

# Künstliche Intelligenz in der Hochschulbildung und das Transparenzproblem: Eine Analyse und ein Lösungsvorschlag

---

Dominikus Herzberg

**Abstract:** *Es ist hochgradig attraktiv, das kognitive Potenzial Künstlicher Intelligenz (KI) in der Hochschulbildung beim Lernen und Lehren auf allen hochschuldidaktischen Ebenen zu den verschiedensten Zwecken heranzuziehen und zu nutzen: von der Lernassistenten über Tutor-systeme bis hin zu Planungssystemen zur Ausrichtung der Curricula. Einmal angenommen, solche KI-Systeme stünden uns morgen in den Hochschulen zur Verfügung: Sollten wir sie einsetzen? Überraschenderweise ist uns die Entscheidung bereits abgenommen. Die Antwort lautet »Nein«, wenn man die sogar rechtlich verankerte Forderung an die Transparenz von KI ernst nimmt, womit die Nachvollziehbarkeit, Erklärbarkeit und Interpretierbarkeit ihres Verhaltens verlangt wird. Dieser Beitrag zeigt auf, dass KI mit dem Transparenz-Prinzip unverträglich ist. Dieses Dilemma muss gesellschaftlich aufgelöst werden. Gleichwohl kann die hochschulische Bildungsforschung Rahmenbedingungen schaffen, die die Intransparenz von KI integriert und nicht ausschließt, indem sie den KI-Einsatz durch eine Bildungsfolgenforschung und wissenschaftliche »Echtzeit«-Methoden begleitet. Es wäre damit denkbar, die Forderung nach Transparenz abzuschwächen oder aufzugeben.*

*It is highly attractive to use the cognitive potential of Artificial Intelligence (AI) in higher education in learning and teaching on all levels of higher education didactics for a wide variety of purposes: from learning assistance to tutoring systems to planning systems for aligning curricula. Let's assume for a moment that such AI systems would be available for us tomorrow in higher education institutions: Should we use them? Surprisingly, the decision is already taken from us. The answer is »no« if we take seriously the even legally enshrined requirement for AI to be transparent, thereby demanding that its behaviour be comprehensible, explainable, and interpretable. This paper shows that AI is incompatible with the transparency principle. This dilemma needs to be resolved by society. Nonetheless, higher educational research can create conditions that integrate rather than exclude the intransparency of AI. And this can be done by accompanying the use of AI with educational impact research and scientific »real-time« methods. It would thus be conceivable to weaken or abandon the demand for transparency.*

**Keywords:** *Künstliche Intelligenz (KI), Transparenz, Bildungsfolgenforschung, Bildungsforschung, Design-Based Research (DBR) / Artificial Intelligence (AI), transparency, educational impact research, educational research.*

## 1 Einleitung

Die Künstliche Intelligenz (KI) ist ein – aus der Sicht der Informatik – altes Forschungsgebiet. Der Computer war kaum erfunden und zu einem kommerziellen Produkt geworden, da entstanden bereits in den 1950er Jahren die ersten Überlegungen zu einem »General Problem-Solver« (Newell, Shaw & Simon, 1959). Problemlösen wurde dabei als ein zielgerichtetes Anwenden von Heuristiken (S. 5) verstanden. Historisch interessant ist, dass damals schon die Zielfindung als Reduktion eines Unterschieds zwischen zwei Zuständen interpretiert wurde (S. 9) – diese Grundidee prägt aktuell als *Gradient Descent* die Lernalgorithmen künstlicher neuronaler Netze (Anderson, 2019). Newell & Simon (1961) waren seinerzeit davon überzeugt, dass ein derartig generelles, heuristisches Problemlöseverfahren das menschliche Denken simuliere. So beeindruckend die Erfolge bis hin zu den Expertensystemen in den 1980er Jahren waren, zeigte sich doch, dass symbolische Regelsysteme, so etwa manifestiert in der Programmiersprache Prolog (Sterling & Shapiro, 1994), über bestimmte Leistungsgrenzen in ihren Fähigkeiten nicht hinaus kamen. Das führte zu einer gewissen Ernüchterung, der Menschheitstraums von einer denkenden Maschine wollte sich nicht erfüllen.

Unabhängig von den Bestrebungen, den Maschinen das Denken beizubringen und mit Qualitäten auszustatten, die über eine programmatische Algorithmik hinausgehen, ist die Vision, Maschinen als Hilfsmittel beim Lehren und Lernen einzusetzen, älter als es der Computer ist. In den 1920er Jahren entwickelte Sidney L. Pressey einen einfachen Automaten für Multiple Choice Tests, und der bekannte Psychologe B.F. Skinner stellte Ende der 1950er Jahre einen Apparat vor, der die Idee des programmierten Lernens umsetzte und Lernende positiv bei ihren Lernbemühungen unterstützte (Skinner, 1958). Das notwendig Repetitive des Lehrens und Lernens, aber auch die Möglichkeiten des individualisierten und stimulierenden Lernens im eigenen Tempo machen die Maschine zu einem reizvollen Partner als Assistenz in der Lehre (ebd.). Die Fortschreibungen dieser Ansätze werden unter dem vielschichtigen Begriff der Educational Technology, kurz EdTech, zusammengefasst (Januszewski & Molenda, 2008).

Zurück zur KI: Neue Fortschritte wurden in den 1990er Jahren mit dem sogenannten Machine Learning (ML) gemacht, als man sich von den symbolischen Regelsystemen entfernte und auf Methoden der Statistik und Wahrscheinlichkeitsrechnung setzte (Langley, 2011). Wie wir heute wissen, kam es erst durch eine Wiederentdeckung der künstlichen neuronalen Netze, gepaart mit der Rechenleistung

heutiger Computer, zu einem Quantensprung in der Nachbildung menschlicher kognitiver Fähigkeiten, dem sogenannten *Deep Learning*. Die Firma DeepMind – ein britisches Startup, das Google 2014 aufgrund der Fortschritte in der Programmierung von KI übernahm –, hat dabei die öffentliche Wahrnehmung durch eine Reihe bahnbrechender Entwicklungen geprägt: In 2016 trafen die künstliche Intelligenz AlphaGo (Silver et al., 2016) und der Weltklasse-Spieler Lee Sedol in einem medial inszenierten Spektakel aufeinander, um fünf Partien des Brettspiels Go gegeneinander zu spielen (Kohs, 2017). Bis dahin galten die kombinatorische Vielfalt der Spielmöglichkeiten und die strategische Komplexität des Spiels als uneinnehmbare Bastion für die KI und Go als die Messlatte für KI schlechthin. AlphaGo gewann 4:1 gegen Sedol. Mittlerweile ist DeepMinds KI unter dem Namen MuZero in der Lage, sich Spiele wie Go und Schach oder ein Atari-Videospiel in wenigen Stunden selbst beizubringen und dann mit einem Spielniveau aufzuwarten, dem Menschen nicht gewachsen sind (Schrittwieser et al., 2020). Das zementiert den Mythos, das dem Menschen jedwede intellektuelle Domäne abgerungen werden kann, es bedarf nicht einmal mehr speziell angepasster Programmierung.

Seit die Forschung und große Tech-Unternehmen so konsequent mit künstlichen neuronalen Netzen arbeiten und sie einsetzen, erobert die KI nicht nur eine menschlich-kognitive Domäne nach der anderen: Deep Learning beginnt zudem, bisherige algorithmische Verfahren erfolgreich abzulösen. Das beflügelt zu mehreren Ansätzen im Hochschulbereich: Zum einen scheint jede Wissenschaft zur Erkenntnisgewinnung vom KI-Einsatz profitieren zu können. Es wird zunehmend erfolgskritisch, KI als Kompetenz und Lehrgebiet in die wissenschaftliche Ausbildung einfließen zu lassen, was eine Diskussion ethischer und datenschutzrechtlicher Aspekte einschließt. Zum anderen befeuert die KI mit ihren Architekturen und Verfahren die Erforschung kognitiver Prozesse und Fähigkeiten und damit auch die Lehr- Lernforschung auf eine ganz neue Weise. Und zu guter Letzt bietet sich KI als Assistenz in der Hochschulbildung an und zwar auf allen Ebenen: auf der Mikroebene z.B. als Lernhilfe, auf der Mesoebene z.B. als Studienverlaufsbegleitung und auf der Makroebene etwa als Schöpfer neuer Studiengänge; mehr dazu im folgenden Kapitel.

Dieser Beitrag befasst sich mit der KI-Assistenz in der Hochschulbildung, was einerseits die »Teaching Machine« von Pressey und Skinner fortschreibt, andererseits weit darüber hinausgeht. Den Möglichkeiten und Chancen stehen Forderungen nach Transparenz, d.h. nach Erklärbarkeit und Nachvollziehbarkeit von KI-Systemen gegenüber, die zwar verständlich und wünschenswert sind, jedoch – so die Argumentation dieses Beitrags – eine inhärente Widerspruchslogik in sich tragen, die sich wissenschaftstheoretisch und am Gegenstand der KI begründen lässt. Auch wenn die Widerspruchslogik nicht auflösbar ist, so heißt das nicht, dass die Bildungs- und Hochschulforschung den KI-Einsatz im Hochschulbereich nicht beglei-

ten und beforschen könnte. Es bedarf einer Bildungsfolgenabschätzung und geeigneter Forschungsmethoden.

## 2 Mit Künstlicher Intelligenz lehren und lernen

Das Whitepaper »Künstliche Intelligenz in der Hochschulbildung«, herausgegeben von de Witt, Rampelt und Pinkwart (2020), bietet eine umfassende und aktuelle Übersicht zu »den Möglichkeiten und Herausforderungen von Künstlicher Intelligenz in Studium und Lehre« (S. 5). Für den KI-Einsatz als Lehr- und Lern-Partner werden eine Reihe von Gründen genannt, die Potenziale vorzeichnen und erkennen lassen. Ein paar Beispiele: Studierende können etwa »ihr eigenes Lernverhalten noch intensiver kennenlernen« und »Lehrende erfahren die Wirksamkeit ihrer didaktischen Methoden«; KI kann nicht nur beim Wissenserwerb und beim Verständnis, sondern auch bei der Ausbildung metakognitiver Fähigkeiten unterstützen; Risiko-Studierende werden identifiziert, »so dass personalisierte Interventionsstrategien vorzeitig entwickelt und Drop-out-Quoten reduziert werden können«; KI kann Studierende bei Aufgaben »durch individualisiertes Feedback unterstützen« und Lehrende können bei Ihrer Lehrstrategie unterstützt werden (S. 11f.).

Am Beispiel der Learning Analytics, also der Auswertung von erhobenen Daten rund um Lernvorgänge zum Zwecke des Verständnisses, der Optimierung und Vorhersage des Lernverhaltens und des Lernerfolgs, lässt sich aufzeigen, wie alle hochschuldidaktischen Ebenen und die Rollen der Lernenden, Lehrenden und der Organisation von KI-Innovationen erreichbar sind, siehe Abb. 1. Die individuelle Ebene des Lernens bis hin zur Entwicklung und Neugestaltung von Studiengängen ist adressierbar.

Abbildung 1: Beispiele für Learning Analytics aus de Witt et al. (2020, Tab. 2, S. 14).

	LERNENDE	LEHRENDE	ORGANISATIONEN
<b>Mikroebene (Lernsequenz)</b>	Adaptive Lernsoftware zur Klausurvorbereitung	Informationen zu Schwierigkeiten des Kurses über eine Aufgabe	Dynamische Bereitstellung von Ressourcen, z.B. Servern
<b>Mesoebene (Semester, Kurs)</b>	Monitoring des eigenen Lernerfolgs über ein Semester	Analyse von Gruppenlernprozessen in großen Onlinekursen	Unterstützung bei der zeitlichen Planung des Lehrangebots
<b>Makroebene (langfristig)</b>	Langfristige ePortfolios, Passung zu Jobprofilen	Weiterentwicklung als Lehrkraft, Erkenntnisse zu Erfolgsfaktoren	Monitoring und Revision von Studiengängen

Die möglichen »Segnungen« des Einsatzes von KI in der Hochschule sind umfassend und übergreifend. Jedes von Hochschuldidaktiker:innen nur erdenkbare Problemfeld scheint mit KI lösbar zu sein. Aber auch das muss konstatiert werden, wie Bates, Cobo, Mariño und Wheeler (2020) in ihrer Untersuchung zeigen: Die Versprechungen des KI- Einsatzes im Bildungssektor sind hoch, »but as yet, little has been achieved« (S. 4). Die Gründe sind vielfältig und haben u.a. damit zu tun, dass viel an technischer und hochschuldidaktischer Expertise, an Rechenleistung und Datenerfassung zusammenkommen müssen. Man darf damit rechnen, dass die KI-Innovationen durch Tech-Giganten von außen an das Bildungssystem herangetragen werden (S. 11f.).

Auch wenn die KI ihren Weg in den Hochschulalltag noch nicht gefunden hat, für den möglichen Einsatz werden neben ethischen und datenschutzrechtlichen Maßstäben darüber hinausgehende Forderungen an solche Systeme formuliert: Dem Ausgeliefertsein von Menschen gegenüber KI-Interventionen möchte man gesetzlich mit Transparenz begegnen (Stichwort »Explainable AI«, was noch Thema sein wird), um eine Nachvollziehbarkeit der KI- Entscheidungen zu gewährleisten (Burchardt, 2020). Ähnlich lautet es in der KI-Strategie der Bundesregierung, dass die »Transparenz, Nachvollziehbarkeit und Überprüfbarkeit der KI-Systeme herzustellen« ist (Bundesregierung, 2018, S. 38). In der Kurzfassung des Abschlussberichts der Enquete-Kommission Künstliche Intelligenz heißt es: Je nach Kritikalität muss über den Einsatz von KI informiert werden und dort, »wo Menschen von den Folgen einer Entscheidung auf Basis eines KI-Systems betroffen sind, müssen sie genügend Informationen erhalten, um ihre Rechte angemessen wahrnehmen und die Entscheidung ggf. infrage stellen zu können« (Deutscher Bundestag, 2020, S. 16).

Die Hochschulen müssen bei einer KI-Assistenz in der Hochschulbildung also stets der Forderung nach Transparenz, d.h. Nachvollziehbarkeit, Erklärbarkeit und Überprüfbarkeit der eingesetzten Systeme nachkommen. Diese Forderung erscheint mehr als vernünftig und verständlich, sieht doch nicht zuletzt die europäische Datenschutzverordnung ein »Recht auf Erklärung« vor. Beispielsweise sollte Studierenden erklärt werden, welche Gründe eine KI für eine personalisierte Studienverlaufsplanung hat. Doch so plausibel das Transparenzprinzip ist: Ist diese Forderung überhaupt erfüllbar? Die folgenden zwei Kapitel sprechen dagegen. Aus wissenschaftstheoretischer Sicht und aus inhärent technologischen Gründen können KI-Systeme nicht transparent sein.

### 3 Wie arbeitet die Wissenschaft, die KIs baut?

Die KI ist als Teilgebiet der Informatik den Technikwissenschaften zuzurechnen (vgl. Herzberg, 2019; Luft, 1988). Zu den Technikwissenschaften gibt es eine ausgear-

beitete Wissenschaftstheorie, die sich gegen die traditionell naturwissenschaftlich, empirisch ausgerichtete Wissenschaftstheorie deutlich absetzt und die den Technikwissenschaften in ihren Eigenarten und Besonderheiten als Wissenschaftsdisziplin gerecht wird (vgl. Kornwachs, 2012, 2018; Poser, 2012). Das beginnt bei einem unterschiedlichen Theoriebegriff: In den Erfahrungswissenschaften (was die Naturwissenschaften einschließt) wird aus einer Theorie eine Hypothese abgeleitet, die sich im Experiment bewähren muss und, abhängig vom Ausgang des Experiments, die Theorie verifiziert. Die Theorie ist somit ein Konstrukt, aus dem sich Wahrheitsbehauptungen generieren lassen, während die experimentelle Ebene vermutete Begründungszusammenhänge bestätigt. In den Technikwissenschaften dagegen ist eine Theorie eine Machbarkeits- bzw. Herstellungsvermutung durch Handlungen. Aus der Theorie werden Regeln abgeleitet, die einem Test unterworfen werden und, je nach Ausgang des Tests, die Effektivität der Regeln belegen und die Theorie validieren. In den Technikwissenschaften geht es demnach nicht um Ursache-Wirkungs-Beziehungen, sondern um Zweck-Mittel-Relationen der Funktionserfüllung (Kornwachs, 2012, S. 25).

Angewendet auf die wissenschaftliche Beschäftigung mit KI bedeutet das: Die KI-Forschung sucht nach informatischen Mitteln zum Zweck des Baus automatisierter und autonomer Lernsysteme und misst den Erfolg in der Effektivität der Zielerreichung. Sofern möglich, ist ein Begründungswissen zur Erklärung und zur planvollen Konstruktion hilfreich und nützlich, aber nicht notwendig. Denn technische Theorien müssen nicht auf »wahren« Theorien und Erklärungsmodellen beruhen, um effektiv zu sein und funktionierende Lösungen hervorzubringen (vgl. König, 2010).

Die KI-Forschung und die Entwicklung von KI-Anwendungen für verschiedenste Einsatzgebiete gerieren sich als Suche nach Lösungen, die funktionieren. Dabei entwickelt sich die KI zusehends zu einer Technikdisziplin, die sich aus einem rapide wachsenden Wissensstand an Modellen und Architekturen des maschinellen und »tiefen« Lernens bedient und ihre Bauten an herausfordernden Szenarien erprobt. Das führt zu einer konstruktiven Beherrschung solcher Systeme und zu einer Vermutungsbehauptung bestimmter Konfigurationen für definierte Einsatzzwecke – so zeigt es sich beispielhaft, wenn man die Entwicklung von AlphaGo zu MuZero studiert. Und selbstverständlich versucht man, die Vorgänge in einem neuronalen Netz in den verschiedenen neuronalen Schichten zu verstehen, um Systeme in Funktionskomponenten zerlegen zu können. Als Google einen Bildklassifizierer für eine solche Untersuchung »invertierte«, generierten die Neuronenschichten bizarr anmutende, traumartige Bilder, was eine Idee davon vermittelte, wie diese KI funktionierte (Mordvintsev, Olah & Tyka, 2015). Das impliziert jedoch nicht, dass man stets begründen könne oder verstünde, warum eine KI so oder so handelt, entscheidet oder sich verhält; geschweige denn, dass eine KI über ihre eigenen Beweggründe Auskunft geben könnte. Und das leitet über zur Explainable AI.

## 4 Vom Fehlen eines Begründungswissens

Die Forschung zur sogenannten Explainable AI (kurz XAI, AI steht für *Artificial Intelligence*) ist so alt wie die KI selbst und befasst sich mit dem Problem der Transparenz, d.h. mit der Frage, ob und inwieweit das Verhalten einer KI nachvollziehbar, erklärbar und interpretierbar ist (Holzinger, 2018). Es geht um die Erfassung von Kontexten, kausales Schlussfolgern, für Menschen geeignete Darbietungsformen und informationelle Verdichtung (ebd.).

Noch recht einfach verhält es sich mit den in der Einleitung erwähnten symbolischen, regelbasierten Ansätzen. Die Schlussfolgerungsketten und Heuristiken sind grundsätzlich protokollierbar und analysierbar. Allerdings sind regelbasierte Systeme wenig robust und für moderne KI-Anwendung eher uninteressant, da sie nicht mit Unsicherheit und widersprüchlichen Daten zurechtkommen (Pearl & Mackenzie, 2019, Kap. 3). Anspruchsvoll und problematisch wird es mit den statistischen und wahrscheinlichkeitsorientierten Verfahren des Machine Learning. Statistische Verfahren erfassen Korrelationen und keine Kausationen. Ohne Kausalbeziehungen gibt es keine kausalen Erklärungen. Und da, formal betrachtet, auch die neuronalen Netze des Deep Learning statistischer Natur sind, stößt man an eine prinzipielle Grenze der Transparenz. Neuronale Netze sind für uns schwarze Boxen, die wir nicht verstehen (vgl. Buder, 2020, S. 19). »Even AlphaGo's programmers cannot tell you why the program plays so well. [...] They simply experimented, and it did.« (Pearl & Mackenzie, 2019) – Dieses Ausprobieren auf der Suche nach effektiven Lösungen steht im Einklang mit einer technikwissenschaftlichen Disziplin, die Technik lösungsorientiert und zweckgebunden konstruiert.

Es bleibt die Aufgabe der Explainable AI, zu erkunden, welche KI-Systeme mit welchem Leistungspotenzial welchen Grad an Transparenz erlauben. Aber je anspruchsvoller die Nachbildung kognitiver Fähigkeiten ist, desto weniger darf man auf Transparenz hoffen. Dyson (2019, S. 39) hat das pointiert als drittes Gesetz der KI formuliert: »The third law states that any system simple enough to be understandable will not be complicated enough to behave intelligently, while any system complicated enough to behave intelligently will be too complicated to understand.« Eine Aussage, die sich auch formal belegen lässt (Alfonseca et al. 2021).

Müssen wir uns damit abfinden, dass wir KI benötigen, die uns die Handlungen von KI erklärt? Das hätte zwar eine humoristische Pointe, wird aber als Forschungsergebnis der XAI nicht zu erwarten sein.

## 5 Perspektivwechsel: Bildungsfolgenforschung und geeignete Forschungsansätze

Für all die hochschuldidaktischen Szenarien, für die ein KI-Einsatz infrage kommt und interessant ist, müssen vermutlich Systeme zum Einsatz kommen, die der Forderung nach Transparenz der KI-Entscheidungen nicht gerecht werden können. Je »brauchbarer« oder »nutzbringender« z. B. Learning Analytics sein sollen (siehe Abb. 1, die alle Ebenen der Hochschule adressiert), desto weniger wird verständlich sein, was den Interventionsvorschlag einer KI jenseits eines Verweises auf statistische Korrelationen kausal, formal-logisch oder anderweitig erkenntnisleitend begründet. Wie dargelegt, wird dieser »Mangel« an Transparenz nicht den Erfolg und die Weiterentwicklung der KI behindern. Die Informatik als Technikwissenschaft arbeitet nach dem Effektivitätsprinzip.

Es ist eine Frage, wie der Konflikt zwischen der Forderung nach Transparenz und dem Bedarf an problemlösender KI gesellschaftlich, ethisch und datenschutzrechtlich aufgelöst werden kann. Eine andere Frage ist, wie die hochschulische Bildungsforschung mit der Intransparenz von KI-Systemen umgeht bzw., so ist es treffender formuliert, umgehen könnte, denn noch ist vom KI-Einsatz in der Hochschule nicht viel zu sehen. Ich sehe eine große Chance für die Bildungsforschung darin, sich im Anwendungskontext der Hochschule in die Diskussion einzubringen mit geeigneten Forschungsmethoden und einem zu etablierenden Forschungszweig, den ich in Anlehnung an die Technikfolgenabschätzung als Bildungsfolgenabschätzung bezeichnen möchte. Welche Folgen hat der KI-Einsatz für die Studierenden, für die Lehrenden, auf das Bildungssystem, auf die mit der Hochschule verzahnte Berufspraxis, die Kultur, die Gesellschaft? Das sind schwierige Fragen von der Art, wie sie die Technikfolgenabschätzung immer wieder und immer wieder neu zu beantworten versucht (vgl. Kornwachs, 1991), selbst wenn ein technisches System intransparent in seinen Wirkmechanismen ist und in den Wechselwirkungen seines Anwendungsfelds und seines Einsatzzwecks nicht vollständig verstanden ist, was bei der Komplexität heutiger Systeme rundweg unterstellt werden muss. Eine Bildungsfolgenforschung sollte in Abgrenzung dazu nicht allein von der Technik kommend argumentieren, sondern den Kontext der Bildung u. a. als gesellschaftliches Konstrukt und die Veränderung der Bildungslandschaft mit in den Blick nehmen und normative Aspekte einbeziehen.

Bildungsfolgenforschung versucht sich an einem Blick in die Zukunft. Es bedarf aber auch Forschungsmethoden, die den Einsatz von KI als Intervention im konkreten Einsatzszenario verstehen und in der Lage sind, – um es in Anlehnung an Mukerji & Mannino (2020) zu formulieren – mit den Auswirkungen in Echtzeit umzugehen: Bildungsforschung zum KI-Einsatz in Echtzeit, um auf eine KI in ihrem Gestaltungseinfluss reagieren zu können. Das ist für eine Bildungs- und Hochschuldidaktikforschung nötig, die das Transparenzproblem integriert und nicht negiert.

Eine Forschungsmethodologie, die sich dafür andient, ist das Design-Based Research (Reinmann, 2017). »Design-Based Research ist ein Forschungsansatz in den Bildungswissenschaften, der sich mit Interventionen im Praxiseinsatz beschäftigt und den gesamten Zyklus einer Intervention in den Blick nimmt, von ihrer Gestaltung bei gleichzeitiger theoretischer Fundierung bis hin zur Umsetzung, Evaluation und Reflektion unter den Bedingungen eines Anwendungsumfeldes vielfältiger Wechselwirkungen und Abhängigkeiten. DBR geht davon aus, dass die Komplexität sozialer Bezüge (inkl. der Involviertheit der Gestaltenden, der Anwendenden und der Betroffenen) und die Kompliziertheit der Wirkmechanismen ein sich herantastendes, erschließendes Vorgehen unumgänglich machen. Ziel ist es, trotz der Unbestimmbarkeit des vollständigen Kontextes und der Unmöglichkeit der Identifikation und Isolation aller Einflussgrößen, sowohl zu praxisrelevanten Handlungsempfehlungen wie auch zu theoriebildenden Erkenntnissen zu gelangen.« (Herzberg & Joller-Graf, 2020). Eine Maschine muss nicht transparenter sein als es ein Mensch sein könnte. In gestaltungsbasierten Forschungsansätzen wird eine KI selbst zum Moment der Gestaltung, was entweder ihre Konfigurierbarkeit und Adaptierbarkeit, ihre Lernfähigkeit und Anpassungsfähigkeit, ihre Vertrauenswürdigkeit und Zuverlässigkeit oder ihre Rolle und Bedeutung in dem entsprechenden Bildungsszenario betrifft.

Die offene Frage ist, ob man angesichts dieser skizzierten Möglichkeiten des Umgangs mit Intransparenz bereit ist, die gesellschaftlich, rechtliche Transparenzforderung an KI-Systeme abzuschwächen oder aufzugeben. Die hier vorgestellte Alternative ist, den KI-Einsatz stattdessen mittels einer Bildungsfolgenforschung und einer forschenden Echtzeitbegleitung gleichsam »unter Aufsicht zu stellen«. Denn es gibt auch aus bildungstheoretischer Sicht gute Gründe dafür, das zu tun. Um mit Sesink (1990) zu argumentieren: Maschinen mit KI-Fähigkeiten sind vollzogene Abstraktionen, die nicht im Lebensvollzug agieren. Dem Menschen wohnt mit seinen Lebensbezügen ein revolutionäres Potenzial inne. »Sie [die Menschen] sind nur bereit, das soziale System, in dem sie leben, anzuerkennen, sofern ihre Lebensverhältnisse in ihm gut aufgehoben sind. Nicht die Systemreproduktion, sondern ihre Lebensreproduktion ist ihnen oberstes Ziel.« (Ebd.) Bildung versteht Sesink als »Entdeckungsreise«, wobei mehr als fraglich ist, ob eine KI dem gerecht werden kann. Und de Witt & Leineweber (2020) ergänzen: »In diesem Sinne setzt Bildung die Differenz zwischen dem Subjekt und seinen sozialen Verhältnissen voraus, Handlungs- und Urteilsfähigkeit setzt die Differenz zwischen Problemen und ihren Lösungen voraus und Intelligenz setzt schliesslich die Differenz zwischen Wissen und Nichtwissen voraus.« (S. 43). Das verweist KI im Bildungskontext auf ihre Plätze und deutet die Notwendigkeit an, KI nicht ohne eine wissenschaftlich reflektierte Begleitung in Hochschulszenarien einzusetzen.

**Dominikus Herzberg:** Technische Hochschule Mittelhessen, Fachbereich Mathematik, Naturwissenschaften und Informatik (MNI); Universität Hamburg, Hamburger Zentrum für Universitäres Lehren und Lernen (HUL); dominikus.herzberg@mni.thm.de

## Literatur

- Alfonseca, M., Cebrian, M., Fernandez Anta, A., Coviello, L., Abeliuk, A. & Rahwan, I. (2021). Superintelligence Cannot be Contained: Lessons from Computability Theory. *Journal of Artificial Intelligence Research*, 70, 65–76.
- Anderson, C. (2019). Gradient Descent. In J. Brockman (Hg.), *Possible Minds. Twenty-five ways of looking at AI* (Kap. 14). New York: Penguin Books.
- Bates, T., Cobo, C., Mariño, O. & Wheeler, S. (2020). Can artificial intelligence transform higher education? *International Journal of Educational Technology in Higher Education*, 17.
- Buder, J. (2020). Wieviel Mensch steckt in der Maschine? *Information – Wissenschaft & Praxis*, 71(1), 17–22.
- Bundesregierung. (2018). *Strategie Künstliche Intelligenz der Bundesregierung*. (Stand: November 2018).
- Bundesministerium Wirtschaft und Energie (BMWi). Zugriff am 12.01.2021. Verfügbar unter: <https://www.bundesregierung.de/breg-de/service/publikationen/strategie-kuenstliche-intelligenz-der-bundesregierung-2018-1551264>.
- Burchardt, A. (2020). Transparenz und Vertrauen schaffen. In *Künstliche Intelligenz in der Hochschulbildung*. *Whitepaper* (S. 39–40). Berlin. KI-Campus.
- de Witt, C. & Leineweber, C. (2020). Zur Bedeutung des Nichtwissens und die Suche nach Problemlösungen. Bildungstheoretische Überlegungen zur Künstlichen Intelligenz. *Medienpädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, (39, Orientierungen), 32–47. <https://doi.org/10.21240/mpaed/39/2020.12.03.X>.
- de Witt, C., Rampelt, F. & Pinkwart, N. (2020). *Künstliche Intelligenz in der Hochschulbildung*. *Whitepaper*. Berlin: KI-Campus. <https://doi.org/10.5281/zenodo.4063722>.
- Deutscher Bundestag. (2020, 27. Oktober). *Kurzfassung des Abschlussberichts der Enquete-Kommission Künstliche Intelligenz*. Enquete-Kommission Künstliche Intelligenz. Zugriff am 12.01.2021. Verfügbar unter: [https://www.bundestag.de/ausschuesse/weitere\\_gremien/enquete\\_ki/sonstige\\_veroeffentlichungen](https://www.bundestag.de/ausschuesse/weitere_gremien/enquete_ki/sonstige_veroeffentlichungen).
- Dyson, G. (2019). The Third Law. In J. Brockman (Hg.), *Possible Minds. Twenty-five ways of looking at AI* (Kap. 4). New York: Penguin Books.

- Herzberg, D. (2019). *Die Bachelorarbeit in der Praktischen Informatik im Spagat zwischen Wissenschaftsanspruch und Wirklichkeit*. Masterarbeit. Universität Hamburg, Hamburg.
- Herzberg, D. & Joller-Graf, K. (2020). Forschendes Lernen mit DBR: Eine methodologische Annäherung. *Impact Free-Journal für freie Bildungswissenschaftler*, (33), 1–16. Zugriff am 29.03.2021. Verfügbar unter: [https://gabi-reinmann.de/wp-content/uploads/2020/11/Impact\\_Free\\_33.pdf](https://gabi-reinmann.de/wp-content/uploads/2020/11/Impact_Free_33.pdf).
- Holzinger, A. (2018). Explainable AI (ex-AI). *Informatik-Spektrum*, 41(2), 138–143.
- Januszewski, A. & Molenda, M. (Hg.). (2008). *Educational Technology. A Definition with Commentary* (2. Aufl.). New York: Routledge.
- Kohs, Greg (Regie). (2017). *AlphaGo – The Movie. Full Documentary* [YouTube]. Verfügbar unter: <https://www.alphagomovie.com/>.
- König, W. (2010). Werte, Wissen und Wissensintegration in den Technikwissenschaften. In K. Kornwachs (Hg.), *Technologisches Wissen. Entstehung, Methoden, Strukturen* (Acatech diskutiert, S. 63–80). Berlin, Heidelberg: Springer.
- Kornwachs, K. (Hg.). (1991). *Reichweite und Potential der Technikfolgenabschätzung*. Stuttgart: Poeschel. Kornwachs, K. (2012). *Strukturen technologischen Wissens. Analytische Studien zu einer Wissenschaftstheorie der Technik*. Berlin: Edition Sigma.
- Kornwachs, K. (2018). *Philosophie für Ingenieure* (3., überarbeitete Auflage). München: Hanser. Langley, P. (2011). The changing science of machine learning. *Machine Learning*, 82(3), 275–279.
- Luft, A. L. (1988). *Informatik als Technik-Wissenschaft. Eine Orientierungshilfe für das Informatik-Studium*.
- Mannheim, Wien, Zürich: BI Wissenschaftsverlag.
- Mordvintsev, A., Olah, C. & Tyka, M. (2015, 17. Juni). *Inceptionism: Going Deeper into Neural Networks*. Google AI Blog, Google. Zugriff am 12.01.2021. Verfügbar unter: <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>.
- Mukerji, N. & Mannino, A. (2020). *Covid-19: Was in der Krise zählt. Über Philosophie in Echtzeit* (Reclams Universal-Bibliothek (Was bedeutet das alles?), Bd. 14053, 4. Aufl.). Ditzingen: Reclam.
- Newell, A., Shaw, J. C. & Simon, H. A. (1959). *Report on a General Problem-Solving Program* (Revision vom 9.2.1959) (P-1584). Rand Corporation. Zugriff am 11.01.2021. Verfügbar unter: [https://home.mis.u-picardie.fr/~furst/docs/Newell\\_Simon\\_General\\_Problem\\_Solving\\_1959.pdf](https://home.mis.u-picardie.fr/~furst/docs/Newell_Simon_General_Problem_Solving_1959.pdf).
- Newell, A. & Simon, H. A. (1961). GPS, A Program that Simulates Human Thought. In H. Billing (Hg.), *Lernende Automaten* (Elektronische Rechenanlagen, Bd. 3, S. 109–124). München: Oldenbourg.
- Pearl, J. & Mackenzie, D. (2019). *The Book of Why. The New Science of Cause and Effect* (Penguin Science). London: Penguin.

- Poser, H. (2012). *Wissenschaftstheorie. Eine philosophische Einführung* (Reclams Universal-Bibliothek, Bd. 18995, 2., überarbeitete und erweiterte Auflage). Stuttgart: Philipp Reclam jun.
- Reinmann, G. (2017). Design-Based Research. In D. Schemme & H. Novak (Hg.), *Gestaltungsorientierte Forschung-Basis für soziale Innovationen. Erprobte Ansätze im Zusammenwirken von Wissenschaft und Praxis* (Berichte zur beruflichen Bildung, S. 49–61). Bielefeld: Bertelsmann.
- Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S. et al. (2020). Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature*, 588(7839), 604–609.
- Sesink, W. (1990). »Künstliche Intelligenz«, Systemreproduktion und Bildung. *Neue Sammlung*, 30(2), 193–207.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G. et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489.
- Skinner, B. F. (1958). Teaching Machines. From the experimental study of learning come devices which arrange optimal conditions for self instruction. *Science (New York, N.Y.)*, 128(3330), 969–977.
- Sterling, L. S. & Shapiro, E. Y. (1994). *The Art of Prolog* (2nd ed.). Cambridge, Mass: MIT Press.