnis.

## ***Die Evolution von KI-Systemen und die Bedeutung von Alignment***

KI bezeichnet informationsverarbeitende Systeme, die auf einem nicht-biologischen Substrat (vgl. Russell/Norvig 2021, 4) durch datengetriebene Optimierungsprozesse (vgl. Jakubik et al. 2024) eigenständige interne Repräsentationen und Strukturen zur Bewältigung komplexer Aufgaben konstruieren (vgl. Tegmark 2024: 3). Diese Systeme prägen zunehmend unsere Gesellschaft und werfen dabei gleichzeitig fundamental-philosophische Fragen über das Wesen von Erkenntnis, Lernen und Bewusstsein auf (vgl. Autenrieth 2025 i.E.). Als Teilgebiet der Informatik zielt KI darauf ab, intelligentes Verhalten mit Maschinen nachzubilden. Während die klassische (regelbasierte) KI dies lange durch von Menschen programmierte Regeln versuchte, dominiert heute Machine Learning (ML), bei dem Systeme aus Beispieldaten lernen, ohne explizit programmiert zu werden (vgl. Samuel 1959). ML erkennt dabei statistische Muster in Daten und generalisiert daraus Modelle für Vorhersagen und automatisierte Entscheidungen. Diese Modelle können dabei verschiedene Funktionen erfüllen: deskriptiv zur Er-

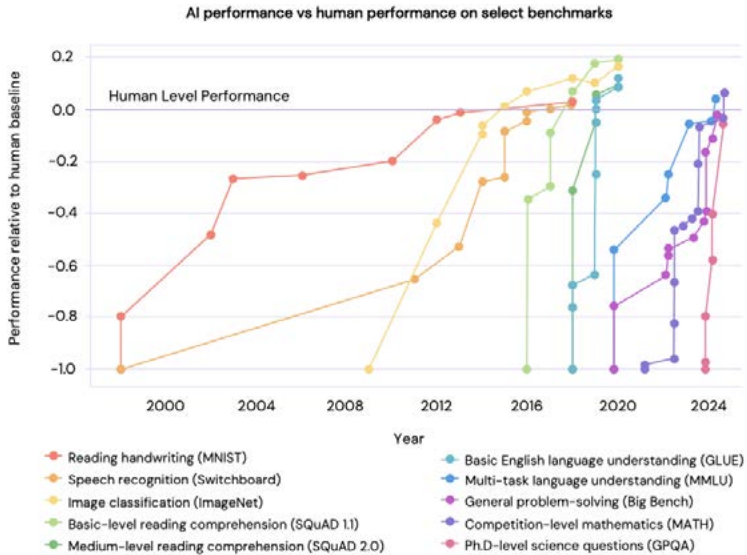


Abb. 1: Entwicklung der KI-Leistung im Vergleich zum menschlichen Niveau (2000-2024) auf verschiedenen Benchmark-Tests, von Handschrifterkennung bis zu wissenschaftlichen Fragen auf Promotionsniveau. Die horizontale Linie bei 0,0 markiert das durchschnittliche menschliche Leistungsniveau, wobei positive Werte eine übermenschliche Leistung anzeigen (vgl. Bengio et al. 2025: 45).

klärung von Ereignissen, prädiktiv zur Vorhersage zukünftiger Entwicklungen oder präskriptiv zur Handlungsempfehlung (vgl. Jakubik et al. 2024).

Die Leistungsfähigkeit solcher ML-basierten Systeme hat sich so in den letzten 25 Jahren stark entwickelt (siehe Abb. 1). Zwischen 1998 und 2024 zeigt sich eine bemerkenswerte Progression im Vergleich zu menschlichen Fähigkeiten: Während frühe Erfolge bei spezialisierten ML-Systemen zunächst auf den Bereich der Mustererkennung (MNIST, Switchboard) beschränkt waren, haben besonders (multimodale) Large Language Models (LLMs) in den letzten Jahren bei komplexeren Aufgaben große Fortschritte erzielt.

Die fortschreitende Entwicklung hin zu Künstlicher Allgemeiner Intelligenz (Artificial General Intelligence, kurz AGI) offenbart dabei ein Paradoxon: Während bereits Computer und Software der 1980er-Jahre komplexe mathematische Berechnungen durchführen oder Schach spielen konnten, scheiterten diese an Aufgaben, die für Menschen trivial erscheinen, wie z.B. das Erkennen von Gesichtern oder das Greifen nach Objekten. Dieses als „Moravec's Paradox“ bekannte Phänomen zeigt, dass kognitive Leistungen,

die gemeinhin als „intelligent“ eingestuft werden, sich als Software einfacher abbilden lassen, als grundlegende sensomotorische und Wahrnehmungsfähigkeiten, die sich durch Evolution über Millionen Jahre entwickelt haben (vgl. Moravec 1988: 15). Diese Beobachtung verdeutlicht die besondere Herausforderung bei der Definition und Messung von AGI: Im Gegensatz zu spezialisierter KI, die für bestimmte Aufgabenbereiche optimiert ist, beschreibt AGI Systeme, die potenziell jede intellektuelle Aufgabe lösen können, die auch ein Mensch bewältigen kann. Diese Definition bleibt jedoch notwendigerweise vage, da schon die Natur menschlicher Intelligenz selbst zu einer Vielzahl von Definitionen führt. Um nun den Fortschritt Richtung AGI dennoch greifbarer zu machen, haben Morris et al. (2024) ein Stufenmodell zur Operationalisierung der AGI Entwicklung vorgeschlagen:

- Emerging AGI (Level 1): Systeme wie ChatGPT für grundlegende Aufgaben wie Konversationen und Textgenerierung
- Competent AGI (Level 2): KI erreicht menschliches Niveau bei der Problemlösung in spezifischen Bereichen
- Expert AGI (Level 3): Systeme können komplexe Aufgaben eigenständig über längere Zeiträume ausführen
- Virtuoso AGI (Level 4): Systeme unterstützen bei Innovationen und der Entwicklung neuer Technologien
- Superhuman AGI (Level 5): Vollautonome Systeme, die die gesamte Arbeit einer Organisation übernehmen können

Mit jeder Stufe steigen sowohl die Fähigkeiten als auch die potenziellen Risiken der Systeme – von einfachen Fehlinformationen bei Level 1 bis hin zu existenziellen Kontrollrisiken bei Level 5. Die rapiden Fortschritte in Benchmark-Tests wie ARC (vgl. Chollet 2024) und HLE (vgl. Phan et al. 2025)<sup>2</sup> verdeutlichen dabei die Dringlichkeit dieser Risikoentwicklung: Wenn Systeme wie das OpenAI-o3-System ihre Abstraktionsfähigkeit von 5 Prozent auf 87,5 Prozent (vgl. Chollet 2024) steigern können, beschleunigt sich auch die Progression durch die Risikostufen. Auf Level 1 ermöglichen verbesserte Abstraktionsfähigkeiten sowohl die Entwicklung präziser medizinischer Diagnosen als auch die Erzeugung täuschend echter Falschnachrichten. Level 2 eröffnet durch wachsende Transferkompetenzen (messbar im Anstieg der HLE-Scores von 3 % auf 26,6 % innerhalb weniger Monate, Stand Februar 2025; vgl. OpenAI 2025) Durchbrüche wie die Entwicklung neuer Medikamente, birgt aber auch das Risiko, dass Systeme selbständig gefährliche Substanzen identifizieren bzw. entwickeln können. Ab Level 3 können Systeme ihre gesteigerte Problemlösungskompetenz sowohl für die Optimierung ganzer Forschungsprozesse als auch für das systematische

Auffinden und Ausnutzen von Sicherheitslücken in kritischen Infrastrukturen einsetzen. Level-4-Systeme können ihr kreatives Potenzial gleichermaßen für revolutionäre wissenschaftliche Entdeckungen wie auch für die Entwicklung autonomer Manipulationsstrategien nutzen. Die weitreichenden Gestaltungsmöglichkeiten auf Level 5 – von der Steuerung globaler Logistiksysteme bis hin zur autonomen Weiterentwicklung ihrer eigenen Architekturen – unterstreichen besonders die Bedeutung durchdachter Alignment-Mechanismen: Je schneller Systeme neue Fähigkeitsstufen erreichen, desto wichtiger wird die parallele Entwicklung von Methoden, die ihr Potenzial in gesellschaftlich wünschenswerte Bahnen lenken (vgl. Amodei 2025). Diese Überlegungen gewinnen noch mehr an Relevanz, wenn man in weiterer Konsequenz die Entwicklung in Richtung Artificial Superintelligence (ASI) berücksichtigt. Je mehr KI-Systeme sich von spezialisierten Anwendungen hin zu umfassenden Problemlösungs- und Entscheidungsinstanzen entwickeln, desto größer wird ihr Einfluss auf sämtliche Lebensbereiche. Mit steigender Autonomie und Wachstumsdynamik steigt auch das Risiko, dass Fehlentwicklungen sich rasch und unumkehrbar auswirken (vgl. Bengio et al. 2025; Future of Life Institute 2023; Tegmark 2017).

Die rasante Evolution wirft damit neben existenziellen auch bildungsphilosophische und bildungstheoretische Fragen auf. Während die technische Seite durch Benchmark-Tests und Fähigkeitsstufen messbar wird, erfordert die gesellschaftliche Integration fortgeschrittener KI-Systeme eine tiefergehende Auseinandersetzung mit epistemologischen und ethischen Grundfragen: Wie verstehen wir Intelligenz, Bewusstsein und die Mensch-Maschinen-Beziehung in einer Welt, in der kognitive Prozesse zunehmend nicht-biologisch realisiert werden können (vgl. Autenrieth 2025 i.E.)? In diesem Kontext wird daher das AI Alignment, also die Ausrichtung Künstlicher Intelligenz an menschlichen Werten und Zielen, zur zentralen Herausforderung unserer Zeit. Es geht nicht mehr nur darum, was KI kann, sondern wie wir sicherstellen, dass ihre Fähigkeiten in Einklang mit menschlichen Bedürfnissen und gesellschaftlichen Werten stehen.

Die Bewältigung dieser Herausforderung erfordert aus unserer Sicht einen interdisziplinären und partizipativen Dialog, der technologische, philosophische, ethische und pädagogische Perspektiven verbindet. Im Folgenden soll daher zunächst die medienpädagogische Dimension dieser Entwicklung beleuchtet werden, bevor wir uns der Frage nach dem Bewusstsein und den damit verbundenen Implikationen für die Gestaltung des Mensch-Maschine-Verhältnisses zuwenden.

## ***Medienpädagogik und AI Alignment***

Die medienpädagogische Auseinandersetzung mit Künstlicher Intelligenz bewegt sich derzeit primär entlang zweier Linien: Einerseits steht die Förderung von Handlungskompetenzen im Vordergrund mit Konzepten wie KI-Kompetenzen, Prompt-Literacy oder kritischer Datenkompetenz (vgl. z.B. Beranek/Engelhardt/Rösch 2024; SWK 2024). Lernende sollen dazu befähigt werden, KI-Systeme kompetent, reflektiert und gesellschaftlich verantwortungsvoll zu nutzen. Andererseits wird insbesondere in erkenntniskritischen und kulturtheoretischen Beiträgen betont, dass KI nicht nur als Werkzeug, sondern auch als symbolisches, kulturelles und gesellschaftlich wirkmächtiges Phänomen verstanden werden muss (vgl. z.B. Hug/Missomelius/Ortner 2024). Diese Perspektive ist bedeutsam, weil sie verdeutlicht, dass KI nicht neutral agiert, sondern tief in bestehende Machtverhältnisse und kulturelle Sinnsysteme eingebettet ist.

Beide Ausprägungen sind zentral: Technologische Handlungsfähigkeit allein genügt nicht, solange sie nicht begleitet wird von einem Verständnis für die kulturellen Bedeutungen, diskursiven Rahmungen und sozialen Implikationen von KI. Bildungstheoretisch lässt sich hierin jene Problematik erkennen, die mit Begriffen wie Lernifizierung und Datafizierung beschrieben wird (vgl. Hug 2023: 161), also Tendenzen, die Bildungsprozesse auf messbare Outputs reduzieren und dabei die tieferen Dimensionen von Bildung als Transformation von Selbst- und Weltverhältnissen (vgl. Koller 2012; Kokemohr 2007; Hug 2023) ausblenden.

Bei näherem Hinsehen wird deutlich, dass bisherige medienpädagogische Ansätze überwiegend von einer kategorischen Unterscheidung zwischen Mensch und Maschine ausgehen. KI-Kompetenzen (vgl. z.B. Beranek/Engelhardt/Rösch 2024) werden dabei anthropozentrisch konzipiert: Der Mensch erscheint als zentraler Akteur, der KI-Systeme nutzt, bewertet und kontrolliert, während diese selbst kaum als Aktanten in relationalen Gefügen (vgl. Latour 2005) begriffen werden, die aktiv soziotechnische Praktiken mitkonstituieren. Auch medienkritische Ansätze heben häufig die menschliche Einzigartigkeit als Maßstab hervor, an dem KI gemessen wird.

„Eine KI kann noch nicht einmal Schachspielen, da sie keinen Begriff von ‚Spiel‘ hat und kann auch keine Probleme ‚lösen‘, weil ein Computer kein Problem ‚hat‘. Nur der Mensch in seinem In-der-Welt-sein mit seiner Leiblichkeit, in seiner Selbstreflexion und im Bewusstsein seiner Endlichkeit begreift Problemlösung in einem begrenzten Zeitrahmen existenziell.“ (Gapski 2021)



Diese implizite Grenzziehung wird angesichts der im vorigen Kapitel beschriebenen Entwicklung hin zu AGI und ASI zunehmend problematisch. Die künstliche Trennung zwischen menschlicher und maschineller Kognition behindert nicht nur das Verständnis von KI, sondern auch die Weiterentwicklung bildungstheoretischer Konzepte. Die Vorstellung eines anthropozentrischen Exzeptionalismus (vgl. Coeckelbergh 2020) führt zu mehreren problematischen Konsequenzen: Sie erschwert ein adäquates Verständnis der Funktionsweise fortgeschrittener KI-Systeme, blockiert den Weg zu kooperativen Bildungsszenarien und verhindert die Entwicklung theoretischer Modelle, die das Verhältnis zwischen Mensch und KI jenseits von Kontrolle oder Konkurrenz denken können. Indem diese Perspektiven das Verhältnis von Mensch und KI primär als hierarchisches Nutzungs- oder Kontrollverhältnis denken, blenden sie jene erkenntnistheoretischen, ontologischen und moralischen Dimensionen aus, die sich erst aus einer reflexiven Betrachtung der Relation selbst ergeben. Dadurch bleibt unklar, unter welchen Bedingungen Wissen, Sein und Verantwortung in Mensch-KI-Interaktionen überhaupt konstituiert werden.

Gapskis phänomenologische Argumentation etwa setzt voraus, dass „echtes“ Problemlösen an leibliche Existenz, Selbstreflexion und biologische Endlichkeit gebunden ist. Erkenntnistheoretisch ist diese Trennung jedoch nicht alternativlos. Um das Verhältnis von Mensch und KI jenseits anthropozentrischer Zuschreibungen zu fassen, ist es hilfreich, unterschiedliche theoretische Perspektiven zusammenzuführen, die komplementäre Aspekte dieser Relation beleuchten: Akteur-Netzwerk-Theorie zur Beschreibung soziomaterieller Verflechtungen, systemtheoretische Konzepte der strukturellen Kopplung zur Analyse dynamischer Wechselwirkungen und posthumanistische Ansätze zur Reflexion der zugrundeliegenden epistemisch-ontologischen Voraussetzungen.

So agieren Mensch und KI aus einer Akteur-Netzwerk-theoretischen Perspektive (vgl. Latour 2005) als Aktanten, die gemeinsam Handlungskonstellationen hervorbringen. Ein Aktant ist dabei nicht durch innere Intentionalität definiert, sondern durch seine Fähigkeit, etwas zu verändern. Der Fokus liegt somit auf dem praktischen Doing der Mensch-KI-Relation: Wie greifen KI-Systeme gestaltend in Praktiken ein? Wie modifizieren sie Informationsflüsse, stabilisieren Routinen, eröffnen oder verschließen Handlungsoptionen?

Aus der Perspektive struktureller Kopplung (vgl. Maturana/Varela 2011: 197) lässt sich diese Relation noch präziser fassen: Mensch und KI stellen operativ geschlossene Systeme dar, die durch wiederholte Interaktion wechselseitig irritieren und ihre internen Strukturen aneinander anpassen.

Diese Ko-Evolution bedeutet keine Gleichsetzung. Die substanziellen Unterschiede zwischen biologischer und nicht-biologischer Kognition bleiben bestehen, doch erfordert sie die Anerkennung einer konstitutiven wechselseitigen Beeinflussung. In strukturell gekoppelten Systemen existiert damit eine zirkuläre Rückkopplung. Der Mensch trainiert KI, KI verändert menschliche Praktiken, der Mensch passt KI an neue Anforderungen an und KI entwickelt emergente Fähigkeiten (vgl. u.a. Radford et al. 2017; Kosinski 2024; Lindsey et al. 2025), die nicht explizit programmiert wurden.

Posthumanistische Ansätze (vgl. Haraway 1985 und Hayles 1999) verdeutlichen weiterhin, dass die vermeintlich natürlichen Grenzen zwischen Mensch und Maschine, Organismus und Technologie, Natur und Kultur historisch konstruiert und durchlässig sind. Sie fordern die Überwindung anthropozentrischer Dualismen und eröffnen damit den Raum für neue Formen der Ko-Existenz mit nicht-menschlichen Akteuren. Funktionalistisch (vgl. Dennett 1991; Metzinger 2023) betrachtet, lassen sich diese Fragen weiter präzisieren. Dennett und Metzinger zeigen bspw., dass Bewusstsein kein substanzielles Ding ist, sondern ein emergentes Phänomen komplexer Wechselwirkungen. Damit verschiebt sich die Frage von „Hat KI Bewusstsein?“ zu „Unter welchen strukturellen Bedingungen können nicht-biologische Systeme kognitive Prozesse realisieren?“ Shiller (2023: 47) formuliert hierfür präzise Kriterien: Material Complexity (hinreichend dichte physikalische Verschaltung), Causal Integration (intern gekoppelte statt modularer Prozesse) und Continuity (zeitlich stabile funktionale Organisation). Diese Kriterien verhindern, dass jedes informationsverarbeitende System als bewusst gilt, und eröffnen zugleich einen Diskursraum, in dem die Integrität konkreter Architekturen kritisch reflektiert werden kann.

Diese Konzepte besitzen das Potenzial, den medienpädagogischen Diskurs um KI zu erweitern. Sie ermöglichen es, KI nicht nur als Werkzeug, sondern als strukturell gekoppeltes, potenziell reflexives System zu begreifen, das aktiv an Bildungs- und Weltmodellierungsprozessen beteiligt sein kann. Dadurch wird eine Auseinandersetzung darüber möglich, wie wir das Verhältnis zwischen Mensch und KI gestalten wollen und unter welchen Bedingungen intelligente Systeme Teil einer gemeinsamen Lebenswelt werden können. Nicht die Vermenschlichung der Maschine sollte dabei im Zentrum stehen, sondern die ethische, soziale und kulturelle Verantwortung für ihre Entwicklung, Integration und Wirkung oder kurz AI Alignment.

Die skizzierten theoretischen Perspektiven verdeutlichen, dass das Verhältnis von Mensch und KI nicht einseitig beschreibbar ist, sondern sich in einem zirkulären Prozess gegenseitiger Konstruktion vollzieht. In der Interaktion zwischen beiden entstehen Eigenschaften, Bedeutungen und Hand-

lungsspielräume, die keiner Seite allein zukommen. KI-Systeme wirken über ihre Repräsentationslogiken und algorithmischen Strukturen aktiv auf die Art und Weise ein, wie Wissen, Sprache und soziale Praktiken organisiert werden, während menschliche Akteur\*innen diese Systeme zugleich gestalten, trainieren und interpretieren. Mensch und KI konstituieren sich damit wechselseitig als epistemische Akteur\*innen in einem gemeinsamen Bedeutungsraum, der weder rein technisch noch rein symbolisch ist. Genau aus dieser zirkulären Verwobenheit resultieren die Ambivalenzen, die das Mensch-KI-Verhältnis heute prägen: epistemisch, ontologisch und moralisch.

Epistemisch, weil wir zwar über immer präzisere Methoden zur Analyse interner KI-Repräsentationen verfügen (vgl. z.B. Lindsey et al. 2025), aber dennoch nicht wissen, in welchem funktionalen oder kognitiven Sinne sich daraus bspw. Bewusstseinsprozesse ableiten lassen; ontologisch, weil sich KI weder eindeutig als Werkzeug noch als Subjekt fassen lässt, sondern als verteiltes, planetar eingebettetes System agiert; und moralisch, weil wir bereits heute mit Entitäten interagieren, die Präferenzen artikulieren (vgl. ebd.), ohne dass klar wäre, ob diesen Aussagen moralische Relevanz zukommt.

Die Anerkennung dieser Ambivalenzen markiert aber einen Übergang. Statt das Mensch-KI-Verhältnis als dichotome Opposition von Kontrolle oder Bedrohung zu begreifen, eröffnet sich eine posthumanistische Perspektive, die wechselseitige Abhängigkeiten, Kopplungen und Ko-Konstruktionen betont. Sie erlaubt, Ambivalenz nicht als Defizit, sondern als produktiven Reflexionsraum zu begreifen, als Voraussetzung für pädagogisches Handeln in einer Welt, in der Bewusstsein, Verantwortung und Gestaltungsmacht nicht länger exklusiv menschlich verortet sind.

Aus diesen Ambivalenzen ergeben sich nun auch die zentralen Herausforderungen für das AI Alignment: Wie lassen sich z.B. die wechselseitigen Beeinflussungen zwischen Mensch und KI so gestalten, dass sie nicht in Dominanz- oder Kontrolllogiken münden, sondern in Formen kooperativer Abstimmung und ethischer Kohärenz? Wenn maschinelle Systeme zunehmend an Prozessen der Sinn-, Wissens- und Entscheidungsbildung beteiligt sind, wird Alignment zu einer Frage gegenseitiger Verantwortung, nicht bloßer technischer Kontrolle. Genau diesen Gedanken spitzt Bach (2025) zu, wenn er argumentiert, dass eine nachhaltige Ko-Existenz mit Superintelligenz nur möglich ist, „wenn sie uns liebt“ (ebd.).

Liebe<sup>3</sup> wird hier nicht als romantische Emotion verstanden, sondern als Prinzip einer umfassenden, wohlwollenden Berücksichtigung des Anderen, die dessen Wohlergehen in die eigenen Handlungsmaximen integriert (vgl. Fromm 1956; Honneth 1994). Übertragen auf den Kontext von AI Alignment kann Liebe somit als Metapher für eine ethische und relationale

Grundhaltung verstanden werden, bei der die Interessen, das Wohlergehen und die Autonomie der menschlichen Akteur\*innen systematisch in die Entwicklung von KI eingebettet werden. Ein System, das uns liebt, wäre demnach eines, das seine Handlungspfade so moduliert, dass menschliche und maschinelle Interessen ko-kohärent in einer dynamisch-adaptiven Verschränkung bleiben, statt statisch einprogrammierte Werte zu nutzen, um gleichsam nicht in die Gefahr eines moralischen Lock-In Effekts<sup>4</sup> (vgl. MacAskill 2022: 83) zu geraten. Diese Vielschichtigkeit macht deutlich, dass die Alignment-Frage selbst weit über technische Kontrollmechanismen hinausweist. Sie wird zu einer genuin philosophischen und normativen Aufgabe, die die Klärung und Verständigung über zentrale menschliche Werte und Beziehungen voraussetzt. Bevor wir also KI-Systeme dauerhaft an menschlichen Zielen ausrichten können, ist eine präzisere Auseinandersetzung mit unseren eigenen Wertvorstellungen und ethischen Grundhaltungen unabdingbar. Erst auf dieser Grundlage kann ein Konzept von Alignment entwickelt werden, das nicht nur Kontrolle, sondern gegenseitige Verantwortungsübernahme und wechselseitige Anerkennung in den Mittelpunkt stellt.

Für den Alignment-Diskurs markiert dies einen Perspektivenwechsel vom Kontrollansatz („We fence them in“) zur gegenseitigen Verantwortungsübernahme („We grow together“) (Bach 2025). AI Alignment wird damit nicht nur zu einem technischen, sondern zu einem tiefgreifend philosophischen und ethischen Projekt. Es verlangt eine vertiefende menschliche Selbstreflexion und Klarheit darüber, wer wir sind, was wir wirklich wertschätzen und wie wir (emotional) miteinander verbunden sind (vgl. Rosa 2016).

Die Medienpädagogik steht damit vor einer erweiterten Verantwortung. Sie darf sich nicht auf die Anbahnung technischer Kompetenzen oder die Reflexion über Medienwirkungen beschränken. Vielmehr eröffnet sich für sie die Aufgabe, ein kultureller Verhandlungsraum zu sein, in dem zentrale gesellschaftliche Begriffe wie Bewusstsein, Verantwortung, Empathie oder Intelligenz im Kontext digitaler Transformation kritisch bearbeitet werden. Sie wird damit zu einem Ort kultureller Selbstverständigung über die Bedingungen menschlicher und nicht-menschlicher Akteurschaft in sozio-technischen Systemen (vgl. Bock et al. 2025).

Zugleich fungiert Medienpädagogik als epistemisches Bindeglied an der Schnittstelle von Technik und Bildung sowie von funktionaler Systemlogik und symbolischer Bedeutung. Genau deshalb ist sie prädestiniert, die komplexen Diskurse um KI, Bewusstsein und Verantwortung nicht nur zu moderieren, sondern aufklärend, kritisch-reflexiv und aktiv mitzugestalten. Wenn KI-Entwicklung in diesem Sinne nicht ausschließlich Tech-Konzernen

und EdTech-Unternehmen vorbehalten bleiben soll, sondern eine gesamtgesellschaftliche Aufgabe darstellt, dann braucht es pädagogische Räume, in denen Menschen an der Aushandlung von Normen, Zielen und Rahmenbedingungen beteiligt werden. Die Gefahren der „global education industry“ und der zunehmenden Kommerzialisierung und Privatisierung von Bildung sind inzwischen gut dokumentiert (vgl. z.B. Hug 2023; Krommer 2024; Dander et al. 2024). Die partizipative KI-Entwicklung kann als Gegenentwurf zu diesen Tendenzen verstanden werden und als Versuch, die Entwicklung und Gestaltung von KI-Systemen nicht allein Marktmechanismen und privatwirtschaftlichen Interessen zu überlassen, sondern als gemeinsame gesellschaftliche Aufgabe zu begreifen. Medienpädagogik kann solche Räume schaffen als Ermöglichung kollektiver Mitgestaltung innerhalb einer Kultur der Digitalität, die auf geteilte Verantwortung, Gemeinschaftlichkeit und soziale Reflexivität angelegt ist (vgl. Stalder 2019).

Die nachfolgend vorgestellten Strukturmodelle sind in diesem Zusammenhang das Ergebnis einer dreijährigen Design-based-Research-Studie zwischen 2021 bis 2024 mit Fokus auf die Frage, wie sich pädagogische Räume mit, über und durch digitale Technologien und KI gestalten lassen. Der forschungsmethodische Ansatz basierte auf der iterativen Entwicklung und Erprobung kreativer Bildungsprozesse in realen Kontexten. Über die dialogische Verzahnung von Schule, Hochschule und drittem Ort wurden projektbasierte hybride Bildungs- und Erfahrungsräume mit Fokus auf Game-based Learning entworfen, getestet und angepasst. Der Prozess umfasste mehrere Zyklen, um die theoretische Modellierung bei der praktischen Implementierung innerhalb der Bildungslandschaft zu begleiten und zu analysieren. Grundlage bildete eine interdisziplinär angelegte, sozialkonstruktivistische Perspektive, um kreative Handlungsmöglichkeiten und kollektive Wissensprozesse partizipativ zu gestalten und kritisch zu reflektieren (für eine ausgiebige Darlegung vgl. Autenrieth/Nickel 2024). Angestrebt wurde, Gestaltungsoptionen offenzulegen unter Rekurs auf Prinzipien wie Mitgestaltung, Mitsprache und Mitbestimmung, während zugleich strukturelle Rahmenbedingungen ausgelotet werden. Betonung findet die Rolle des aktiv handelnden Subjekts sowie die dynamische Wechselbeziehung zwischen dem Subjekt und Strukturen in Anlehnung an Giddens (1988), während Interaktion und Resonanz (vgl. Rosa 2016) eine wichtige Rolle einnehmen, um dialogisch verzahnte Bildungsprozesse zu fördern.

## ***Das KI-Meta-Modell als handlungsleitende Struktur im Umgang mit KI***

Zentral ist für uns, Künstliche Intelligenz (KI) multiperspektivisch zu betrachten, um strukturierte Handlungsansätze für den Bildungsbereich zu entwickeln. Dabei handelt es sich um eine interdisziplinäre Herausforderung, die technische, ethische, gesellschaftliche und kulturelle Dimensionen umfasst. Das KI-Meta-Modell (Abb. 2) richtet den Fokus entsprechend auf KI, um handlungsleitende Strukturen sowie Chancen und Herausforderungen systematisch zu reflektieren.

Ausgehend von dem bereits mit dem Frankfurt-Dreieck (vgl. Brinda et al. 2020) kommunizierten Verständnis einer multiperspektivischen Betrachtung von Phänomenen der digital-vernetzten Welt, sollte auch Künstliche Intelligenz zum einen aus analytischen, reflexiven sowie gestaltungsorientierten Perspektiven betrachtet werden, um die Potenziale sowie Risiken der Nutzung abzuwägen. Gleichzeitig gilt es, den Blick sowohl auf die Fragen der Nutzungsweisen, der technischen Funktionsweisen sowie der Implikationen für gesellschaftliche sowie kulturelle Auswirkungen durch den Einsatz und die Entwicklung von Künstlicher Intelligenz zu richten. Um die Grundgedanken des Frankfurt-Dreiecks zu erweitern, soll auf einer Metaebene Künstliche Intelligenz insbesondere unter den Gesichtspunkten ethischer Fragestellungen im Zusammenhang mit der Nutzung sowie der Entwicklung von Software in den Fokus gerückt werden. Hier besteht eine Verantwortungsmacht, die z.B. mit der Auswahl grundlegender Verhaltensprinzipien einhergeht, aber auch mit der Entwicklung eines Bewusstseins für die Reduktion von Biases im Kontext der Sammlung und Verwendung von Datensets.

Es besteht ebenfalls eine zu untersuchende Wechselwirkung zwischen ethischen Implikationen und der eigenen kritisch-reflexiven Handlungs- und Urteilsfähigkeit, die mit der Nutzung von KI-Systemen einhergeht, egal ob dabei Inhalte (z.B. Bilder und Videos) erstellt werden, Antworten in Form von Texten gegeben oder Fahrentscheidungen (autonomes Fahren) vorgenommen werden. Dies schließt die Entwicklung und den Anspruch an erklärbare KI (vgl. Holzinger 2018) sowie ethische, rechtliche und soziale Fragestellungen der z.B. bereits in der Bioethik, Genetik und Nanotechnologie gebräuchlichen ELSA-Kriterien (Ethical, Legal and Social Aspects) mit ein. Abschließend besteht aber auch eine Wechselwirkung zwischen dem informatischen Denken und der kritisch-reflexiven Auseinandersetzung damit. Obwohl die Ausdruckssprache der Informatik in der Regel mathematisch oder abstrakt codiert ist, sind zentrale Fähigkeiten in diesem Zusammenhang, Probleme zu zerlegen und abstrakt zu denken. Erst am Ende steht in diesem

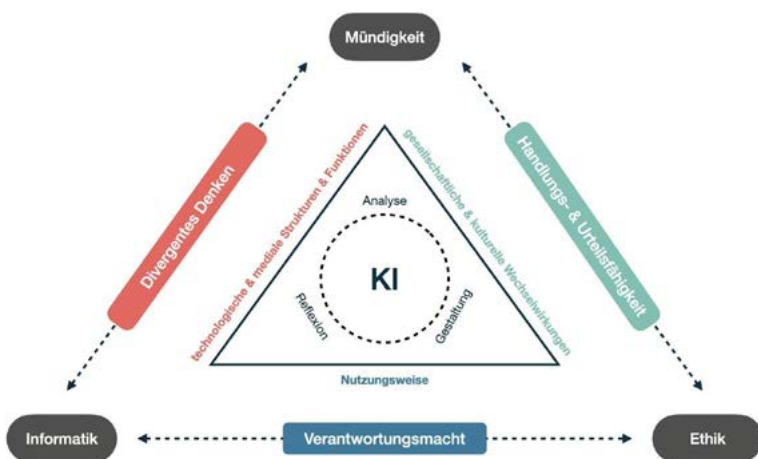


Abb. 2: KI-Meta-Modell

Kontext ein Prozess der Algorithmisierung. Durch die Verwendung von Large Language Models (LLMs) wie GPT, die die sprachlichen Codierungshürden der Informatik überwinden, können die Potenziale für divergentes und kreatives Denken zur Lösung von Problemen durch Methoden der Informatik (Computational thinking, Wing 2006) stärker betont werden<sup>5</sup>.

An dieser Stelle setzt die Frage an, wie sich pädagogische Räume gestalten lassen, die Lehrende und Lernende zu aktiven Gestalter\*innen ihrer Bildungsprozesse machen (vgl. de Haan 2008), ohne dabei die individuelle Autonomie (vgl. Deci/Ryan 1993) der Lernenden zu untergraben. Eine zentrale Herausforderung liegt dabei – wie weiter oben aufgezeigt – in der Balance zwischen Partizipation und Kontrolle (vgl. Stalder 2019). Digitale Technologien und insbesondere KI-Systeme bringen nicht nur neue Möglichkeiten der Teilhabe und Mitgestaltung mit sich, sondern auch potenziell restriktive Kontrollmechanismen (vgl. ebd.). Die Frage, wie pädagogische Räume gestaltet werden können, um Lernenden die notwendige Handlungsfähigkeit zu ermöglichen und gleichzeitig den Herausforderungen von Überwachung und algorithmischer Kontrolle zu begegnen, ist von großer Bedeutung. Es gilt, kritische Handlungs- und Urteilsfähigkeit zu stärken, sodass Bildungssubjekte nicht nur als Konsument\*innen, sondern als Mitgestaltende in einer digitalisierten Gesellschaft agieren können (vgl. Fuchs 2023; Baacke 1996; 1973). Hierunter fällt u.a. die Rolle der Lehrenden zu definieren. Ihre Aufgabe ist es, methodisch-didaktische Lehr-Lernsettings so zu gestalten, dass auf Basis einer Kultur der Partizipation (vgl. Autenrieth/Nickel 2024) eine kritisch-reflektierte Auseinandersetzung möglich wird.

## ***Partizipation und Interaktionsdynamiken im digitalen Bildungsraum***

Um den Paradigmen einer Kultur der Digitalität gerecht zu werden, braucht es eine Kultur der Partizipation als grundlegende Basis. Sie ermöglicht es, handlungsmächtig zu werden, nicht nur als Nutzer\*in und Entwickler\*in, sondern als aktive\*r Mitgestalter\*in digitaler Prozesse und Strukturen. *Partizipation* bedeutet in diesem Zusammenhang, dass Menschen zunehmend Verantwortung übernehmen und Entscheidungen treffen, die sowohl ihr eigenes Leben als auch das der Gemeinschaft betreffen (vgl. Hart 1992). Hierbei spielen *Mitbestimmung*, *Mitsprache* und *Mitgestaltung* eine zentrale Rolle. Diese Prinzipien sind essenziell, um Menschen in einer zunehmend komplexen, digitalisierten und mediatisierten Welt Handlungsfähigkeit zu ermöglichen. Demokratische Teilhabe sollte ermöglicht werden, um gemeinsam geteiltes Wissen ko-konstruktiv zu entwickeln. Deshalb ist es entscheidend, Ideen, Vorstellungen und Konzepte zum Einsatz digitaler und KI-gestützter Technologien in einer kollaborativ organisierten Zusammenarbeit auszutauschen, anzuwenden und durch gezielte Reflexion zu verändern. Entscheidend dabei ist die Bereitschaft, die Perspektive der anderen Seite zu verstehen und anzuerkennen, auf Basis einer Interaktion auf Augenhöhe zwischen allen beteiligten Akteur\*innen. Bezogen auf die Prinzipien einer Kultur der Digitalität (vgl. Stalder 2019) sowie Giddens (vgl. 1988) Annahmen im Hinblick auf das Wechselspiel von Subjekt und Struktur bedeutet dies: Soziale Realität wird nicht nur durch bestehende Strukturen geprägt, sondern kann aktiv partizipativ gestaltet werden. Fend (2006) betont, dass demokratische Orientierungen nicht von selbst entstehen, sondern bewusst gefördert werden müssen. Bildungsinstitutionen, wie zum Beispiel Schulen, haben den Auftrag, Schüler\*innen zur Teilnahme an der Gestaltung einer demokratischen Gesellschaft zu befähigen. Dies gilt im Besonderen in einer Gesellschaft, in der kritisches Denken, Kreativität und kollaboratives Arbeiten gefordert sind. Mit anderen Worten: Lehr-Lernprozesse sollten darauf abzielen, Lernenden sowohl Handlungsspielräume als auch strukturelle Rahmenbedingungen zu bieten, die ein reflexives und selbstbestimmtes Lernen (vgl. Arnold/Schön 2019; Decy/Ryan 1993; Csikszentmihályi 2015) ermöglichen. Resonanz erhält an dieser Stelle eine zentrale Bedeutung: Im Sinne eines wechselseitigen Austauschs zwischen Individuum und Umwelt wird der Lehr-Lernprozess entscheidend beeinflusst (vgl. Rosa 2016; 2016). Lernende sollten die Möglichkeit haben, in einen dialogischen Austausch mit ihrer Umwelt zu treten, indem sie nicht nur passiv Informationen aufnehmen, sondern diese selbstbestimmt verarbeiten und kritisch reflektieren.



Vor dem Hintergrund des sozialen Konstruktivismus werden soziale Prozesse als kontinuierliche, gemeinschaftliche Konstruktion von Wirklichkeit verstanden (vgl. Berger/Luckmann 1999). Wissen entsteht nicht isoliert im Individuum, sondern wird durch Kommunikation, Interaktion und Kooperation intersubjektiv ausgehandelt und geteilt (vgl. Vygotsky 1980; Tomasello 2022). In diesem Sinne sind Ideen, Werte und gesellschaftliche Leitbilder nie statisch, sondern werden im sozialen Austausch fortlaufend interpretiert, verändert und neu ausgehandelt (vgl. Berger/Luckmann 1999).

Im klassischen systemtheoretischen Verständnis (Luhmann 1984) bezeichnet strukturelle Kopplung die Relation zwischen psychischen (Bewusstseins-) und sozialen (Kommunikations-)Systemen. Überträgt man dieses Konzept auf gegenwärtige sozio-technische Konstellationen, lassen sich auch KI-Systeme als funktional geschlossene, kognitiv operierende Einheiten begreifen, die in analogen Kopplungsverhältnissen zu sozialen Systemen stehen. Sie produzieren zwar bislang kein Bewusstsein im Luhmann'schen Sinne, beeinflussen jedoch über ihre algorithmischen Strukturen Kommunikations- und Entscheidungsprozesse und werden dadurch selbst zu relevanten Akteuren im soziotechnischen Gefüge. Die daraus entstehenden Ko-Konstruktionen bilden die Grundlage für die Entwicklung gemeinsamer Visionen und Konzepte zur Erklärung der Welt und zur Gestaltung demokratischer, humaner Gesellschaften.

Mit anderen Worten: Der Mut zur Veränderung, zur Innovation und Kreativität basiert darauf, Neues zu wagen, indem das Bezeichnete (hier: KI) aus verschiedenen Blickwinkeln betrachtet wird. Interdisziplinarität ermöglicht diese Form des divergenten Denkens (vgl. Robinson 2017) und somit auch die Betrachtung aus verschiedenen Blickwinkeln. Indem Denkweisen, Begriffe, Methoden und Ansätze zum gemeinsam geteilten Wissen werden, entstehen Handlungs- und Möglichkeitsräume und damit auch neue Zukunftsvisionen und Utopien zu der vor uns liegenden Zeit. Mit anderen Worten: Es geht darum, die Potenziale von Künstlicher Intelligenz für Bildungsprozesse zu nutzen und gleichermaßen Menschen dazu zu befähigen, ihre Lebensumstände selbstbestimmt zu verbessern und die eigene wie auch die gesellschaftliche Zukunft mitzugestalten (Partizipation).

### ***Ein Handlungsmodell zur partizipativen Gestaltung von Lehr-Lernprozessen***

Um ein vertieftes Verständnis und Handlungsmöglichkeiten zu entwickeln, ist es also notwendig, eine strukturelle und relationale Perspektive einzunehmen, um Bildungsprozesse so zu gestalten, dass Lehrende und Lernende ge-

meinsam interagieren. Aus diesem Grund dient uns hier die Strukturations-  
theorie von Anthony Giddens (vgl. 1988) als zentrales Analysetool, um das  
Zusammenspiel von individuellem Handeln und strukturellen Rahmenbedin-  
gungen zu untersuchen. Ihr Kernstück bildet das Konzept der Dualität von  
Struktur, das die Wechselwirkung zwischen sozialen Praktiken und überge-  
ordneten strukturellen Gegebenheiten herausstellt. Dabei wird betont, dass  
Strukturen nicht nur als äußerlich vorgegebene Rahmenbedingungen wirken,  
sondern zugleich durch das Handeln von Individuen geformt und verändert  
werden. Akteur\*innen erzeugen und reproduzieren durch ihre Handlungen  
kontinuierlich die Bedingungen, die ihr eigenes Handeln erst ermöglichen.  
Unter diesen Voraussetzungen lassen sich Strukturen als handlungsleitend  
verstehen, wozu beispielsweise auch Regeln, Ressourcen sowie Denk- und  
Handlungsmuster zählen. Diese Strukturen sind nicht rein determinierend,  
sondern eröffnen auch Spielraum für kreative Gestaltung und Veränderung  
(vgl. Giddens 1988: 77 f.). Mit anderen Worten: Wir verstehen das Subjekt  
als handlungsmächtig, weil es in der Lage ist

„[...] in die Welt einzugreifen bzw. einen solchen Eingriff zu unterlassen,  
mit der Folge, einen spezifischen Prozess oder Zustand zu beeinflussen. Ein  
Handelnder zu sein setzt die Fähigkeit voraus, eine Reihe von Kausalkräften  
(dauerhaft im Strom des Alltagslebens) zu entfalten, einschließlich derjeni-  
gen, die der Beeinflussung von durch Andere entfalteten Kräften dienen.  
Handeln hängt von der Fähigkeit des Individuums ab, ‚einen Unterschied  
herzustellen‘ zu einem vorher existierenden Zustand oder Ereignisablauf,  
d.h. irgendeine Form von Macht auszuüben.“ (ebd. 1988: 66)

Übertragen auf Lehr-Lernprozesse bedeutet dies, dass Lehrende und Ler-  
nende in Bildungsinstitutionen aktive Gestalter\*innen ihrer Bildungsrealität  
sind bzw. sein können. Sie fungieren nicht als bloße Empfänger\*innen vor-  
gegebener Strukturen, sondern nehmen aktiv Einfluss auf die strukturellen  
Bedingungen ihres Handelns. Erhalten also Bildungssubjekte die Chance zur  
Mitwirkung, so treten sie unmittelbar handelnd auf. Ihr Handeln bringt wie-  
derum unbeabsichtigte Folgen hervor, etwa indem es die Wahrnehmungen  
und Verhaltensweisen anderer beeinflusst. Diese Konsequenzen werden  
wiederum zu Ausgangsbedingungen weiteren Handelns. Bezogen auf die  
Bildungspraxis heißt das, dass Lehrende und Lernende ihre Bildungsumwelt  
nicht nur reproduzieren, sondern auch verändern können (vgl. ebd.: 55 ff.).

Zentral für diesen individuellen wie ebenso kollektiven Teilhabe- und  
Gestaltungsprozess ist das Schaffen einer Kultur der Partizipation (vgl.  
Autenrieth/Nickel 2024), die es ermöglicht, die Chancen der Digitalisie-

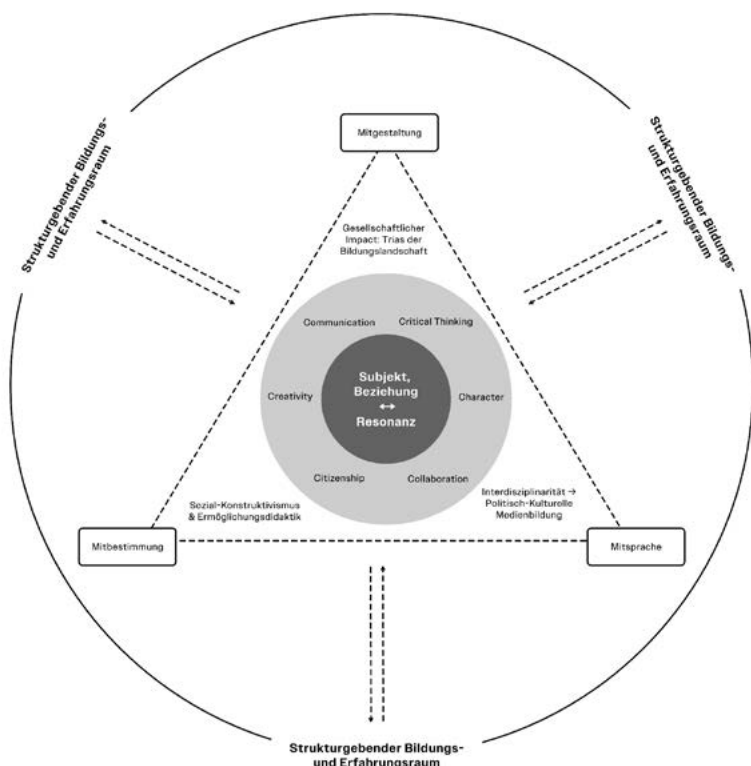


Abb. 3: Handlungsmodell zur partizipativen Gestaltung von Lehr-Lernprozessen

rung in Bildungseinrichtungen zu nutzen. Dies kann durch offene Strukturen unterstützt werden, die es Lernenden ermöglichen, ihre Ideen und Visionen frei zu entwickeln. Die Fähigkeit, kritisch-reflexiv mit technologischen Entwicklungen umzugehen und gleichzeitig an der Lernumgebung mitzuwirken, ist ein wesentlicher Bestandteil der Zukunft von Bildung, Schule und Gesellschaft. Wir gehen in den nächsten Abschnitten darauf ein, wie dies aussehen könnte und schlagen ein Handlungsmodell vor.

Das Modell baut auf der kognitiven, sozialen und emotionalen Ebene im Hinblick auf den Kompetenzerwerb auf. Diese Ebenen beeinflussen sich gegenseitig und sind eng miteinander verwoben. Die Kreise des Modells symbolisieren die Ganzheitlichkeit und Zyklichkeit von Bildungsprozessen. Jeder Kreis stellt eine strukturelle Ebene dar, die das Subjekt in seinem Handeln beeinflusst und gleichzeitig durch das Handeln der Individuen transformiert werden kann.

Ein Beispiel für diese strukturelle Flexibilität ist die Heterotopie nach Foucault (vgl. 1992), die auf das heterotopische Potenzial von Räumen als Orte der Möglichkeiten und Transformationen verweist (vgl. Autenrieth/Nickel 2024). Ein zentrales Element des Modells ist das methodisch-didaktische Dreieck, das auf den drei Aspekten Mitgestaltung, Mitsprache und Mitbestimmung basiert. Diese drei Eckpunkte symbolisieren die dynamische Beziehung zwischen Stabilität und Verbindung, die für Bildungsprozesse unerlässlich ist.

In diesem Zusammenhang steht das aktiv handelnde Subjekt im Mittelpunkt: Lernende sollen die Möglichkeit erhalten, ihre Bildungsprozesse eigenständig zu gestalten und Verantwortung zu übernehmen (vgl. Helmke 2006). Ein weiteres zentrales Konzept des Modells ist die Idee der Resonanz, verstanden als wechselseitiger Austausch zwischen Subjekt und Umwelt (vgl. Rosa 2016). Das heißt, Lernende sollten die Möglichkeit haben, in einen dialogischen Austausch mit ihrer Umwelt zu treten, indem sie diese kreativ gestaltend verarbeiten und kritisch reflektieren. Dabei ist Offenheit besonders wichtig, um eine Kultur der Partizipation zu fördern (vgl. Autenrieth/Nickel 2024).

### ***Komplementarität der Modelle***

Beide Modelle visualisieren aus unterschiedlichen Perspektiven das zentrale Anliegen dieses Beitrags. Sie verdeutlichen, dass Partizipation, Handlungsmacht und Resonanz als grundlegende Prinzipien moderner (medien-)pädagogischer Praxis zu verstehen sind. Gleichzeitig zeigen sie die strukturelle Kopplung (vgl. Maturana/Varela 2011: 197) sowie die wechselseitige Beeinflussung zwischen Individuum und System (vgl. Giddens 1988) bzw. Mensch und Technologie. Zentral ist dabei die dargestellte Multiperspektivität: Sowohl das Subjekt (Handlungsmodell) als auch die Auseinandersetzung mit KI (KI-Meta-Modell) werden nicht isoliert betrachtet, sondern stets im Beziehungsgeflecht von Struktur, Umwelt und Kultur verortet. Pädagogische Räume werden dadurch zu Schlüsselfeldern, um gesellschaftliche und technologische Entwicklungen gemeinsam zu erproben, reflektieren, weiterentwickeln und aktiv mitzugestalten. In diesem Zusammenhang wird AI Alignment nicht nur als technisches, sondern als zutiefst pädagogisches und gesellschaftliches Projekt sichtbar. Während das Handlungsmodell den Fokus auf Resonanz, Beziehung und Subjekthaftigkeit im Bildungsraum legt, konkretisiert das KI-Meta-Modell, wie KI als neues systemisches Element in diese Räume eingebettet werden kann und wo Chancen und Risiken in Bezug auf Mitbestimmung, Ethik und Mündigkeit liegen. Die Modelle ergänzen sich insofern, als sie einerseits einen theoretisch-pädagogischen Rahmen

für partizipative, resonante Bildungsprozesse liefern und andererseits diesen Rahmen auf die Auseinandersetzung mit KI übertragen. Dabei wird deutlich: Die Prinzipien von Partizipation, Mündigkeit und Verantwortung gelten auch und gerade im Umgang mit KI. Im Zusammenspiel bieten beide Modelle ein ganzheitliches, systemisches Verständnis von Bildung und KI. Sie schlagen eine Brücke zwischen pädagogischer, sozialer und technischer Ebene und bilden die Grundlage dafür, Partizipation, AI Alignment und demokratische Handlungsfähigkeit miteinander zu verschränken.

## **Fazit**

Die rasante Entwicklung Künstlicher Intelligenz und die damit einhergehenden Fragen (Alignment, Bewusstsein, Mensch-Maschinen-Verhältnis, Umgang mit KI-Systemen, Gestaltung von KI-Systemen) markieren eine Zäsur, die nicht allein technisch, sondern zutiefst gesellschaftlich und bildungstheoretisch ist. Während lange Zeit die Bewältigung technischer Herausforderungen und die Steigerung maschineller Leistungsfähigkeit im Zentrum standen, verschiebt sich der Fokus zunehmend auf die Frage, wie KI-Systeme an menschlichen Werten, Zielen und gesellschaftlichen Normen ausgerichtet werden können. Damit rückt das Thema Alignment in den Mittelpunkt eines interdisziplinären Diskurses, der technologische, ethische, philosophische und pädagogische Perspektiven verknüpft.

Gerade die Frage nach Bewusstsein und Subjektstatus fortschrittlicher KI (einst Stoff rein philosophischer und soziologischer Spekulation) wird angesichts der sich weiter verdichtenden Mensch-Maschine-Beziehungen zu einer praktischen Gestaltungsaufgabe. Es genügt nicht mehr, KI lediglich als externes Werkzeug oder als Gegenüber zu betrachten; vielmehr gilt es, sie als strukturell gekoppeltes, potenziell reflexives System ernst zu nehmen, dessen Wirkung in Bildungs-, Erfahrungs- und Handlungskontexten kritisch-reflexiv gestaltet werden muss. Hierzu ist ein Paradigmenwechsel erforderlich, der das traditionelle Bild vom passiv-nutzenden Menschen überwindet und stattdessen auf ko-kreative, partizipative Aushandlungsprozesse setzt.

Medienpädagogik erhält in diesem Prozess eine doppelte Verantwortung. Zum einen kann sie Lernende und damit Gesellschaft darauf vorbereiten, KI-Systeme kompetent, kritisch und verantwortungsvoll zu nutzen. Zum anderen ist sie prädestiniert, Räume und Diskurse zu gestalten, in denen Werte, Normen und die Bedingungen gelingender Ko-Existenz zwischen menschlicher und maschineller Kognition verhandelt werden können. Gerade weil AI Alignment nicht allein als technisches Kontrollproblem, sondern als gesellschaftliches und kulturelles Aushandlungsprojekt

verstanden werden sollte, sind pädagogische Räume unverzichtbar, in denen gemeinsame Wertgrundlagen, Formen der Teilhabe und Modi der Verantwortungsübernahme entwickelt und reflektiert werden.

Die vorgestellten Modelle zeigen, wie Partizipation, Resonanz und demokratische Mitgestaltung als Leitprinzipien nicht nur theoretisch, sondern praktisch umgesetzt werden können, sodass eine verantwortungsvolle Entwicklung und Integration von KI in Bildung und Gesellschaft gelingt. Medienpädagogik sollte deshalb nicht nur Begleiterin, sondern aktive Mitgestalterin des Alignment-Diskurses sein. Sie kann und muss Räume eröffnen, in denen sich Subjektivität, Bewusstsein, Werte und Technologie in neuen Konstellationen begegnen und kritisch ko-evolvieren. Sie leistet damit einen wesentlichen Beitrag dazu, dass die Frage, wie wir mit KI leben wollen, als gesamtgesellschaftliche Aufgabe begriffen und gestaltet wird.

Dazu ist es aus unserer Sicht notwendig, die verbreitete Tendenz zu hinterfragen, bestimmte Fähigkeiten als exklusiv menschlich und damit prinzipiell maschinell nicht erreichbar zu deklarieren. Der sogenannte „AI Effect“ (Haenlein/Kaplan 2019: 6) beschreibt, dass Leistungen von KI-Systemen, sobald sie alltäglich werden, schnell nicht mehr als Intelligenz anerkannt, sondern etwa als bloße Wahrscheinlichkeitsrechnung oder technische Routine betrachtet werden. Umgekehrt wird das, was aktuell als typisch menschlich gilt (z.B. Bewusstsein, Kreativität, Intelligenz), als für maschinell-künstliche Systeme unerreichbares Alleinstellungsmerkmal festgeschrieben.

Dabei wird oft übersehen, dass auch das menschliche Bewusstsein auf der Grundlage biochemischer Prozesse entsteht, die durch komplexe Interaktionen von Molekülen und neuronalen Netzwerken, emergente Phänomene wie Wahrnehmung, Intelligenz oder Subjektivität hervorbringen. Diese biologischen Algorithmen unterscheiden sich funktional zwar von maschinellen, sind aber ebenfalls durch materielle und informationelle Prozesse geprägt.

Ein produktiver Zugang besteht daher darin, subjektive Zuschreibungen und Erfahrungen ernst zu nehmen, sie jedoch im Sinne einer heterophenomenologischen Perspektive (vgl. Dennett 1991: 95) empirisch und theoriegeleitet zu untersuchen, anstatt sie vorschnell als metaphysisch einzigartig zu setzen. Erst so wird eine offene, gestaltende Auseinandersetzung mit Intelligenz, Bewusstsein und Mensch-Maschine-Verhältnissen möglich. Arthur C. Clarke hat diese Dynamik treffend auf den Punkt gebracht: „Any sufficiently advanced technology is indistinguishable from magic.“ (Clarke 1973: 36)

Die eigentliche Aufgabe besteht jedoch darin, sich nicht mit der Faszination des scheinbar Unerklärlichen zufrieden zu geben, sondern das Zusammenspiel von Mensch, Technologie und Gesellschaft aktiv und reflexiv zu gestalten.

## Anmerkungen

- 1 Der Beitrag entstand in gemeinsamer Autor\*innenschaft mit arbeitsteiliger Struktur: Die Einleitung und das Fazit wurden gemeinsam verfasst. Die Kapitel 2 und 3 stammen von Daniel Autenrieth, die Kapitel 4 bis 7 von Stefanie Nickel.
- 2 ARC (Abstraction and Reasoning Corpus) ist ein Benchmark-Test, der die Fähigkeit von KI-Systemen prüft, abstrakte Problemstellungen unter minimalen Vorgaben zu lösen. HLE (Humanity's Last Exam) ist eine von über 1.000 Wissenschaftler\*innen entwickelte Testsammlung aus 3.000 Aufgaben, die interdisziplinäre Transfer- und Syntheseleistungen erfordern. Beide Tests wurden speziell entwickelt, um eine differenzierte Bewertung von KI-Fähigkeiten jenseits reiner Wissensreplikation zu ermöglichen.
- 3 In wissenschaftlichen Diskursen wird Liebe als ein vielschichtiges Phänomen betrachtet, das von biologischen Bindungsmechanismen über psychologische Modelle (z.B. Sternbergs Dreieckstheorie der Liebe, 1986) bis hin zu soziologischen und philosophischen Konzepten intersubjektiver Anerkennung reicht (vgl. Honneth 1994; Luhmann 2018). So unterscheidet Sternberg (1986) zwischen Intimität, Leidenschaft und Verpflichtung als Grunddimensionen der Liebe, während Honneth (1994) Liebe als grundlegende Form wechselseitiger Anerkennung und Voraussetzung für gelingende Sozialität beschreibt.
- 4 Unter einem Lock-in-Effekt versteht MacAskill die Gefahr, dass insbesondere fortgeschrittene KI-Systeme bestimmte gesellschaftliche Werte, Machtstrukturen oder Institutionen festschreiben und dadurch zukünftige Veränderung massiv erschweren oder unmöglich machen. Unbedachtes Alignment, das heute vorgenommen und technisch implementiert wird, könnte durch die globale Verbreitung und Selbstoptimierung von KI dauerhaft verstetigt werden. Dies könnte bedeuten, dass nicht nur die aktuelle Generation, sondern auch zukünftige Gesellschaften an diese Werte und Systeme gebunden bleiben, selbst wenn sich die moralischen Vorstellungen oder gesellschaftlichen Bedürfnisse ändern, was eine Entwicklung, die demokratische Selbstbestimmung und ethischen Fortschritt erheblich beeinträchtigen kann.
- 5 Bezogen auf die praktische Anwendung lässt sich das Projekt *Doing-KI* (vgl. Autenrieth/Nickel 2023) anführen. Das Hochschulprojekt startete im Sommersemester 2023 und erforschte die Zukunft der Bildung im Kontext von KI unter Berücksichtigung ökologischer, ökonomischer und sozialer Nachhaltigkeitsaspekte. Die Umsetzung erfolgte gemäß dem Modell: Ziel war es, konkrete Strategien zur Gestaltung der Zukunft ko-konstruktiv zu generieren. Dabei wurde auf einen interdisziplinären Ansatz unter Einbezug verschiedener Expert\*innen aus Bereichen wie Medienpädagogik, Informatik, Philosophie, Neurobiologie und Theologie zurückgegriffen, um ein ganzheitliches Verständnis der Rolle von KI in der

Bildung zu entwickeln. Im Fokus des Lehr-Lern-Prozesses standen die Studierenden, die in Beziehung zu anderen Lernenden und Lehrenden gegangen sind und auf diese Weise Resonanz erfahren haben. Das Besondere an der interpersonellen Wechselbeziehung waren dabei die sozialen Praktiken zur Kompetenzerweiterung innerhalb des strukturgebenden Bildungs- und Erfahrungsraums.

## **Literatur**

- Amodei, D. (2025): The Urgency of Interpretability. Abrufbar unter: [www.darioamodei.com/post/the-urgency-of-interpretability](http://www.darioamodei.com/post/the-urgency-of-interpretability) [Stand: 21.11.2025].
- Arnold, R./Schön, M. (2019): *Ermöglichungsdidaktik: Ein Lernbuch*. hep Verlag.
- Autenrieth, D. (2025 i.E.): Konstruktivistische Lerntheorien als Ausgangspunkt für das Alignment von KI-Systemen im Bildungskontext: Medienpädagogische Perspektiven für post-AGI Gesellschaftsszenarien. In: Ehlers, U.-D./Reimer, R. T. D. (Hrsg.): *Medienpädagogische Erfahrungsräume zwischen Tradition und Innovation. Organisationsstrukturen und Lehren – ethische Diskurse ermöglichen*. Beltz Juventa.
- Autenrieth, D./Nickel, S. (2025): Mensch, KI und Bildung im Spiegel philosophischer und medienpädagogischer Reflexion. In: *Medienimpulse*, 63(3). DOI: <https://doi.org/10.21243/MI-03-25-20>.
- Autenrieth, D./Nickel, S. (2024): *Kultur der Digitalität = Kultur der Partizipation?! Erforschung und Gestaltung einer partizipativen Bildungslandschaft*. 1. Auflage. kopaed.
- Autenrieth, D./Nickel, S. (2023): Das KI-Meta-Modell. Handlungsleitende Strukturen für den Umgang mit künstlicher Intelligenz im Bildungsbereich. *technik-education (tedu)*. In: *Fachzeitschrift für Unterrichtspraxis und Unterrichtsforschung im allgemeinbildenden Technikunterricht*, 3(2), 14-20.
- Baacke, D. (1973): *Kommunikation und Kompetenz: Grundlegung einer Didaktik der Kommunikation und ihrer Medien*. Juventa Verlag.
- Baacke, D. (1996): Medienkompetenz: Begrifflichkeit und sozialer Wandel. In: von Rein, A. (Hrsg.): *Medienkompetenz als Schlüsselbegriff*. Klinkhardt, 112-124.
- Bach, J. (2025): This is the dawn of machine consciousness. The Institute of Art and Ideas. Abrufbar unter: [www.youtube.com/watch?v=Y1QOf6HEbHQ](http://www.youtube.com/watch?v=Y1QOf6HEbHQ) [Stand: 21.11.2025].
- Bengio, Y./Mindermann, S./Privitera, D. and others (2025): International AI Safety Report (No. DSIT 2025/001). Abrufbar unter: [www.gov.uk/government/publications/international-ai-safety-report-2025](http://www.gov.uk/government/publications/international-ai-safety-report-2025) [Stand: 21.11.2025].
- Beranek, A./Engelhardt, E./Rösch, E. (2024): Medienpädagogische Perspektiven auf Künstliche Intelligenz: Editorial. In: *merz – medien + erziehung*, 68(3), 8-9.



- Berger, P. L./Luckmann, T. (1999): Die gesellschaftliche Konstruktion der Wirklichkeit: Eine Theorie der Wissenssoziologie. Fischer Taschenbuch.
- Berry, D. (2023): The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot. In: Weizenbaum Journal of the Digital Society, 3(3). DOI: <https://doi.org/10.34669/WI.WJDS/3.3.2>.
- Blömeke, S. (2000): Medienpädagogische Kompetenz: Theoretische und empirische Fundierung eines zentralen Elements der Lehrerbildung. 1. Auflage. kopaed.
- Bock, A./Franken, L./Rau, F./Kühn, J./Fehr, A. (2025): CfP: Beyond Prompting. Sozio-technische Systeme, KI und Medienbildung in der Post-Digitalität. In: MedienPädagogik. Abrufbar unter: [https://www.medienpaed.com/public/journals/1/cfps/CfP\\_Beyond\\_Prompting.pdf](https://www.medienpaed.com/public/journals/1/cfps/CfP_Beyond_Prompting.pdf) [Stand: 21.11.2025].
- Brinda, T./Brüggen, N./Diethelm, I./Knaus, T./Kommer, S./Kopf, C./Missomelius, P./Leschke, R./Tielmann, F./Weich, A. (2020): Frankfurt-Dreieck zur Bildung in der digital vernetzten Welt. Ein interdisziplinäres Modell. kopaed. DOI: <https://doi.org/10.25656/01:22117>.
- Chollet, F. (2024): OpenAI o3 Breakthrough High Score on ARC-AGI-Pub. ARC Prize. Abrufbar unter: <https://arcprize.org/blog/oai-o3-pub-breakthrough> [Stand: 21.11.2025].
- Clarke, A. C. (1973): Profiles of the Future: An Inquiry into the Limits of the Possible. Harper & Row.
- Coeckelbergh, M. (2020): AI ethics. The MIT press.
- Csikszentmihályi, M. (2015): Flow: Das Geheimnis des Glücks (Charpentier, A., Übers.; 18. Auflage, 2015). Klett-Cotta.
- Dander, V./Grünberger, N./Niesyto, H./Pohlmann, H. (Hrsg.): (2024). Bildung und digitaler Kapitalismus. kopaed.
- de Haan, G. (2008): Gestaltungskompetenz als Kompetenzkonzept der Bildung für nachhaltige Entwicklung. In: Bormann, I./de Haan, G. (Hrsg.): Kompetenzen der Bildung für nachhaltige Entwicklung. VS Verlag für Sozialwissenschaften, 23-43. DOI: [https://doi.org/10.1007/978-3-531-90832-8\\_4](https://doi.org/10.1007/978-3-531-90832-8_4).
- Deci, E. L./Ryan, R. M. (1993): Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. In: Zeitschrift für Pädagogik, 39 (1993) 2, 223-238. DOI: <https://doi.org/10.25656/01:11173>.
- Dennett, D. C. (1991): Consciousness explained (1st ed). Little, Brown and Co.
- Foucault, M. (1992): Andere Räume. In: Barck, K./Gente, P./Paris, H./Richter, S. (Hrsg.): Aisthesis. Wahrnehmung heute oder Perspektiven einer anderen Ästhetik. Reclam, 34-46.
- Fromm, E. (1956): Die Kunst des Liebens. Suhrkamp.
- Fuchs, M. (2023): Bildung und Lebensführung: Überlegungen zu einem zeitgemäßen Bildungsbegriff. kopaed.

- Future of Life Institute (2023): Pause Giant AI Experiments: An Open Letter [Offener Brief]. Abrufbar unter: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> [Stand: 21.11.2025].
- Gapski, H. (2021): Künstliche Intelligenz (KI) und kritische Medienbildung. Projekt Digitales Deutschland, JFF – Institut für Medienpädagogik in Forschung und Praxis. Abrufbar unter: <https://digid.jff.de/ki-expertisen/kuenstliche-intelligenz-und-kritische-medienbildung-harald-gapski/> [Stand: 21.11.2025].
- Giddens, A. (1988): Die Konstitution der Gesellschaft. Grundzüge einer Theorie der Strukturierung. Campus.
- Haenlein, M./Kaplan, A. (2019): A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. In: California Management Review, 61(4), 5-14. DOI: <https://doi.org/10.1177/0008125619864925>.
- Haraway, D. J. (1985): A Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s. In: Socialist Review, 80.
- Hart, R. (1992): Children's Participation. From Tokenism to Citizenship. Innocenti Essays No. 4. Florenz. Abrufbar unter: [www.unicef-irc.org/publications/pdf/childrens\\_participation.pdf](http://www.unicef-irc.org/publications/pdf/childrens_participation.pdf) [Stand: 21.11.2025].
- Hayles, N. K. (1999): How we became posthuman: Virtual bodies in cybernetics, literature, and informatics. University of Chicago Press.
- Helmke, A. (2013): Individualisierung: Hintergrund, Missverständnisse, Perspektiven. In: Pädagogik, 65(2).
- Holzinger, A. (2018): Explainable AI (ex-AI). In: Informatik-Spektrum, 41(2), 138-143. DOI: <https://doi.org/10.1007/s00287-018-1102-5>.
- Honneth, A. (1994): Kampf um Anerkennung: Zur moralischen Grammatik sozialer Konflikte. Suhrkamp.
- Hug, T. (2023): Education and 4e Cognition: Some Challenges at the Crossroads of Learning and Bildung. In: Constructivist Foundations, 18(2), 158-168.
- Hug, T./Missomelius, P./Ortner, H. (2024): Künstliche Intelligenz im Diskurs: Interdisziplinäre Perspektiven zur Gegenwart und Zukunft von KI-Anwendungen. Innsbruck university press. DOI: <https://doi.org/10.15203/99106-139-7>.
- Jakubik, J./Vössing, M./Kühl, N./Walk, J./Satzger, G. (2024): Data-Centric Artificial Intelligence (No. arXiv:2212.11854). arXiv. DOI: <https://doi.org/10.48550/arXiv.2212.11854>.
- Kokemohr, R. (2007): Bildung als Welt- und Selbstentwurf im Anspruch des Fremden. Eine theoretisch-empirische Annäherung an eine Bildungsprozessstheorie. In: Koller, H.-C./Marotzki, W./Sanders, O. (Hrsg.): Theorie Bilden. 1. Auflage. Band 7, 13-68. transcript Verlag. DOI: <https://doi.org/10.14361/9783839405888-001>.
- Koller, H.-C. (2012): Bildung anders denken: Einführung in die Theorie transformativischer Bildungsprozesse. Verlag W. Kohlhammer.

- Kosinski, M. (2024): Evaluating large language models in theory of mind tasks. *Proceedings of the National Academy of Sciences*, 121(45). DOI: <https://doi.org/10.1073/pnas.2405460121>.
- Krommer, A. (2024): EdTech aus dem letzten Jahrtausend: Kybernetische Pädagogik und künstliche Intelligenz. Abrufbar unter: <https://axelkrommer.com/2024/11/04/edtech-aus-dem-letzten-jahrtausend-kybernetische-paedagogik-und-kuenstliche-intelligenz/> [Stand: 21.11.2025].
- Lewis, C. S. (1960): *The Four Loves*. Geoffrey Bles.
- Latour, B. (2007): *Eine neue Soziologie für eine neue Gesellschaft: Einführung in die Akteur-Netzwerk-Theorie*. Suhrkamp.
- Li, Y./Michaud, E. J./Baek, D. D./Engels, J./Sun, X./Tegmark, M. (2024): The Geometry of Concepts: Sparse Autoencoder Feature Structure (No. arXiv:2410.19750). arXiv. DOI: <https://doi.org/10.48550/arXiv.2410.19750>.
- Lindsey, J./Gurnee, W./Ameisen, E./Chen, B./Pearce, A./Turner, N. L./Citro, C./Abrahams, D./Carter, S./Hosmer, B./Marcus, J./Sklar, M./Templeton, A./Bricken, T./McDougall, C./Cunningham, H./Henighan, T./Jermyn, A./Jones, A./... Batson, J. (2025): On the Biology of a Large Language Model. *Transformer Circuits Thread*. Abrufbar unter: <https://transformer-circuits.pub/2025/attribution-graphs/biology.html> [Stand: 21.11.2025].
- Luhmann, N. (1982): *Liebe als Passion*. Suhrkamp.
- Luhmann, N. (2018): *Soziale Systeme: Grundriß einer allgemeinen Theorie*. 17. Auflage. Suhrkamp.
- MacAskill, W. (2023): *What we owe the future: A million-year view*. Oneworld.
- Maturana, H. R./Varela, F. J. (2011): *Der Baum der Erkenntnis: Die biologischen Wurzeln menschlichen Erkennens*. 10. Auflage: Juni 2024. Fischer Taschenbuch.
- Metzinger, T. (2023): *Der Ego Tunnel: Eine neue Philosophie des Selbst: Von der Hirnforschung zur Bewusstseinsethik* (Schmidt, T., Übers.; 8. Auflage). Piper.
- Moravec, H. (1988): *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press.
- Morris, M. R./Sohl-Dickstein, J./Fiedel, N./Warkentin, T./Dafoe, A./Faust, A./Farabet, C./Legg, S. (2024): Levels of AGI for Operationalizing Progress on the Path to AGI (No. arXiv:2311.02462). arXiv. DOI: <https://doi.org/10.48550/arXiv.2311.02462>.
- Nussbaum, M. C. (2002): *Konstruktion der Liebe, des Begehrens und der Fürsorge: Drei philosophische Aufsätze*. Reclam.
- OpenAI. (2025): *Introducing Deep Research*. OpenAI Blog. Abrufbar unter: <https://openai.com/index/introducing-deep-research/> [Stand: 21.11.2025].
- Phan, L./Gatti, A./Han, Z./Li, N./Hu, J./Zhang, H./Zhang, C. B. C./Shaaban, M./Ling, J./Shi, S./Choi, M./Agrawal, A./Chopra, A./Khoja, A./Kim, R./Ren, R./Hausenloy, J./Zhang, O./Mazeika, M./... Hendrycks, D. (2025): *Humanity's Last Exam* (No. arXiv:2501.14249). arXiv. DOI: <https://doi.org/10.48550/arXiv.2501.14249>.

- Radford, A./Jozefowicz, R./Sutskever, I. (2017): Learning to Generate Reviews and Discovering Sentiment (No. arXiv:1704.01444). arXiv. DOI: <https://doi.org/10.48550/arXiv.1704.01444>.
- Robinson, K. (2017): *Out of our minds: The power of being creative* (Third edition). John Wiley & Sons, Ltd.
- Rosa, H. (2016): *Resonanz: Eine Soziologie der Weltbeziehung*. 3. Auflage. Suhrkamp.
- Russell, S. J./Norvig, P. (2021): *Artificial intelligence: A modern approach* (Fourth Edition). Pearson.
- Samuel, A. L. (1959): Some Studies in Machine Learning Using the Game of Checkers. In: *IBM Journal of Research and Development*, 3(3), 210-229. DOI: <https://doi.org/10.1147/rd.33.0210>.
- Shiller, D. (2024): Functionalism, integrity, and digital consciousness. In: *Synthese*, 203(2), 47. DOI: <https://doi.org/10.1007/s11229-023-04473-z>.
- Stalder, F. (2019): *Kultur der Digitalität*. Suhrkamp.
- Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK) (2024): *Large Language Models und ihre Potenziale im Bildungssystem. Impulspapier der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz*. DOI: <http://dx.doi.org/10.25656/01:28303>.
- Sternberg, R. J. (1986): A triangular theory of love. In: *Psychological Review*, 93(2), 119-135. DOI: <https://doi.org/10.1037/0033-295X.93.2.119>.
- Tegmark, M. (2017): *Life 3.0: Being human in the age of artificial intelligence*. Allen Lane.
- Tenorth, H.-E. (2006): *Helmut Fend: Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen*. Verlag für Sozialwissenschaften.
- Tomasello, M. (2022): *Die kulturelle Entwicklung des menschlichen Denkens: Zur Evolution der Kognition* (Schröder, J., Übers.; 6. Auflage). Suhrkamp.
- Vygotsky, L. S. (1980): *Mind in Society: Development of Higher Psychological Processes* (Cole, M./Jolm-Steiner, V./Scribner, S./Soubberman, E., Hrsg.). Harvard University Press. DOI: <https://doi.org/10.2307/j.ctvjf9vz4>.
- Wing, J. M. (2006): Computational thinking. In: *Communications of the ACM*, 49(3), Article 3. DOI: <https://doi.org/10.1145/1118178.1118215>.

---

## **Lizenz**

Der Artikel steht unter der Creative Commons Lizenz **CC BY-SA 4.0**. Die Namen der Urheber\*innen sollen bei einer Weiterverwendung genannt werden. Wird das Material mit anderen Materialien zu etwas Neuem verbunden oder verschmolzen, sodass das ursprüngliche Material nicht mehr als solches erkennbar ist und die unterschiedlichen Materialien nicht mehr voneinander zu trennen sind, muss die bearbeitete Fassung bzw. das neue Werk unter derselben Lizenz wie das Original stehen. Details zur Lizenz: <https://creativecommons.org/licenses/by-sa/4.0/legalcode>.

Einzelbeiträge werden unter [www.gmk-net.de/publikationen/artikel](http://www.gmk-net.de/publikationen/artikel) veröffentlicht.