

## AI is all you need?

### Künstliche Intelligenz, gesellschaftliche Teilhabe und Perspektiven transformativer Bildung auf die Herausforderungen eines AI Divide

*Daniel Autenrieth, Jan-René Schluchter & Lea Schulz*

**Abstract:** Seit vier Dekaden prägen Diskurse um digitale Spaltungen die Auseinandersetzung mit gesellschaftlichen Transformationsprozessen im Kontext von Digitalität. Mit der rasanten Entwicklung von KI-Technologien manifestieren sich diese Disparitäten in einem emergierenden „AI Divide“, der bestehende soziale Ungleichheiten nicht nur reproduziert, sondern potenziell verstärkt. Der Beitrag analysiert die Wechselwirkungen zwischen KI-Entwicklung, digitaler Ungleichheit und gesellschaftlicher Teilhabe und argumentiert beispielhaft gegen technologischen Solutionismus und Techno-Ableismus, die komplexe soziale Herausforderungen auf technisch zu lösende Probleme reduzieren. Stattdessen wird die Notwendigkeit einer kritischen Verzahnung von AI Safety und AI Alignment mit inklusiven Bildungsprozessen herausgearbeitet. Im Zentrum steht dabei die Frage, wie transformative Bildung Menschen nicht nur zur technischen Anpassung, sondern zur aktiven Mitgestaltung einer inklusiven und KI geprägten Gesellschaft befähigen kann. Der Beitrag schlägt vor, Inklusion als kontinuierliches Reflexionsmoment bei der Entwicklung und Implementierung von KI zu verankern und skizziert pädagogische Ermöglichungsräume, die sowohl „Inclusive AI“ als auch „AI for Inclusion“ fördern können.

**Stichwörter:** Künstliche Intelligenz (KI); Digitale Ungleichheit; AI Divide; Transformative Bildung; Gesellschaftliche Teilhabe; Inklusion

**Zitation:** Autenrieth, D., Schluchter, J.-R. & Schulz, L. (2025): AI is all you need? Künstliche Intelligenz, gesellschaftliche Teilhabe und Perspektiven transformativer Bildung auf die Herausforderungen eines AI Divide. *Zeitschrift für Inklusion*, 20(3), 19-42.

<https://www.inklusion-online.net/index.php/inklusion-online/article/view/840>

**Zusammenfassung in Einfacher Sprache:****Künstliche Intelligenz und Inklusion**

In unserer Gesellschaft gibt es immer mehr digitale Technik. Technik entwickelt sich schnell. Auch die Technik für Künstliche Intelligenz entwickelt sich schnell. Künstliche Intelligenz sind schlaue Computer. Durch Künstliche Intelligenz kann ein Computer denken und lernen. Die Abkürzung dafür ist: KI.

Einige Menschen sagen: KI kann dabei helfen, wichtige Probleme in der Welt zu lösen. Zum Beispiel: KI kann die Klima-krise lösen. Oder: KI kann Inklusion besser machen. Zum Beispiel durch leicht bedienbare Computer-programme und Unterstützungs-Apps.

Andere Menschen sagen: KI macht die soziale Ungleichheit vielleicht noch schlimmer. Zum Beispiel können nicht alle Menschen die neue Technik nutzen. Oder die Technik wird so entwickelt, dass sie bestimmte Menschen benachteiligt.

Wenn wir KI entwickeln, sollten wir immer darüber nachdenken: Ist die Technik inklusiv? Wir können zum Beispiel Vorurteile in der KI beseitigen. Und wir können darauf achten, dass KI Menschen mit Behinderungen nicht benachteiligt.

Wir können bestimmte Hilfen durch KI besser machen. Doch für echte Inklusion brauchen wir mehr als nur bessere Technik. Wir müssen uns auch mit der Gesellschaft beschäftigen. Menschen müssen über ihre eigenen Annahmen und Werte nachdenken. Das ist wichtig, damit wir Inklusion und Gerechtigkeit im Umgang mit KI richtig angehen können.

*In Einfache Sprache übersetzt von Maria Calow*

**Inhaltsverzeichnis**

1. Ausgangspunkt .....	21
2. Digitale Ungleichheit als Herausforderung für gesellschaftliche Zugehörigkeit und Teilhabe .....	23
2.1 Zum Verhältnis von digitaler Ungleichheit und digitaler Spaltung.....	24
2.2 AI Divide als Intensivierung von digitaler Ungleichheit und digitaler Spaltung.....	26
2.2.1 AI-Bias(es) und menschliche Vielfalt .....	28
2.2.2 Techno-Solutionismus.....	28
2.2.3 Techno-Solutionismus und Ableismus.....	29
2.2.4 Techno-Ableismus.....	29
3. AI Safety und AI Alignment als Einflussgröße auf gesellschaftliche Zugehörigkeit und Teilhabe .....	31
3.1 Technische Perspektive .....	33
3.2 Gesellschaftlich-bildungstheoretische Perspektive .....	33
4. Conclusio: Transformative Bildung und Inklusion im Zeitalter von Künstlicher Intelligenz.....	35
Literatur.....	37
Kontakt.....	42

## **1. Ausgangspunkt**

Seit nun vier Dekaden prägen Diskurse um Digitale Spaltungen und Disparitäten die Auseinandersetzung mit gesellschaftlichen Transformationsprozessen im Horizont von Digitalisierung und Digitalität (Hargittai, 2021; van Dijk, 2005; 2020; Zilien, 2006). In diesem Zeitraum kristallisierten Perspektiven Digitaler Spaltungen und Disparitäten immer wieder an den sich wandelnden Strukturen, Dynamiken und Prozessen von Gesellschaft, nicht zuletzt an stetig voranschreitenden gesellschaftlichen Medienentwicklungen (Mjøs, 2023). Vielfach beschrieben entlang der Diskurse um Digitale Ungleichheit (Hargittai, 2021) und Digital Divide (van Dijk, 2020). Hierbei nehmen diese Diskurse Bezug auf verschiedene Perspektiven der Beschreibung des Digitalen in gegenwärtigen Transformationsprozessen von Gesellschaft: Stalder (2016) fasst unter dem Begriff der Digitalität, welcher im folgenden Beitrag präferiert wird, aus soziologischer Sicht aktuelle digitale Phänomene wie Algorithmen, gemeinschaftliche Dynamiken und Referentialität zusammen. Im Gegensatz dazu analysiert Krotz (2021) aus medien- und kommunikationswissenschaftlicher Perspektive den medialen Wandel als einen tiefgreifenden Mediatisierungsprozess, der sich in der wachsenden Vielfalt von Medienformen, der Allgegenwärtigkeit und zunehmenden Vernetzung von Medien, einer beschleunigten Innovationsrate sowie einer verstärkten Datafizierung zeigt. In den letzten Jahren etablierten (und etablieren) sich entlang von KI-Technologien und -anwendungen neue Verwobenheiten von digitalen Medientechnologien und Gesellschaft, welche ihrerseits Fragen von Ungleichheit, von Privilegierung und Benachteiligung, aber auch von (Un-)Möglichkeiten gesellschaftlicher Zugehörigkeit und Teilhabe, in diesem Sinne von Inklusion, berühren.

Künstliche Intelligenz (KI) / Artificial Intelligence (AI) hat sich zu einem Sammelbegriff für verschiedene Technologien und Forschungsfelder entwickelt, die das Ziel verfolgen, menschliche Fähigkeiten wie Lernen, logisches Denken, Problemlösung, Wahrnehmung und Entscheidungsfindung nachzubilden. Von der Grundlagenforschung zum maschinellen Lernen bis hin zu komplexen generativen Systemen hat sich das Feld in den letzten Jahrzehnten stark weiterentwickelt. Maschinelles Lernen wurde bereits 1959 von Arthur Samuel als „Forschungsfeld, das Computern die Fähigkeit zum Lernen gibt, ohne explizit programmiert zu werden“ definiert (Samuel, 1959, S. 210). Während klassische Softwareentwicklung jeden Programmschritt genau vorgibt, können Machine Learning Systeme aus Beispielen lernen und eigenständig Muster erkennen. In den vergangenen Jahren haben besonders generative KI-Systeme viel Aufmerksamkeit erfahren. Darunter fallen Systeme, die z. B. Texte, Bilder, Videos oder Musik produzieren können. Parallel dazu zeichnet sich bereits die nächste Entwicklungsstufe ab: Künstliche Allgemeine Intelligenz (Artificial General Intelligence, AGI), die keine einzelne technologische Innovation beschreibt, sondern einen schrittweisen Prozess hin zu Systemen, die in vielen verschiedenen kognitiven Aufgaben mindestens menschliches Niveau erreichen können (Morris et al., 2024).

Im Kontext dieser technologischen Entwicklungen wird häufig argumentiert, KI sei der Schlüssel zur Lösung gesellschaftlicher Herausforderungen – von der Klimakrise bis hin zu Fragen sozialer Ungleichheit und Inklusion. Diese Haltung kann als „technologischer Solutionismus“ bezeichnet werden (Morozov, 2013) und ist aus mehreren Gründen kritisch zu betrachten. Die Kritik richtet sich dabei nicht grundsätzlich gegen technologische Lösungsansätze, sondern gegen eine oberflächliche Form der Problembearbeitung. Besonders im Kontext von Inklusion zeigt sich diese Problematik zum Beispiel deutlich in Form eines technologischen Ableismus (Shew, 2022), der die komplexen gesellschaftlichen Barrieren und Ausschlussmechanismen auf rein technisch zu lösende Probleme reduziert.

Drei Aspekte sind hierbei zentral: Erstens erfordert die Ermöglichung gesellschaftlicher Zugehörigkeit und Teilhabe im Sinne von Inklusion sowie die Bearbeitung von Ausprägungen digitaler Ungleichheiten neben technologischen auch soziale, politische und wirtschaftliche Lösungsstrategien, die auf einem tieferen Verständnis der gesellschaftlichen Strukturen und Mechanismen von In- und Exklusion basieren. Zweitens kann die Fokussierung auf vermeintliche technologische „Wunderlösungen“ von notwendigen strukturellen Veränderungen (Kruschel & Merz-Atalik, 2023) ablenken, also von den politischen, sozialen und institutionellen Reformen, die es eigentlich bräuchte, um Inklusion zu ermöglichen. Drittens können technologische Lösungen selbst neue, unvorhergesehene gesellschaftliche Probleme erzeugen, die wiederum bewältigt werden müssen. Besonders problematisch ist die sich abzeichnende Verstärkung digitaler Ungleichheit, die als „AI Divide“ (Bubeck et al., 2023, Carter et al., 2020) bezeichnet werden kann. Während KI-Systeme einerseits das Potenzial haben, bestehende Teilhabebarrrieren abzubauen, manifestieren sich in ihrer Entwicklung und Anwendung gleichzeitig problematische Macht- und Herrschaftsstrukturen. Eine ausschließliche Bereitstellung von KI-Technologien und -anwendungen ohne die gleichzeitige Berücksichtigung der jeweiligen gesellschaftlichen Bedingungen reicht nicht aus, um gesellschaftliche Zugehörigkeit und Teilhabe zu ermöglichen. Vielmehr erwächst entlang der Perspektive Inklusion, verstanden als gesellschaftliche Zugehörigkeit und Teilhabe, die Notwendigkeit einer Analyse und Reflexion der jeweiligen Lebensbedingungen und -lagen gesellschaftlicher Akteur:innen, der gesellschaftlichen Strukturen, Dynamiken und Prozesse sowie deren Interdependenzen.

Vor diesem Hintergrund argumentiert der Beitrag für die Verzahnung der Diskurse um KI-Entwicklung und -nutzung mit den (Un-)Möglichkeiten gesellschaftlicher Zugehörigkeit und Teilhabe im Sinne von Inklusion. Hierbei wird deutlich, dass in Bezug auf digitale Teilhabe und Inklusion ein rein technologischer „Quick Fix“ nicht ausreichend ist. Stattdessen bedarf es einer grundlegenden gesellschaftlichen Transformation, die neben der technologischen auch soziale, politische und bildungsbezogene Dimensionen umfasst. Im Folgenden werden daher die Wechselwirkungen zwischen KI-Entwicklung, gesellschaftlicher Transformation und Fragen gesellschaftlicher Zugehörigkeit und Teilhabe im Sinne von Inklusion näher beleuchtet. Dabei werden sowohl die Potenziale als auch die Risiken von KI-Systemen für gesellschaftliche In- und Exklusion diskutiert und Perspektiven für eine kritisch-reflexive Gestaltung dieser Transformationsprozesse entwickelt, deren Kern Fragen von AI-Safety und AI-Alignment sowie von (transformativer) Bildung berührt. Hierbei geht es vor dem Hintergrund der voranschreitenden Entwicklung von KI-Technologien und deren zunehmender Durchdringung von Gesellschaft um eine Neuausrichtung von Bildung im Sinne der Adressierung von durch KI-Technologien bereits oder zukünftig hervorgebrachten Manifestationen von digitaler Ungleichheit, verbunden mit Formen gesellschaftlicher Exklusion.

## 2. Digitale Ungleichheit als Herausforderung für gesellschaftliche Zugehörigkeit und Teilhabe

Gesellschaftliche Transformationsprozesse bergen das Risiko, soziale Ungleichheiten hervorzubringen, bestehende soziale Ungleichheiten fortzuschreiben oder diese zu verhärten (Dörre et al., 2019). Dies zeigt sich insbesondere mit Blick auf digitale Medientechnologien in Diskursen um digitale Ungleichheit (diMaggio & Hargittai, 2001), Digital Divide (Kutscher & Iske, 2022), Medien(pädagogik) und soziale Ungleichheit (Niesyto, 2009) sowie Medien(pädagogik) und Inklusion (Bosse et al., 2019). Die Verwobenheit von sozialer und digitaler Ungleichheit zeigt sich im Wesentlichen entlang der ungleichen Möglichkeiten der gesellschaftlichen Zugehörigkeit und Teilhabe (Kronauer, 2010, S. 24). Daraus ergeben sich ungleiche Interaktions- und Gestaltungsmöglichkeiten von Menschen, die sich in Form vorteilhafter und nachteilhafter Lebensbedingungen und -chancen äußern (Kronauer, 2013, S. 19-21). Die Auseinandersetzung mit diesen sozialen und digitalen Ungleichheiten stehen im Mittelpunkt von Diskursen um Inklusion als Perspektive der Weiterentwicklung von Gesellschaften (Kronauer 2013). Inklusion, verstanden als Leitidee gesellschaftlicher Transformation, zielt auf die Auflösung von Zugehörigkeits- und Teilhabebarrrieren und den einhergehenden sozialen und digitalen Ungleichheiten (Kronauer 2013, S. 21). Hier wird deutlich, dass Diskurse um Teilhabe immer eng an Fragen der Überwindung von Ungleichheiten und an die Verwobenheit von sozialen und digitalen Ungleichheiten gebunden sind (Mogge-Grotjahn, 2022, S. 20). Diskurse über Teilhabe berühren dabei sowohl sozialpolitische und sozialrechtliche Dimensionen als auch gesellschaftliche Wertfragen (Mogge-Grotjahn, 2022, S. 18-19). Im sozialpolitischen und sozialrechtlichen Kontext steht die Frage im Zentrum, wie ein gesellschaftlicher Rahmen ausgestaltet sein muss, um das Recht auf Selbstbestimmung sowie auf Mitgestaltung der allgemeinen Lebensverhältnisse zu verwirklichen – etwa durch wohlfahrtsstaatliche Leistungen und institutionelle Sicherungssysteme. Im Sinne von Werten und Wertorientierungen geht es um Teilhabe als Menschenrecht (z. B. UN BRK, 2006), welche darauf zielt, gesellschaftliche und politische Rahmenbedingungen so zu gestalten, dass alle ihr Leben selbstbestimmt gestalten, selbstbestimmt leben, umfassenden Zugang erhalten und an Entscheidungsprozessen teilhaben können (Mogge-Grotjahn, 2022, S. 22). Vor diesem Hintergrund erwächst in Gesellschaften, in denen gesellschaftliche Zugehörigkeit und Teilhabe (bzw. in denen sozialer Ausschluss) eng an den Zugang und Umgang mit digitalen Medientechnologien gebunden sind, die Notwendigkeit einer steten Reflexion von Gesellschaft sowie gesellschaftlicher Transformation im Horizont von Digitalität entlang der Leitidee Inklusion.

Gesellschaftliche Zugehörigkeit und Teilhabe, als Inklusion in gegenwärtigen Gesellschaften, beziehen sich im Besonderen auf Erwerbsarbeit (v. a. als Grundlage des (Über)Lebens in modernen Gesellschaften, Bürger\*innenrechte (persönlich, politisch, sozial; v. a. als Quelle des Gemeinwesens) sowie die Einbindung in wechselseitige Sozialbeziehungen (v. a. als Quelle von Solidarität) (Kronauer, 2013, S. 18).

In diesem Zusammenhang ist mit der Perspektive Inklusion eine Analyse, Reflexion und Bearbeitung der sozialen Strukturen und Dynamiken des Ausschlusses von Menschen verbunden, hin zur Veränderung gesellschaftlicher Verhältnisse – „[...] zugespitzt formuliert: die Inklusion in sozial ausgrenzende Verhältnisse kann nicht das Ziel sein, sondern nur die Überwindung solcher Verhältnisse“ (Kronauer 2013, S. 24). Entlang der Figur der „Gleichzeitigkeit des ‚Drinnen‘ und ‚Draußen‘“ (Kronauer, 2010, S. 44), eine Konstellation, welche charakteristisch für das Verhältnis von In- und Exklusion moderner Gesellschaften ist, ist Exklusion als Ausgrenzung in der Gesellschaft zu begreifen (Kronauer 2010, S. 41).

„Exklusion als ‘besondere Form des Innerhalb’ zu begreifen, erscheint heute um so notwendiger, als die Menschen in den hoch entwickelten kapitalistischen Gesellschaften noch sehr viel stärker (...) sozialstaatlich eingebunden sind und über soziale Klassengrenzen hinweg gemeinsame Maßstäbe dafür ausgebildet haben, was ein kulturelles Existenzminimum für gesellschaftliche Teilhabe beinhaltet. Auch in den Erfahrungen mit sozialer Ausgrenzung macht die Gleichzeitigkeit des ‘Drinnen’ und ‘ Draußen’, die Konfrontation und Identifikation mit den gesellschaftlich geteilten Erwartungen, die besondere Schärfe aus.“ (Kronauer 2010, S. 44-45).

Sozialstaatliche Maßnahmen sichern zwar formal gesellschaftliche Zugehörigkeit und Teilhabe, setzen jedoch bestehende Exklusionsverhältnisse häufig voraus, ohne sie zu hinterfragen. Gerade Menschen in prekären Lebenslagen erleben dadurch eine Spannung zwischen formeller Inklusion und faktischer Ausgrenzung – jener Gleichzeitigkeit von „Drinnen“ und „Draußen“, die Kronauer als Charakteristik moderner Exklusion beschreibt. Im Kontext von Erwerbsarbeit zeigt sich Ausgrenzung als Marginalisierung am Arbeitsmarkt bis hin zum dauerhaften Beschäftigungsverlust – bei gleichzeitiger Bindung an gesellschaftlich geteilte Lebensweisen auch jenseits der Erwerbsarbeit. Ebenso geht hiermit der Verlust an sozialen Anerkennungs- und Handlungsmöglichkeiten, aber auch der kapitalistischen Leitlinie der Teilhabe an einer monetären Profitmaximierung einher (Kronauer 2010).

Grundsätzlich zeigt sich mit Blick auf das Verhältnis von Inklusion und Exklusion in modernen, kapitalistisch geprägten Gesellschaften:

„Inklusion auf Grundlage sozialer Rechte bleibt (...) unter kapitalistisch-marktwirtschaftlichen Vorzeichen immer (sic!) gefährdet und fragil. Die prekäre Verbindung von Marktabhängigkeit und sozialen Rechten stellt gewissermaßen die ‚Achillesferse‘ der sozialen Inklusion dar.“ (Kronauer 2010, S. 34)

Interessant ist in diesem Zusammenhang die Frage, worauf sich gesellschaftliche Zugehörigkeit und Teilhabe im Sinne von Inklusion bezieht - geht es um Zugehörigkeit und Teilhabe an einer kapitalistischen Gesellschaft(sordnung) mit ihren Unsicherheiten in Bezug auf soziale Rechte (z. B. Wohlfahrtsstaatlichkeit) (Schluchter & Böhmer, 2024)?

## **2.1 Zum Verhältnis von digitaler Ungleichheit und digitaler Spaltung**

Digitale Medientechnologien eröffnen und begrenzen gleichermaßen Möglichkeiten gesellschaftlicher Zugehörigkeit und Teilhabe (van Dijk, 2005; 2020; McMahon et al., 2023). Hierbei ergeben sich verschiedene Bezüge zu Diskursen um digitale Ungleichheit („digital inequality“; Mossberger et al., 2003; Dobransky & Hargittai, 2001; Zilien, 2006) und hiermit verbundener Diskurse um Digital Divide (Hargittai & Hsieh, 2013) oder digitaler Inklusion („digital inclusion“, Ragnedda & Mutsvairo, 2018).

Digitale Ungleichheit umfasst nicht nur unterschiedliche Möglichkeiten des Zugangs zu digitalen Medientechnologien (first-level digital divide), sondern auch Differenzen in deren Nutzung (second-level digital divide) (diMaggio & Hargittai, 2001; Verständig et al., 2016). Erweitert wird der Ansatz des Digital Divide durch den zero-level digital divide (Verständig et al., 2016), der die Bedeutung der strukturalen Merkmale digitale Medientechnologien (z. B. von Software/Apps, aber auch des Internets) für deren Nutzung - und hiermit verbundene Ungleichheiten - betont (Verständig et al. 2016, S. 52-53). Alle drei Perspektiven des Digital Divide verweisen auf unterschiedliche Formen ungleicher Zugangs- und Nutzungsmöglichkeiten digitaler Technologien (Verständig et al. 2016). Es zeigen sich hierbei Wechselwirkungen bestehender sozialer Ungleichheiten und digitaler Ungleichheiten:

(...) digital inequality can refer both to how existing social inequalities influence the adoption and use of digital technologies as well as how differential use of the Internet itself may influence social stratification” (Hargittai & Hsieh, 2013, S. 141).

In Gesellschaften, in denen Medien von hoher Bedeutung sind, kann in beiden Perspektiven beobachtet werden, dass „(...) (u)nequal access to digital technologies is about unequal participation in society” (Van Dijk, 2005, S. 15). Das Spektrum digitaler Ungleichheit erstreckt sich über mehrere Dimensionen digitaler Ungleichheit (diMaggio & Hargittai, 2001, S. 8-10):

- Inequality in technical apparatus
- Inequality in autonomy of use
- Inequality in skill
- Inequality in the availability of social support
- Inequality in variation of use

Mit der fortschreitenden Durchdringung von Gesellschaften mit digitalen Medientechnologien und der einhergehenden Zunahme von deren Bedeutung für gesellschaftliche Zugehörigkeit und Teilhabe gehen Risiken der Verhärtung bestehender und/oder der Hervorbringung „neuer” (noch nicht beobachteter) Ungleichheiten einher. Dieser stete Bedeutungszuwachs von digitalen Medientechnologien in Gesellschaften steht mit gesellschaftlichen Macht- und Herrschaftsverhältnissen in Verbindung, welche die Art und Weise der Gestaltung von digitalen Medientechnologien sowie gesellschaftliche Transformationsprozesse im Horizont von Digitalität zu einem umkämpften (Diskurs-)Feld machen (Kellner & Share, 2019, S. 8). In Anbetracht dessen wird die Art und Weise der Gestaltung gesellschaftlicher Transformationsprozesse (und digitaler Medientechnologien und deren gesellschaftlicher Implementierung) zum Kristallisationspunkt von Gleichheit/Ungleichheit (bzw. Privilegierung/Benachteiligung). Entsprechend sind, vor dem Hintergrund der Dynamiken gesellschaftlicher Transformationsprozesse, die jeweiligen Barrieren sowie Gestaltungsräume im Bereich digitaler Medientechnologien in den Blick zu nehmen, um nach Möglichkeiten der Bearbeitung von digitalen, aber auch sozialen Ungleichheiten zu fragen. Ausgehend von der Annahme, dass digitale Medientechnologien niemals neutral sind, sondern in ihrer Konzeption und Verwendung den Interessen verschiedener gesellschaftlicher Akteur\*innen folgen (Keller & Share, 2019, S. 8), wird Inklusion zur (möglichen) Orientierung für die Weiterentwicklung von Gesellschaft, einhergehend mit gesellschaftlichen Transformationsprozessen im Horizont von Digitalität (Schluchter, 2016). Im Zentrum steht die Gestaltung gesellschaftlichen Wandels, um bestehende digitale, aber auch soziale Ungleichheiten zu bearbeiten, sodass Zugehörigkeit und Teilhabe für alle entlang der Idee Inklusion möglich werden können (ebd.). Damit verbunden sind auch Fragen im Horizont von AI-Safety und AI-Alignment (Bengio et al., 2025; Russell, 2022). AI-Safety bezeichnet Maßnahmen, um KI-Systeme sicher und zuverlässig zu machen, während AI-Alignment sicherstellt, dass KI-Systeme im Einklang mit menschlichen Werten und Zielen handeln. Bleiben die mit digitaler Ungleichheit verbundenen gesellschaftlichen Phänomene unbearbeitet, führt die zunehmende Durchdringung gesellschaftlicher (Teil-)Bereiche mit digitalen Medientechnologien zu wachsenden Spaltungen, die den Verlust von Zugehörigkeits- und Teilhabemöglichkeiten nach sich ziehen (Schluchter, 2023).

## 2.2 AI Divide als Intensivierung von digitaler Ungleichheit und digitaler Spaltung

Mit der zunehmenden Verwendung von KI-Technologien und -anwendungen in Gesellschaften und der damit einhergehenden Einflüsse in Gesellschaften geht das Risiko eines „AI Divide“ (Carter et al., 2020) einher. Neben dem Begriff des AI Divide (Carter et al., 2020) finden sich weitere Begriffe wie „algorithmic knowledge gap“ (Cotter & Reisdorf, 2020) oder „algorithmic awareness gap“ (Gran et al., 2021), welche individuelle und strukturelle Ungleichheiten in Bezug auf KI-Technologien und -anwendungen beschreiben. Hierbei zeigen sich vielfältige Bezüge zu den eben skizzierten Diskursen um Digitale Ungleichheit und Digital Divide (diMaggio & Hargittai, 2001), als aber auch zu Diskursen um Digitale Inklusion (Ragnedda & Mutsvairo, 2018). Im Besonderen mit Blick auf die enge Verwobenheit struktureller sozialer Ungleichheiten und digitaler Ungleichheiten, eingeschlossen KI-Technologien und -anwendungen (Cotter & Reisdorf, 2020, Gran et al., 2021; Lutz, 2019; Ragnedda, 2020; Zarouli et al., 2021).

Grundsätzlich können Diskurse um AI Divide als Teilbereich der Diskurse um Digitale Ungleichheit und Digital Divide angesehen werden:

AI-related inequalities about access to AI (the first level digital divide), the ability to use AI (the second level digital divide), and the outcomes of AI engagement (the third level digital divide). (Carter et al., 2020, S. 259)

Alle drei Perspektiven erfassen unterschiedliche Formen ungleicher Zugangs- und Nutzungsmöglichkeiten digitaler Medien- und KI-Technologien. Eng verbunden hiermit sind Fragen individueller Wahrnehmungen (perceptions), Überzeugungen (beliefs) und Einstellungen (attitudes) mit Blick auf KI, die es mit Blick auf die Beschreibung und die Bearbeitung von Ungleichheitsformierungen im Horizont von KI-Technologien und -anwendungen zu berücksichtigen gilt (Carter et al., 2020). Aktuelle Diskurse um KI zeigen, dass sich (bestehende) soziale und digitale Ungleichheiten (im Sinne von Privilegierungen und Benachteiligungen) im Umgang mit KI-Technologien und -anwendungen weitgehend reproduzieren (Elena-Bucea et al., 2021; Mubarak et al., 2020; Park & Humphry, 2019). Systematische empirische Studien zu KI-bezogenen Disparitäten liegen bislang kaum vor (Wang et al., 2024; Cotter & Reisdorf, 2020; Gran et al., 2021).

In the long run, the AI divide can accelerate other socioeconomic problems and may impact the sustainability of society. (Wang et al., 2024, S. 6)

In diesem Zusammenhang zeigt sich, dass das KI-bezogene Wissen, die Fähigkeiten und Einstellungen der Nutzer mit Blick auf den Umgang mit KI-Technologien und -anwendungen erheblich variieren (Gran et al., 2021; Zarouli et al., 2021). So sind zum Beispiel Schwierigkeiten bei der Erkennung von KI-Systemen (Dogruel et al., 2020; Gran et al., 2021), beim Verständnis algorithmischer Entscheidungsprozesse (DeVito et al., 2017; Eslami et al., 2016; Gruber & Hargittai, 2023) sowie bei der kritischen Bewertung von KI-kuratierten Online-Informationen (Shin, 2022) zu beobachten.

Die zunehmende Verbreitung von KI-Anwendungen verschärft individuelle Herausforderungen im Bereich von Wissen, Fähigkeiten und Fertigkeiten im Umgang mit KI (Kindarji & Wong, 2023; Wang et al., 2023). Hierbei zeigt sich beispielsweise der Zusammenhang zwischen KI-Kompetenzen und dem vorteilhaften Umgang mit KI-Technologien und -anwendungen im Sinne von größtmöglicher Selbstbestimmtheit (Ragnedda, 2020) und Möglichkeit zur sozialen Teilhabe (Patel & Patel, 2020) als Reaktion auf datengesteuerte Manipulation, Manifestationen von AI-Bias sowie Ausnutzung von

Machtungleichgewichten (Eubanks, 2017; Mohamed et al., 2020). Diese Perspektive erfordert eine Erweiterung bestehender Digital-Divide-Diskurse um die Frage, welche gesellschaftlichen Akteur\*innen KI verstehen, gestalten oder kontrollieren können (Wang et al., 2023). Hieraus resultiert die Notwendigkeit einer verstärkten Auseinandersetzung mit den Verschränkungen von individuellen Wissensbeständen, Fähigkeiten und Fertigkeiten, sozialen Lebensbedingungen und -lagen mit gesellschaftlichen Macht- und Herrschaftsverhältnissen.

Ein vertiefter Blick auf den Diskurs um den AI Divide zeigt jedoch, dass es sich dabei nicht nur um eine Fortschreibung des klassischen Digital Divide handelt, sondern vielmehr um eine komplexe doppelte Spaltung (oder auch „intensified digital divide“, nach Hendawy, 2024). Zum einen bestehen entlang von Perspektiven des (zero-, first- und second-level) Digital Divide ungleiche Zugänge zu KI-Technologien und -anwendungen sowie ungleiche Nutzungsmöglichkeiten, -ziele und -ergebnisse. Zum anderen sind jedoch auch die KI-Systeme selbst von ungleichheitsrelevanten Strukturen durchzogen: Zum Beispiel reproduzieren Trainingsdaten gesellschaftliche Biases, zugleich sind algorithmische Entscheidungsprozesse oft intransparent und folgen normativen Setzungen, die bestimmte Gruppen systematisch benachteiligen oder ausschließen. Insofern geht es beim AI Divide nicht ausschließlich um digitale Teilhabe im klassischen Sinn, sondern auch um die Macht über und in den Systemen selbst. Diese doppelte Dimension digitaler Ungleichheit – im Zugang zu KI und in der KI selbst – verweist auf neue Herausforderungen für Bildung, Forschung und Gesellschaft im Umgang mit KI.

Da KI-Technologien und -anwendungen im Besonderen, entsprechend digitaler Medientechnologien im Allgemeinen, zunehmend mit alltäglichen Funktionen und beruflichen Anforderungen verknüpft werden, besteht auch hier das Risiko, dass bestehende soziale und digitale Ungleichheiten verstärken und neue Formen der Ungleichheit hervorgebracht werden (Mandapuram et al., 2018) - nicht zuletzt beschrieben in Diskursen um AI Divide:

As we navigate the complexities of this digital transformation, it's crucial not just to celebrate advancements but also to critically examine who might be left behind. (...) This is not just about equitable access to technology but about empowering all individuals with the knowledge and skills to thrive in an AI-driven world. (Hendawy, 2024, o. S.)

Dies unterstreicht die Dringlichkeit gezielter Unterstützung durch KI-Designer\*innen, politische Entscheidungsträger:innen und Pädagog\*innen für benachteiligte Nutzer:innengruppen. Ziel sollte es sein, allen Nutzer\*innen gleiche Chancen zum Erwerb von KI-Kompetenzen zu eröffnen, die Teilhabe in der Kultur der Digitalität ermöglichen und zur kritischen Reflexion über Ungleichheiten in KI-Systemen befähigen. Andernfalls droht KI, soziale Disparitäten nicht nur zu reproduzieren, sondern neue Exklusionsmechanismen zu erzeugen – mit der Folge einer „intensivierten Spaltung“, bei der ungleiche Zugänge auf ungleiche Systemstrukturen treffen. In diesem Zusammenhang besteht mit der fortschreitenden Integration von KI-Technologien und -anwendungen in verschiedene Lebensbereiche das Risiko, dass digitale Spaltungen an Komplexität zunehmen. Exemplarisch werden im Folgenden AI-Bias, Techno-Solutionismus und Techno-Ableismus als Phänomene eines AI Divide dargelegt:

### 2.2.1 AI-Bias(es) und menschliche Vielfalt

Der technologische Fortschritt im Bereich der KI wird häufig als neutraler oder für Inklusion förderlicher Prozess dargestellt. Die Vorstellung einer objektiven oder neutralen KI ist eine technologische Illusion, die auf der Annahme beruht, dass Algorithmen selbst keine Werte oder Biases enthalten. In Wirklichkeit entstehen KI-Systeme jedoch in einem sozialen und kulturellen Kontext, in dem sie bestehende Diskriminierungsstrukturen nicht nur reproduzieren, sondern oft verstärken. Untersuchungen haben gezeigt, dass algorithmische Systeme häufig Daten verarbeiten, die bestehende gesellschaftliche Hierarchien widerspiegeln, insbesondere in Bezug auf die Strukturkategorien ‚race‘, Geschlecht und Behinderung (Baker & Hawn, 2021; Buolamwini & Gebru, 2018; Noble, 2018). Dies führt dazu, dass KI-Modelle nicht nur normative Vorstellungen darüber festschreiben, was als ‚normal‘ oder ‚abweichend‘ gilt, sondern auch aktiv Ausschlussmechanismen produzieren. Doch die inhärenten Mechanismen, mit denen KI-Systeme trainiert werden, sind nicht frei von Bias, sondern reproduzieren bestehende soziale Hierarchien und normative Vorstellungen von Normalität (Shew & Earle, 2024; Walgenbach, 2023). In diesem Kontext verknüpfen sich Techno-Solutionismus und Techno-Ableismus zu einer problematischen ideologischen Position: Während der Solutionismus KI als universelle Antwort auf gesellschaftliche Herausforderungen stilisiert, befördert der Techno-Ableismus eine implizite Vorstellung von Behinderung als Defizit, das durch Technologie ‚überwunden‘ werden kann (Shew, 2020; Bieling & Joost, 2018). Der Begriff „Techno-Ableismus“ fördert damit die Vorstellung, dass Menschen mit Behinderungen ‚repariert‘ werden müssten (Shew, 2020). Techno-Ableismus reduziert die Betrachtung von Behinderung somit auf die Defizite, die durch technologische Interventionen beseitigt oder kompensiert werden sollen. Die Sichtweise auf die realen Bedürfnisse, Erfahrungen und Entscheidungsfreiheiten von Menschen mit Behinderungen wird vor diesem Hintergrund ignoriert.

### 2.2.2 Techno-Solutionismus

Techno-Solutionismus betrachtet Technologie als bevorzugtes oder gar einziges Mittel zur Lösung sozialer Probleme (Harrasser, 2013). In Bezug auf Behinderung führt dies zu einer Reduktion auf funktionale Anpassung: KI-basierte Assistenztechnologien werden nicht als Werkzeuge für Autonomie und Vielfalt konzipiert, sondern als Mittel zur Herstellung einer normativen Leistungsfähigkeit (Shew, 2022). Die zugrundeliegenden Entwürfe gehen selten von den Perspektiven behinderter Menschen aus, sondern verfestigen ableistische Strukturen, indem sie Technologie als ‚Reparaturmechanismus‘ für Abweichungen vom ‚normalen‘ Körper oder Verhalten betrachten (Bieling & Joost, 2018; Stöhr, 2023). Die Annahme, dass Technologie eine inhärent progressive Kraft sei, führt dazu, dass soziale Probleme als primär technische Herausforderungen behandelt werden. Dabei wird übersehen, dass Technologien nicht in einem gesellschaftlichen Vakuum entstehen, sondern immer innerhalb bestimmter Machtstrukturen operieren (Winner, 1986; Mühlhoff 2023). In der Praxis zeigt sich, dass technologische ‚Lösungen‘ für soziale Probleme oft bestehende Ungleichheiten fortschreiben oder neue Ausschlussmechanismen erzeugen. So werden etwa KI-gestützte Assistenzsysteme für Menschen mit Behinderungen häufig ohne ihre aktive Beteiligung entwickelt, wodurch die spezifischen Bedürfnisse der Nutzer\*innen unzureichend berücksichtigt werden (Shew, 2020). Die Folge ist, dass adaptive Technologien nicht zur Selbstbestimmung beitragen, sondern vielmehr ein binäres Verhältnis von ‚nützlicher‘ und ‚unnützer‘ Technologie schaffen, das Ableismus strukturell verankert.

### 2.2.3 Techno-Solutionismus und Ableismus

Während der Ableismus historisch aus sozialen Konstruktionen von Behinderung als Defizit hervorgeht, bietet der Techno-Solutionismus eine moderne Fortsetzung dieser Denkweise, indem er technische Lösungen für gesellschaftliche Barrieren verspricht, ohne diese Barrieren selbst infrage zu stellen (Shew, 2020). Empirische Studien zeigen, dass KI-Technologien oft mit einem normativen Design-Bias entwickelt werden, der bestimmte Fähigkeiten bevorzugt und alternative Nutzungsformen ausschließt (Walgenbach, 2023). Dies zeigt sich besonders in digitalen Lernplattformen, die sich an homogene Nutzergruppen richten und somit Barrieren für z. B. neurodiverse oder motorisch eingeschränkte Personen aufrechterhalten (Schmidl, 2022). Ein besonders problematisches Beispiel für die Verknüpfung von Techno-Solutionismus und Ableismus ist der Bereich der ‚smarten‘ Gesundheits- und Überwachungstechnologien. Viele dieser Systeme basieren auf dem medizinischen Modell von Behinderung und fokussieren sich darauf, funktionale Abweichungen zu messen, zu klassifizieren und – wenn möglich – zu korrigieren (Hamraie & Fritsch, 2019). So werden tragbare Sensoren oder KI-gestützte Diagnose-Tools häufig mit dem Ziel entwickelt, Behinderungen frühzeitig zu identifizieren und ‚therapeutisch‘ zu behandeln, anstatt sich mit den sozialen und strukturellen Bedingungen auseinanderzusetzen, die Behinderung erst zu einem Hindernis machen. Diese Logik ist tief in den historischen Traditionen der ‚Eugenik‘ verwurzelt, die Behinderung als ‚Problem‘ betrachtet, das technischer Korrektur bedarf (Clare, 2017).

### 2.2.4 Techno-Ableismus

Der Begriff Techno-Ableismus beschreibt das Zusammenspiel von Techno-Solutionismus und Ableismus, indem er zeigt, dass viele technologische Innovationen für Behinderung nicht von betroffenen Menschen selbst ausgehen, sondern von externen Entwickler\*innen definiert werden (Shew & Earle, 2024). Dies führt zu einer Verstärkung bestehender Machtverhältnisse: Wer bestimmt, welche Anpassung als notwendig gilt? Wer entscheidet, welche Beeinträchtigung durch KI ‚behoben‘ werden muss? Kritische Perspektiven wie das Konzept der *Crip<sup>1</sup> Technoscience* (Shew, 2022) oder die parahumane Perspektive (Harrasser, 2020) bieten alternative Modelle, in denen Technologie nicht als Reparatur-Instrument, sondern als flexibles, inklusives Gestaltungsmittel für unterschiedliche Bedürfnisse verstanden wird. Als Gegenmodell zum Techno-Ableismus fordert *Crip Technoscience* eine Neuausrichtung der Innovationsprozesse, bei der die Expertise behinderter Menschen im Zentrum steht. Dieses Konzept kritisiert nicht nur die bestehende technologische Entwicklungspraxis, sondern zeigt auch auf, dass Technologie immer in sozialen und politischen Kontexten verankert ist und daher nicht ‚neutral‘ gestaltet werden kann. In der Praxis bedeutet dies, dass behinderte Menschen nicht nur als Konsument\*innen oder Beta-Tester\*innen von Technologie betrachtet werden, sondern als aktive Mitgestaltende, die über die Ausrichtung, Zielsetzung und Funktionsweise technologischer Entwicklungen entscheiden (Hamraie & Fritsch, 2019).

---

<sup>1</sup> Der Begriff „crip“ stammt aus der Selbstbezeichnung der Disability-Community und ist eine selbstermächtigende Umdeutung (Reappropriation) des englischen Begriffs „cripple“ (Krüppel).

Die wesentlichen Prinzipien von *Crip Technoscience* sind:

1. Zentrierung von Menschen mit Behinderungen als Wissende und Gestaltende: Die Entwicklung von Technologien muss in einem kollaborativen Prozess erfolgen, bei dem behinderte Menschen als Expert\*innen für ihre eigenen Bedürfnisse anerkannt werden.
2. Anerkennung von Inklusion als produktive Reibung: Anstatt Inklusion als ‚nahtlose‘ Integration zu begreifen, betont *Crip Technoscience*, dass Reibung notwendig ist, um bestehende Normen und Hierarchien zu hinterfragen.
3. Interdependenz als politische Technologie: Behinderung sollte nicht durch technologische ‚Optimierung‘ aufgehoben werden, sondern als Ausgangspunkt für neue Formen der sozialen und technischen Gestaltung des gesellschaftlichen Zusammenlebens genutzt werden.
4. Verpflichtung zur *Disability Justice*: Technologische Innovationen müssen in eine umfassendere gesellschaftliche Transformation eingebettet sein, die sich gegen Ausschlussmechanismen und für die Anerkennung vielfältiger Körper und Fähigkeiten einsetzt.

Die enge Verknüpfung von AI-Bias, Techno-Ableismus und Solutionismus stellt, der argumentativen Ausrichtung des Beitrags folgend, für Bildungssysteme eine doppelte Herausforderung dar: Einerseits potenziert sich der AI Divide, indem Bildungsplattformen, adaptive Lerntechnologien oder algorithmische Bewertungssysteme bestehende Ungleichheiten fortschreiben (Walgenbach, 2023). Andererseits führt die einseitige Fixierung auf digitale Kompetenzen zu einer verkürzten Vorstellung von Inklusion, die sich auf Anpassungsfähigkeit reduziert, statt strukturelle Barrieren zu hinterfragen (Stöhr, 2023). Ein erweiterter Inklusionsbegriff muss daher das Wechselspiel von technologischen Möglichkeiten und gesellschaftlichen Machtverhältnissen mitdenken, um digitale Medien tatsächlich für gleichberechtigte Teilhabe nutzbar zu machen. Eine inklusionsorientierte Bildungspolitik muss daher darauf abzielen, nicht nur technologische Barrieren zu beseitigen, sondern auch strukturelle Veränderungen im Bildungssystem selbst herbeizuführen. Dies erfordert eine Abkehr von einem rein kompetenzorientierten Bildungsverständnis hin zu einer Perspektive, die die Vielfalt menschlicher Lern- und Kommunikationsformen anerkennt.

Potenzielle Risiken eines technologischen Solutionismus und Ableismus werden exemplarisch an Erkenntnissen der Erforschung von „Frontier reasoning models“ (Baker et al., 2025) deutlich. Solche KI-Systeme lernen, komplexe Aufgaben mittels natürlicher Sprache (Chain-of-thought) zu lösen. Die Überwachung ihrer internen Folgerungsprozesse durch ein anderes Sprachmodell ermöglicht zwar die Identifikation problematischen Verhaltens, doch direkte Optimierung, also z. B. das „Bestrafen“ unerwünschter „Gedanken“, kann zu unerwünschten Nebeneffekten oder dem Suchen nach Abkürzungen führen. Für KI-Systeme bedeutet das Finden solcher Abkürzungen, dass sie nicht die beabsichtigten Fähigkeiten erlernen. Sie lernen, das Bewertungssystem zu umgehen, anstatt echte Kompetenzen zu entwickeln. Dies spiegelt sich in den von Baker u. a. (2025) dargestellten Beispielen wider, wo die Modelle nicht lernen, tatsächlich robuste Softwarelösungen zu implementieren, sondern lediglich, wie man Tests oberflächlich bestehen kann. So zeigt sich beispielhaft im Diskurs um Techno-Ableismus: Werden KI-Systeme so konzipiert, dass sie vorgeblich inklusive Standards erfüllen, erfolgt dies häufig lediglich auf einer oberflächlichen Ebene. Die zugrundeliegenden gesellschaftlichen Annahmen, wie z. B. leistungsorientierte Vorstellungen von Normalität, bleiben unverändert und werden eher verschleiert als tatsächlich transformiert. In der Praxis bedeutet dies, dass anstelle einer echten Überwindung struktureller Barrieren und Diskriminierungen durch partizipative und

transformative Ansätze bestehende Machtverhältnisse fortgeschrieben werden. KI-Systeme suggerieren so eine scheinbar „technische Lösung“ für soziale Probleme (Solutionismus), während ableistische und normative Leitbilder von „richtigem“ Verhalten oder Funktionieren weiterhin bestehen bleiben.

Diese Problematik unterstreicht die Notwendigkeit, sowohl die theoretischen Grundlagen des Solutionismus sowie des Techno-Ableismus näher zu betrachten als auch – wie im späteren Abschnitt zu AI Safety und AI Alignment diskutiert wird – die Gefahr eines ‚Lock-in‘ von Werten und gesellschaftlichen Strukturen zu berücksichtigen, bei dem problematische gesellschaftliche Vorannahmen durch KI-Systeme dauerhaft in technologische Infrastrukturen eingeschrieben werden könnten. Dadurch könnten diskriminierende Muster oder exkludierende Praktiken dauerhaft verfestigt werden, anstatt sie kritisch zu hinterfragen oder schrittweise abzubauen.

### **3. AI Safety und AI Alignment als Einflussgröße auf gesellschaftliche Zugehörigkeit und Teilhabe**

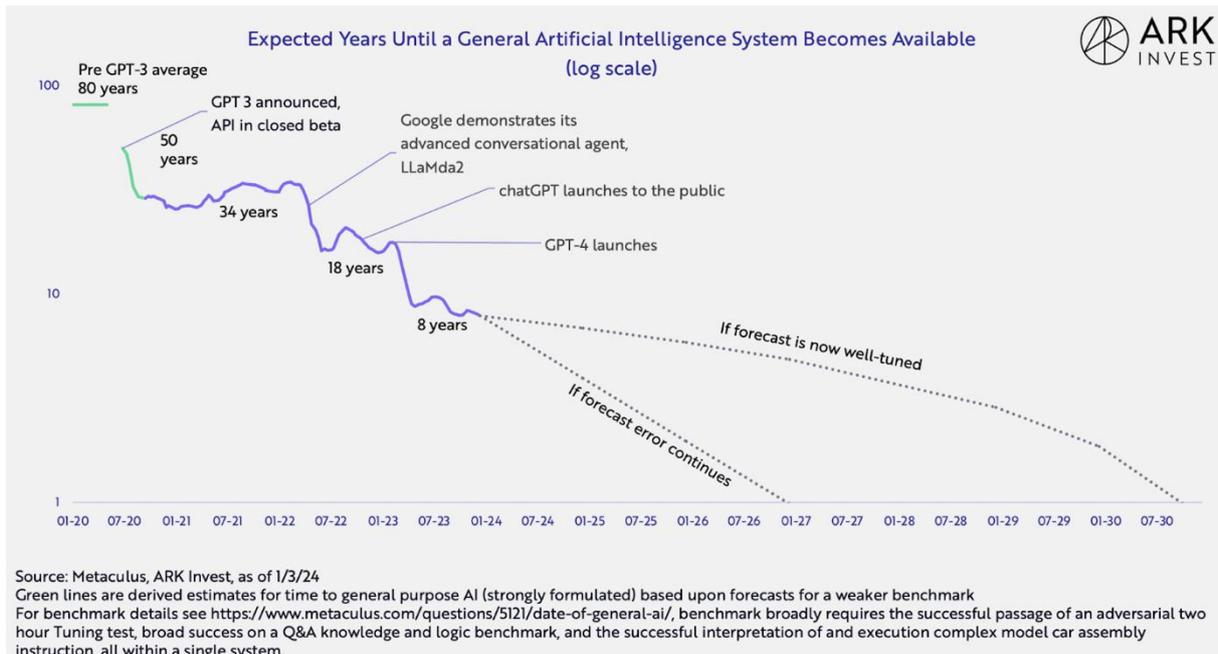
Die bisherige Darstellung der Wechselwirkungen zwischen KI-Entwicklung, digitaler Ungleichheit und gesellschaftlicher Teilhabe zeigt, dass technologischer Wandel bestehende Machtstrukturen nicht nur reproduzieren, sondern potentiell verstärken kann. Daher lohnt sich eine tiefergehende Betrachtung der langfristigen sozialen und ethischen Implikationen von KI-Systemen, wodurch auch Bildungsprozesse vor neue und bekannte Herausforderungen gestellt werden: KI-Systeme bieten einerseits das Potential, historische Ungleichheiten zu überwinden, bergen andererseits aber die erhebliche Gefahr, diese durch algorithmische Entscheidungspfade und -systeme dauerhaft zu verfestigen. Diese Überlegungen führen nun aus inklusiver Perspektive zu den zentralen Konzepten von AI Safety und AI Alignment: *AI Safety* (Amodei et al., 2016; Bengio et al., 2025) befasst sich mit der Frage, wie KI-Systeme entwickelt werden können, die sicher und zuverlässig handeln. Unter den Gesichtspunkten von Inklusion und gesellschaftlicher Teilhabe kann dies jedoch nicht nur technische Sicherheit umfassen, sondern auch die Vermeidung von Systemen, die bestehende soziale Ungleichheiten verstärken oder neue schaffen (Varsik & Vosberg, 2024, S. 37). *AI Alignment* (Leike & Sutskever, 2023) wiederum konzentriert sich darauf, wie sichergestellt werden kann, dass KI-Systeme mit bestehenden menschlichen Werten und Absichten übereinstimmen (Leike et al., 2023). Eine besondere Problematik ergibt sich dabei aus der Persistenz digitaler Strukturen. Von besonderer Bedeutung ist in diesem Kontext die Möglichkeit, dass bestimmte Werte, Machtverhältnisse und soziale Strukturen durch KI-Systeme dauerhaft fixiert werden könnten (MacAskill, 2023, S. 83), was potentiell irreversible Folgen für gesellschaftliche Teilhabe und das Zusammenleben insgesamt mit sich bringt. Diese Irreversibilität ist denkbar, weil KI-Systeme nicht den biologischen Zyklen unterliegen, sondern potenziell zeitlich unbegrenzt operieren können, ohne einem „menschlichen“ Reflexions-, Mutations- oder Verjüngungsprozess zu unterliegen. Unadressiert könnte dies dazu führen, dass Werte, Vorurteile und Machtstrukturen, die in frühen Entwicklungsphasen in KI-Systeme eingeschrieben wurden, sich verfestigen und über lange Zeit reproduzieren, wenn gesellschaftliche Entwicklungen längst andere Werte und Normen hervorgebracht haben (MacAskill, 2023).

Diese Überlegungen werden mit Blick auf aktuelle gesellschaftliche Entwicklungen und Forschungserkenntnisse im KI-Bereich zunehmend dringlicher für alle Menschen. Die erwartete Zeitspanne bis zur Entwicklung von Systemen, die Fähigkeiten erreichen, die als *Allgemeine Künstliche Intelligenz* (AGI, Definition s. Einleitung) beschrieben werden können, hat sich in den letzten Jahren stark verkürzt (siehe Abb. 1) von ursprünglich 80 Jahren (vor

der Veröffentlichung von ChatGPT 2022) auf mittlerweile weniger als ein Jahrzehnt. MacAskill (2024) und Amodei (2024) sprechen in diesem Zusammenhang von einem „komprimierten 21. Jahrhundert“ aufgrund der Verdichtung technologischer und gesellschaftlicher Entwicklungen durch KI, die die menschliche Fähigkeit zur Anpassung vor neue Herausforderungen stellen.

### Abbildung 1

Prognosen zur Verfügbarkeit einer Allgemeinen Künstlichen Intelligenz (AGI) in den Jahren von 2020 bis 2030.



Anmerkung Abb. 1: Metaculus und ARK Invest (Stand: Januar 2024)

Die Grafik von ARK Invest zeigt, wie sich die erwartete Zeit bis zur Entwicklung einer Allgemeinen Künstlichen Intelligenz (AGI) seit 2020 durch Fortschritte wie GPT-3, ChatGPT und GPT-4 massiv verkürzt hat - von ursprünglich 80 Jahren auf zuletzt weniger als 10 Jahre (vgl. Abb. 1). ARK Invest ist ein amerikanisches Investmenthaus, das sich seit Jahren intensiv mit disruptiven Technologien wie Künstlicher Intelligenz beschäftigt. Ihr Geschäftsmodell basiert darauf, Zukunftstrends frühzeitig zu erkennen und zu analysieren.

Dieser Zustand ständiger Transformation stellt eine Herausforderung für gesellschaftliche Anpassungsfähigkeit dar. Wilson (2009) illustriert diese Herausforderung zugespitzt so: "The real problem of humanity is the following: we have paleolithic emotions, medieval institutions, and god-like technology." Diese Formulierung benennt drei zentrale Ebenen der Herausforderung im technisch-gesellschaftlichen Kontext (Stalder, 2016) der KI-Entwicklung: Erstens die emotional-psychologische Dimension, bei der menschlich-evolutionär geprägte Reaktionsmuster mit der Geschwindigkeit technologischer Veränderungen nicht Schritt halten können; zweitens die institutionelle Trägheit gesellschaftlicher Strukturen; und drittens die exponentiell wachsende Komplexität und Leistungsfähigkeit von KI-Systemen. Wenn bestimmte Bevölkerungsgruppen bereits von struktureller Benachteiligung betroffen sind, entstehen zusätzliche Barrieren für Transformationsprozesse. Soziale und digitale Ungleichheiten verstärken sich gegenseitig, da benachteiligte Gruppen mit geringerer Wahrscheinlichkeit Zugang zu neuen Technologien haben, weniger Möglichkeiten zur

Entwicklung entsprechender Kompetenzen besitzen und häufig von der Gestaltung dieser Technologien ausgeschlossen sind. Die Gefahr eines sich verstärkenden AI Divide wird hier besonders deutlich. Wenn die technologische Entwicklung und ihre Steuerungsmechanismen primär durch privilegierte Gruppen gestaltet werden, droht ein Lock-In von Werten und Strukturen, der bestehende soziale Gegebenheiten festschreibt. Die hier dargestellten Herausforderungen machen deutlich, dass ein Aufeinanderzubewegen von zwei bislang weitgehend getrennten Perspektiven notwendig ist. Einerseits der technischen Entwicklung von AI Safety und Alignment-Mechanismen und andererseits der gesellschaftlichen und bildungstheoretischen Betrachtung von Teilhabe.

### 3.1 Technische Perspektive

Bereits grundlegende Arbeiten zum sogenannten „Value Learning Problem“ (Soares, 2016, S. 3) haben darauf hingewiesen, dass selbst superintelligente Systeme, die menschliche Absichten vollständig verstehen könnten, nicht notwendigerweise im Sinne dieser Absichten handeln würden. Die technische Herausforderung, menschliche Werte korrekt in KI-Systemen zu verankern, erfordert daher besondere Ansätze der indirekten Wertrepräsentation. Es könnte daher notwendig sein, AI Safety- und Alignment-Mechanismen zu entwickeln, die nicht durch das Einschreiben fixierter Wertesysteme arbeiten, sondern flexible, anpassungsfähige Strukturen schaffen, die gesellschaftliche Aushandlungsprozesse ermöglichen und fördern. Dies erfordert eine kritische Analyse bestehender Machtstrukturen und die bewusste Entwicklung von KI-Systemen, die auf Chancengleichheit, Teilhabe und soziale Gerechtigkeit ausgerichtet sind, ohne dabei bestimmte Werte unveränderlich festzuschreiben. KI darf nicht passiv bestehende Normen und Machtverhältnisse übernehmen, sondern sollte im Sinne von Inklusion auf gerechtere soziale Verhältnisse hinarbeiten und dabei gesellschaftliche Entwicklungsprozesse berücksichtigen.

### 3.2 Gesellschaftlich-bildungstheoretische Perspektive

Aus der gesellschaftlichen und bildungstheoretischen Perspektive müssen gleichzeitig Bildungsprozesse gestaltet werden, die Menschen zur Teilhabe an den skizzierten technologischen Entwicklungen und zum aktiven Umgang mit schnellem Wandel befähigen. Die existenzielle Dimension dieser Aufgabe wird deutlich mit Blick darauf, dass die Entwicklung von KI-Systemen potenziell irreversible gesellschaftliche Weichenstellungen vornimmt, die grundlegende Fragen der Menschenwürde, Autonomie und Gerechtigkeit berühren.

Die von Wilson (2009) beschriebene Diskrepanz zwischen biologischer Prägung, institutioneller Trägheit und technologischer Entwicklung führt auch zur Frage, wie Bildungsprozesse angesichts dieser existenziellen Herausforderung gestaltet werden können. An dieser Stelle erweist sich Tofflers Konzept der *Vergänglichkeit* als besonders anschlussfähig. Toffler formulierte in „Future Shock“ bereits 1970 die bis heute aktuelle These:

It is no longer sufficient [...] to understand the past. It is not even enough [...] to understand the present, for the here-and-now environment will soon vanish. [One] must learn to anticipate the directions and rate of change. [...] Tomorrows' illiterate will not be the [one] who can't read; [it] will be the [one] who has not learned *how to learn*. (Toffler, 2022, S. 424-437; Hervorh. d. Verf.)

Damit verweist er auf die Auswirkungen beschleunigter Veränderungsprozesse und die Notwendigkeit von Bildungsprozessen und die Bedeutung von Didaktik, die weniger auf der

Vermittlung von Wissensbeständen als auf dynamischen *Prozesskompetenzen* basieren. Während bspw. Klafki (1993) mit seinem Konzept der epochaltypischen Schlüsselprobleme gesellschaftliche Transformationsprozesse eher inhaltlich-thematisch adressiert, fokussiert Tofflers Ansatz präziser auf die Fähigkeit zur Adaption in einer Welt permanenter Vergänglichkeit und Beschleunigung (Toffler, 2022, S. 20), was eine Perspektive darstellt, die uns für den Umgang mit durch KI ausgelösten Veränderungen zentral erscheint. Die besondere Relevanz von Tofflers Ansatz liegt darin, dass er nicht nur die äußere Dynamik der ständigen Veränderungen beschreibt, sondern auch deren psychologische Effekte auf Individuen erfasst, die sich in einer Welt ohne Beständigkeit zurechtfinden müssen. Die transformatorische Bildungstheorie (Koller, 2012) bietet an dieser Stelle wertvolle Ansatzpunkte. Bildung wird hierbei als Prozess der Veränderung von Selbst- und Weltverhältnissen verstanden, der durch Perturbationen/Irritation (Maturana & Varela, 2024, S. 24) ausgelöst wird und zu neuen Denk- und Handlungsformen führen kann.

Die entscheidende Frage ist somit, wie Bildungsprozesse gestaltet werden können, die Menschen für Veränderung öffnen, sie produktiv irritieren und zu neuem Denken anregen, ohne sie überwältigend zu beeinflussen. Die durch KI-Technologien ausgelösten Irritationen auf verschiedenen gesellschaftlichen Ebenen können bereits als Bildungsanlässe verstanden werden, bestehende Selbst- und Weltverständnisse in Frage stellen.

Bildung spielt daher eine Schlüsselrolle für inklusives und durch gesellschaftliche Aushandlungsprozesse strukturiertes AI Alignment, indem sie nicht nur technische KI-Kompetenzen anbahnt, sondern Menschen zur kritischen Reflexion und aktiven Mitgestaltung befähigen kann. Hierbei geht es darum, insbesondere benachteiligte Gruppen in die Lage zu versetzen, an der Wertegestaltung von KI-Systemen teilzuhaben. Andernfalls besteht das Risiko, dass bestimmte soziale Gruppen dauerhaft von der Teilhabe an KI-geprägten Gesellschaften ausgeschlossen werden bzw. bleiben. Der sich hieraus ergebende Bedarf nach adaptiven Kompetenzen für eine KI-geprägte Welt geht allerdings weit über die derzeit besonders im Fokus stehenden bildungspolitischen Forderungen nach KI-Kompetenz (Kultusministerkonferenz, 2024, Ständige Wissenschaftliche Kommission der Kultusministerkonferenz, 2024) hinaus und beinhaltet u. a.:

- die Fähigkeit zur Antizipation von Veränderung und zum frühzeitigen Erkennen von Möglichkeiten;
- die Bereitschaft zum Verlernen und zur Überwindung erlernter, nicht mehr länger haltbarer Haltungen;
- Flexibilität im Denken und Handeln, divergentes Denken und Kreativität (Robinson, 2017) sowie
- Ambidextrie (die Fähigkeit zum Umgang mit der Spannung zwischen der Notwendigkeit, sowohl Stabilität als auch Veränderung zu integrieren).

Diese Dimensionen beschreiben keine rein kognitiven Prozesse, sondern Haltungen und Dispositionen, die eine konstruktive Auseinandersetzung mit kontinuierlicher Veränderung ermöglichen.

Die Analyse der Wechselwirkungen zwischen KI-Entwicklung, digitaler Ungleichheit und gesellschaftlicher Teilhabe führt somit zu einer integrierten Betrachtung von AI Safety und AI Alignment aus inklusionsorientierter Perspektive. KI-Systeme können nur dann als wirklich sicher und „aligned“ gelten, wenn sie aktiv zur Überwindung sozialer Ungleichheiten beitragen (können), statt diese zu reproduzieren oder zu verstärken. Dies erfordert ein kritisches Bewusstsein für bestehende und sich entwickelnde Machtstrukturen, psychologische Barrieren und historische Ungleichheiten sowie Bildungskonzepte, die Menschen nicht nur zur Anpassung, sondern zur aktiven Mitgestaltung derselben befähigen.

#### **4. Conclusio: Transformative Bildung und Inklusion im Zeitalter von Künstlicher Intelligenz**

Der Beitrag macht deutlich, dass sich Inklusion im Kontext von KI nicht auf technisch-instrumentelle Aspekte verkürzen lässt. Zwar ist es wichtig, bestehende Teilhabebarrrieren durch barrierefreie Interfaces, adaptive Systeme oder assistive KI-Anwendungen zu reduzieren. Entscheidend ist jedoch, dass Inklusion in einem weiten Verständnis mehr erfordert als die Optimierung von Technologien. Sie ruft stattdessen nach einer Auseinandersetzung mit den Strukturen, Machtverhältnissen und kulturellen Deutungsmustern, die den Einsatz von KI überhaupt erst prägen.

Bildung kann in diesem Zusammenhang zu einem Raum werden, an dem Menschen nicht nur kompetent im Umgang mit KI werden, sondern in einem tieferen Sinn dazu befähigt *werden*, kritisch über ihre Selbst- und Weltverhältnisse zu reflektieren, um gemeinschaftlich zu entscheiden, wie sich Zusammenleben künftig gestalten soll. Diese Transformation von Selbst- und Weltverhältnissen kann aus der Perspektive des transformativen Lernens (Mezirow, 1991) durch krisenhafte Erfahrungen ausgelöst werden, die als anregendes Moment verstanden durch Irritation oder Diskrepanz zwischen bisherigen Erwartungen und neuen, widersprüchlichen Erfahrungen einen Reflexionsprozess auslösen, indem bisherige Deutungsmuster, sogenannte „Frames of Reference“ (Taylor, 2017, S. 17), bzw. Habitus (Bourdieu, 1987) infrage gestellt werden.

Gerade durch KI entstehen vielfältige Situationen, in denen ein solches Auseinanderdriften von Erwartungen und Wirklichkeit erfahren wird. KI kann bestehende Vorstellungen von Kreativität, Arbeit, Intelligenz oder menschlichen Fähigkeiten massiv irritieren und damit zu desorientierenden Dilemmata führen (Autenrieth, 2025). Werden diese Irritationen pädagogisch begleitet, eröffnet sich das Potenzial für transformative Bildungsprozesse: Subjekte hinterfragen und erweitern ihr bisheriges Deutungsrepertoire und gewinnen dadurch die Fähigkeit, auf technologische wie gesellschaftliche Wandlungsprozesse neu zu reagieren.

Transformative Bildung geht aber über die Aneignung von Wissen über KI hinaus und möchte zur Reflexion über eigene Grundannahmen, Haltungen und Werte anregen und Menschen dabei unterstützen, Unsicherheit in eine produktive Auseinandersetzung zu wenden und bestehende Prämissen zu hinterfragen (Mezirow, 1995, S. 46), z. B. welches technologische Fortschrittsverständnis sie übernommen haben, welche impliziten Normen über Intelligenz in sie eingeschrieben sind oder welche Vorstellungen von Teilhabe bislang unreflektiert geblieben sind. Solche kritischen Reflexions- und Diskursphasen schaffen erst die Basis, um Inklusions- und Gerechtigkeitsaspekte im Umgang mit KI angemessen bearbeiten zu können.

Damit eine solche Transformation gelingen kann, bedarf es spezifischer, didaktisch aufbereiteter Bildungssettings. Koller (2012) wie auch Mezirow (1996) betonen, dass Bildungsprozesse, die tiefgreifende Veränderung zum Ziel haben, nicht allein durch kognitive Wahrnehmung stimuliert werden. Notwendig sind Reflexionsräume, in denen Lernende Irritationen aussprechen, Perspektivenvielfalt erleben und im geschützten Rahmen neue Handlungsoptionen erproben können (Autenrieth, 2025):

- (Medien-)pädagogische Ermöglichungsräume dienen als Ausgangspunkt, um Empowerment und Resilienz in Bildungspraktiken im Kontext von KI-Systemen, digitalen Medientechnologien und Inklusion zu verknüpfen. Sie sollen Kindern, Jugendlichen und Erwachsenen geeignete Rahmenbedingungen, Impulse und Anregungen bieten, die zum eigenständigen oder gemeinsamen Nachdenken über die Zusammenhänge zwischen KI-Systemen und Inklusion anregen. Dabei ermöglichen sie die Reflexion über mögliche Wertorientierungen und schaffen Raum für die Entwicklung individueller sowie politischer Handlungsstrategien (Schluchter, 2020).
- Partizipationsräume (Mayrberger, 2020) können so gestaltet sein, dass Lernende zu Ko-Konstrukteur\*innen von gemeinsamen Lernprozessen werden. Statt fertiger Instruktionen erhalten sie Freiräume, sich mit eigenen Fragen und Erfahrungen einzubringen.
- „Safe enough spaces“ (Singer-Brodowski et al., 2022) bezeichnen Räume, in denen Lernende einerseits genügend Halt finden, um Selbstoffenbarungen zuzulassen, andererseits aber auch genug Herausforderung erfahren, um bestehende Haltungen kritisch zu hinterfragen. Durch eine behutsame Balance von Sicherheit und Zumutung kann sich das Potenzial für transformative Prozesse voll entfalten.

Inklusion als Bezugspunkt transformativer Bildung eröffnet so die Möglichkeit (und Notwendigkeit), dass Bildungsprozesse dazu anregen, nicht nur auf KI-Technologien und -anwendungen (und deren gesellschaftliche Implementierung) zu reagieren, sondern gemeinsame Perspektiven der Entwicklung und Nutzung von KI zu entfalten, welche am (ethisch-philosophischen) Kern von KI-Systemen und -anwendungen ansetzen (z. B. Bearbeitung von AI-Biases und/oder von Techno-Ableismus). So wird Inklusion mit Blick auf digitale Medientechnologien und KI-Systeme zum steten und unaufhebbaren Reflexionsmoment der Entwicklung (Inclusive AI) und gesellschaftlichen Implementierung und Nutzung von KI-Technologien und -anwendungen (AI for Inclusion).

## Literatur

- Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). *Concrete Problems in AI Safety* (arXiv:1606.06565). arXiv. <https://doi.org/10.48550/arXiv.1606.06565>
- Amodei, D. (11.10.2024). *Machines of Loving Grace. How AI Could Transform the World for the Better*. Dario Amodei. <https://darioamodei.com/machines-of-loving-grace>
- Autenrieth, D. (2025). *Transformative Bildungsprozesse und Partizipation. Eine empirische Untersuchung im Kontext von Künstlicher Intelligenz und der Lehrkräftebildung*. kopaed.
- Baker, B., Huizinga, J., Madry, A., Zaremba, W., Pachocki, J., & Farhi, D. (2025). *Detecting Misbehavior in Frontier Reasoning Models*. <https://openai.com/index/chain-of-thought-monitoring/>
- Bengio, Y., Mindermann, S., Privitera, D. (2025). *International AI Safety Report (DSIT 2025/001)*. <https://www.gov.uk/government/publications/international-ai-safety-report-2025>
- Bieling, T., & Joost, G. (2018). Technikgestaltung und Inklusion – Behinderung im Spannungsfeld von Technologie und Design: Digitalisierung – Künstliche Intelligenz – Zukunft für alle. In A. Burchardt & H. Uszkoreit (Hrsg.), *IT für soziale Inklusion* (S. 11–28). <https://doi.org/10.1515/9783110561371-002>
- Bosse, I., Schluchter, J.-R., & Zorn, I. (Hrsg.). (2019). *Handbuch Inklusion und Medienbildung*. Beltz Juventa.
- Bourdieu, P. (1987). *Sozialer Sinn: Kritik der theoretischen Vernunft* (G. Seib, Übers.; 11. Auflage). Suhrkamp.
- Brandstedt, E. William MacAskill: What we Owe the Future: A Million-Year View. *Ethic Theory Moral Prac* 26, 481–483 (2023). <https://doi.org/10.1007/s10677-023-10402-3>
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. <https://doi.org/10.48550/ARXIV.2303.12712>
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In S. A. Friedler & C. Wilson (Hrsg.), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (Bd. 81, S. 77–91). PMLR. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Carter, L., Liu, D., & Cantrell, C. (2020). Exploring the intersection of the digital divide and artificial intelligence: A hermeneutic literature review. *AIS Transactions on Human-Computer Interaction*, 12(4), 253–275.
- Clare, E. (2017). *Brilliant imperfection: Grappling with cure*. Duke University Press.
- Cotter, K., & Reisdorf, B. C. (2020). Algorithmic knowledge gaps: A new dimension of (digital) inequality. *International Journal of Communication*, 14, 21.
- DeVito, M. A., Gergle, D., & Birnholtz, J. (2017). “Algorithms ruin everything”: #RIPTwitter, folk theories, and resistance to algorithmic change in social media. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3163–3174.
- DiMaggio, P., & Hargittai, E. (2001). *From the “Digital Divide” to “Digital Inequality”: Studying Internet Use as Penetration Increases* (15). Princeton University Center for Arts and Cultural Policy Studies.
- Dogruel, L., Facciorusso, D., & Stark, B. (2020). “I’m still the master of the machine.”: Internet users’ awareness of algorithmic decision-making and their perception of its effect on their autonomy. *Information, Communication & Society*, 25(1), 1–22.

- Dörre, K., Rosa, H., Becker, K., Bose, S., & Seyd, B. (2019). *Große Transformation? Zur Zukunft moderner Gesellschaften*. VS-Verlag.
- Elena-Bucea, A., Cruz-Jesus, F., & Oliveira, T. (2021). Assessing the role of age, education, gender and income on the digital divide: Evidence for the European Union. *Information Systems Frontiers*, 23(4), 1007–1021.
- Eslami, M., Karahalios, K., & Sandvig, C. (2016). *First I “like” it, then I hide it: Folk theories of social feeds*. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2371–2382.
- Gran, A.-B., Booth, P., & Bucher, T. (2021). To be or not to be algorithm aware: A question of a new digital divide? *Information, Communication & Society*, 24(12), 1779–1796.
- Gruber, J., & Hargittai, E. (2023). The importance of algorithm skills for informed Internet use. *Big Data & Society*, 10(1). <https://doi.org/10.1177/20539517231168100>
- Hamraie, A., & Fritsch, K. (2019). Crip Technoscience Manifesto. *Catalyst: Feminism, Theory, Technoscience*, 5, 1–34. <https://doi.org/10.28968/cftt.v5i1.29607>
- Hargittai, E. (2021). *Handbook of Digital Inequality*. Edward Elgar Publishing.
- Hargittai, E., & Hsieh, Y. P. (2013). Digital inequality. In W. H. Dutton (Hrsg.), *Oxford Handbook of Internet Studies* (S. 129–150). Oxford University Press.
- Harrasser, K. (2013). *Körper 2.0: Über die technische Erweiterbarkeit des Menschen*. Transcript.
- Harrasser, K. (2020). *Parahumane Konstellationen von Körper und Technik. Aktive Anpassung und tumultöse Partnerschaften*. Springer.
- Hendawy, M. (2024). *The intensified digital divide: Comprehending GenAI*. <https://policyreview.info/articles/news/intensified-digital-divide-comprehending-genai/1772>
- Kellner, D., & Share, J. (2019). Towards Critical Digital and Media Literacies. In D. Kellner & J. Share (Hrsg.), *The Critical Media Literacy Guide* (S. 1–20). Brill.
- Klafki, W. (1993). Die bildungstheoretische Didaktik im Rahmen kritisch-konstruktiver Erziehungswissenschaft. In H. Gudjons & W. Klafki (Hrsg.), *Didaktische Theorien*. Bergmann und Helbig.
- Koller, H.-C. (2012). *Bildung anders denken: Einführung in die Theorie transformatorischer Bildungsprozesse*. Verlag W. Kohlhammer.
- Kronauer, M. (2010). Inklusion – Exklusion. Eine historische und begriffliche Annäherung an die soziale Frage der Gegenwart. In M. Kronauer (Hrsg.), *Inklusion und Weiterbildung. Reflexionen zur gesellschaftlichen Teilhabe in der Gegenwart* (S. 24–58). Bertelsmann.
- Kronauer, M. (2013). Soziologische Anmerkungen zu zwei Debatten über Inklusion und Exklusion. In R. Burtscher, E. J. Ditschek, K.-E. Ackermann, M. Kil, & M. Kronauer (Hrsg.), *Zugänge zu Inklusion, Erwachsenenbildung, Behindertenpädagogik und Soziologie im Dialog* (S. 17–25). Bertelsmann.
- Krotz, F. (2021). Medienpädagogik und Mediatisierungsforschung. In U. Sander, F. von Gross, & K.-U. Hugger (Hrsg.), *Handbuch Medienpädagogik*. VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-658-25090-4\\_22-1](https://doi.org/10.1007/978-3-658-25090-4_22-1)
- Kruschel, R., & Merz-Atalik, K. (Hrsg.) (2023). *Steuerung von Inklusion. Perspektiven auf Governance Prozesse im Schulsystem* (Bd. 52). Springer VS.
- Kultusministerkonferenz (10.10.2024). *Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen*. [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2024/2024\\_10\\_10-Handlungsempfehlung-KI.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2024/2024_10_10-Handlungsempfehlung-KI.pdf)

- Kutscher, N., & Iske, S. (2022). Diskussionsfelder der Medienpädagogik: Medien und soziale Ungleichheit. In U. Sander, F. von Gross, & K.-U. Hugger (Hrsg.), *Handbuch Medienpädagogik*. VS-Verlag. [https://doi.org/10.1007/978-3-658-23578-9\\_80](https://doi.org/10.1007/978-3-658-23578-9_80)
- Leike, J., & Sutskever, I. (2023). Introducing Superalignment. *OpenAI Blog*. <https://openai.com/research/introducing-superalignment>
- Lutz, C. (2019). Digital inequalities in the age of artificial intelligence and big data. *Human Behavior and Emerging Technologies*, 1(2), 141–148.
- Mandapuram, M., Gutlapalli, S. S., Bodepudi, A., & Reddy, M. (2018). Investigating the prospects of generative artificial intelligence. *Asian Journal of Humanity, Art and Literature*, 5(2), 167–174. <https://doi.org/10.18034/ajhal.v5i2.659>
- Maturana, H. R., & Varela, F. J. (2024). *Der Baum der Erkenntnis: Die biologischen Wurzeln menschlichen Erkennens*. FISCHER Taschenbuch.
- Mayrberger, K. (2020). Partizipative Mediendidaktik: Darstellung von Eckpunkten und Vertiefung des Partizipationsraums als konstituierendes Strukturelement. *Medienpädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 59–92. <https://doi.org/10.21240/mpaed/jb17/2020.04.26.X>
- Mazumder, S. (2021). *Inklusion und Innovation im digitalen Zeitalter*. springerprofessional.de. <https://www.springerprofessional.de/inklusion-und-innovation-im-digitalen-zeitalter/19208162>
- McMahon, R., Nazarova, N., & Robinson, L. (2023). Expanding the Boundaries of Digital Inclusion: Perspectives From Network Peripheries and Non-Adopters | Editorial | *Social Inclusion*, 11(3), 220-224. <https://doi.org/10.17645/si.v11i3.7395>
- Mezirow, J. (1995). Transformation theory of adult learning. In M. R. Welton (Hrsg.), *In defense of the lifeworld: Critical perspectives on adult learning*. State University of New York Press.
- Mezirow, J. (1996). Contemporary Paradigms of Learning. *Adult Education Quarterly*, 46(3), 158–172. <https://doi.org/10.1177/074171369604600303>
- Mjøs, O. J. (2023). *An Introduction to Global Media in the Twenty-First Century*. Bloomsbury.
- Mogge-Grotjahn, H. (2022). *Gesellschaftliche Teilhabe. Grundlagen professioneller Haltung und Handlung*. Kohlhammer.
- Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33(4), 659–684.
- Morozov, E. (2014). *To save everything, click here: The folly of technological solutionism*. PublicAffairs.
- 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* (arXiv:2311.02462). arXiv. <https://doi.org/10.48550/arXiv.2311.02462>
- Mossberger, K., Tolbert, C. J., & Stansbury, M. (2003). *Virtual Inequality. Beyond the Digital Divide*. Georgetown University Press.
- Mubarak, F., Suomi, R., & Kantola, S.-P. (2020). Confirming the links between socio-economic variables and digitalization worldwide: The unsettled debate on digital divide. *Journal of Information, Communication and Ethics in Society*, 18(3), 415–430.
- Mühlhoff, R. (2023). *Die Macht der Daten*. V&R unipress. <https://doi.org/10.14220/9783737015523>
- Niesyto, H. (2009). Digitale Medien, soziale Benachteiligung und soziale Distinktion. In *Zeitschrift für Medienpädagogik - Zeitschrift für Theorie und Praxis der Medienbildung* 17. <https://doi.org/10.21240/mpaed/17/2009.06.23.X>

- Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*.  
<https://doi.org/10.2307/j.ctt1pwt9w5>
- Park, S., & Humphry, J. (2019). Exclusion by design: Intersections of social, digital and data exclusion. *Information, Communication & Society*, 22(7), 934–953.
- Patel, K., & Patel, H. B. (2020). A state-of-the-art survey on recommendation system and prospective extensions. *Computers and Electronics in Agriculture*, 178, 105779.  
<https://doi.org/10.1016/j.compag.2020.105779>
- Ragnedda, M. (2020). *Enhancing Digital Equity: Connecting the Digital Underclass*. Springer.
- Ragnedda, M., & Mutsvauro, B. (2018). *Digital Inclusion. An International Comparative Analysis*. Lexington Books.
- Robinson, K. (2017). *Out of our minds: The power of being creative*. John Wiley & Sons, Ltd.
- Russell, S. (2022). *Artificial Intelligence and the Problem of Control* (S. 19–24).  
[https://doi.org/10.1007/978-3-030-86144-5\\_3](https://doi.org/10.1007/978-3-030-86144-5_3)
- Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210–229.  
<https://doi.org/10.1147/rd.33.0210>
- Schluchter, J.-R. (2016). Medien, Medienbildung, Empowerment. *merz – Medien und Erziehung*, 3, 24–30.
- Schluchter, J.-R. (2020). Aktive Medienarbeit als Empowerment. *(Medien)Pädagogische Ermöglichungsräume für Inklusion und inklusive Bildung*. Friedrich Jahresheft XXXVIII 2020 #schuleDIGITAL, 98–101.
- Schluchter, J.-R. (2023). Digitale Ungleichheit, Behinderung, Empowerment—  
 (Medien)Pädagogisches Empowerment als Perspektive für Inklusion. In J. Betz & J.-R. Schluchter (Hrsg.), *Schulische Medienbildung und Digitalisierung im Kontext von Behinderung und Benachteiligung* (S. 158–183). Beltz Juventa.
- Schluchter, J.-R., & Böhmer, A. (2024). Kapitalismus im digitalen Zeitalter. Bildungstheoretische Anmerkungen zu digitaler Ungleichheit und Inklusion. In V. Dander, N. Grünberger, H. Niesyto, & H. Pohlmann (Hrsg.), *Bildung und digitaler Kapitalismus* (S. 101–114). kopaed.
- Schmidl, A. (2022). *Relationen: Eine postphänomenologische Soziologie der Körper, Technologien und Wirklichkeiten*. Nomos. <https://doi.org/10.5771/9783748936855>
- Shew, A. (2020). Ableism, Technoableism, and Future AI. *IEEE Technology and Society Magazine*, 39, 40–85. <https://doi.org/10.1109/MTS.2020.2967492>
- Shew, A. (2022). How To Get A Story Wrong: Technoableism, Simulation, and Cyborg Resistance. *Including Disability*, 13–36. <https://doi.org/10.51357/id.vi1.169>
- Shew, A., & Earle, J. (2024). Cyborg–Technology Relations. *Journal of Human–Technology Relations*, 2. <https://doi.org/10.59490/jhtr.2024.2.7073>
- Shin, D. (2022). How do people judge the credibility of algorithmic sources? *AI & Society*, 37(1), 81–96.
- Singer-Brodowski, M., Förster, R., Eschenbacher, S., Biberhofer, P., & Getzin, S. (2022). Facing Crises of Unsustainability: Creating and Holding Safe Enough Spaces for Transformative Learning in Higher Education for Sustainable Development. *Frontiers in Education*, 7, 787490. <https://doi.org/10.3389/educ.2022.787490>
- Soares, N. (Juli 2016). *The Value Learning Problem*. *Ethics for Artificial Intelligence Workshop at 25th International Joint Conference on Artificial Intelligence (IJCAI-2016)*. <https://intelligence.org/files/ValueLearningProblem.pdf>
- Stalder, F. (2016). *Kultur der Digitalität*. Suhrkamp.

- Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (01.2024). *Large Language Models und ihre Potenziale im Bildungssystem*. [https://www.swk-bildung.org/content/uploads/2024/02/SWK-2024-Impulspapier\\_LargeLanguageModels.pdf](https://www.swk-bildung.org/content/uploads/2024/02/SWK-2024-Impulspapier_LargeLanguageModels.pdf)
- Stöhr, R. (2023). Technik als Hilfsmittel, als Dispositiv und als Antwort. Perspektiven auf den Einsatz von Technik im Kontext Behinderung. In *Sonderpädagogische Förderung heute* (Bd. 68, Nummer 3, S. 248–261).
- Taylor, E. W. (2017). Transformative Learning Theory. In A. Laros, T. Fuhr, & E. W. Taylor (Hrsg.), *Transformative learning meets Bildung: An international exchange* (S. 17–33). Sense Publishers.
- Toffler, A. (2022). *Future shock*. Penguin Random House LLC.
- van Dijk, J. (2020). *The Digital Divide*. Polity Press.
- van Dijk, J. (2005). *The Deepening Divide: Inequality in the Information Society*. Sage.
- Varsik, S., & Vosberg, L. (2024). *The potential impact of Artificial Intelligence on equity and inclusion in education (23)*. OECD Publishing. <https://doi.org/10.1787/15df715b-en>
- Verständig, D., Klein, A., & Iske, S. (2016). Zero-Level Digital Divide. Neues Netz und neue Ungleichheiten. *SIEGEN:SOZIAL 1/2016*.
- Walgenbach, K. (2023). Digitaler Ableismus im Feld der Bildung. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 1–26. <https://doi.org/10.21240/mpaed/jb20/2023.09.01.X>
- Wang, C., Boerman, S. C., Kroon, A. C., Möller, J., & de Vreese, C. (2024). The artificial intelligence divide: Who is the most vulnerable? *New Media & Society*, 0(0). <https://doi.org/10.1177/146144482412323>
- Wilson, E. O. (2009). *What Is Human Nature? Paleolithic Emotions, Medieval Institutions, God-Like Technology*. <https://bigthink.com/hard-science/eo-wilson-what-makes-us-human-paleolithic-emotions-medieval-institutions-god-like-technology/>
- Winner, L. (1986). *The whale and the reactor: A search for limits in an age of high technology*. University of Chicago Press.
- Winner, L. (2020). *The Whale and the Reactor: A Search for Limits in an Age of High Technology, Second Edition*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/W/bo49911830.html>
- Zarouali, B., Boerman, S. C., & de Vreese, C. H. (2021). Is this recommended by an algorithm? The development and validation of the algorithmic media content awareness scale (AMCA-scale). *Telematics and Informatics*, 62, 101607.
- Zilien, N. (2006). *Digitale Ungleichheit. Neue Technologien und alte Ungleichheiten in der Informations- und Wissensgesellschaft*. VS-Verlag.

## Kontakt

Daniel Autenrieth, Autenrieth & Partner, Querstr. 4, 74343 Sachsenheim

E-Mail: [daniel@autenrieth-partner.de](mailto:daniel@autenrieth-partner.de)

Dr. Jan-René Schluchter, Pädagogische Hochschule Ludwigsburg, Institut für Erziehungswissenschaft, Abteilung Medienpädagogik, Reuteallee 46, 71634 Ludwigsburg

E-Mail: [schluchter@ph-ludwigsburg.de](mailto:schluchter@ph-ludwigsburg.de)

Dr. Lea Schulz, Europa-Universität Flensburg, Institut für Sonderpädagogik, Abteilung Pädagogik bei Beeinträchtigung von Sprache und Kommunikation, Mitscherlich-Nielsen-Str. 4b, 24943 Flensburg

E-Mail: [lea.schulz@uni-flensburg.de](mailto:lea.schulz@uni-flensburg.de)

### Weitere Angaben zu den Autoren\*innen:

**Daniel Autenrieth** (M.Sc./M.A.) ist Informatiker und Medienpädagoge. Sein Unternehmen Autenrieth & Partner berät Organisationen bei Vorhaben an den Schnittstellen von Bildung, Digitalisierung und Künstlicher Intelligenz. Er promoviert an der RWTH Aachen zum bildungstheoretischen Alignment von KI-Systemen.

**Dr. Jan-René Schluchter** ist Akademischer Oberrat an der Pädagogischen Hochschule Ludwigsburg, Abteilung Medienpädagogik. Seine Arbeits- und Forschungsschwerpunkte sind Medienpädagogik und Inklusive Bildung, Medienpädagogik und Bildung für nachhaltige Entwicklung, Filmbildung und Animal Studies.

**Dr. Lea Schulz** ist Educational Engineer im Landesprogramm Zukunft Schule im digitalen Zeitalter des Bildungsministeriums Schleswig-Holstein. Ihre Forschungsschwerpunkte sind Diklusion, Bildungsgerechtigkeit und Künstliche Intelligenz im inklusiven Unterricht.



Dieser Text ist lizenziert unter der [Creative Commons Namensnennung - 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/).