

Algorithmische Alchemie – die sozio-technische Reproduktion sozialer Ungleichheit im Bildungssystem

Tobias Röhl & Matthias Kirchner

Zusammenfassung Algorithmische Systeme versprechen, die Bewertung schulischer Leistung zu objektivieren und Bildungsangebote zu individualisieren. Statt subjektiver Urteile von Lehrpersonen sollen Algorithmen entscheiden, wie Freitextaufgaben und Essays zu bewerten sind und wie Aufgaben zugeteilt werden. Doch die Intransparenz der Technik erzeugt unbemerkt neue Formen der Ungleichheit. Die Datensätze und ihre algorithmische Auswertung unterliegen einem *algorithmic bias*, der bestehende Bildungsungleichheiten nicht bloß reproduziert, sondern sogar verstärkt. Wir diskutieren diesen Umstand anhand zweier Anwendungsfälle: adaptive Lernsysteme und automatisierte Bewertung. Wir betrachten die angewandten algorithmischen Systeme im Kontext ihrer praktischen Wirkzusammenhänge und plädieren abschließend für eine Aufwertung der pädagogischen Profession angesichts der geschilderten Risiken.

Schlüsselwörter Ungleichheit, Bildung, algorithmic bias, pädagogische Profession, data literacy

Algorithmic alchemy – socio-technical reproduction of social inequality in education

Abstract Algorithmic systems promise to make grading objective and individualize education. Instead of subjective preferences of teachers algorithms seemingly base their decisions about grading and allocating tasks on seemingly objective data. Yet, due to their nontransparent nature new forms of inequality are unwittingly introduced. Datasets and their algorithmic analysis are subject to algorithmic bias, thus not only reproducing but enhancing existing educational inequalities. Two cases for algorithmic bias in education are discussed: intelligent tutoring systems and automated grading. We situate algorithmic systems in their practical context and argue for a strengthening of the teacher profession in light of the portrayed risks.

Keywords inequality, education, algorithmic bias, educational profession, data literacy

1 Einleitung

Die Bildungssoziologie befasste sich schon sehr früh damit, dass Bildungssysteme in meritokratischen Gesellschaften nicht nur nach Leistung selektieren, sondern auch daran beteiligt sind, soziale Ungleichheiten zu reproduzieren (Peisert 1967). Teils hebt man hervor, dass Eltern aus den unteren Schichten die Kosten einer weiterführenden Bil-

dung für ihre Kinder als zu hoch bewerten (Boudon 1974; Becker 2000). Teils sieht man im an die Kinder vererbten „kulturellen Kapital“ die Grundlage dafür, dass Kinder aus ärmeren Familien in der bürgerlichen Institution Schule weniger Chancen auf Erfolg haben (Bourdieu/Passeron 1971; Georg 2005). Für Bourdieu und Passeron (1971) besteht der entscheidende Punkt darin, dass man den Einfluss sozio-kultureller Bewertungen erkennt und Bildungserfolg auf „Begabung“ zurückführt. Allenfalls gebe es wirtschaftliche Hindernisse zu überwinden:

„Zu meinen, wenn man allen gleiche wirtschaftliche Mittel bereitstelle, gäbe man auch allen, sofern sie die unerlässliche ‚Begabung‘ mitbrächten, gleiche Chancen für den Aufstieg in die höchsten Stufen der Bildungshierarchie, hieße in der Analyse der Hindernisse auf halben Wege stehenbleiben und übersehen, daß die an Prüfungskriterien gemessenen Fähigkeiten weit mehr als durch natürliche ‚Begabung‘ (die hypothetisch bleibt, solange sich der unterschiedliche schulische Erfolg auf andere Ursachen zurückführen läßt) durch die mehr oder minder große Affinität zwischen den kulturellen Gewohnheiten einer Klasse und den Anforderungen des Bildungswesens oder dessen Erfolgskriterien bedingt sind.“ (Bourdieu/Passeron 1971: 40)

Erfolg in Schule und Hochschule ist damit Ergebnis einer Bewertung, die kulturelle Passung attestiert. Dabei ist eine „symbolische Alchemie“ (Bourdieu 1992: 152) am Werk, die den eigentlichen Mechanismus verschleiert und Erfolg so als „Begabung“ individualisiert.

Jüngst schicken sich nun algorithmische Systeme an, derartige und andere Probleme im Bildungswesen zu lösen. Algorithmische Systeme – so das Versprechen – können pädagogische Tätigkeiten automatisieren, die bislang als nicht automatisierbar galten, und sie dadurch gerechter, effizienter und objektiver gestalten. Konkret betrifft dies insbesondere zwei Bereiche: die Auswahl und Zuweisung von Aufgaben an Schüler*innen sowie die Bewertung schulischer Leistungen. Im ersten Fall sollen Lernende mit adaptiven Lernsystemen bzw. *intelligent tutoring systems* (Kulik/Fletcher 2016) am Rechner individuell Aufgaben bearbeiten und erhalten durch die Systeme neue Aufgaben, die ihrem jeweils ermittelten Leistungsniveau entsprechen. Der zweite Fall bezieht sich auf *automated grading*-Systeme (Hussein/Hassan/Nassef 2019), die in der Lage sind, auch Freitextaufgaben und Aufsätze zu bewerten. Auf der einen Seite steht das Versprechen auch in Zeiten großer Klassen und großer Stoffmengen individualisierte Bildungsangebote zu erhalten, auf der anderen Seite die Hoffnung einer gerechten und objektiven Bewertung in einer Welt verdichteter Leistungsprüfungen.

Der Beitrag hinterfragt diese Hoffnungen und zeigt auf, wie sich die gutgemeinten Ansprüche der Entwickler*innen bisweilen ins Gegenteil verkehren können (siehe hierzu auch Macgilchrist 2019). Statt den Einfluss askriptiver Merkmale (wie Klasse und Geschlecht) im Bildungswesen zu reduzieren, bergen algorithmische Systeme die Gefahr in sich, bestehende soziale Ungleichheiten zu verschärfen. Es bleibt bei einer „ständische[n] Kanalisierung der Bildungschancen“ (Vester 2006), die hier durch eine *algorithmische Alchemie* gekennzeichnet ist. Wir nehmen dabei eine praxistheoretische und

techniksoziologisch informierte Perspektive ein. Damit stellen wir uns gegen einen die Macht der Algorithmen überhöhenden Technikdeterminismus, ohne die Risiken der technischen Möglichkeiten aus dem Blick zu nehmen.

Um diese Argumentation zu entfalten, entmystifiziert der Beitrag zunächst die technikdeterministische Haltung rund um algorithmische Systeme (2.). Daran anschließend erarbeitet er an zwei Anwendungsfeldern, inwiefern algorithmische Systeme die in sie gesetzten Hoffnungen enttäuschen (3.). Vor dem Hintergrund der Bourdieuschen Reproduktionsthese kann eine neue Form *algorithmischer Alchemie* beschrieben werden, die auf technische Rationalität und die individualisierte Arbeit an Defiziten setzt (4.). Abschließend diskutiert der Beitrag, was dies für die pädagogische Praxis und die Rolle der Lehrer*innen bedeutet (5.).

2 Zur Wirkmacht algorithmischer Systeme

Im öffentlichen Diskurs ist oft von der Macht der Algorithmen die Rede.¹ Mensch und Maschine stehen sich gegenüber und ersterer hat das Nachsehen. In der sozialwissenschaftlichen Forschung hat man verschiedentlich darauf hingewiesen, dass die Macht solcher Systeme nicht in der Technik allein liegt (Beer 2017; Boyd/Crawford 2012; Bucher 2018). Stattdessen sind die technischen Systeme Teil von „Algorithmenkulturen“ (Seyfert/Roberge 2017) und wie andere Technik auch mit menschlicher Praxis verwoben (Rammert/Schulz-Schaeffer 2002; Latour 2000). Ihre Macht liegt also weniger in ihnen selbst begründet als in der Bedeutung, die wir ihnen praktisch wie symbolisch zukommen lassen. So kann der Verweis auf ein algorithmisches System beispielsweise als argumentative „Ressource“ dazu dienen, ein Vorgehen zu rechtfertigen (Ziewitz 2017).

Wie andere technische Systeme sind algorithmische Systeme dementsprechend nicht losgelöst von den sozialen Umständen und Praktiken zu denken, in denen sie zum Einsatz kommen (Lange/Lenglet/Seyfert 2019; Perrotta/Selwyn 2020). Um nutzbringend eingesetzt werden zu können, bedürfen sie der menschlichen Kontrolle und ihre Einsatzgebiete müssen durch Menschenhand entsprechend ausgewählt und vorbereitet werden – etwa, wenn Böden und Beleuchtung eines Flughafens so angepasst werden, dass die Systeme Gepäckstücke, Mobiliar und Menschen besser voneinander unterscheiden können (Neyland/Möllers 2017). Man müsste daher im Fall des maschinellen Lernens von „menschengestütztem maschinellen Lernen“ (Mühlhoff 2019) sprechen. Die KI-Forscherin und Sozialwissenschaftlerin Kate Crawford hält dementsprechend fest:

„AI is neither *artificial* nor *intelligent*. Rather, artificial intelligence is both embodied and material, made from natural resources, fuel, human labor, infrastructures, logistics, histories, and classifications. AI systems are not autonomous, rational, or able to discern any-

1 Siehe etwa <https://www.deutschlandfunk.de/mensch-und-maschine-die-macht-der-algorithmen-100.html> [12. 09. 2022].

thing without extensive, computationally intensive training with large datasets or pre-defined rules and rewards.“ (Crawford 2021: 9; Hv. i. O.)

KI und andere algorithmische Systeme sind also keine aus sich heraus operierende Technik, sondern eingebunden in menschliche Praxis. Natürliche Ressourcen wie etwa Seltene Erden sind Bestandteil der Computertechnik. Ohne die Einrichtung und Wartung von Dateninfrastrukturen ist der Betrieb zunehmend vernetzter algorithmischer Systeme nicht mehr denkbar (Kitchin 2014). Die Datensätze wiederum müssen durch Menschen vorbereitet, klassifiziert und bereinigt werden (Pink et al. 2018). Algorithmische Systeme bauen auf menschlicher Arbeit auf und sind dementsprechend mit sozialen Annahmen, Interessen und Machtasymmetrien durchsetzt (Beer 2017). Und genau hierin liegt der Ansatzpunkt für eine sozialwissenschaftliche Betrachtung dieser Technik und für eine differenzierte Kritik ihrer sozialen Wirkmacht wie sie beispielsweise Vertreter*innen der *critical data studies* im Allgemeinen (Dalton/Taylor/Thatcher 2016; Iliadis/Russo 2016) wie auch für den Bildungsbereich im Besonderen (Hartong/Nikolai 2021; Williamson 2017; Decuyper/Grimaldi/Landri 2021) praktizieren. Entsprechend geht es im Folgenden darum, algorithmische Systeme in praktische Wirkzusammenhänge zu stellen. Durch diese Einbettung werden die technikdeterministischen Versprechen zur Wirkmacht algorithmischer Systeme sozial kontextualisiert und relationiert.

3 Individualisierung und Objektivierung schulischer Bewertungen?

Im Bildungsbereich lassen sich verschiedene Anwendungen für algorithmische Systeme finden (für einen Überblick siehe Luckin et al. 2016): beispielsweise die automatische Erkennung von Handschriften auf Aufgabenblättern, virtuelle Agenten und Chatbots, mit denen Schüler*innen interagieren können, die Auswertung großer Datenmengen, um lernförderliche Faktoren zu identifizieren und Probleme zu diagnostizieren (kritisch hierzu Allert/Richter 2020), die automatisierte Bewertung von Prüfungen inklusive Freitextaufgaben und Aufsätzen sowie adaptive Lernsysteme, die das Niveau von Aufgaben an den Wissensstand der Lernenden anpassen.

Die beiden letztgenannten Anwendungsfelder dienen hier als Beispiele, um zu zeigen, wie sich in algorithmischen Systemen bestehende Reproduktionsmechanismen digital teils fortführen, teils rekonfigurieren. Im Sinne einer Relationierung technischer Systeme gilt es nachzuzeichnen, wie sich Technik in ganz unterschiedliche Funktionszusammenhänge stellen lässt – und dies bisweilen auch jenseits der ursprünglichen Intention der Entwickler*innen. Dabei können sich einerseits hehre Ziele wie die intendierte Reduktion sozialer Ungleichheit ins Gegenteil verkehren oder scheinbar neutrale Diagnoseinstrumente zu laufbahnnentscheidenden Bewertungsinstrumenten werden.

3.1 „Superlehrkräfte“? Adaptive Lernsysteme und die Individualisierung von Bildungsangeboten

Als wiederkehrende Motivation und Verkaufsargument für algorithmische Systeme im Bildungsbereich gilt die individualisierte Begleitung und Bewertung von Lernenden: Kein Kind müsse sich mehr langweilen und keines sei je überfordert, da erstens das Niveau immer perfekt abgestimmt sei und zweitens die „Lehrkraft“ immer verfügbar sei. Die Hoffnung ist es, „[j]edem Kind seine Superlehrkraft“ (TA-SWISS 2020: 14) zur Verfügung stellen zu können. Dies geht einher mit einem allgemeinen Ruf nach einer Individualisierung von Bildungsangeboten, um so beispielsweise der wachsenden Heterogenität der Lernenden zu begegnen (kritisch hierzu Ricken/Thompson/Casale 2016; Kron et al. 2022).

Adaptive Lernsysteme (*intelligent tutoring systems*) versprechen, genau diesem Wunsch nach einem individualisierten Bildungsangebot nachzukommen. Lernen ist hier als „Regelkreis“ konzipiert, „in dem die Darbietung des Lehrstoffs eng an den Lernprozess und die Performanz der Lernenden gekoppelt wird“ (Kerres u. a. 2023: 2). Adaptive Lernsysteme bestehen dementsprechend aus zwei Elementen: Zum einen bearbeiten die Schüler*innen selbständig Aufgaben am Rechner und erhalten jeweils Rückmeldung zu ihren Eingaben. Zum anderen ermitteln die Systeme den Wissensstand der Schüler*in und passen so das Niveau der Aufgaben an. Hierzu ein Beispiel eines adaptiven Lernsystems aus der Schweiz: der „Lernpass plus“.² Der Lehrmittelverlag St. Gallen sowie der Lehrmittelverlag Zürich entwickeln die Online-Plattform gemeinsam. Grundlage ist ein zusammen mit der Pädagogischen Hochschule St. Gallen entwickelter Aufgabenpool für die Fächer Mathematik, Deutsch, Französisch und Englisch in den Klassen 7 bis 9. Ein regelmäßig durchgeführter „Orientierungstest“ ermittelt die jeweiligen individuellen Kompetenzen der Schüler*innen. Davon ausgehend kann das System Aufgaben nach Schwierigkeitsgrad individuell zusammenstellen. Dadurch sollen die Schüler*innen eine jeweils für sie passende Förderung erhalten. Darüber hinaus können Lehrpersonen aber auch „Standortbestimmungen“ vornehmen, bei der die gesamte Klasse die gleichen Aufgaben erhält, um so den Leistungsstand ihrer Schüler*innen miteinander zu vergleichen.

Die Entwickler*innen versprechen sich davon nicht nur eine maßgeschneiderte Förderung, sondern auch eine gesteigerte Motivation seitens der Schüler*innen. Den Lehrer*innen verheißen sie eine Entlastung von „Vor- und Nachbereitungsarbeiten“ und damit mehr „Zeit für beratende Gespräche“.³ Sie stehen damit auch für eine neue Lehr- und Lernkultur, bei der Lehrer*innen nicht als Wissensautorität, sondern als Lernbegleitung auftreten sollen (King 1993). Gleichzeitig tritt das Produkt nicht nur als Diagnosetool und Lernhilfe auf, sondern auch als Bewertungsinstrument, das „einen Beitrag zur Objektivierung der Leistungsbeurteilung“ leistet.

2 <https://lernpassplus.ch/> [07.09.2022].

3 <https://lernpassplus.ch/lernpass/#ziel> [07.09.2022].

Ein weiteres Beispiel zeigt jedoch, wie schnell aus den Diagnosemöglichkeiten der adaptiven Lernsysteme Bewertungsinstrumente werden können, die ein „parasitäres“ (Neyland 2012) Eigenleben jenseits des ursprünglichen Anwendungskontexts (hier: Schulunterricht) führen und mit anderen Bewertungspraktiken „verwoben“ (Meier/Peetz/Waibel 2016: 309) sein können. Wir bleiben in der Schweiz: Der sogenannte „Stellwerktest“ (auch eine Entwicklung der bereits genannten Lehrmittelverlage) wurde ebenfalls als Diagnosetool und adaptives Lernsystem eingeführt. Er hat in der Deutschschweiz mittlerweile eine große Verbreitung gefunden und gehört in einigen Kantonen zum festen Repertoire der schulischen Diagnostik. Es handelt sich um ein „standardisiertes Testsystem, das losgelöst vom besuchten Schultyp schulische Kernkompetenzen vergleichbar“⁴ machen soll. Anders als Lernpass plus ist er jedoch längst zu einem wichtigen Bewertungsinstrument auch außerhalb der Schule avanciert. Viele Ausbildungsbetriebe verlangen von den Jugendlichen, dass sie die Ergebnisse des Tests einreichen. Die Bildungsdirektion des Kantons Zürich muss eigens daran erinnern, dass der

„Stellwerk-Test (...) ein schulisches Förderinstrument [ist]. Das heisst, er ist nicht für die Lehrstellen-Suche gemacht worden. Trotzdem kann es vorkommen, dass ein Lehrbetrieb als Beilage zu einer Lehrstellen-Bewerbung auch dieses Stellwerk-Profil verlangt. Dabei geht es um ein besseres Bild deiner schulischen Entwicklung (...).“⁵

Die Unternehmen verlassen sich lieber auf den standardisierten Test eines überregional agierenden Anbieters statt auf die Zeugnisnoten einer lokalen Schule. Den Zeugnisnoten haftet der Ruf an, kein verlässliches Abbild der Leistungen der Schüler*innen abzugeben. Sie gelten als das Ergebnis subjektiv gefärbter, sozial ‚verunreinigter‘ Bewertungspraktiken (Filer 2000; Kalthoff 1996). Insbesondere die Auswahl von Antwortenden und die Zuweisung von Fragen im Unterrichtsgespräch folgt beispielsweise impliziten Mustern der Unterscheidung von Schüler*innen (Kalthoff 2000). Anders der Stellwerktest. Ein Aufgabenpool ist ausschließlich dem „Testen“ und nicht dem Lernen vorbehalten. Er enthält ein standardisiertes Aufgabenset, dessen Aufgaben zufällig – und damit ohne direkte menschliche Einwirkung – jede*n gleichermaßen treffen können. Dementsprechend verweisen die Entwickler*innen auf ihrer Homepage auf zahlreiche Maßnahmen, die der Standardisierung des Tests und damit der Vergleichbarkeit dienen – etwa, dass Lehrer*innen darauf achten, dass keine unerlaubten Hilfsmittel benutzt werden, und sie nur bei technischen Fragen unterstützen dürfen.⁶

Wir haben es hier dementsprechend mit einer Verschiebung von der Diagnose hin zur Bewertung zu tun. Problematisch ist dies vor dem Hintergrund von „Matthäuseffekte[n]“ (Merton 1968) im Zusammenhang mit digitalen Medien (Saleh/Sanders 2014; Patterson/Patterson 2017; Doroudi/Brunskill 2019). Von selbstorganisierten Formen des Lernens – wie wir sie bei adaptiven Lernsystemen vorfinden – profitieren ins-

4 <https://stellwerk.ch/wp-content/uploads/2021/02/Stellwerk-Flyer-2021-1.pdf>; S. 4 [13. 09. 2022].

5 <https://berufswahl.zh.ch/berufswahlfahrplan/stellwerk-test/> [15. 09. 2022].

6 <https://stellwerk.ch/stellwerk/#rahmenbedingungen> [14. 09. 2022].

besondere die leistungsstärkeren Schüler*innen, während die schwächeren unter ihnen stärkere Unterstützung seitens der Lehrkraft benötigen (Selwyn 2019: 95). Es kann also relativ unbemerkt zu einer Verschärfung bestehender Ungleichheiten kommen.

3.2 Objektivierte Bewertung? Automatisierte Bewertungssysteme

Historisch sind die heute noch gebräuchlichen schulische Zensuren als ein Versuch zu werten, dem „Ideal der Aperspektivität“ dadurch nahezukommen, dass man die „generierten Aussagen von den Subjekten“ loslöst (Lindenhayn 2018: 119): Zensuren sollten über den lokalen Kontext hinaus Vergleichbarkeit ermöglichen und nicht nur innerhalb der Klasse Leistungen bewerten. In der Forschung zum schulischen Bewerten kennt man freilich verschiedene Effekte, die diesem Bild eines objektiven Urteils zuwiderlaufen (Filer 2000). Vier Effekte sind hier relevant:

1. *Reaktivität der Messung*: Die sozialwissenschaftliche Bewertungsforschung kennt zahlreiche Beispiele für reaktive Effekte, bei denen die Messung selbst das Messergebnis verzerrt, denn „people change their behavior in reaction to being evaluated, observed, or measured“ (Espeland/Sauder 2007: 1). In der Bildungsforschung kennt man beispielsweise, das Phänomen des „teaching to the test“ (Popham 2001). Die Messung selbst verändert wie und was gelehrt und gelernt wird. Dies kann so weit gehen, dass Schülerinnen und Schüler gezielt auf standardisierte Tests (wie sie etwa in den USA üblich sind) vorbereitet werden. Im Extremfall misst der Test, dass man den Test beherrscht – mehr aber auch nicht.
2. *(Vor-)Urteilshärtungen*: Eine weitere Fehlerquelle findet sich im „confirmation bias“ (Nickerson 1998): Wenn jemand ein Urteil fällt, neigt er oder sie dazu, nach Indizien zu suchen, die dieses Urteil härten; solche, die es entkräften oder revidieren werden hingegen ignoriert. Solche Effekte lassen sich auch in der Schule beobachten (Zaborowski/Meier/Breidenstein 2011). Dies zeigt sich beispielsweise darin, dass Lehrpersonen Schulaufsätze danach vorsortieren, wie sie die bisherigen Leistungen ihrer Schülerinnen und Schüler einschätzen (Kalthoff 1996). Auch im Unterrichtsgespräch richtet sich die Auswahl und Adressierung von Schülerinnen und Schülern an den bereits getroffenen Kategorisierungen (Kalthoff 2000). In informellen wie formellen Gesprächen im Lehrerzimmer werden solche Urteile über Schülerinnen weiter gehärtet (Kalthoff/Dittrich 2016).
3. *Soziale Verzerrungen*: Bildungssoziologische Studien haben schon früh gezeigt, dass die Beurteilung schulischer Leistungen oftmals soziale Ungleichheiten reproduziert (Peisert 1967). Ein zentraler Mechanismus ist dabei die Abwertung von der bürgerlichen Kultur abweichender Muster, die als defizitär gelten (Bourdieu/Passeron 1971). Bei schulischen Prüfungen kann dies testexterne oder testinterne Merkmale betreffen. Mit testexternen Merkmale sind Zuschreibungen gemeint, die jenseits der schriftlichen dokumentierten Leistungen liegen: also ein Wissen der Bewertenden um Schichtzugehörigkeit, Geschlecht und ethnischer Herkunft der Bewerteten, das

sich nicht aus der Prüfungsleistung selbst ableiten lässt. So ist beispielsweise belegt, dass die Angabe von Namen, die auf eine niedrige soziale Herkunft schließen lassen, bei der Bewertung zu einer schlechteren Note führen kann (Kaiser 2010). Zu test-internen Merkmalen, die zu sozialen Verzerrungen führen können, zählen hingegen in erster Linie sprachliche Codes (Bernstein 1975), die aus schulischer Sicht als nicht passend gelten und deshalb per se schlechter bewertet werden.

4. *Reliabilitätsdefizit*: Schulische Leistungsbewertung ist immer auch abhängig von einer Reihe von situativen Faktoren und weicht damit erheblich vom Ideal „mechanischer Objektivität“ (Daston/Galison 2007: 121–133) ab. Dieses – messtheoretisch gesprochen – Reliabilitätsdefizit zeigt sich unter anderem darin, dass die Reihenfolge mündlicher Prüfungen oder die Tageszeit einen Einfluss auf die Bewertung haben kann (Kalthoff 1996).

Die (teil)automatisierte Bewertung durch algorithmische Systeme verspricht bei dreien dieser Punkte Abhilfe. In den Vereinigten Staaten finden entsprechende Systeme bereits große Verbreitung. In mehr als 20 Bundesstaaten setzt man *automated grading* ein.⁷ Auch KI-Systeme kommen selbstredend nicht darum herum, dass die Bewerter ihr Verhalten an die Messung anpassen (Punkt 1). Sie versprechen aber vorurteilsfrei (Punkte 2 und 3) und reliabel (Punkt 4) Urteile über schulische Leistungen fällen zu können. Amerikanische Anbieter werben u. a. damit, dass ihre Systeme „konsistent“ und ohne Beeinflussung durch „externe Faktoren“ bewerteten:

„Consistency – Teachers work hard, but they’re constantly fighting an uphill battle against factors that can affect their scoring ability, such as fatigue, distractions and time constraints. Essay grading software is immune to these hurdles, allowing for a consistent scoring standard for every reviewed essay.“⁸

„The Copyleaks Grading Tool is credible, not biased, and can assign a grade through an automatic artificial intelligence algorithm that learns exactly what should be included in the student essay, according to your requirements. The Grading tool is not affected by any external factors, only determining a grade based on the actual quality of the work.“⁹

Erste Beobachtungen strafen diese Versprechen der Hersteller jedoch Lügen und legen nahe, dass sich auch hier soziale Verzerrungen („algorithmic bias“; Akter et al. 2021) wiederfinden. Denn die Datensätze, mit denen ein Bewertungstool trainiert wird, sind auch von den sozialen Vorannahmen der Bewertenden aus Fleisch und Blut durchsetzt. Es ist also davon auszugehen, dass sich das Problem eines sozialen Bias dadurch nicht lösen lässt.

7 <https://www.vice.com/en/article/pa7dj9/flawed-algorithms-are-grading-millions-of-students-essays> [31. 03. 2023].

8 <https://www.intellimetric.com/blogs/why-essay-grading-software-is-smarter-than-all-of-us> [30. 07. 2021].

9 <https://copyleaks.com/education/ai-grading> [30. 07. 2021].

Folgender in der Presse diskutierter Fall aus dem Vereinigten Königreich illustriert dies:¹⁰ Weil man die schulischen Abschlussprüfungen 2020 während der Coronapandemie nicht ordnungsgemäß durchführen konnte, bat man zunächst die Lehrer*innen darum, voraussichtliche Abschlussnoten ihrer Schüler*innen zu schätzen. Allerdings fielen dadurch die Ergebnisse im Schnitt positiver aus als in den bisherigen Jahren. Ein vom *Department of Education* eingesetzter Algorithmus sollte nun diese wahrgenommene Schieflage korrigieren. Grundlage der Korrektur waren unter anderem die Prüfungsergebnisse der jeweiligen Schule aus den vergangenen Jahren. Nun wurden insbesondere Schüler*innen an schlecht ausgestatteten staatlichen Schulen schlechter bewertet. Diejenigen, die auf Privatschulen gingen, waren weniger stark durch die Abwertung betroffen. Der eingesetzte Algorithmus normalisierte und reproduzierte eine soziale Schieflage, die es schon vorher gab. Teile der Schülerschaft protestierten, auf ihren Schildern standen Slogans wie „*teachers know my potential, algorithms do not*“ oder „*my grades not my postcode*“. Das *Department of Education* lenkte schließlich nach einigen Tagen des Protests ein und zog die Entscheidung zurück.

4 Algorithmische Alchemie

Wie lassen sich diese Fälle vor dem Hintergrund des bildungssoziologischen Befunds einer Reproduktion sozialer Ungleichheiten einordnen? Die Befürworter*innen einer teilweisen algorithmischen Automatisierung der Bildung setzen oftmals auf einen „*technological solutionism*“ (Selwyn 2019: 18), der Lernen vor allem als individuelles Erreichen quantifizierbarer Ziele mittels Technik operationalisiert („*achievement technologies*“; Chang 2019: 34; siehe auch Macgilchrist/Allert/Bruch 2020). Die Schüler*innen lernen unabhängig voneinander an individuell zugewiesenen Aufgaben und in ihrem eigenen Tempo. Die Probleme der Bildung – etwa die Reproduktion sozialer Ungleichheiten – lassen sich aus dieser Perspektive heraus mit technischen Mitteln lösen. Dabei lässt sie den gesellschaftlichen Kontext von Bildung außen vor. Es geraten dementsprechend nicht nur die negativen Nebenfolgen der Nutzung von Bildungsmedien und -technik aus dem Blick, sondern auch gesellschaftliche Probleme wie etwa die finanzielle Unterversorgung bestimmter Schulen (Macgilchrist 2019). Aus praxistheoretischer wie techniksoziologischer Perspektive verkennt ein solcher Technikdeterminismus also die Einbettung von Technik in pädagogische und andere Praktiken (Schatzki 2010; Rammert 2007). Technik und andere materielle Entitäten sind Teil menschlicher Praxis, erst in ihrem Zusammenspiel ergibt sich ihre Wirkmacht.

Dementsprechend lassen sich die beiden angeführten Fälle neu einordnen. Der britische Korrekturalgorithmus konnte zunächst nur deshalb so einen skandalträchtigen Unterschied machen, weil das Bildungsministerium und mit ihm zumindest ein Teil

10 Siehe etwa <https://www.sueddeutsche.de/politik/coronavirus-grossbritannien-schulen-noten-1.5001145> [01.11.2021].

der Öffentlichkeit an seine Effizienz und Objektivität glaubten. Er schien als zeitsparende wie gerechte Maßnahme, um die Fehlbarkeit und Milde menschlicher Lehrer*innen zu bändigen. Schließlich sprachen die eingespeiste Datenmenge und der Mythos einer „mechanischen Objektivität“ (Daston/Galison 2007: 121–133) für diese technische Maßnahme. Das sahen die betroffenen Schüler*innen allerdings anders. Erst der massive Protest führte zu einer Rücknahme der Maßnahmen. Es zeigt sich, dass algorithmische Systeme nicht automatisiert Entscheidungen treffen, sondern in gesellschaftliche Entscheidungsprozesse eingebunden sind und dort als „Ressource“ (Ziewitz 2017) dienen, um Argumenten Gewicht und Legitimität zu verleihen. Dementsprechend deutlich musste die Gegenrede in Form von Straßenprotesten erfolgen. Verschiedene Vorstellungen von Bildung treffen aufeinander (Horvath/Steinberg/Frei 2023; Horvath/Steinberg 2023): Die Befürworter*innen der algorithmischen Korrektur berufen sich auf eine „industrielle“ Logik, bei der Produktivität und Effizienz im Vordergrund stehen. Die Kritiker*innen sehen hingegen einerseits allgemeine „staatsbürgerliche“ Gerechtigkeitsprinzipien in Gefahr und setzen andererseits auf freie Entfaltung von Subjekten und ihren Potentialen („*teachers know my potential*“). Jede Position bringt dabei nicht nur andere normative Ordnungen ins Spiel, sondern darüber hinaus auch damit verbundene Medien: algorithmische Systeme auf der einen, mündliches Gespräch und augenscheinliche Einschätzung auf der anderen.

Im Fall der adaptiven Lernsysteme lassen sich ähnliche Argumente finden. Ausbildungsbetriebe sehen in der standardisierten Diagnostik des Stellwerktest eine Möglichkeit, lokalen wie idiosynkratischen Bewertungspraktiken zu entkommen. Obwohl die Bildungspolitik und Lehrmittelverlage betonen, dass es sich lediglich um ein „Förderinstrument“ handle, verwenden ihn Betriebe *und* Schüler*innen anders. Sie sehen darin ein wichtiges Instrument für die berufliche Selektion. Anders als im Fall der algorithmischen Korrektur im Vereinigten Königreich bleiben Proteste allerdings aus, allenfalls Betrugsversuche wurden öffentlich.¹¹ Woran liegt das? Wer eine einmal ausgesprochene Bewertung in Form einer Zensur zu einem späteren Zeitpunkt nach unten „korrigiert“, riskiert, dass dies als Eingriff markiert ist. Das *Department of Education* war so als Akteur einer Entscheidung (Einsatz der algorithmischen Korrektur) deutlich sichtbar. Die angeführten Matthäuseffekte im Fall der adaptiven Lernsysteme lassen sich hingegen aus Sicht der Schüler*innen eher auf mangelnde individuelle Kompetenzen zurückführen. Denn das System behandelt alle Schüler*innen gleichermaßen auf Grundlage ihrer Eingaben. Dass bei ihren Eingaben selbstredend soziale Unterschiede bestehen, wird ausgeblendet. Und gerade weil sich das System in erster Linie als Diagnose- und nicht als Bewertungsinstrument ausgibt, bürgt es für ein objektives Urteil. Statt als Akteur aufzutreten, sind die Entwickler*innen somit lediglich als die Betreiber*innen einer (als neutral geltenden) Plattform zu fassen. So bestätigen letztlich auch Schüler*innen, die versuchen, das System auszutricksen, dessen Gültigkeit.

11 <https://www.watson.ch/digital/analyse/671374749-so-einfach-haben-schueler-den-stellwerk-test-gehackt-mit-folgen> [15.09.2022].

Was ändert sich nun an den eingangs geschilderten Mechanismen der Reproduktion sozialer Ungleichheiten im Bildungssystem? Algorithmische Systeme versprechen dem sozio-kulturellen Mechanismus – der „symbolischen Alchemie“ Bourdieus – ein Schnippchen zu schlagen. Wie die aufgeführten Beispiele gezeigt haben, kann dies nur bedingt gelingen. In der automatisierten Bewertung sind die Trainingsdatensätze selbst durchsetzt mit vergangenen sozio-kulturellen Bewertungen.¹² Und auch bei den adaptiven Lernsystemen ist die Auswahl und Gestaltung der Aufgaben selbstredend Ergebnis menschlicher Praxis, die darüber bestimmt, welche Antwort als adäquat gilt. Sie weisen den „schlechten“ Schüler*innen „einfache“ Aufgaben zu und ihren „guten“ Mitschüler*innen die „schwierigen“ Aufgaben – auch hier findet eine Grenzziehung qua Klassifikation statt (Bowker/Star 1999). Es ist also weiterhin davon auszugehen, dass Schule eine letztlich bildungsbürgerliche Institution bleibt, die auf Grundlage von Klassifikationsakten nach entsprechenden Passungen sucht.

Es kommt aber nun zu zwei Verschiebungen, die eine neue Form *algorithmischer Alchemie* darstellen. Erstens sind es nun nicht mehr Lehrer*innen, denen man zutraut Leistung zu messen, sondern technische Verfahren, die sich auf einen „*trust in numbers*“ (Porter 1996) berufen können. Dies zeichnet sich bereits jenseits von algorithmischen Systemen in dem Bedarf an standardisierten Prüfungen und Testverfahren ab – so ist etwa auch die Schulleistungsuntersuchung PISA der OECD letztlich der Versuch ein standardisiertes Messverfahren jenseits idiosynkratischer nationaler Notensysteme zu finden, um so Vergleichbarkeit herzustellen (Gorur 2011). Zweitens geht es insbesondere bei adaptiven Lernsystemen um ein anderes Verständnis von Begabung. Schüler*innen, die ein an ihre Kompetenzen angepasstes Bildungsangebot erhalten, können selbst an ihren Defiziten arbeiten und so zum Bildungserfolg geführt werden. Sowohl die Idee der Begabung wie die der individuellen Kompetenzen verorten Leistung und Erfolg im Individuum. Während aber Begabung als angeborener fester Bestandteil des Individuums gilt, kann man an seinen Kompetenzen arbeiten. Beim Vorhandensein der entsprechenden technischen Mittel (adaptive Lernsysteme u. a.) ist Misserfolg damit gemäß dieser Perspektive auf mangelnden Willen zurückzuführen und wird somit zum individuell zurechenbaren Scheitern (Gromala 2022).

Die algorithmische Alchemie baut dabei auf der Intransparenz der Systeme auf (Rudin/Radin 2019). Die Funktionsweise der algorithmischen Systeme bleibt undurchsichtig und kann gerade deshalb so wirkmächtig auftreten und unbemerkt soziale Effekte zeitigen. Selbst die Entwickler*innen können die Verarbeitung der Daten oft kaum mehr erklären – und wenn sie es könnten, unterlassen sie es aus Angst vor Konkurrenz. Für die Endnutzer*innen (Lehrer*innen und Schüler*innen) ist die Funktionsweise der algorithmischen Systeme ohnehin in der Regel eine Blackbox. Sie geben etwas hinein und erhalten eine Ausgabe, ohne genau zu wissen, wie dies zustande kommt. Das kann zu Misstrauen führen:

12 Massimo Airoidi spricht dementsprechend davon, dass Algorithmen einen „machine habitus“ (2022) aufweisen, da auch sie gesellschaftliche Strukturen inkorporieren und dadurch an ihrer Reproduktion beteiligt sind.

„Parents and school administrators may find it difficult to trust AI technologies used to influence or make decisions about student learning. Mistrust can stem from the refusal of companies to disclose their algorithms, which they argue are trade secrets, or from the ‚black box problem‘ which occurs when an algorithm’s complexity renders its processes inscrutable even to developers.“ (Lu/Harris 2018: 2)

Dies führt zu verschiedenen Problemen: Erstens kann eine solche Intransparenz dem Gerechtigkeitssinn der Schüler*innen widersprechen. Bewertete fordern ein Recht auf die Begründung der Bewertung, eine Rechtfertigung ein (Boltanski/Thévenot 2007; Meier/Peetz/Waibel 2016). Damit verbunden sind, zweitens, schulische Leistungsurteile oft Ziel von Einsprüchen seitens der Schülerschaft und ihrer Eltern (Lüders 2001). An wen sich die Einsprüche im Fall algorithmischer Bewertung richten müssen, erscheint unklar. Drittens ist ein solcher Algorithmus für „formative Beurteilung“ (Smit 2009) unbrauchbar: Für die Zwecke einer differenzierten, lernförderlichen Rückmeldung reichen die intransparenten, rein „summativen“ Beurteilungen nicht aus. Damit einher geht die Befürchtung, dass Lehrpersonen für etwas verantwortlich gemacht werden, dessen Zustandekommen sie selbst nur bedingt überblicken können (Abbass 2019). Die algorithmische Alchemie enttäuscht also gerade die normativen Erwartungen an Bildungsinstitutionen und erschwert zugleich die Kritik an Schief lagen, da deren menschliche Adressat*innen hinter der Wirkweise einer scheinbar unfehlbaren technischen Rationalität verschwinden können.

5 Der Lehrberuf zwischen De-Professionalisierung und Aufwertung

Was bedeuten nun die aufgezeigten Risiken für die pädagogische Praxis der pädagogischen Profession? Was passiert, wenn algorithmische Systeme zum „Motor“ und Daten zum „Treibstoff“ der Bildung (Selwyn et al. 2020: 2) werden? In der erziehungswissenschaftlichen Diskussion um den Lehrberuf ist die Frage nach der nötigen Professionalität bzw. der Professionalisierung zentral (siehe etwa Helsper/Tippelt 2011; Combe/Helsper 1996). Der Einsatz von algorithmischen Systemen in der Schule stellt diese Professionalität – wie in anderen Feldern auch – nun teilweise in Frage (Susskind/Susskind 2017). Insbesondere der Stellenwert menschlicher Expertise und Autonomie als zentrale Merkmale professionellen Handelns sind durch KI in besonderer Weise herausgefordert. Wenn Aufgaben (wie die Bewertung schulischer Leistungen) übernommen werden, die zum Kerngeschäft des Berufs der Lehrer*in gehören, berührt dies sowohl die Expertise als auch die Autonomie von Pädagog*innen.

Was bedeutet dies für die pädagogische Professionalität? Einige Autor*innen befürchten eine De-Professionalisierung des Lehrer*innenberufs (Selwyn 2019: 96–99). Im Extremfall hieße dies, dass Lehrpersonen zum technischen Servicepersonal der algorithmischen Systeme degradiert würden. Arbeitssoziologische und historische Forschungen zur Industriearbeit haben solche Prozesse des De-Skilling in unterschiedli-

chen Feldern beschrieben und erforscht: etwa in der Industriearbeit (Braverman 1974). Einher gehen solche Prozesse mit einer Verflachung der Ausbildung sowie sinkenden Löhnen. Auf der anderen Seite steht die Hoffnung, dass Automatisierung zur Verschiebung und Aufwertung beruflicher Tätigkeiten führen kann, weil diese Prozesse Überwachung durch kompetente Arbeitstätige benötigen (schon früh hierzu Hirschhorn 1984). Zudem können Zeitgewinne entstehen, wenn Routineaufgaben technisch automatisiert werden.

Dementsprechend sehen manche Autor*innen einen Wandel der Rolle der Lehrperson in Richtung einer Lernbegleitung. Statt der zentralen Wissensautorität eines „*sage on the stage*“ ermögliche KI die Ausübung der Rolle eines „*guide on the side*“, der Lernenden unterstützt (Susskind/Susskind 2017: 60; siehe auch King 1993). Gleichwohl bleibt offen, inwiefern Lehrer*innen als Lernbegleitende in zunehmend individualisierte und dadurch asynchrone Lernprozesse eingreifen können (Breidenstein/Rademacher 2017). Das Lernverständnis der vorgestellten algorithmischen Systeme versteht Lernen als einzelnen Erfahrungsprozess (siehe hierzu auch Deckert-Peaceman/Scholz 2022). Ein solches individualisiertes Verständnis von Lernen unterschlägt dabei auch die Rolle der Lehrer*innen und ihrer Beziehungsarbeit beim Lernen (Ricken/Thompson/Casale 2016).

Da sich zentrale pädagogische Tätigkeiten wie die Bewertung schulischer Leistungen früher oder später wohl automatisieren lassen, stellt sich weniger die Frage, ob KI die Lehrpersonen ersetzen kann, sondern ob sie dies sollte (Selwyn 2019). Die hier skizzierten Risiken sprechen nun gerade dafür, die Position der pädagogischen Profession zu stärken und Lehrpersonen so zu schulen, dass sie einerseits über entsprechende Kompetenzen im Umgang mit Daten verfügen und andererseits weiterhin mit den Lernenden in Beziehung treten können.

Statt also Lehrer*innen abzuschaffen und zu de-professionalisieren, verlangt der Einsatz algorithmischer Systeme geradezu die Aufwertung der Profession. Sie müssen „*humans on the loop*“ (Mellamphy 2021) sein, die sicherstellen, dass algorithmische Systeme stets durch kompetente menschliche Expertise gerahmt, begleitet und überwacht werden. Gerade dort, wo Maschinen Kontrolle übernehmen, müssen Menschen sie zurückerhalten: „*[I]n cybernetic settings workers must control the controls*“ (Hirschhorn 1984: 2). Letztendlich sollten algorithmische Systeme in der Bildung gerade nicht zu einem Kontrollverlust und zu einer De-Professionalisierung, sondern zu einem Wandel der Rolle der Lehrer*in führen: Von der Wissensautorität zu Datenexpert*innen. Dies fordert Kompetenzen, die über eine einfache „*data literacy*“ (Grillenberger/Romeike 2018) hinausgehen und die Infrastrukturen und Kontexte berücksichtigen, in denen Daten entstehen und zum Einsatz kommen (Gray/Gerlitz/Bounegru 2018): Zu welchem Zweck wurden Daten von wem erhoben und wie könnten sie gegebenenfalls weiterverwendet werden? Welche blinden Flecke weisen die erhobenen Daten und ihre Analyse jeweils auf?

Nur so ist gewährleistet, dass die Gefahr einer verschärften, algorithmischen Reproduktion sozialer Ungleichheiten gebannt werden kann. Algorithmische Systeme und deren Datenanalysen bergen Risiken. Die neuen Techniken sind kein unfehlbares All-

heilmittel für die Herausforderungen in der schulischen Bildung. Nur wer das begreift, kann verantwortungsvoll mit ihnen umgehen.

Literatur

- Abbass, Hussein A. (2019): Social Integration of Artificial Intelligence: Functions, Automation Allocation Logic and Human-Autonomy Trust, in: *Cognitive Computation* 11, 159–171.
- Airolidi, Massimo (2022): *Machine Habitus. Toward a Sociology of Algorithms*. Cambridge.
- Akter, Shahriar/McCarthy, Grace/Sajib, Shahriar/Michael, Katina/Dwivedi, Yogesh K./D'Ambra, John/Shen, K. N. (2021): Algorithmic bias in data-driven innovation in the age of AI, in: *International Journal of Information Management* 60, 102387.
- Allert, Heidrun/Richter, Christoph (2020): Learning Analytics: subversive, regulierende und transaktionale Praktiken, in: Stefan Iske/Johannes Fromme/Dan Verständig/Katrin Wilde (Hg.), *Big Data, Datafizierung und digitale Artefakte*, Wiesbaden, 5–35.
- Becker, Rolf (2000): Klassenlage und Bildungsentscheidungen, in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 52, 450–474.
- Beer, David (2017): The social power of algorithms, in: *Information, Communication & Society* 20, 1–13.
- Bernstein, Basil (1975): Soziale Schicht, Sprache und Sozialisation, in: Detlef C. Kochan (Hg.): *Sprache und kommunikative Kompetenz. Theoretische und empirische Beiträge zur sprachlichen Sozialisation und Primärsprachendidaktik*. Stuttgart, 43–65.
- Boltanski, Luc/Thévenot, Laurent (2007): *Über die Rechtfertigung. Eine Soziologie der kritischen Urteilkraft*. Hamburg.
- Boudon, Raymond (1974): *Education, Opportunity and Social Inequality: Changing Prospects in Western Society*. New York.
- Bourdieu, Pierre (1992): *Rede und Antwort*. Frankfurt a. M.
- Bourdieu, Pierre/Passeron, Jean-Claude (1971): *Die Illusion der Chancengleichheit. Untersuchungen zur Soziologie des Bildungswesens am Beispiel Frankreichs*. Stuttgart.
- Bowker, Geoffrey C./Star, Susan Leigh (1999): *Sorting things out: classification and its consequences*. Cambridge, MA.
- Boyd, Danah/Crawford, Kate (2012): Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon, in: *Information, Communication & Society* 15, 662–679.
- Braverman, Harry (1974): *Labour and Monopoly Capital. The Degradation of Work in the Twentieth Century*. New York.
- Breidenstein, Georg/Rademacher, Sandra (2017): *Individualisierung und Kontrolle: Empirische Studien zum geöffneten Unterricht in der Grundschule*. Wiesbaden.
- Bucher, Taina (2018): *If... Then: Algorithmic Power and Politics*. New York.
- Chang, Ethan (2019): Beyond workforce preparation: contested visions of 'twenty-first century' education reform, in: *Discourse: Studies in the Cultural Politics of Education* 40, 29–45.
- Combe, Arno/Helsper, Werner (Hg.) (1996): *Pädagogische Professionalität: Untersuchungen zum Typus pädagogischen Handelns*. Frankfurt a. M.
- Crawford, Kate (2021): *Atlas of AI*. New Haven.
- Dalton, Craig M/Taylor, Linnet/Thatcher, Jim (2016): Critical Data Studies: A dialog on data and space, in: *Big Data & Society* 3, <https://journals.sagepub.com/doi/10.1177/2053951716648346> [17.05.2023].
- Daston, Lorraine/Galison, Peter (2007): *Objektivität*. Frankfurt a. M.

- Deckert-Peaceman, Heike/Scholz, Gerold (2022): Individualisierung und Distanzunterricht in der Grundschule – am Beispiel der Lernplattform „Anton“, in: Thomas Kron/Felix Naglik/Yvonne Nehl/Michael Röhrig (Hg.): *Individualisierung und Schule*, Weinheim, 68–82.
- Decuyper, Mathias/Grimaldi, Emiliano/Landri, Paolo (2021): Introduction: Critical studies of digital education platforms, in: *Critical Studies in Education* 62, 1–16.
- Doroudi, Shayan/Brunskill, Emma (2019): Fairer but Not Fair Enough On the Equitability of Knowledge Tracing, in: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge, LAK19: The 9th International Learning Analytics & Knowledge Conference*. Tempe AZ USA: ACM, 335–339.
- Espeland, Wendy Nelson/Sauder, Michael (2007): Rankings and Reactivity. How Public Measures Recreate Social Worlds, in: *American Journal of Sociology* 113, 1–40.
- Filer, Ann (Hg.) (2000): *Assessment: Social practice and social product*. London.
- Georg, Werner (2005): Die Reproduktion sozialer Ungleichheit im Lebenslauf, in: *Zeitschrift für Soziologie der Erziehung und Sozialisation* 25, 178–197.
- Gorur, Radhika (2011): ANT on the PISA trail: following the statistical pursuit of certainty, in: *Educational Philosophy and Theory* 43, 76–93.
- Gray, Jonathan/Gerlitz, Carolin/Bounegru, Liliana (2018): Data infrastructure literacy, in: *Big Data & Society* 5, 1–13.
- Grillenberger, Andreas/Romeike, Ralf (2018): Developing a theoretically founded data literacy competency model, in: *Proceedings of the 13th Workshop in Primary and Secondary Computing Education, WiPSCE '18: Workshop in Primary and Secondary Computing Education*. Potsdam Germany: ACM, 1–10.
- Gromala, Lisa (2022): Individualisierung an Schulen im Spannungsfeld von Interaktion, Organisation und Gesellschaft, in: Thomas Kron/Felix Naglik/Yvonne Nehl/Michael Röhrig (Hg.): *Individualisierung und Schule*, Weinheim, 58–67.
- Hartong, Sigrid/Nikolai, Rita (2021): Warum es unabdingbar ist, Dateninfrastrukturen in der Bildungssteuerung stärker kritisch in den Blick zu nehmen, in: *Zeitschrift für Pädagogik* 67, 317–322.
- Helsper, Werner/Tippelt, Rudolf (Hg.) (2011): *Pädagogische Professionalität*. Weinheim.
- Hirschhorn, Larry (1984): *Beyond Mechanization. Work and Technology in a Postindustrial Age*. Cambridge, MA.
- Horvath, Kenneth/Steinberg, Mario (2023): Social classification and the changing boundaries of learning. A neopragmatic perspective on social sorting in digital education, in: *Learning, Media and Technology*, online first, 1–15.
- Horvath, Kenneth/Steinberg, Mario/Frei, Andrea Isabel (2023): Bridging inquiry and critique: a neopragmatic perspective on the making of educational futures and the role of social research, in: *Learning, Media and Technology*, online first, 1–14.
- Hussein, Mohamed Abdellatif/Hassan, Hesham/Nassef, Mohammad (2019): Automated language essay scoring systems: a literature review, in: *PeerJ Computer Science* 5, e208.
- Iliadis, Andrew/Russo, Federica (2016): Critical data studies: An introduction, in: *Big Data & Society* 3, <https://journals.sagepub.com/doi/10.1177/2053951716674238> [17. 05. 2023].
- Kaiser, Astrid (2010): Vornamen: Nomen est omen? Vorerwartungen und Vorurteile in der Grundschule, in: *Schulverwaltung. Nordrhein-Westfalen* 21, 58–59.
- Kalthoff, Herbert (1996): Das Zensurenpanoptikum. Eine ethnographische Studie zur schulischen Bewertungspraxis, in: *Zeitschrift für Soziologie* 25, 106–124.
- Kalthoff, Herbert (2000): „Wunderbar, richtig“. Zur Praxis mündlichen Bewertens im Unterricht, in: *Zeitschrift für Erziehungswissenschaft* 3, 429–446.
- Kalthoff, Herbert/Dittrich, Tristan (2016): Unterscheidung und Härtung. Bewertungs- und Notenkommunikation in Lehrerzimmer und Zeugniskonferenz, in: *Berliner Journal für Soziologie* 26, 459–483.

- Kerres, Michael/Buntins, K./Buchner, Jutta/Drachsler, H./Zawacki-Richter, O (2023): Interaktive, adaptive und künstlich-intelligente Lernprogramme: Potenziale für das digitale Lernen ausloten, in: Claudia De Witt/C. Gloerfeld/S. E. Wrede (Hg.): Künstliche Intelligenz in der Bildung, Wiesbaden (im Erscheinen).
- King, Alison (1993): From Sage on the Stage to Guide on the Side, in: *College Teaching* 41, 30–35.
- Kitchin, Rob (2014): *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London.
- Kron, Thomas/Naglik, Felix/Nehl, Yvonne/Röhrig, Michael (Hg.) (2022): *Individualisierung und Schule*. Weinheim.
- Kulik, James A./Fletcher, J. D. (2016): Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review, in: *Review of Educational Research* 86, 42–78.
- Lange, Ann-Christina/Lenglet, Marc/Seyfert, Robert (2019): On studying algorithms ethnographically: Making sense of objects of ignorance, in: *Organization* 26, 598–617.
- Latour, Bruno (2000): Ein Kollektiv von Menschen und nicht-menschlichen Wesen. Auf dem Weg durch Dädalus' Labyrinth, in: *Die Hoffnung der Pandora*, Frankfurt a. M., 211–264.
- Lindenhayn, Nils (2018): *Die Prüfung. Zur Geschichte einer pädagogischen Technologie*. Wien.
- Lu, Joyce J/Harris, Laurie A (2018): Artificial Intelligence (AI) and Education, in: Congressional Research Service, 3, <https://crsreports.congress.gov/product/details?prodcode=IF10937> [17. 05. 2023].
- Luckin, Rose/Holmes, Wayne/Griffiths, Mark/Corcier, Laurie B/Pearson (Firm)/University College, London (2016): *Intelligence unleashed: an argument for AI in education*, <https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/innovation/Intelligence-Unleashed-Publication.pdf> [12. 02. 2021].
- Lüders, Manfred (2001): Probleme von Lehrerinnen und Lehrern mit der Beurteilung von Schülerleistungen, in: *Zeitschrift für Erziehungswissenschaft* 4, 457–474.
- Macgilchrist, Felicitas (2019): Cruel optimism in edtech: when the digital data practices of educational technology providers inadvertently hinder educational equity, in: *Learning, Media and Technology* 44, 77–86.
- Macgilchrist, Felicitas/Allert, Heidrun/Bruch, Anne (2020): Students and society in the 2020s. Three future ‚histories‘ of education and technology, in: *Learning, Media and Technology* 45, 76–89.
- Meier, Frank/Petz, Thorsten/Waibel, Désirée (2016): Bewertungskonstellationen. Theoretische Überlegungen zur Soziologie der Bewertung, in: *Berliner Journal für Soziologie* 26, 307–328.
- Mellamphy, Nandita Biswas (2021): Humans „in the Loop“?: Human-Centrism, Posthumanism, and AI, in: *Nature + Culture* 16, 11–27.
- Merton, Robert K. (1968): The Matthew Effect in Science, in: *Science* 159, 56–63.
- Mühlhoff, Rainer (2019): Menschengestützte Künstliche Intelligenz. Über die soziotechnischen Voraussetzungen von „deep learning“, in: *Zeitschrift für Medienwissenschaft* 2019, 56–64.
- Neyland, Daniel (2012): Parasitic accountability, in: *Organization* 19, 845–863.
- Neyland, Daniel/Möllers, Norma (2017): Algorithmic IF ... THEN rules and the conditions and consequences of power, in: *Information, Communication & Society* 20, 45–62.
- Nickerson, Raymond S. (1998): Confirmation Bias: A Ubiquitous Phenomenon in Many Guises, in: *Review of General Psychology* 2, 175–220.
- Patterson, Richard/Patterson, Robert M. (2017): Computers and productivity: Evidence from laptop use in the college classroom, in: *Economics of Education Review* 57, 66–79.
- Peisert, Hansgert (1967): *Soziale Lage und Bildungschancen in Deutschland*. München.
- Perrotta, Carlo/Selwyn, Neil (2020): Deep learning goes to school: toward a relational understanding of AI in education, in: *Learning, Media and Technology* 45, 251–269.
- Pink, Sarah/Ruckenstein, Minna/Willim, Robert/Duque, Melisa (2018): Broken data. Conceptualising data in an emerging world, in: *Big Data & Society* 5, 1–13.
- Popham, W. James (2001): Teaching to the test, in: *Educational Leadership* 58, 16–20.
- Porter, Theodore M. (1996): *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton.

- Rammert, Werner/Schulz-Schaeffer, Ingo (2002): Technik und Handeln. Wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt, in: Werner Rammert (Hg.): Können Maschinen handeln? Frankfurt a. M., 11–64.
- Ricken, Norbert/Thompson, Christiane/Casale, Rita (Hg.) (2016): Die Sozialität der Individualisierung. Paderborn.
- Rudin, Cynthia/Radin, Joanna (2019): Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition, in: Harvard Data Science Review 1, <https://hdsr.mitpress.mit.edu/pub/f9kuryi8/release/6> [21.10.2021].
- Saleh, Amany/Sanders, Heath (2014): The Wolf in Sheep's Clothing: The Matthew Effect in Online Education, in: International Journal of Sociology of Education 3, 26–50.
- Selwyn, Neil (2019): Should Robots Replace Teachers? AI and the Future of Education. Cambridge.
- Selwyn, Neil/Hillman, Thomas/Eynon, Rebecca/Ferreira, Giselle/Knox, Jeremy/Macgilchrist, Felicitas/Sancho-Gil, Juana M. (2020): What's next for Ed-Tech? Critical hopes and concerns for the 2020s, in: Learning, Media and Technology 45, 1–6.
- Seyfert, Robert/Roberge, Jonathan (Hg.) (2017): Algorithmuskulturen: Über die rechnerische Konstruktion der Wirklichkeit. Bielefeld.
- Smit, Robbert (2009): Die formative Beurteilung und ihr Nutzen für die Entwicklung von Lernkompetenz: Eine empirische Studie in der Sekundarstufe. Baltmannsweiler.
- Susskind, Richard E./Susskind, Daniel (2017): The future of the professions: how technology will transform the work of human experts. Oxford.
- TA-SWISS (2020): Wenn Algorithmen für uns entscheiden. Chancen und Risiken der künstlichen Intelligenz. Zürich.
- Vester, Michael (2006): Die ständische Kanalisierung der Bildungschancen. Bildung und soziale Ungleichheit zwischen Boudon und Bourdieu, in: Werner, Georg (Hg.), Soziale Ungleichheit im Bildungssystem. Eine empirisch-theoretische Bestandsaufnahme, Konstanz, 13–54.
- Williamson, Ben (2017): Big data in education: the digital future of learning, policy and practice. Thousand Oaks.
- Zaborowski, Katrin Ulrike/Meier, Michael/Breidenstein, Georg (Hg.) (2011): Leistungsbewertung und Unterricht. Ethnographische Studien zur Bewertungspraxis in Gymnasium und Sekundarschule. Wiesbaden.
- Ziewitz, Malte (2017): A not quite random walk. Experimenting with the ethnomethods of the algorithm, in: Big Data & Society 4, 1–13.

Prof. Dr. Tobias Röhl, Pädagogische Hochschule Zürich, Lagerstr. 2, CH-8090 Zürich.
E-Mail: tobias.roehl@phzh.ch

Dr. Matthias Kirchner, Pädagogisches Hochschuleinstitut NMS Bern, Waisenhausplatz 29,
CH-3011 Bern. E-Mail: matthias.kirchner@phnmsbern.ch