Wandel und Kontinuität von Wissenschaft durch KI. Zur aktuellen Veränderung des Wissenschafts- und Technikverständnisses

Jan C. Schmidt

4.1 Einleitung

Diagnosen und Beschreibungen von Veränderungsprozessen sind zentral für deren Gestaltung. Im Folgenden soll erörtert werden, ob und in welcher Hinsicht sich das Wissenschaftsverständnis¹ durch KI- und Machine Learning-Verfahren verändert: Steht eine methodologische Revolution vor der Tür oder findet sich eine Kontinuität, freilich mit modifizierten Mitteln und Instrumenten?

Wenn man fragt, ob sich empirische Wissenschaft durch aktuelle Entwicklungen der Informatik und Mathematik wandelt, setzt man voraus, dass Verfahren, Mittel und Instrumente, d. h. dass Technik(en) und technikbasierte Methoden der Wissenschaft nicht nur äußerlich sind, sondern deren innere Konstitution betreffen und im Kern relevant sind. Die Anerkennung dieser Eingangsprämisse² legt überhaupt erst

Dieses Kapitel entstand durch intensive Diskussion und Zusammenarbeit mit der interdisziplinären Projektgruppe „Digitale Arbeitswelten“, der Carl Friedrich Gethmann, Peter Buxmann, Julia Distelrath, Bernhard G. Humm, Stephan Lingner, Verena Nitsch, Jan C. Schmidt und Indra Spiecker genannt Döhmann angehörten.

¹Wissenschaft wird im Sinne von *Science*, d. h. als empirische Wissenschaften, unter Einschluss empirischer Subdisziplinen der Sozial-, Kultur-, Geistes- und Humanwissenschaften verstanden. Damit fallen auch neuere Modeströmungen wie *Digital Humanities* unter dieses Begriffskonzept.

²Wissenschaftsphilosophisch ist nicht unumstritten, welchen Stellenwert Technik(en) im Rahmen von Wissenschaft spielen – doch wird die Prämisse seit einige Jahrzehnten, zumindest von der Mehrheit der Wissenschaftsphilosophen, akzeptiert.

J. C. Schmidt (✉)

Department of Social Sciences, Hochschule Darmstadt, Darmstadt, Deutschland

E-Mail: Jan.Schmidt@h-da.de

einen Untersuchungsbedarf nahe. Schaut man dann detaillierter hin, ist für die Frage nach einem Wandel das Verständnis von Wissenschaft ausschlaggebend. Veränderungsthese sind abhängig von Referenzsystemen, bezüglich derer eine Veränderung untersucht wird. Allerdings ist das Referenzsystem, also das, was als Wissenschaft angesehen wird, seit jeher umstritten: Der Begriff „Wissenschaft“ hat keine einheitliche Bedeutung und ist Gegenstand vielfältiger Kontroversen.

Um nicht von vorneherein zentrale Aspekte eines wissenschaftlichen Wandels auszuschließen, entwirft dieses Kapitel ein plurales Verständnis von Wissenschaft.³ Ein solch plurales Referenzsystem, das sich pragmatisch auf Ziele und mit diesen zusammenhängenden Erfolgs- und Leistungsmerkmalen bezieht, wird durch Äußerungen von Wissenschaftlern und von Wissenschaftsphilosophen gestützt. Peter Janich (1997, S. 62) spricht von einem „eindrucksvollen technischen, prognostischen und Erklärungs-Erfolg“ der Wissenschaft. Janichs Dreier-Typologie lässt sich in *einem* Punkt weiter entwickeln, nämlich um den *Prüf- und Testerfolg* von Aussagen, Theorien oder Modellen an empirischen Daten oder Experimenten. Die anderen drei Erfolgskriterien können im Folgenden übernommen werden: Janichs *Erklärungserfolg* nimmt deduktiv-gesetzeshafte und mechanistische Erklärungen in den Blick, d. h. die Zurückführung von neuen Phänomenen auf allgemeine Theorien oder die Angabe Phänomen-erzeugender Ursachen. Ferner umfasst der *technische Erfolg* von Wissenschaft menschliche Interventionen in Gegebenes und die handelnde Herstellung reproduzierbarer Ereignisse. Hier geht es um die Handlungsseite von Wissenschaft. Prototyp ist das Experiment, doch dieses Erfolgskriterium zeigt sich allgemeiner in technischen Systemen, Produkten und Prozessen. Der *prognostische Erfolg* weist wiederum auf die mathematische Möglichkeit der Vorwegnahme von Zukunft hin; ihm ist eine Zeitstruktur eingraviert.

Im Rahmen der vier Erfolgs- und Leistungsmerkmale, in denen sich verschiedene Ziele von Wissenschaft widerspiegeln, können vier (nicht disjunkte) Wissenschaftsverständnisse identifiziert werden (Abschn. 4.2). Die zu entfaltende These ist, dass sich einige Aspekte dieser vier Wissenschaftsverständnisse durch KI- und Machine Learning-Verfahren wandeln, andere nicht. Für die Beantwortung der Frage eines Wandels ist also das jeweils investierte Wissenschaftsverständnis von zentraler Bedeutung (Abschn. 4.3). Um weitergehend den Erfolg der neuen informatischen Verfahren einschätzen zu können, sind ferner jene Objektsysteme in den Blick zu nehmen, an denen diese ihre Leistungsfähigkeit zeigen: komplexe, nichtlineare, mitunter instabile und oftmals selbstorganisationsfähige Objektsysteme in Natur, Technik, Wirtschaft und Gesellschaft, zu denen herkömmliche Wissenschaft aus prinzipiellen Gründen kaum einen methodischen Zugang fand. Die Pointe der KI- und Machine-Learning-Verfahren ist, wie gezeigt wird, dass sie die für diese Objektsysteme charakteristische (verborgene) schwache Kausalität oder Regelmäßigkeit für praktische Zwecke nutzbar machen können, ohne diese erkennen (und in Gesetzesform offenlegen) zu müssen. So wird deutlich: Nicht Korrelation, sondern schwache Kausalität ist der eigentliche Angriffspunkt der KI- und Machine-Learning-Verfahren (Abschn. 4.4).

³ Siehe hierzu auch die Darstellung und Begründung in Schmidt (2008, S. 19 ff.).

Allerdings hat die erfolgreiche methodische und objektseitige Erweiterung der Wissenschaft ihren Preis. Nicht nur die neu bzw. vertieft zugänglichen Objektsysteme sind komplex und nichtlinear, sondern auch die neuen informatischen Verfahren sind es. Sie sind, wie dargelegt wird, in ihrer Struktur und Dynamik eine Black Box – weithin eigendynamisch, unübersichtlich, intransparent, opak, schwer interpretierbar. Damit sind epistemische Risiken der Qualitätssicherung und des intersubjektiven Geltungsausweises der Resultate verbunden (Abschn. 4.5). Wenn KI- und Machine-Learning-Verfahren nun weithin opak bleiben und wenn sie selbst eine so große Komplexität, Nichtlinearität, Instabilität und Selbstorganisationsfähigkeit aufweisen, dann ist zu fragen, ob das eigentlich noch die Art von Technik (und von technischen Instrumenten, Werkzeugen, Verfahren und Mittel) ist, die wir lebensweltlich kennen und die wir bisher eingesetzt haben. Es wird gezeigt, dass ein qualitativ anderer Techniktyp, eine *nachmoderne Technik*, im Entstehen ist. Dieser kann (i. w. S.) Autonomie und Entscheidungsfähigkeit, Produktivität und Adaptivität zugeschrieben werden. Dabei sind im Kern der nachmodernen Technik KI- und Machine-Learning-Verfahren am Werke (Abschn. 4.6). Schließlich wird eine Einordnung der informatischen Verfahren in allgemeine Technologietrends vorgenommen, wobei diese Verfahren in unterschiedlicher Hinsicht als paradigmatisch angesehen werden können (Abschn. 4.7). Das soll anhand von drei prominenten Sondierungsbegriffen dargelegt werden, nämlich: Ermöglichungstechnologie, Konvergenztechnologie sowie Technowissenschaft.

Zusammengenommen wird dafür argumentiert, dass sich das Regime der modernen Wissenschaft, in dem alle vier Wissenschaftsverständnisse balanciert koexistieren, durch ein nachmodernes Regime erweitert und ergänzt wird. Im nachmodernen Regime treten die vier Wissenschaftsverständnisse in eine veränderte, nicht-balancierte Relation. Es werden andere Ziele verfolgt: Statt der (zumindest partiellen) Erkenntnis-, Theorie- und Erklärungsorientierung tritt eine (starke und explizite) Orientierung auf (technischen und gesellschaftlichen) Nutzen und Anwendbarkeit. In welchem Verhältnis zukünftig diese beiden komplementären epistemischen Regime stehen werden, ist bis dato nicht entschieden. Zu vermuten ist, dass eine weitere Verschiebung in Richtung des nachmodernen Regimes vonstattengeht. Die damit verbundene mögliche Verdrängung des modernen Regimes gilt es, kritisch in den Blick zu nehmen.

4.2 Vier leitende Wissenschaftsverständnisse – das Referenzsystem

4.2.1 Wissenschaft als theoriebasiertes Prognoseverfahren

Seit der griechischen Antike zielt der Mensch darauf ab, Zukunft vorwegzunehmen und zukünftige Ereignisse vorauszusehen. Während einst das Orakel von Delphi angerufen wurde, tritt an diese Stelle später die Wissenschaft. Durch wissenschaft-

liches Wissen soll der Zufall gebändigt und die Natur berechenbar werden, so die erkenntnisleitende Hoffnung. Wer Zukünftiges kennt, kann sein heutiges Handeln darauf einrichten. Pierre-Simon Laplace, Physiker und Mathematiker, spitzt dieses Ideal im frühen 19. Jahrhundert als Höhepunkt einer mechanistischen Weltansicht zu. Wenn wir alles über Natur wissen würden – alle Gesetze, Anfangs- und Randbedingungen –, dann sei Zukunft und Vergangenheit universell berechenbar, insbesondere liege uns die Zukunft vollständig vor Augen. Auch wenn es uns Menschen aufgrund kognitiver Limitierungen nicht vergönnt sein sollte, die Fähigkeit universeller Berechenbarkeit zu erlangen, so komme diese Möglichkeit zumindest einem omnipotenten Geist oder Dämon zu. Denn Natur sei von Gesetzmäßigkeit, stärker noch: von deterministischer Gesetzmäßigkeit, durchzogen und als solche im Prinzip berechenbar. Auguste Comte überträgt dieses Ideal des Laplace'schen Dämon von der Natur auf sozial-gesellschaftliche Systeme und begründet die Sozialwissenschaften. Sein berühmtes Diktum lautet: Wissen, um vorherzusehen; vorhersehen, um handeln zu können.⁴

Wissenschaft wird vielfach über das Leistungsmerkmal erfolgreicher Prognosen qualifiziert.⁵ Carl Friedrich v. Weizsäcker (1974, S. 122) sieht im „Schluss auf die Zukunft“ die „eigentliche Pointe“ der Wissenschaft. Dabei können Prognosen verschiedene Funktionen erfüllen. Es macht einen Unterschied, ob man in erfolgreichen Prognosen Ziel, Produkt und Höhepunkt von Wissenschaft sieht, so wie Laplace es mit seinem Dämon tat. Oder ob diese lediglich ein Mittel und Medium darstellen, um Aussagen induktiv zu verallgemeinern oder um Phänomene zu diagnostizieren, etwa den Planeten Neptun, die Sonnenfinsternis im Jahre 1919, das Higgs-Boson oder eine seltene Krankheit. Oder um Aussagen zu deduzieren und um Testbarkeit von Theorien im Sinne falsifizierender Verfahren zu ermöglichen, Theorien anzuerkennen, und Intersubjektivität von Aussagen sicherzustellen. „Jede Nachprüfung“, so Wolfgang Stegmüller (1987, S. 266), „vollzieht sich über beobachtbare Voraussagen.“⁶ Ferner können noch weitreichendere, deutlich zu unterscheidende Motive eine Rolle spielen, wenn man sich auf Prognose bezieht. Erfolgreiche Prognosen stellen nicht nur aus wissenschaftsinterner Perspektive ein zentrales Motiv dar, wie soeben beschrieben. Vielmehr gibt es auch wissenschafts-externe Motive, nämlich Prognosen als wissenschaftsbasierte Ermöglichung von menschlichen Handlungen allgemein, etwa zum planenden Umgang mit Naturereignissen wie einst Pest, Erdbeben, Dürrekatastrophen oder wie heute mit SARS-CoV.

Prognostizierbarkeit wird also stets mit wissenschaftlicher Erkenntnis, d. h. mit Wissen,⁷ in Verbindung gebracht. Erkenntnis manifestiert sich in Theorien, Modellen oder Gesetzmäßigkeiten, welche Aussagen über das jeweilige Gegenstandsfeld

⁴An der Programmatik von Comtes' Sozialphysik ist freilich innerhalb der Sozial- und Humanwissenschaften eine Diskussion entbrannt, die sich bis in den so genannten Positivismusstreit der Soziologie in den 1960er-Jahren erstreckt.

⁵Die unterschiedlichen Verständnisweisen und Konzepte von Prognose bzw. Berechenbarkeit werden untersucht in Schmidt (2004).

⁶Dies sagt Stegmüller bezüglich Duhem und Quine.

⁷Genauer, siehe Kap. 3: Wissen kann als Resultat des Erkenntnisprozesses verstanden werden.

und den jeweiligen Phänomenbereich treffen. Mit anderen Worten, ein Gegenstandswissen wird als notwendig angesehen. Zwar mögen Prognosen mitunter auf statistischen, nicht auf deterministischen Gesetzmäßigkeiten aufbauen und somit probabilistisch sein, immer jedoch basieren sie auf einem Gegenstandswissen und sind in diesem Sinne theoriebasiert. Insofern war Auguste Comtes Diktum für moderne Wissenschaft im umfassenden Sinne treffend: um vorhersehen zu können, ist Wissen unabdingbar.

4.2.2 Wissenschaft als Experimentierhandeln und Interventionsform zur Herstellung von Reproduzierbarkeit

Das epochal Neue der wissenschaftlichen Moderne, die in der Spätrenaissance eingeleitet wurde und mit Namen wie Kopernikus und Kepler, mit Galilei, Bacon und Descartes verbunden ist, ist das Experiment. Wissenschaft ist seit dem 17. Jahrhundert nicht nur eine passiv-kontemplative Beobachtungswissenschaft desjenigen, was von selbst da ist, oder eine kognitiv-theoriebezogene Prognosewissenschaft. Vielmehr ist sie, wie Francis Bacon zu Beginn der Moderne im Jahre 1620 ausführt, eine aktive, intervenierende Experimentalwissenschaft (Bacon 1999). Ohne technische Apparate und Instrumente ist moderne Wissenschaft unmöglich. Natur zeigt sich gerade in technisch präparierten, experimentellen Settings in ihrem Kern, nämlich als (Natur-)Gesetz. Technik ist nicht als Gegenüberstellung zu Natur zu verstehen, wie in der Antike und im Mittelalter, sondern Technik ist das, was naturgesetzmäßig möglich ist: Technik ist zweckmäßig realisierte Naturgesetzmäßigkeit. Das hat zu einer „Revolution der Denkart“ geführt, wie Kant (1989, S. xii) sagt – eine neue Sichtweise, die das aristotelische Natur- wie Technikverständnis modifiziert und zu einer veränderten Verhältnisbestimmung führt. So ist moderne Wissenschaft eng mit der Technikentwicklung verbunden. Für das Wissenschaftsverständnis rückt das technische Handeln und die experimentelle Handlungspraxis in den Mittelpunkt (Hacking 1983).

Wissenschaftliches Wissen wird dabei nicht nur durch (und an) Technik hervorgebracht, sondern manifestiert sich auch *in* Technik, d. h. in technischen Produkten und Verfahren: Technik ist Werk von Wissenschaft. Für Bacon (1999, S. 81) „ergänzen“ sich „Wissen und menschliches Können“, Theorie und Technik. Bacons wirkungsgeschichtlich prägende Programmatik zu Beginn der Moderne kennzeichnet auch Galileis und Torricellis Laborexperimente und ihre empirisch-messende Systematik der präparierten Naturbefragung. Sie sind die ersten Experimentatoren im modernen Sinn. Vico, Hume, Kant⁸ und Mill heben dann hervor, dass man nur dasjenige verstanden habe, was man machen, kontrollieren und erzeugen

⁸ Insbesondere Kant bezieht sich in der Vorrede der *Kritik der reinen Vernunft* (B xii) direkt auf Bacon.

könne. Die experimentelle Konstruktion ist die Bedingung der Möglichkeit, dass etwas *als etwas* erkannt werden kann: Die Naturforscher begriffen, so Kant, „daß die Vernunft nur das einsieht, was sie selbst nach ihrem Entwurfe hervorbringt.“ (Kant 1989, B XIIIff.)⁹ Der technisch-experimentelle Zugang erhält ein Natur- sowie Naturwissenschafts-konstitutives Moment.¹⁰

Wissenschaftliche Tat-Sachen sind Sachen menschlicher Tat. Damit verbunden sind normative Kriterien, wie Tat-Sachen her- und sicherzustellen sind. Gernot Böhme und Wolfgang van den Daele (1977, S. 189) heben hervor, dass das „methodische Ideal“ der Wissenschaft „die regelmäßige Tatsache [ist], die die Bedingungen enthält, unter der ihre Beobachtung für jedermann und jederzeit wiederholbar ist.“ Jürgen Mittelstraß (1998, S. 106) sieht in der „Reproduzierbarkeit“ „eine allgemeine wissenschaftliche Norm“, die „die Kontrollierbarkeit wissenschaftlicher Aussagen“ durch die damit verbundenen Personen-, Zeit- und Ortsinvarianzen sicherstelle. Für Karl Popper (1989, S. 54) steht fest, dass „nichtreproduzierbare Einzelereignisse [...] für die Wissenschaft bedeutungslos“ sind – und Popper verweist auf die Notwendigkeit „experimenteller Testfähigkeit“.

Moderne Wissensgesellschaften gehen davon aus, dass wissenschaftliches Wissen notwendig ist, um erfolgreich handeln zu können: „Was bei der Betrachtung als Ursache erfasst ist, dient bei der Ausführung als Regel“, so schon Bacon (1999, S. 81). Man könnte behaupten, nichts ist so handlungsmächtig wie wissenschaftliches Wissen, nichts ist so praktisch wie eine gute Theorie.

4.2.3 *Wissenschaft als stringente Prüfmethode und als kritischer Garant intersubjektiver Geltung*

Neben die beiden Handlungsseiten von Wissenschaft – d. h. neben die technisch-experimentelle sowie die mathematisch-prognoseermöglichende Handlungsseite – treten zwei weitere Verständnisweisen. Gelegentlich werden diese als grundlegender angesehen, insofern sie stärker theoretische Dimensionen in den Mittelpunkt rücken. Es handelt sich um das *Test- und Prüfbarkeitsverständnis* von Wissenschaft einerseits sowie das *Erklärbarkeitsverständnis* andererseits; zu letzterem verweisen wir auf das nächste Unterkapitel, ersteres behandeln wir im Folgenden.

Kein anderer Zugang zur Wirklichkeit prüft Aussagen so personenunabhängig, so hartnäckig, so vorurteilsfrei, so kritisch: Wissenschaft basiert auf methodischem

⁹Weiter heißt es: „Die Vernunft muß mit ihren Prinzipien, nach denen allein übereinkommende Erscheinungen für Gesetze gelten können, in einer Hand, und mit dem Experiment, das sie nach jenen ausdachte, in der anderen, an die Natur gehen, zwar um von ihr belehrt zu werden, aber nicht in der Qualität eines Schülers [...], sondern eines bestellten Richters, der die Zeugen nötigt, auf Fragen zu antworten, die er ihnen vorlegt.“ (Kant 1989, B XIII–B XIV).

¹⁰Allerdings können wir heute Bacon nicht mehr folgen: Dass wissenschaftlich-technischer Fortschritt stets human-gesellschaftlichen Fortschritt induziert, gilt angesichts der Ambivalenz von moderner Wissenschaft und Technik als naiv.

Zweifel, so programmatisch René Descartes im 17. Jahrhundert. Über drei Jahrhunderte später versteht der Wissenschaftsforscher Robert K. Merton, ganz ähnlich, Wissenschaft als „organisierten Skeptizismus“. Ausgehend von Descartes und der rationalistischen Traditionslinie betont Karl Popper, dass Wissenschaft der „kritischen“ Methode bedürfe, in deren Mittelpunkt Prüfbar- bzw. Testbarkeit steht (Popper 1989, S. 77 f.). Prüfbarkeit ermöglicht eine Falsifikation von Aussagen und Theorien,¹¹ womit Bedingung der Intersubjektivität her- und sichergestellt werden können. Bei Popper bildet sich hieraus der Kern für sein normatives Abgrenzungskriterium: Wissenschaftliches Wissen ist ein solches, das an der Empirie scheitern können muss.¹²

Dass mit der Prüfbarkeit zunächst allgemein eine deduktive, also ableitend-folgernde Beziehung verbunden ist, ist unverkennbar: Prüfen kann man nur, was schon da ist. Während die Genese von Aussagen und Theorien als unsicher, mitunter als spekulativ gilt, weshalb zumeist von Hypothesen und Vermutungen gesprochen wird, wird durch die Prüfung eine Widerlegung oder Bewährung ermöglicht. Popper setzt analog zu Bacon voraus, dass es Entscheidungsexperimente („*experimentum crucis*“) gibt, welche eine Aussage oder Theorie eindeutig als falsch ausweisen und diese verwerfen. Im positiven Fall hat sich die Theorie zunächst bewährt. Neben Poppers deduktiv ausgerichtete Methodologie spielt Prüfbarkeit in anderen Traditionslinien, etwa der empiristischen¹³ eine zentrale Rolle. Auch Rudolf Carnap (1936) verwendet Prüfbarkeit als Kriterium, um eine Aussage als wissenschaftlich zu bezeichnen. Popper und Carnap können sich auf Vorläufer stützen. Ernst Mach (1917, S. 465) hat im späten 19. Jahrhundert gefordert, dass dort, „wo weder eine Bestätigung noch eine Widerlegung ist, [...] die Wissenschaft nichts zu schaffen [hat].“

Im Zentrum dieses Wissenschaftsverständnisses findet sich also die Idee der Zusammenführung von Theorie und Empirie: Je „näher“ die auf Basis von Theorien deduktiv gewonnenen Aussagen oder die durch induktive Verallgemeinerung entwickelten mathematischen Gesetzmäßigkeiten den empirischen Daten sind, d. h. je genauer die Repräsentation, desto mehr spricht für eine Theorie. Wissenschaft wird mithin als kritisches Prüfverfahren verstehbar, durch welche die Theorie- mit der Empirieseite verbunden wird.

¹¹ Popper selbst verwendet „Prüfbarkeit“ und „Falsifizierbarkeit“ vielfach synonym. Für Popper ist „Prüfbarkeit [...] Falsifizierbarkeit“, wobei „es Grade der Prüfbarkeit [gibt]“, welche zu unterscheiden sind und riskante Hypothesen durch hohe Prüfbarkeitsgrade gekennzeichnet werden können (Popper 1989, S. 77 f.). Also: „Ein empirisch-wissenschaftliches [Aussagen-] System muss an der Erfahrung scheitern können.“ (Popper 1989, S. 15) Bei Popper ist von Bestätigung im Sinne von Bewährung bzw. „corroboration“ und bei Carnap im Sinne von „confirmation“ die Rede.

¹² Damit tritt Prüfbarkeit methodologisch an die Stelle einer unerreichbaren metaphysischen Grundlegung wissenschaftlichen Wissens.

¹³ Diese sind induktivistisch bzw. wahrscheinlichkeitslogisch ausgerichtet.

4.2.4 *Wissenschaft als Grundlage des Weltverständnisses und als Erklärungslieferant*

Wissenschaft tritt an, Phänomene zu erklären. Erklärt werden Phänomene unter Verwendung von Theorien, Modellen oder Gesetzen – man spricht allgemein von theorie-, modell- oder gesetzesbasierter Erklärung (Kitcher und Salmon 1989). Das zu erklärende Phänomen oder genauer: die zu erklärende Aussage wird mit etwas Allgemeinem, wie einem Gesetz, in Verbindung gebracht.

Bevor eine Erklärung vorgenommen wird, muss das zu erklärende Phänomen mathematisch beschrieben werden. Hierzu werden Kriterien wie Sparsamkeit, Einfachheit und Verdichtung in Anschlag gebracht: Wissenschaft eliminiert Redundanzen in der Beschreibung der Phänomene. Diese Grundhaltung wird schon im frühen 14. Jahrhundert von Wilhelm von Ockham formuliert. Das Ockham'sche Rasiermesser entwickelt sich zu einem zentralen Prinzip der modernen Wissenschaft. René Descartes schreibt im 17. Jahrhundert der Wissenschaft entsprechend ins Stammbuch, dass sie „verwinkelte und dunkle Propositionen stufenweise auf einfachere zurückführen“ solle. Schließlich kann der Physiker Heinrich Hertz (1963, S. xxv.) Ende des 19. Jahrhunderts ausrufen: Alle Wissenschaftler „sind einstimmig darin, dass es die Aufgabe [...] sei, die Erscheinungen der Natur auf die einfachen Gesetze [...] zurückzuführen.“

Je mehr Phänomene einheitlich und sparsam-ökonomisch im Rahmen einer Theorie beschrieben werden können, je größer die Kompression der Information über das betrachtete Objektsystem, desto höher gilt deren reduktive Beschreibungsleistung. Reduktiv meint, dass eine solche Theorie die Phänomene auf das Wesentliche zu reduzieren vermag. Die Reduktion findet via mathematischer Gesetze statt, die den Strukturkern von Theorien darstellen. Diese Theorien und ihre Gesetze beschreiben dann nicht nur die Phänomene, vielmehr erklären die Theorien die Phänomene auch in einer bestimmten Hinsicht, weshalb man von deduktiv-gesetzeshaften Erklärungen spricht. Aus einer solchen Theorie lassen sich – das ist der Anspruch an dieses Leistungsmerkmal – alle zentralen Eigenschaften der Phänomene wiedergewinnen.¹⁴ In der reduktiven Erklärungsstruktur der Wissenschaft zeigt sich nach Ernst Mach eine Denkökonomie, welche als normatives Prinzip für das Forschungshandeln gefasst werden kann: Nach diesem Ökonomieprinzip hat Wissenschaft „Erfahrung zu ersetzen oder zu ersparen durch Nachbildung und Vorbildung von Tatsachen in Gedanken.“ (Mach 1917, S. 457) Für Kant ist darüber hinaus „die Ersparung der Prinzipien nicht bloß ein ökonomischer Grundsatz der Vernunft, sondern inneres Gesetz der Natur“, womit er auf die regulative Idee der Einheit hinweist, „nach welcher jedermann voraussetzt, diese Vernunftseinheit sei

¹⁴ Dass Wissenschaft in diesem Sinne erfolgreich ist, sieht man daran, dass beispielsweise in der Physik drei der vier fundamentalen Gesetze in einem vereinheitlichten Korpus vorliegen. Prominente Beispiele für Erklärungen sind mit den Namen Newton, Maxwell, Einstein oder auch Darwin verbunden. Newton konnte durch seine drei Gesetze, verbunden mit dem Gravitationsgesetz und speziellen Anfangs- und Randbedingung, die Bewegung von Planeten beschreiben, also diese Phänomene der Planetenbewegung ableiten, also u. a. die Kepler Gesetze.

der Natur selbst angemessen“ (Kant 1989, B 678/A 650). Vereinheitlichung ist dann nicht nur Vereinfachung, sondern spiegelt sich konstitutiv in der Natur, nämlich als Einheit der Natur.

Der Bezug zu Gesetzen und Theorien findet sich nicht nur in dem bisher diskutierten deduktiv-nomologischen Erklärungsverständnis, sondern auch in funktionalistischen oder dispositionellen Konzepten, die sich stärker an den Technikwissenschaften, der Biologie oder den Sozialwissenschaften orientieren. Ähnliches gilt hinsichtlich mechanistischer Erklärung: Diese erklärt ein Phänomen durch Angabe der Beiträge der Elemente bzw. Komponenten oder allgemeiner durch jene der Teilsysteme, die an der Erzeugung oder Hervorbringung des Phänomens ursächlich beteiligt sind. Es kann im mechanischen (oder algorithmischen) Sinne von Ursache und Wirkung gesprochen werden. Auch hier wird auf kausale Gesetzmäßigkeiten, allerdings weniger auf Theorien, sondern auf mechanische, d. h. faktisch-materielle bzw. -energetische Erzeugung(leistung) Bezug genommen: B, das Phänomen (Wirkung), wird immer (ursächlich) hervorgerufen durch A, die bewirkende Komponente (Ursache). Dabei kann die kausale Gesetzmäßigkeit deterministisch oder probabilistisch sein (Bunge 1987).

Zusammengenommen kann man sagen, dass Wissenschaft darauf abzielt, Phänomene zu erklären. Notwendige Bedingung für eine Erklärung ist die Angabe von grundlegenden Theorien oder Modellen oder, in abgeschwächten Varianten, von Gesetzmäßigkeiten oder Regelmäßigkeiten.

4.2.5 Zwischenfazit: Vielheit und Einheit wissenschaftlicher Disziplinen

In den vier Leistungsmerkmalen – die das Referenzsystem bilden, das im Anschluss zur Untersuchung von Veränderungen durch KI- und Machine Learning-Verfahren verwendet werden soll – spiegeln sich, wie angedeutet, vier verschiedene Ziele moderner Wissenschaft wider. Bei so viel Pluralität verwundert es nicht, dass unterschiedliche Wissenschaftsdisziplinen verschiedene Leistungsmerkmale als zentral ansehen und andere als sekundär. Technikwissenschaftler präferieren das Ziel der Prognose- und (Re-)Produktionsleistung (ad 1 und ad 2), während Naturwissenschaftler (Physiker, Chemiker, Biologen) traditionell eher die Prüf- und die Erklärungsleistung in den Mittelpunkt stellen (ad 3 und ad 4). Bei Human-, Sozial- und Wirtschaftswissenschaftlern fällt dies, je nach Feld, in dem sie forschend unterwegs sind, jeweils unterschiedlich aus. Ein Kognitions- und Neuropsychologe, der über den Spracherwerb von Kindern arbeitet, zielt auf Erklärungen, während ein Markt-, Medien- und Konsumpsychologe Prognosen über Kaufentscheidungen vornehmen möchte. – Disziplinäre Schwerpunkte bedeuten aber nicht, dass nicht auch die jeweils anderen Leistungsmerkmale eine Rolle spielen.¹⁵

¹⁵ Herkömmliche Technik- und Ingenieurwissenschaften nutzen beispielsweise Leistungsmerkmale wie (3) und (4), um (1) zu realisieren.

Trotz der pluralen Vielheit disziplinärer Präferenzen liegt eine Einheit wissenschaftlicher Disziplinen empirischer Forschung – von der Physik über die Biologie, die Lebenswissenschaften und Medizin, die Technikwissenschaften und Informatik, bis hin zur Psychologie und den Wirtschafts- und Sozialwissenschaften – darin, dass im Prinzip alle vier Leistungsmerkmale als normative Anforderung an Wissenschaftshandeln anerkannt werden. Die Merkmale sind zudem nicht disjunkt, vielmehr können – je nach Präferenz und Position – Ableitungs- und Folgerungsbeziehungen jeweils ausgewiesen und argumentativ begründet werden, so dass eine gewisse Familienähnlichkeit der Leistungsmerkmale gegeben ist.

4.3 KI und Machine Learning als Katalysator eines Wandels der vier Wissenschaftsverständnisse?

4.3.1 *Prognostizieren ohne Wissen: KI stärkt das prognoseorientierte Wissenschaftsverständnis – ohne Gegenstandswissen*

Im Folgenden soll das erste (oben in Abschn. 4.2 genannte) Wissenschaftsverständnis in den Blick genommen werden, nämlich dasjenige, das das Leistungsmerkmal der Prognose in den Mittelpunkt stellt. Dies passt vortrefflich zu KI- und Machine Learning-Verfahren, insofern der Rekurs auf die Prognoseleistung eine typische Kennzeichnung der neuen informatischen Verfahren darstellt.¹⁶

Die Prognoseorientierung ist terminologisch insbesondere in *Predictive Analytics* und der Entwicklung von algorithmischen (statistischen) Prognosemodellen verankert. Predictive Analytics verwendet Machine Learning-Verfahren, zumeist die des überwachten, teilweise auch des unüberwachten oder des Reinforcement-Lernens, zur Erstellung von algorithmischen Prognosemodellen. Informatischer Kern dieser Verfahren bilden, je nach Aufgabenstellung: Künstliche Neuronale Netze, Genetische Algorithmen, Support Vector Machines, Random Forest, Regressions-, Nächste-Nachbarn-, Gradienten- und Clusteranalyse-Verfahren, allgemein auch Verfahren des Data Mining und der nichtlinearen Zeitreihenanalyse. Eingangsdaten sind die Ereignisse der Vergangenheit sowie jeweils spezifische Rand- und Anfangsbedingungen.¹⁷ Hieraus werden Muster, d. h. Regelmäßigkeiten, extrahiert und durch Kenngrößen beschrieben, welche sich auf interessierende Eigenschaften und Ordnungsstrukturen des Objektsystems beziehen.

¹⁶ Ein zweites, zentrales Leistungsmerkmal sind Klassifikationen und Identifikationen von Mustern bzw. allgemeiner: von Strukturen und Signaturen in Daten (s. u.). Damit verbunden sind im allgemeinen Sinne Kenngrößen. Vielfach werden diese zur Prognose verwendet. Prognostiziert werden beispielsweise das Eintreten von Krankheiten oder Mortalitätsraten, Umsätze, Ankunftszeiten von Flugzeugen, Wetter etc.

¹⁷ Letztere spezifizieren das betrachtete System oder den Gegenstand, sie werden auch als (System-)Parameter bezeichnet.

So werden Wahrscheinlichkeiten des Eintretens eines bestimmten Ereignisses oder eines Ensembles von Ereignissen prognostiziert, d. h. ein Muster, eine Struktur bzw. Signatur. Musterprognosen sind nicht auf bildlich-grafische Prognosen beschränkt, sondern sie sind umfassender: Muster von Daten- oder Zeitreihen, also Ordnungsstrukturen, wobei diese Prognosen auf Mustererkennung und auf Klassifikationen von Regelmäßigkeiten aufbauen, die entweder vorgeben oder während des Lernverfahrens selbst gebildet werden. Dabei handelt es sich um statistische bzw. probabilistische Prognosen, die auf Basis einer Datenanalyse vorgenommen werden, die sich Korrelation zunutze machen. Es sind also keine deterministischen oder modellbasiert-stochastischen Prognosen, die theoretischer Fundamente und Kenntnisse über zugrundeliegende Gesetze bedürften.

In der öffentlichen sowie der sozialwissenschaftlichen Diskussion wird zumeist allein auf Korrelationen geschaut. Das ist zwar nicht falsch, aber einseitig. Es wird übersehen, dass um valide Aussagen über Zukünftiges treffen zu können, Korrelationen nicht hinreichend sind. Zufällige Korrelationen, wie etwa die Korrelation der Storchenpopulation und der Geburtenrate von Kindern in den 50er-Jahren in Deutschland, helfen zu prognostischen Zwecken nicht. Vielmehr müssen im Hintergrund deterministische oder zumindest probabilistische Kausalgesetze vorliegen; es muss eine gewisse Kausalität gegeben sein. Die Pointe der KI- und Machine-Learning-basierten Prognoseverfahren ist, dass sie wetten, dass Kausalität – wenn auch in einem schwachen Sinne – vorliegt und dass man diese kausalen Regelmäßigkeiten nutzen kann, ohne sie explizit kennen zu müssen.¹⁸

Derartige datengetriebenen theoriefreien Prognosen¹⁹ stehen im Fokus des programmatischen Zugangs des Informatikers Leo Breiman, der eine Differenzierung von zwei Zugängen in Informatik und Statistik vorgeschlagen hat, die er als „two cultures in the use of statistical modeling to reach conclusions from data“ bezeichnet (Breiman 2001, S. 199). Der traditionellen „data modeling culture“ stehe die „algorithmic modeling culture“ gegenüber, wie sie für Machine Learning-Verfahren charakteristisch ist. Mit der traditionellen Wissenschaftskultur der „Datenmodellierung“

¹⁸ Diese datengetriebenen theoriefreien Prognosen weisen eine gewisse Nähe zur induktivistischen Sichtweise auf, nach welcher aus vergangenen Ereignissen auf zukünftige geschlossen wird. Allerdings ist das nicht der *volle* Induktivismus, der darauf abzielt, aus einer endlichen Anzahl von Messdaten auf *alle* zukünftigen zu schließen und damit allgemeine Gesetze zu gewinnen. Auch wenn induktive Schlüsse seit dem 18. Jahrhunderten und den Zeiten des Philosophen David Hume als logisch problematisch gelten und man bekanntlich von einem Induktionsproblem spricht, sind sie in den empirischen Wissenschaften weit verbreitet. In logischer (deduktiver) Hinsicht ist dieser Schluss als unzureichend zu bezeichnen, dennoch findet sich dieses Schlussverfahren in den Wissenschaftsdisziplinen vielfach: es dient traditionell der Erkenntniserweiterung, etwa auch zur Hypothesengenerierung.

¹⁹ Es geht hier im Kern um theoriefreie probabilistische Prognosen. Dies sind so genannte Ensemble-Prognosen zu verstehen. Dieser Begriff ist in der Meteorologie hinsichtlich der numerischen Wettervorhersage wohletabliert. Allgemein gehören probabilistische Prognosen zu den so genannten probabilistischen Klassifikationen, welche ermöglichen, auf einer Wahrscheinlichkeitsgröße als Eingabe Aussagen über Klassen von Wahrscheinlichkeiten zukünftiger Ereignisse vorzunehmen, d. h. von Ensembles. Man spricht auch von Ensemble-Prognose, welche mit so genannten Monte-Carlo-Verfahren verbunden sind.

ist bei Breiman gemeint, dass Statistiker üblicherweise voraussetzen, dass der Typ von Daten, den sie vorfinden, von einer Art „deterministic or stochastic data model“ sei. Damit unterstellen sie, dass die Daten durch Prozesse erzeugt wurden, die im Prinzip aus linearen oder logistischen Funktionen aufgebaut sind, wenn auch aus infiniten. Sie schlussfolgern, dass die Daten durch *ein* entsprechendes, klug gewähltes mathematisches Modell nachgebildet werden können. Dieses gefittete Modell verwenden sie für Anwendungszwecke – doch dessen prognostische Leistungsfähigkeit ist vielfach gering, auch wenn das Modell Kriterien wie die der Repräsentation, d. h. der Nachbildung der Daten der Vergangenheit exzellent erfülle.²⁰

Breiman argumentiert für einen anderen Weg. Denn nicht Repräsentation, sondern Prognostizierbarkeit stellt für ihn das zentrale Leistungsmerkmal von Wissenschaft dar: Es gelte, „Daten zu nutzen, um zu prognostizieren“ (Breiman 2001, S. 214). Doch die traditionelle Wissenschaftskultur der repräsentierenden „Datenmodellierung“ habe dieses Ziel nicht hinreichend im Blick. Ein zu dieser Kultur komplementärer Zugang ist die „algorithmische Modellierungskultur“, die Breiman favorisiert. Sie basiert auf „Machine Learning“, so sagt Breiman (2001, S. 200) explizit.²¹ Sie mache keine einschränkende Voraussetzung und behandle den Prozess der Datenerzeugung als unbekannt oder als Black Box. Sie zielen nicht auf repräsentierende Nachbildung der Daten im Modell oder in einer Theorie, sondern auf Prognosen, wofür ein Algorithmus trainiert wird. Anders als das Daten-Modell, was gefittet wird, stellt ein Algorithmus einen rückgekoppelt-iterativen Prozess dar, um gegebene Daten iterativ darzustellen und diesen für Prognosen zu trainieren. Dieser Zugang ist, was die Repräsentation des Gegenstandsfelds oder, allgemeiner, der Objektseite angeht, als theoriefrei zu bezeichnen – und damit voraussetzungsarm. Vor dem Hintergrund von Breimans Gegenüberstellung identifiziert Wolfgang Pietsch (2013) einen „Paradigmenwechsel in der Statistik“ (vgl. Pietsch 2021). Ähnliches hat Thomas Lengauer (2019) herausgestellt, wobei er sich spezieller auf das Machine Learning-Verfahren der Support-Vector-Machine bezieht. Mit Breiman, Pietsch und Lengauer kann man sagen, dass sich nicht nur die Rechner-, Sensor-, Speicher- und Netztechnologie in den letzten Dekaden verbessert hat, sondern es finden sich Veränderungen im mathematischen Kern der (theoretischen) Informatik und der verwendeten Statistik – sowie der damit verbundenen Wissenschaftskultur.

So ermöglichen KI- und Machine Learning-Verfahren probabilistische Prognosen, ohne dass Theorien vorliegen und Modelle gefittet werden, die das Objektsystem repräsentierend beschreiben. Wenn Auguste Comte im 19. Jahrhundert forderte: Wissen, um vorausszusehen, so kann man heute sagen, es scheint gar nicht mehr notwendig zu sein, Wissen zu erwerben. Prognosen sind auch ohne

²⁰ Gelegentlich wird hier von „Overfitting“ gesprochen. Dieser Begriff ist etwas irreführend. Im Kern geht es aber darum, dass die Daten in *einem* Modell, d. h. durch *eine* mathematische Funktion bzw. *eine* Formel, freilich eine sehr komplizierte (im Prinzip mit infiniten Termen), nachgebildet werden.

²¹ Heutzutage würde man auch auf KI allgemein verweisen.

Gegenstandswissen (theoriefrei) möglich, Daten sind hinreichend. Natürlich sind dafür andere Wissenstypen entscheidend, nämlich mathematisch-statistisches Wissen, Wissen über Datenanalyse, Algorithmen und Numerik sowie über rechner-technische Implementierungen. Traditionell hatte Wissenschaft ihre prognostische Kraft aus geprüftem und bewährtem (theoriebasiertem) Gegenstandswissen gewonnen, wie vorne beschrieben. Wenn sich das Gegenstandswissen gar in deterministischen Gesetzen als Kern von kausalen Theorien ausdrücken ließ, dann konnten deterministische Prognosen vorgenommen werden. Diese galten, und gelten bis heute, als besonders präzise. Ein Höhepunkt dieser Sichtweise war der bereits erwähnte Laplace'sche Dämon: Wenn wir die Gesetze sowie die Anfangs- und Randbedingungen über ein uns interessierendes Gegenstandsfeld kennen würden, dann könnten wir die Zukunft dieses Gegenstandsfelds eineindeutig vorausberechnen; in abgeschwächter Form kann dies auch für probabilistische Prognosen gesagt werden.²²

Trotzdem ist zu vermuten, dass auch zukünftig Theorien und mathematische Gesetzmäßigkeiten der „Goldstandard“ für die gute Prognoseleistung von Wissenschaft bleiben werden. Allerdings finden sich angesichts von Machine Learning-Verfahren Möglichkeiten, einen „Silberstandard“ weitreichend zu entwickeln und zu nutzen, um auch jenseits von Theorien und der Kenntnis von Gesetzen prognostische Kraft entfalten zu können, etwa wenn es um komplexe Objektsysteme geht.²³ Ob dieser neue Silberstandard bisherige Wissenschaft ergänzt oder, weitergehend, ersetzt, ist noch nicht entschieden. Zusammengefasst kann man sagen, dass ein Wissenschaftsverständnis, das das Leistungsmerkmal der Prognostizierbarkeit in den Mittelpunkt stellt, durch die aktuelle informatische Entwicklung eine Stärkung erfährt. Geht man weitergehend davon aus, dass Handlungen und Planungen auf Prognoseoptionen zur Antizipation von Zukunft basieren, dann werden gleichzeitig die Handlungsmöglichkeiten vergrößert.²⁴

²² Nun ist angesichts von Machine Learning-Verfahren die Konditional-Bedingung, d. h. das Gegenstandswissen für Prognosen entbehrlich. Genaugenommen muss man aber sagen, dass die Konditional-Bedingung ersetzt werden durch die Annahme einer hinreichenden Datenqualität sowie einer gewissen Stabilität des Gegenstandsfeldes: die Daten müssen hinreichend aussagekräftig sein und es darf sich zeitlich zwischen der Situation, in denen die Daten aufgenommen wurden und der uns interessierenden Zukunft keine exzeptionellen Ereignisse, Bursts oder Katastrophen, also keine strukturelle Systemveränderung, eingetreten sein. Ansonsten hätten wir es mit einem anderen Gegenstandsfeld zu tun, und die prognostische Kraft wäre deutlich reduziert. Diese Situation gleicht einem Erdbeben oder Tsunami, was das betrachtete System verändert. Ferner, siehe allgemein zu den Voraussetzungen der Anwendungen Kap. 2.

²³ Dieses Argument wird in Abschn. 4.4 ausgeführt.

²⁴ Allerdings ist damit auch eine entscheidende Veränderung gegenüber bisherigen Prognoseverständnissen markiert: Prognosen ohne Theorie sind anders zu verstehen als jene, die auf Basis einer grundlegenden Theorie aufsetzen.

4.3.2 Handlungsermöglichung ohne Experiment: KI folgt keinem experimentbasiertem Wissenschaftsverständnis – verstärkt dennoch Handlungsoptionen

Seit Beginn der modernen Wissenschaft und Francis Bacons Programmatik steht das systematisch durchgeführte Experiment im Zentrum wissenschaftlicher Wissensproduktion. Es dient der Erzeugung von neuen, mitunter unerwarteten Phänomenen sowie der kritisch-deduktiven Überprüfung von Aussagen oder Theorien. Kurzum, Erkenntnishandeln ohne Experimentierhandeln ist nicht möglich. Diese These über den engen Zusammenhang von Erkenntnisgewinn und Experimentieren bleibt auch angesichts der Verbreitung von KI- und Machine Learning-Verfahren prinzipiell bestehen.

Doch was ist, so mag man fragen, wenn Erkenntnisgewinn gar nicht mehr primäre Rahmenorientierung oder vorrangiges Ziel eines Forschungsfeldes oder einer Wissenschaftsdisziplin ist? – Dann, so kann gesagt werden, entfällt im Prinzip die Notwendigkeit des Experimentierens. Für Teile der aktuellen und zukünftigen Wissenschaften, gerade für anwendungsorientierte Disziplinen wie Technik- und Ingenieurwissenschaften oder einige Lebenswissenschaften und die Medizin, reduziert sich durch KI- und Machine-Learning-Verfahren der Bedarf, Experimente durchzuführen. Diese Wissenschaftsdisziplinen zielen in erster Linie darauf ab, Klassifikationen vorzunehmen und Prognosekraft zur Verfügung zu stellen, um durch Diagnose von Gegenwärtigem und durch Antizipation von Zukünftigem bessere Planungen, Entscheidungen und Handlungen zu ermöglichen. Beobachten – und die auf Beobachtung basierende Messung – ist oftmals hinreichend; hier setzen KI- und Machine Learning-Verfahren an.

Experimente gelten demgegenüber als aufwendig, als ressourcen-, zeit- und kostenintensiv – oder, je nach Forschungskontext, als schwer realisierbar oder als unmöglich. Unmöglich sind Experimente gelegentlich in der Humanmedizin, etwa aus ethischen Gründen. Schwer realisierbar sind sie bei bestimmten Objektsystemen, insofern eine Kontrolle von Anfangs- und Randbedingungen, insbesondere eine Isolierung des Experimentalsystems von der Umwelt nicht möglich ist. Das kann an der schieren Größe liegen, wie in der Astrophysik oder der Kosmologie, man denke an den Kosmos. Aber es kann auch von den spezifischen Eigenschaften des Objektsystems herrühren, die mit der Schnittstellen-Problematik in Verbindung stehen: Bei dieser weiß man nicht, was zum System und was zu seiner Umwelt gehört. Damit ist unklar, was Störungen einerseits und was Systemdynamiken andererseits sind.²⁵ Zu diesen Objektsystemen gehören Klima- und Wettersysteme,²⁶ aber auch Ökosysteme sowie biologische und medizinische Systeme. Dazu gesellen sich komplexe technische Systeme, von komplexen Produktionsprozessen über Kraftwerke bis hin zu konkreten Bauteilen (wie z. B. Flugzeugflügel). Berücksichtigt man zudem gesellschaftliche und ökonomische Ver-

²⁵ Janich (1997) spricht von „Störungsbeseitigungswissen“.

²⁶ Siehe auch Kap. 2 (Gesamtbuch) sowie hier in Abschn. 4.4 zur Klimaforschung.

wendungskontexte, wird ein systematisches Experimentieren weiter erschwert. Alle soeben genannten Systeme sind gekennzeichnet durch Komplexität, Nichtlinearität und mitunter Instabilität/Sensitivität.²⁷ Mit ihnen ist ein Experimentierhandeln erschwert und eine Reproduktion der Ereignisse ist oft unmöglich. An diesen Systemen wird aufgrund dieser multiplen Problematik primär gemessen und beobachtet.

Damit ist nicht das Ende des Experiments in den Wissenschaften gemeint. Denn überall dort, wo Experimente möglich und hilfreich sind, werden weiter welche durchgeführt. Hilfreich ist eine auf Experimenten basierende systematische Datenerzeugung etwa auch dort, wo diese als Rohdaten Eingang finden in KI- und Machine Learning-Verfahren, beispielweise für Trainings im Rahmen des überwachten Lernens oder in Simulationen. Daten, die aus einem Experiment stammen, sind gut kontrollierbar, ihre Qualität ist zumeist besser einschätzbar und sie eignen sich als Input für diese Verfahren besser als nicht-experimentelle, d. h. reine Beobachtungs-Datenreihen. – Neben diesem Experimentieren zur Datengenerierung gibt es eine zweite, komplementäre Verbindungsrichtung zwischen Experimenten und den neuen informatischen Verfahren: Letztere dienen oftmals ihrerseits als Grundlage für neue Experimente. Teilchenphysiker trennen durch Machine Learning-Verfahren Signal- und Rauschanteil in Datenreihen und erhalten Hinweise auf besondere Ereignisse („bursts“). Diese Hinweise können zu Hypothesen verdichtet werden, was eine Grundlage für neue Experimente auf der Suche nach einer Physik jenseits des Standardmodells darstellt.²⁸ Gleiches gilt für KI-basierte Analysen von Gensequenzen, die dann Detailuntersuchungen und molekularbiologische Experimente nahe legen können, etwa im Rahmen von CRISPR/Cas-Methoden. Somit wird die Interventionsleistung durch datengetriebene Verfahren erhöht, ein Ziel, das auch Bacon verfolgte.

So kann man sagen: Bezügliches des Bacon'schen Programms findet sich angesichts aktueller KI- und Machine Learning-Verfahren sowohl ein Wandel als auch eine Kontinuität: *Einerseits* ein Wandel, insofern das Experiment in datengetriebener Forschung nicht mehr als zentrales Mittel der Forschung gilt; *andererseits* eine Kontinuität, ja vielleicht Verstärkung, insofern durch KI- und Machine Learning-Verfahren Wissenschaft die Interventionskräfte, Experimentieroptionen und Handlungsmöglichkeiten des Forschers vergrößern, und Wissenschaft verstärkt einer Nutzens- und Verwertungsorientierung folgen kann. Damit wird das Bacon'sche Programm sowohl beendet – hinsichtlich des Experimentfokus als *das* methodische Zentrum – als auch bestärkt – bezüglich der Handlungsermöglichung und der Nutzenorientierung.

Neben die Beendigung *und* die Bestärkung des Bacon'schen Programms findet sich ein *dritter* Aspekt, die Ersetzung. Man kann fragen, was an die Stelle des Experimentierhandelns tritt. In gewissem Rahmen sind dies Computerexperimente. Doch man kann noch einen anderen Aspekt hervorheben. Was bei Bacon kaum eine Rolle spielte und auch in aktuellen wissenschaftsphilosophischen Zugängen zu Wissenschaft eher randständig ist, wird von KI- und Machine-Learning-Verfahren

²⁷ Siehe auch Abschn. 4.4–4.6 sowie Schmidt (2015).

²⁸ Siehe auch Kap. 2.

hervorgehoben: Wahrnehmungs- und Beobachtungsprozesse. Konstitutiv für Wissenschaft allgemein ist nicht allein das Experimentieren, sondern auch Wahrnehmen und Beobachten, d. h. nicht allein Intervention, sondern auch Kontemplation. Dabei meint Wahrnehmen stets ein unterscheiden: Etwas wird *als etwas Spezifisches* wahrgenommen, ansonsten bliebe es eine amorphe Masse oder eine beliebige („verrauschte“) Datenreihe. Mit anderen Worten: Muster, Ordnungsstrukturen und Regelmäßigkeiten werden identifiziert und klassifiziert, um ein Phänomen als solches wahrzunehmen. In einem Pixelbild wird ein Muster identifiziert, welches die Struktur einer Katze oder eines Karzinoms hat.

So liegt in datengetriebenen Verfahren eine gewisse phänomenologische Grundorientierung vor, was bemerkenswert ist. Denn die meisten Verständnisweisen von moderner Wissenschaft reflektieren Wahrnehmen und Beobachten nicht eigens: Moderne Erkenntnistheorien umfassen nur äußerst selten Wahrnehmungstheorien.²⁹ Dabei hatte Kant Wahrnehmungsprozesse im Blick, als er den Begriff der „Einbildungskraft“ entwickelte, nämlich jene „blinde, obgleich unentbehrliche Funktion der Seele“, die konstitutiv ist für Erkenntnis (Kant 1989 A 78/B 103). Zwar sind KI- und Machine-Learning-Verfahren nicht als eine „Funktion der Seele“ anzusehen; doch mit ihrer „Pattern Recognition“ knüpfen sie an jene erkenntnistheoretischen Traditionen an, die heutzutage randständig geworden sind. Sie erweitern das Wissenschaftsverständnis, indem sie Muster-Wahrnehmungen in den methodischen Mittelpunkt stellen.

4.3.3 Prüfen auf komplementären Wegen: KI relativiert das Ideal des testbasierten Wissenschaftsverständnisses – ermöglicht aber neue phänomenologische Prüfformen

Wissenschaft stellt intersubjektive Nachvollziehbarkeit ihrer Aussagen her durch personen-, zeit- und ortsinvariantes Prüfen von Theorien, Modellen oder Gesetzen an empirischen Daten. Unabhängig davon, ob man annimmt, dass diese Prüfung falsifizierend/verwerfend oder verifizierend/bewährend zu verstehen ist, entscheidend ist, dass theoriebezogene Modelle und empirische Daten zusammengeführt werden, um kriteriengeleitet zu begründeten und nachvollziehbaren Urteilen über Annahme oder Verwerfen von Aussagen zu gelangen.

Nun liegt, wie ausgeführt, das Ziel datengetriebener Forschung nicht in Entwicklung und Prüfung von Theorien oder Modellen. Diese können zwar Mittel zum Ziel von Prognosen als Grundlage von Handlungen, mithin von Interventionen, sein. Aber wenn dieses Ziel auch aufwands-, zeit- und ressourcenarm, d. h. theoriefrei möglich ist, wird man in den angewandten Wissenschaften und in der nutzenbezogenen Anwendungspraxis darauf verzichten. Angesichts der Entwicklungen

²⁹ Dies überlassen heutige Erkenntnistheoretiker gerne der Psychologie und Neuropsychologie. Dabei hatten bspw. positivistische und später neopositivistische Traditionslinien (zunächst) eine sensualistische Sinnesdatentheorie entworfen und verfolgt.

von KI und Machine Learning hat Chris Anderson (2008) vor über 10 Jahren ein „Ende der Theorie“ diagnostiziert und gegenüber der bisherigen Wissenschaft behauptet: „There is now a better way. [...] We can stop looking for models [, theories, interpretable laws or hypothesis]. We can analyze the data without hypotheses about what it might show. We can [...] let statistical algorithms find patterns where science cannot.“ Nach Anderson sind Hypothesen und, weitgehend, Theorien und Modelle in zukünftiger Wissenschaft schlicht überflüssig. Damit ist ein Prüfen derselben nicht mehr möglich – denn man hat nichts, was überprüft werden könnte. Insofern sich Wissenschaft in ihrem Selbstverständnis stets als Theorie-prüfend verstand und sich hieran sowohl falsifikationistische wie auch verifikationistische Wissenschaftskonzepte angeschlossen haben, muss man sagen, dass in datengetriebener Forschung auch Falsifikationismus und Verifikationismus verschwinden.

Allerdings gilt diese Diagnose nur hinsichtlich stark anwendungsorientierter Wissenschaftsdisziplinen. In anderen Wissenschaftsdisziplinen bleibt das Prüfen, basierend auf einer Theorie- und Erkenntnisorientierung, bestehen – auch wenn diese Disziplinen mitunter die neuen informatischen Verfahren als Mittel, Methoden und Instrumente verwenden: KI- und Machine Learning-Verfahren spielen in der Genese von Erkenntnis eine unterstützende, funktionale Rolle, etwa wenn man es mit komplexen Objektsystemen zu tun hat; sie ersetzen allerdings Theorien und theoriebezogene Rechtfertigung und Geltungsausweis wissenschaftlicher Aussagen nicht.³⁰ Das gilt etwa für prominente Forschungsfelder wie die Hochenergie- und Teilchenphysik und die Kosmologie sowie die Klimaforschung. KI- und Machine Learning-Verfahren tragen dazu bei, (1) Regelmäßigkeiten, d. h. Muster, Strukturen und Objekte aus gemessenen oder Simulations-basierten Datenreihen zu identifizieren und zu (re-)konstruieren sowie, hierauf aufbauend, (2) Modelle, Aussagen, Hypothesen zu entwickeln und auch numerische Simulationen durch Aufbereitung von Daten zu ermöglichen.

So hat die Verwendung von KI- und Machine-Learning-Verfahren in der Astrophysik dazu geführt, Existenz sowie Eigenschaften von 50 neuen Exoplaneten durch Analyse von älteren Datenreihen festzustellen. In diesen Daten wurden Regelmäßigkeiten gefunden, nachdem durch KI-Verfahren eine Signal-Rauschen-Diskriminierung vorgenommen wurde, also gehaltvolle Informationen einerseits und Hintergrundrauschen andererseits getrennt werden konnten. Ein solcher Existenzhinweis von bisher nicht wahrgenommenen Objekten wird in der Astrophysik und Kosmologie als „Validierung“ bezeichnet (ad 1). In der Klimaforschung tragen KI- und Machine-Learning-Verfahren dazu bei, die Anzahl der Zustandsgrößen, d. h. der Zustandsraumdimensionen aus Zeitreihen zu bestimmen, um auf dieser Basis Zustandsraummodelle zu entwickeln, die lokale, raumzeitliche Simulation von Klima- oder Wettermodellen ermöglichen (ad 2).³¹

³⁰ Man kann dann sagen, wie es der Physiker und Philosoph Hans Reichenbach vor 90 Jahren ausführt, dass sich zwar der Entdeckungszusammenhang („context of discovery“) verändern mag, nicht aber der Rechtfertigungszusammenhang („context of justification“).

³¹ Siehe auf Kap. 2 sowie hier Abschn. 4.4.

Ferner findet sich auf Basis von KI- und Machine-Learning-Verfahren (3) gar eine ergänzende, alternative Form des Prüfens, die in der modernen Wissenschaft eher randständig war. Es handelt sich um eine Prüfform von Modellen oder Theorien durch phänomenologischen Mustervergleich, d. h. durch Kenngrößenvergleich mit empirischen Messdaten. Eine solche Prüfform wird bei komplexen, nicht-linearen, mitunter instabilen Objektsystemen notwendig. Denn traditionelle Prüfformen, welche auf klassischer Fehlerabschätzung und ihrer mathematisch-statistischen Konvergenztheoreme aufbauen, greifen nicht. Einerseits muss das zu prüfende Modell überhaupt erst durch numerische Verfahren, etwa durch Simulationen, zugänglich gemacht werden, um seine raumzeitlichen Phänomene zu zeigen; denn die mathematischen Gesetzmäßigkeiten des Modells sagen als solche noch gar nichts. Die durch Simulation offengelegten raumzeitlichen Phänomene bilden dann die Grundlage, um Klassifikationen vorzunehmen und Muster zu identifizieren, genauer gesagt: um Kenngrößen zu bilden. *Andererseits* werden analog auf Seiten des Objektsystems Messdaten herangezogen und auf dieser Basis ebenfalls Muster identifiziert und Kenngrößen entwickelt.³² Mit anderen Worten: Durch KI- und Machine Learning-Verfahren können sowohl aus dem modellbasierten Simulationssystem wie aus dem realen Objektsystem Muster und Ordnungsstrukturen extrahiert werden, die durch Kenngrößen darstellbar werden. Diese Kenngrößen dienen der Vergleichsoption von Modell- und Objekt-Verhalten – und damit der Prüfung. Man spricht hier von einer phänomenologischen Muster- oder Kenngrößen-Prüfung.³³

Zusammengenommen fällt die Einschätzung hinsichtlich der Veränderung des prüfbasierten Wissenschaftsverständnisses zweigeteilt aus. *Einerseits* ist das zentrale Ziel datengetriebener Forschung gerade nicht die Entwicklung und anschließende Prüfung von Theorien, Modellen oder Gesetzen – die Notwendigkeit des Prüfens entfällt. *Andererseits* spielen KI- und Machine Learning-Verfahren als neue Instrumente und erweiterte Mittel zur Prüfung eine zunehmend bedeutendere Rolle im Feld komplexer, dynamischer, nichtlinearer Objektsysteme – von der Physik und Klimawissenschaft über die Lebenswissenschaften, Biologie und Medizin bis hin zu den Technik- und Ingenieurwissenschaften. Diese neueren Prüfinstrumente stellen Muster, Ordnungsstrukturen und spezifische (Komplexitäts-)Kenngrößen in den Mittelpunkt: Sie sind damit i. w. S. phänomenologisch ausgerichtet.

³² Das ist etwa notwendig bei komplexen und nichtlinearen Objektsystemen wie dem Klimasystem oder metabolischen Systemen in Biologie, Ökologie oder Medizin.

³³ Noch in einer anderen Hinsicht verändert sich durch KI- und Machine Learning-Verfahren nicht nur die Wissensgenese. In einigen Feldern der Mathematik und theoretischen Physik haben sich KI und Machine Learning als Beweisverfahren etabliert. Ein Beispiel ist der „Beweis“ des Vier-Farben-Theorem. Ein analytischer Beweis ist bislang nicht gefunden worden, aber ein numerisch-algorithmischer. Nur durch Verwendung eines Computers war es möglich, die notwendige Vielfalt der Varianten durchzuspielen. Nimmt man dieses Beispiel als paradigmatisch, muss man sagen, dass sich die Rolle des wissenschaftlichen Wissensarbeiters und Forschers zu verändern beginnt.

4.3.4 Kausalität ohne Theorie: KI folgt keinem erklärungsbezogenem Wissenschaftsverständnis – verbleibt dennoch im Horizont von Kausalität

Traditionell zielt Wissenschaft darauf ab, Phänomene durch Angabe von fundamentalen Theorien, Modellen oder Gesetzen zu erklären. Damit weist Wissenschaft, was Erklärungen angeht, eine deduktive Ableitungsstruktur auf. Gesetze werden als deterministische oder probabilistische (Kausal-)Gesetze verstanden.

Nun ist es nicht vorrangiges Ziel angewandter Forschung, etwa der Technikwissenschaften oder der Medizin, Erklärungen zu liefern, sondern Handlungen zu ermöglichen, wobei Erklärungen ein Mittel hierfür sein können. KI- und Machine Learning-Verfahren weisen damit eine zu den Zielen angewandter Wissenschaften strukturelle Verwandtschaft auf, wie ausgeführt. Vordergründig scheinen hierfür Korrelationen hinreichend zu sein, während Kausalität nicht als notwendig erachtet wird, wie Chris Anderson (2008) unter dem Titel „The End of Theory: The Data Deluge Makes the Scientific Method Obsolete“ behauptet: „Correlation is enough. [...] Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.“ Einige gehen weiter, wie David Chandler (2015, S. 837 f.), und sehen in kritischer Absicht eine „Welt ohne Kausalität“ heraufziehen, d. h. eine Wissenschaft und Gesellschaft, in der aufgrund des allgemeinen Datenfokus’ („Datafication“) niemand Interesse an der Entdeckung kausaler Zusammenhänge und der Entwicklung kausaler Erklärungen hat: „The promise is that, with high levels of data generation and developments in computational analysis, the world (coded through datafication) can begin to speak for itself without its (more than) fallible human interpreter.“

Doch mit dieser pointierten Zuspitzung von Chandler, Andersen und anderen ist Vorsicht geboten.³⁴ Zwar bleibt richtig, dass Erklärungen nicht im Zentrum stehen. Allerdings kann damit nicht gemeint sein, dass Kausalität keine Rolle mehr spielt, im Gegenteil. Man kann sogar sagen, dass KI- und Machine Learning-Verfahren darauf abzielen, deterministische oder probabilistische Kausalgesetze nutzbar zu machen, allerdings ohne diese offenlegen und mathematisch beschreiben zu müssen oder zu können. Die Wissenschaftsphilosophin Nancy Cartwright (1983, S. 22) stellt (in anderem Zusammenhang) heraus: „[C]ausal laws cannot be done away with, for they are needed to ground the distinction between effective strategies and ineffective ones.“ Diese Position entspricht einem handlungstheoretischen oder interventionalistischen Kausalitätskonzept. Demnach können wir uns der Kausalität „genauso sicher sein wie unseren Fähigkeiten, etwas zu tun.“ Mit anderen Worten: „p ist eine Ursache von q bedeutet, dass ich q herbeiführen könnte, wenn ich p tun könnte.“ (Wright 1991, S. 76/75) Für das Herbeiführen oder Erzeugen eines Phäno-

³⁴ Eine weitreichende Kritik findet sich bei Gary Smith (2018) in seinem Werk „The AI Delusion“: „Far too many intelligent and well-meaning people believe that number-crunching is enough. We do not need to understand the world. We do not need theories. It is enough to find patterns in data. Computers are really good at that, so we should turn our decision-making over to computers.“

mens sind Korrelationen allein nicht genug; sie sind zu schwach, um Handlungen zu ermöglichen und Interventionen erfolgreich durchführen zu können. Denn Korrelationen können zufällig sein, wie die Storchpopulation und die Geburtenrate von Kindern in einigen Regionen in den 1950er-Jahren zeigt oder wie der Konsum von Schokolade und die Anzahl der Nobelpreisträger in einem Land. Vergrößert oder verringert man die Population der Störche, hat dies keine Wirkung auf die Geburtenrate von Kindern; gleiches hinsichtlich des Schokoladenkonsums und den Nobelpreisträgern.³⁵

Machine Learning-Verfahren zielen demgegenüber – insofern sie auf effektive Handlungsstrategien aus sind – darauf ab, jene Korrelationen zu erfassen, die nicht zufällig sind, sondern wo zumindest eine schwache Form der Gesetzmäßigkeit, also der Kausalität, zugrunde liegt. Auch wenn es im strengen Sinne für Kausalitätsdiagnosen keine logisch zwingenden Argumente geben kann,³⁶ tragen KI- und Machine Learning-Verfahren dazu bei, Wahrscheinlichkeitsaussagen auf dem Kontinuum zwischen reinem Zufall einerseits und Kausalgesetz andererseits zu ermöglichen. Damit kann man sagen, Korrelation ist der umfassendere, Kausalität der speziellere Begriff: Aus Kausalität folgt Korrelation, die Umkehrung gilt nicht.

Insofern KI und Machine Learning, neben anderem, als Verfahren zur Erkennung von Regelmäßigkeiten, d. h. zur Muster-, Signatur- oder Ordnungserkennung in Daten und damit zur Identifikation schwach-kausaler Abhängigkeiten gelten können, können sie auch als Mittel angesehen werden, um Mechanismen zu erkennen und mechanistische Erklärungen vorzubereiten.³⁷ Dieser Erklärungstyp kennzeichnet ein Phänomen als erklärt, wenn man die erzeugenden Ursachen angeben kann, die das Phänomen hervorbringen (siehe Abschn. 4.2).³⁸ Natürlich ist eine kontroverse Detaildiskussion über notwendige und hinreichende Bedingungen, über

³⁵ Das heißt interveniert man und vergrößert oder verkleinert man die Storchpopulation, ohne dass sich die Geburtenrate verändert, so liegt kein kausaler Zusammenhang vor (Falsifikation). – Es sollte noch eine weitere Problematik der Kausalitätsthematik zur Sprache kommen, die in der wissenschaftlichen Praxis eine Herausforderung darstellt – und auf die der eben genannte Interventionalismus zu reagieren versucht. Die Problematik wird mitunter unter dem Stichwort der „Drittvariable“ oder „kontingentes Zuschreibungsproblem von Ursache und Wirkung“ thematisiert. In der wissenschaftlichen Praxis findet man oftmals Korrelationen zwischen zwei Datenreihen. So korrelieren beispielsweise sexuelle Aktivität und Lebenserwartung (wie in Studien gezeigt, aber insbesondere oft in der Regenbogenpresse dargestellt). Daraus wird vielfach vorschnell (a) die Zuschreibung einer kausalen Abhängigkeit und, weitergehend, (b) eine (zeitliche) Sukzession, d. h. eine zweifelhafte Identifikation von Ursache und Wirkung im Sinne eines einseitig-sukzessiven Ursache-Wirkungs-Verhältnisses: Sexuelle Aktivität (Ursache) → höhere Lebenserwartung (Wirkung). Dabei wird übersehen, dass eine so genannte Drittvariable vorliegen könnte, etwa die der Vitalität oder auch allgemein der Gesundheit. Steigert man (durch Intervention) die Vitalität, so führt dies ursächlich sowohl zu einer höheren sexuellen Aktivität als auch zu einer höheren Lebenserwartung.

³⁶ Es gilt selbstverständlich das Hume'sche Induktionsproblem, nach dem ein solcher Schluss im logischen Sinne nicht statthaft ist.

³⁷ Man sieht hier recht schön eine strukturelle Verwandtschaft von handlungstheoretischen und interventionalistischen Kausalitätskonzepten mit mechanistischen Erklärungstypen.

³⁸ Damit ist eine deutliche Nähe zu handlungstheoretischen und interventionistischen Kausalitätsverständnissen gegeben.

Tun- oder Unterlassenshandlungen sowie über den Stellenwert von fundamentalen oder phänomenologischen Gesetzen als Bestandteil dieses Erklärungstyps innerhalb der Wissenschaftsphilosophie entfacht, was hier nicht weiter zu interessieren braucht. Ein Phänomen, etwa der trockene Husten oder das Fieber eines Menschen gilt in der Humanmedizin genau dann als erklärt, wenn man eine erzeugende (oder unterlassende) Ursache dafür angeben kann und dies über große Fallzahlen reproduzierbar ist, beispielsweise das Vorliegen des Virustyps SARS-CoV-2, welches durch einen positiven PCR-Test nachgewiesen wird. Gleiches gilt hinsichtlich der Nachweisproblematik der Wirkung eines potenziellen Impfstoffs. Auch wenn man die Details des komplexen kausalen Ursache-Wirkungs-Mechanismus (noch) nicht verstanden hat, so dass keine fundamentalen, theoriefundierten Erklärungen möglich sind, geht die Medizin davon aus, dass (probabilistische oder deterministische) Kausal-Gesetze im Hintergrund wirken, auch wenn man derzeit weithin nur Korrelationen und Sukzessionen zwischen den beiden Ereignissen – Virustyps SARS-CoV-2 und trockener Husten – empirisch unter Verwendung von Machine-Learning-Verfahren findet. Trotz der kontroversen Diskussion, wie schwach oder stark diese mechanistischen Erklärungen sind, ist die Praxisrelevanz derartiger Mechanismen in den Technikwissenschaften sowie in der Medizin unbestritten.

Zusammengenommen gilt, weder die Identifikation von Regelmäßigkeiten, also die Musteridentifikation und -Klassifikationen noch die sich diese zunutze machen den Prognosen können als hinreichend für jene in den Wissenschaften wohletablierten, theoriebezogenen, starken Erklärungstypen angesehen werden. Sollten zukünftig KI- und Machine Learning-Verfahren erkenntnisorientierte Wissenschaften unterstützen und ergänzen, dann werden weiterhin starke Erklärungen geliefert werden, die das Weltverständnis vertiefen. Sollten sich allerdings die datengetriebenen Verfahren in den Vordergrund schieben, ist damit gleichermaßen das Interesse an starken Erklärungen reduziert.

4.4 Von den Verfahren und Methoden zu den Objekten: Schwache Regelhaftigkeit nutzen – komplexe Objektsysteme werden zugänglich

Um den weitreichenden Erfolg von KI- und Machine-Learning-Verfahren einordnen zu können, sollte man über obige vier methodologische Aspekte hinausgehen – und die Analyse vertiefen. Dazu sind jene Objektsysteme und Gegenstandsfelder in den Blick zu nehmen, an denen die neuen datengetriebenen Verfahren ihre Leistungsfähigkeit zeigen.³⁹ Zu fragen ist, ob diese Objektsysteme auf einer grundlegenden Ebene Gemeinsamkeiten aufweisen.

³⁹ Philosophisch gesprochen sollen im Folgenden über methodologische Untersuchungen hinaus (im weitesten Sinne) „ontologische“ und epistemologische Fragen erörtert werden.

Ausgangspunkt ist die Beobachtung, dass durch KI- und Machine-Learning-Verfahren Wissenschaft einen verbesserten und vertieften Zugang zu jenen Objektsystemen in Natur, Technik, Wirtschaft und Gesellschaft erlangt, die bislang wissenschaftlich schwer zugänglich waren. Damit rücken auch solche Objektsysteme in den Fokus von Wissenschaft, an denen bislang eine Forschung nicht als vielversprechend angesehen wurde.⁴⁰

Ein beispielhaftes Objektsystem, um den Erfolg der neuen informatischen Verfahren zu beleuchten, ist das Klimasystem (Storch 2019; Karpatne et al. 2019). Zwar sind Atmosphärenphysik, Meteorologie, Geophysik und Physikalische Chemie der Atmosphäre wohletablierte Teildisziplinen der exakten Naturwissenschaften und damit grundlegend für die aktuelle Klimaforschung. Doch ohne neuere Entwicklungen in Informatik und Mathematik sowie in der informatischen Technik (Rechner-, Sensor-, Speicherungs- und Netztechnologie) hat die Klimaforschung nur einen beschränkten Zugang zum Klimasystem, auch wenn sie im Prinzip alle zugrunde liegenden Phänomene und allgemeinen Gesetzmäßigkeiten kennt.⁴¹ Das Klimasystem ist aufgrund der hohen Komplexität, d. h.

- der vielen Zustandsgrößen und der Hochdimensionalität des Zustandsraumes,
- der Nichtlinearität, Instabilität und Sensitivität der grundlegenden Prozesse und der weitreichenden raumzeitlichen Kopplungen teils ausgedehnter Systemelemente (schwache Kausalität) (Lorenz 1963),
- der hohen Dynamik und starken Selbstorganisations- (und Emergenz-)Fähigkeit,
- der umfassenden raumzeitlichen Abhängigkeit der Phänomene und jeweiligen Skalierung(seigenschaft)en,
- der reduzierten Isolierbarkeit des Systems von der Umwelt (Separabilitätsproblematik), verbunden mit der Nichtdefinierbarkeit der Systemgrenzen (Schnittstellenproblematik) und der somit im Detail epistemisch schwer beschreibbaren Offenheit des Systems (Energie, Stoff und/oder Information) von seinen Umwelten,
- und folglich einer weitgehenden Nichtkontrollierbarkeit der Rand- (und Anfangs-)Bedingungen (Kontrollproblematik),

schwer zugänglich für moderne Naturwissenschaft. Dazu kommt, dass man mit dem Klimasystem nicht, wie in traditionellen Laborwissenschaften, experimentieren kann. Zusammengenommen limitieren Komplexität, Nichtlinearität und Sensitivität/Instabilität den üblichen methodologischen Zugriff, insbesondere hinsichtlich der Merkmale wie theoriebasierter Berechenbarkeit, experimenteller Reproduzierbarkeit, empirischer Testbarkeit und deduktiver Erklärbarkeit (Schmidt

⁴⁰ Dies ist besonders auffällig im Feld der so genannten *Digital Humanities*, aber auch in vielen Bereichen der Sozial- und Humanwissenschaften (z. B. der Psychologie). Im Folgenden soll aber die Objekterweiterung auch im Feld der Natur- und Lebenswissenschaften beleuchtet werden und dafür ein systematisches Argument vorgelegt werden.

⁴¹ Es handelt sich hier um die Hydrodynamik und Kontinuumsmechanik mit den so genannten Navier-Stokes-Gleichungen (partielle Differenzialgleichungen) sowie den entsprechenden Grundgleichungen der (Statistischen) Thermodynamik.

2008). Spezieller finden sich auch Begrenzungen bezüglich Modellierung und Simulation sowie der Szenario- und Prognose-Entwicklung. Diese Erkenntnisse haben in den letzten Jahrzehnten dazu geführt, den systemischen Charakter des Klimasystems stärker zu berücksichtigen, so dass die aktuelle Klimaforschung eingebunden ist in die so genannte „Earth Systems Science“.

Mit diesen Charakteristika steht das Klimasystem nicht allein da, im Gegenteil. Dass große Klassen von Objektsystemen in Natur, Technik, Wirtschaft und Gesellschaft in ihrem Kern diese Charakteristik aufweisen, legen seit einige Dekaden Komplexitäts- und Selbstorganisationstheorien, im Verbund mit der Theorie nichtlinearer dynamischer Systeme, Synergetik, Theorien dissipativer Strukturbildung und des Hyperzyklus, Chaos-, Katastrophen- und Autopoiesis-Theorien, u. a. nahe (Jackson 1989; Haken 1980; Prigogine 1992; Mainzer 1996; Schmidt 2015). Diese Objektsysteme sind komplex und nichtlinear, oftmals sind sie zudem dynamisch und strukturell instabil bzw. sensitiv, wodurch auf unterschiedlichen Skalenebenen raumzeitlich verteilte Selbstorganisations- und dynamische Stabilisierungsprozesse ermöglicht werden. An instabilen Punkten oder Grenzflächen, den so genannten Tipping Points, können sich Systemdynamiken strukturell verändern. So kann der für den Wärmetransport wichtige Golfstrom abreißen, es können Brücken brechen und einstürzen, Staus vermeintlich aus dem Nichts entstehen, der Tod metabolischer Systeme eintreten, die Börse kollabieren, öffentliche Meinungen kippen, Bürgerprotestbewegungen entstehen, Produkte von Konsumenten plötzlich gemieden werden, etc. Domino-, Schneeball-, Schwarm- und Schmetterlingseffekte treten auf: Ein Flügelschlag eines Schmetterlings in Südamerika kann einen Tornado in Nordamerika erzeugen – ein Phänomen, das als sensitive Abhängigkeit bezeichnet wird: Kleine Ursache, große Wirkung, wobei ähnliche Ursachen ganz unterschiedliche Wirkungen nach sich ziehen können (Lorenz 1989). Diese Systeme sind durch eine schwache Form der Gesetzmäßigkeit oder Regelhaftigkeit bestimmt, die man treffend als *schwache Kausalität* bezeichnet.⁴²

Derartige Charakteristika sind auch für metabolische Systeme der Lebenswissenschaften, der Biologie, Ökologie und Medizin generisch (Mainzer 1996, 2014). Diese Systeme sind komplex, dynamisch, nichtlinear, sie sind im Detail oftmals instabil und im Allgemeinen dadurch besonders dynamisch-adaptiv stabil, sie sind offen und selbstorganisationsfähig, ihnen kann „Autonomie“ zugeschrieben werden. Ihre Systemdynamiken werden von (probabilistischen oder deterministischen) schwachen Kausalgesetzen bestimmt – allerdings hat auch die beste Wissenschaft mit diesem Typ von Gesetzen prinzipielle Probleme, diese aufzufinden und zu rekonstruieren.⁴³ In vielen, vielfach auch anwendungsrelevanten

⁴² Dies gilt unabhängig davon, ob es sich um deterministische oder um probabilistische (Kausal-) Gesetze handelt.

⁴³ Diese epistemischen Probleme sahen schon Newton (Mond-/Planetentheorie), Hume (Erdbeben), Maxwell (allgemeine Maxime der exakten Naturwissenschaften; Zugweichen/Bifurkationen), Duhem (Nutzlosigkeit mathematischer Deduktionen in bestimmten Feldern der Physik), Einstein (Allgemeine Relativitätstheorie/Einstein Kosmos/Kosmologische Konstante), und einige andere (Navier und Stokes, Poincaré, Prantl, u. a.).

Fällen weiß man wenig über Ursache-Wirkungs-Zusammenhänge, man hat es mit einer Black Box zu tun. Für die Medizin liegt hier eine Herausforderung, denn für Diagnose und Therapie vieler Krankheiten ist nichts so zentral wie ein kausales Wissen über Ursache-Wirkungs-Zusammenhänge. Die in der hohen Komplexität wurzelnde Herausforderung wird gerade in der aktuellen Covid-Pandemie deutlich – von der Analyse von Infektionsformen und Übertragungswegen, über die Einschätzung adäquater Schutzmaßnahmen und entsprechender Medikamente, bis hin zur Suche sowie zur Validierung und Zulassung von Impfstoffen. Und auch in den Technikwissenschaften gelten komplexe technische Anlagen, Produktionsanlagen und Produktionsprozesse (z. B. Automobilproduktion) als komplexe Systeme, in denen vielfach auch schwache Kausalität vorherrscht (Moon 1998). Gleiches findet sich in den Human-, Sozial- und Wirtschaftswissenschaften. Die Psychotherapieforschung, die von Wissenschaftsdisziplinen wie Psychologie und medizinischer Psychiatrie betrieben wird, kennt diese Systemcharakteristika; sie verwendet Charakteristika der Theorien nichtlinearer dynamischer Theorien, um den Therapieerfolg zu monitoren und therapeutische Interventionen zu verbessern (Haken und Schiepek 2006). In den Finanzwissenschaften gibt es neuere Zugänge zu fluktuierenden Finanzdaten, etwa zu Kursen von Aktien, Wertpapieren, Derivaten, Hedgefonds, um unter Verwendung von KI- und Machine-Learning-Verfahren den Prognosehorizont zu erweitern. Unter dem Sammelbegriff der „Econophysics“ hat sich in diesem Feld seit über 20 Jahren eine interdisziplinäre Forschungsrichtung von Mathematikern, Informatikern, Physikern, Wirtschafts- und Finanzwissenschaftlern etabliert, um diese komplexen Objektsysteme, die Instabilität/Sensitivität und damit schwache Kausalität aufweisen, zu prognostizieren und, wo möglich, verstehen zu können (McCauley 2004; Mantegna und Stanley (2000).⁴⁴

Vor diesem Hintergrund können KI- und Machine Learning-Verfahren als Ansätze angesehen werden, methodologisch mit der Komplexitätsproblematik umzugehen und einen Zugang zu komplexen Objektsystemen zu erlangen (Pathak et al. 2018a, b; Dubois et al. 2020; Fan et al. 2020). Wesentliches Ziel ist, Formen der Regelmäßigkeit, d. h. verborgene Muster oder Abhängigkeiten in diesen Objektsystemen zu finden, also schwache Kausalität zu diagnostizieren, um dies prognostisch nutzen zu können – auch wenn man die (deterministischen oder probabilistischen) Gesetze prinzipiell nicht explizit kennen kann. Und selbst wenn man die Gesetze kennen würde, könnte man aufgrund der Nichtlinearitäten und Instabilitäten keine weitreichenden (theorie- und gesetzesbasierte) Prognosen vornehmen. Das mag zunächst paradox erscheinen, doch ist es eine Einsicht der Theorien nichtlinearer dynamischer Systeme.

Ein solcher Zugang zu komplexen Objektsystemen, der heute durch KI- und Machine Learning-Verfahren beschritten wird, hat eine eigene Historie, die sich im

⁴⁴ In Kosmologie und Astronomie sind instabile Systemdynamiken üblich. Mittels KI-Verfahren, welche Daten für Simulationen (u. a. Monte-Carlo) bereitgestellt haben, ist die Geometrie unseres heutigen Sonnensystems aus der modellierten Akkretions- und Kollisionsgeschichte einer protoplanetaren Wolke erstaunlich gut rekonstruiert worden, ohne im Einzelnen trajektorienorientiert (instabile) Mehrkörperdynamiken berechnen zu müssen.

Umfeld von Komplexitäts- und Selbstorganisationstheorien, d. h. der Theorien nichtlinearer dynamischer Systeme entwickelte hat. In den 1960er- und 1970er-Jahren versuchten Mathematiker und Physiker, versteckte Regelhaftigkeiten in mechanischen Glücksspielen, wie Roulette, aus akustisch Sensordaten zu identifizieren und prognostisch zu nutzen (Crutchfield 1994). Um dies zu ermöglichen, wurden eigens neue Verfahren entwickelt, um die versteckte, schwache Kausalität offenzulegen (Takens 1985; Sauer et al. 1991; Theiler et al. 1992; Abarbanel 1996).⁴⁵ Mit anderen Worten: Einige Verfahren, die heute im Machine Learning etabliert sind, wie Falsche Nächste Nachbarn, Surrogate Datenanalyse inklusive Surrogate Gradient Learning und Gradientenverfahren, oder allgemein Bootstrapping oder Data Mining finden sich schon hier. Gleiches gilt für Modellierungs-, Analyse- und Trainingsverfahren im Umfeld Zellulärer Automaten mit entsprechender Rekursivität (Pathak et al. 2018b).

Zusammenfassend kann man sagen: Während herkömmliche (moderne) Wissenschaften bei komplexen, nichtlinearen, selbstorganisationsfähigen Objektsystemen epistemische Limitierungen aufweisen – nämlich: Grenzen der mathematischen Prognostizierbarkeit, der experimentellen (Re-)Produzierbarkeit, der empirischen Prüfbarkeit sowie der Erklärbarkeit (Mainzer 1996; Schmidt 2015) –, so ermöglichen KI- und Machine-Learning-Verfahren einen verbesserten und erweiterten Zugang zu diesen Systemen. Im Zentrum dieses Zugangs liegt ein neuer und kluger Umgang mit der für diese Systeme charakteristischen schwachen Regelhaftigkeit („schwache Kausalität“) und den instabil-sensitiven Systemdynamiken. Machine-Learning-Verfahren beruhen auf der Wette, dass man in den entsprechenden Daten solche Regelhaftigkeiten, Muster und Strukturen findet, welche eine zugrunde liegende verborgene Ordnung dieser Daten beschreibt und diese für praktische Ziele sodann nutzbar gemacht werden kann. Dieser Zugang zur schwachen Kausalität und ihrer Nutzung zu klassifikatorisch-diagnostischen und prognostischen Zwecken war schon kennzeichnend für die seit den 1970er-Jahren entwickelten Komplexitäts- und Selbstorganisationstheorien. Im Rahmen dieser Konzepte wurde Verfahren, die heute im Umfeld von KI und Machine Learning etabliert sind, entwickelt und getestet. Allerdings war die Rechnerleistung damals noch nicht hinreichend und es lagen nicht so viele hoch aggregierte Daten vor. Seither konnten die Daten-Verarbeitungsprozesse, -speicherungskapazitäten und -verfügbarkeiten gesteigert werden. Was damals auf den Weg gebracht wurde, zeigt sich erst in der letzten Dekade, nämlich als Erfolg der KI- und Machine-Learning-Verfahren in ihrem Zugriff auf komplexe Objektsysteme.⁴⁶

Die neuen datengetriebenen Verfahren können als Katalysatoren einer weiteren Verwissenschaftlichung gesehen werden – auch wenn sich verändert, was

⁴⁵ Hieraus wurden z. B. Anwendung in der Medizin abgeleitet, etwa zur Prognose von bestimmten Typen von Herzinfarkten sowie von Schizophrenien. Zudem wurde das Riechsystem (von Hasen u. a.) entsprechend untersucht und Wahrnehmungsmuster klassifiziert (Freeman und Skarda 1985).

⁴⁶ Dieser Zusammenhang wird seit einigen Jahren in KI- und ML-Communities anerkannt – und hieraus werden interdisziplinäre Forschungsfragen entwickelt, die für die Zukunft von KI und Machine Learning als relevant eingeschätzt werden (Pathek et al. 2018b).

Wissenschaft ist. Der wissenschaftliche Zugriff weitet sich aus und bezieht sich auf immer komplexere Objektsysteme, die sich bisher der wohletablierten (modernen) Wissenschaft aufgrund spezifischer Objekteigenschaften entzogen haben.

4.5 Die Black Box und ihre Intransparenz

Bisher wurde dargelegt, dass KI- und Machine-Learning-Verfahren als methodische Ansätze zu verstehen sind, um einen Umgang mit komplexen (nichtlinearen, selbstorganisationsfähigen, teilweise instabilen) Objektsystemen zu ermöglichen. Vor diesem Hintergrund kann man weitergehen und eine Vertiefung der Analyse dieser Verfahren versuchen. Es stellt sich die Frage, wie transparent – d. h. wie interpretier-/erklärbar, wie vorausberechenbar, wie reproduzierbar und damit wie kontrollierbar – Struktur und Verhalten der KI- und Machine Learning-Verfahren selbst ist. Zur Beantwortung dieser Frage werden nun die Verfahren selbst als (nicht-materielle) Objektsysteme betrachtet, die hinsichtlich von Transparenz zu untersuchen sind: Für die folgenden Ausführungen wird also ein Perspektivenwechsel notwendig, nämlich die Verfahren, Instrumente und Mittel als Objektsysteme anzusehen.⁴⁷

Genau genommen weiß man schon einiges über diese Objektsysteme. Denn ein nahe verwandtes Gegenstandsfeld, nämlich das der Computersimulationen, ist hinsichtlich der Frage nach Transparenz gut untersucht.⁴⁸ Computersimulationen gelten – neben Empirie (Beobachtung, Experiment) und Theorie (Modell, Gesetz) – mittlerweile als dritte Säule wissenschaftlicher Wissensproduktion (Humphreys 1991; Küppers und Lenard 2005; Winsberg 2010; Gramelsberger 2010). Unter dem Stichwort „Opazität“ wird diskutiert, ob sich eine Intransparenz des Geneseprozesses wie des Geltungsausweises von Wissen, erzeugt durch die reduzierte analytische Nachvollziehbarkeit von Computersimulationsläufen und deren Resultate, in die Wissenschaft einschleichen. Eng verbunden sind Probleme der Reproduzierbarkeit der Resultate und des intersubjektiven Geltungsausweises von Aussagen (Kaminski et al. 2018).

Bezogen auf Computersimulationen meint der Wissenschaftsphilosoph Paul Humphreys (2004), dass der Forscher vom Zentrum des Forschungsprozesses sukzessive an den Rand gedrängt werde – allerdings ohne, dass dieser vollständig ersetzt werden könnte. Die Verdrängung gehe einher mit einer zunehmenden epistemischen Opazität, so Humphreys, also einer durch die erfolgreiche und breite

⁴⁷ Das ist in den Technikwissenschaften üblich.

⁴⁸ Für eine Einschätzung und Beurteilung der Transparenz bzw. Opazität von Forschung unter Verwendung von KI- und Machine-Learning-Verfahren kann einiges aus der Diskussion um Computersimulationen übernommen werden. Denn einerseits haben diese Verfahren eine strukturelle Ähnlichkeit zu Computersimulationen, insofern sie ebenfalls auf Algorithmen basieren, andererseits sind für die Entwicklung mathematischer Modelle, die Computersimulationen zugrunde liegen, Dimensions-, Parameter- und Anfangsbedingungs-Schätzungen notwendig, welche heutzutage oftmals durch KI- und Machine Learning-Verfahren vorgenommen werden.

Verwendung computernumerischer Verfahren und algorithmischer Instrumente erzeugte Intransparenz: Der Forscher sei kognitiv nicht mehr in der Lage zu überblicken, was in Simulationen geschehe und wie die Ergebnisse produziert werden. Zwar könne er die jeweiligen algorithmischen Einzelschritte nachverfolgen. Doch bliebe ihm verborgen, wie und warum bei einem konkreten Simulationslauf eines komplexen Modells aus einem bestimmten Inputwert ein bestimmter Output folgt. Unerwartetes und Überraschungen können auftreten, d. h. in gewisser Hinsicht kann Neues entstehen. Zudem sei oftmals die Reproduktion der Output-Ereignisse eingeschränkt, da die Rechenprozesse nichtlinear sind und sensitiv reagieren. Eine Beschreibung, Interpretation oder Erklärung dessen, was in der „Black Box“ der computernumerischen Simulation vor sich gehe, sei kaum möglich, womit Probleme der Validierung verbunden seien. Der Forscher habe kein Kriterium zur Hand, um zwischen richtigem und falschem Resultat zu unterscheiden oder gar Zweifel an einem Resultat adäquat begründen zu können. Was dann noch als intersubjektiver Geltungsausweis anzusehen sei, sei offen.

Humphreys Analyse kann man verallgemeinern und vertiefen: Als Quelle epistemischer Opazität sind nicht nur Computersimulationen anzusehen, sondern allgemein algorithmisch-informatische Systeme ab einem bestimmte Komplexitätsgrad. Schließlich bezieht sich Humphreys' Begründung für Opazität auf hohe Rechengeschwindigkeit und große Komplexität⁴⁹ von Simulationen.⁵⁰ Beides gilt auch für KI- und Machine Learning-Verfahren. Genetische Algorithmen und Support Vector Machines sowie (versteckte) Markov Ketten/Modelle und insbesondere Künstliche Neuronale Netze, welche aufgrund ihrer Struktur grundlegend für KI- und Machine Learning-Verfahren sind, sind komplexe, nichtlineare, selbstorganisationsfähige und mitunter instabile Systeme.

Bei Künstlichen Neuronalen Netzen liegt eine wesentliche Quelle der Nichtlinearität⁵¹ in der so genannten *Spike*-Struktur, die durch das Verhalten der Einzel-

⁴⁹ Mit Komplexität ist nicht nur Kompliziertheit (z. B. Anzahl der Elemente oder Komponenten) gemeint, sondern zusätzlich die Dynamik des Systems.

⁵⁰ Man kann Humphreys' Ansatz weiter fundieren und fragen, worin die große Komplexität von Simulationen wurzelt. Man wird auf Objektsysteme schauen müssen (s. o., vorangegangenes Unterkapitel), deren Modelle simuliert werden. Zu den Charakteristika der Objektsysteme, die in modellbasierten Simulationen repräsentiert werden, gehören: Komplexität, Nichtlinearität, Instabilität/Sensitivität, Selbstorganisationsfähigkeit – und mithin: schwache Kausalität. Diese Charakteristika stellen eine prinzipielle, durch die Objektsysteme induzierte Quelle der Opazität dar. Ein analytischer, rein mathematischer Zugang mit Papier und Bleistift zu diesen Modellen ist unmöglich: das nackte Modell sagt noch nichts aus. Um zu erkennen, muss man geschehen lassen, d. h. man muss numerisch simulieren und berechnen. Angesichts der (nichtlinearen und sensitiv-instabilen) Struktur der Objektsysteme kann man sagen: „Es gibt Beweise dafür, dass die mathematische Opazität [in Computersimulationen] nicht aufgehoben werden kann.“ (Kaminski et al. 2018, S. 275)

⁵¹ Von einer „extreme nonlinearity of deep neural networks“ sprechen Goodfellow et al. (2015). Auch Weinan (2017) spricht von „high-dimensional nonlinear functions used in machine learning“. Und Bertels et al. (2001) sehe im „Backpropagation Paradigm“ neuronaler Netze eine Quelle für Instabilität und weitergehend für regelbehaftetes Chaos.

komponenten oder der Neurone generiert wird (Mainzer 1996).⁵² Neurone feuern ab einer bestimmten Inputgröße bzw. einem bestimmten Erregungsniveau. Wird die Inputgröße nicht erreicht, ist das entsprechende Neuron nicht aktiv und leitet kein Signal weiter. Dieses Verhalten des Neurons entspricht in mathematischer Hinsicht einer starken Form der Nichtlinearität. Darüber hinaus sind in Künstlichen Neuronalen Netze Feedback-Schleifen eingebaut, so genannte Back-Propagationen, ohne die KI-, Machine- und insbesondere Deep-Learning-Verfahren nicht möglich wären. Mit Back-Propagationen werden Parameter, wie beispielsweise die Anzahl der Neuronenebenen, ihre Gewichte und Wechselwirkungen, jeweils iterativ (in großer Geschwindigkeit entsprechend der jeweiligen Prozessorleistung des Computersystems) justiert, ohne dass der Forscher Hand anlegen muss oder etwas davon mitbekommt. Zudem gibt es in vielen Typen Künstlicher Neuronaler Netze versteckte Neuronenschichten („hidden layers“), welche auf- und abgebaut werden. Sie sind typischerweise aus nichtlinearen Funktionsklassen aufgebaut. In vielen Anwendungsfällen ist die zeitliche Iterationslänge, d. h. die Anzahl der Parameter- und Systemgrößen-Updates nicht festgelegt. Aufgrund all dieser Nichtlinearitäten ist es bei Neuronalen Netzen nicht möglich, Standardverfahren der Fehlerfortpflanzung und der Fehlerberechnung zu implementieren.⁵³

Was für Neuronale Netze hier beispielhaft für eine Vielzahl informatischer Verfahren dargelegt wurde, kennzeichnet – grundlegender als eine alleinige Reflexion von Anwendungsfeldern – den mathematisch-informatischen Kern von (nicht-symbolischer) KI und Machine-Learning. Dieser Kern kann charakterisiert werden durch Stichworte wie Komplexität, Nichtlinearität, Selbstorganisationfähigkeit und Sensitivität/Instabilität. Insofern KI- und Machine-Learning-Verfahren als (iterative) nichtlineare dynamische Systeme mit diesen Charakteristika anzusehen sind, die zudem kontinuierlich Updates von einer sich verändernden Systemumwelt (z. B. via Sensoren) als Inputgrößen erhalten, gelten für sie (als Objektsysteme) – analog zu dem, was in den vorangegangenen Kapiteln gezeigt wurde – Limitierungen. Sie sind begrenzt (1) prognostizierbar, (2) reproduzierbar, (3) prüf-/testbar und (4) interpretier-/erklärbar (vgl. allgemein: Mainzer 1996; Schmidt 2008).⁵⁴ Diese vier Limitierungen können als vier (nicht-disjunkte) Dimensionen epistemischer Opazität von KI- und Machine-Learning-Verfahren verstanden werden. Sie sind nicht eliminierbar – auch wenn Wege eines reflektierten Umgangs möglich sind. So

⁵² Meist handelt es sich um so genannte McCulloch-Pitts-Neuronen oder um kleine Modifikationen davon. Klassisch werden die Hebb'schen Lernregeln verwendet.

⁵³ Siehe auch Kap. 2 der Gesamtstudie zur Einführung in die mathematisch-informatischen Aspekte der KI- und Machine-Learning-Verfahren.

⁵⁴ Mit anderen Worten: Das Verhalten vieler Verfahren ist (1) *ex ante* nicht antizipierbar – mithin ist es nicht extern prognostizierbar, ohne dass das Verfahren selbst die iterative Berechnung durchführt. Es ist (2) oftmals nicht robust und nicht reproduzierbar – bei kleinsten Störungen werden andere Ergebnisse erreicht (s. u.). Und (3) das Verhalten ist hinsichtlich seiner Güte und Validität i. A. nicht überprüfbar – es ist nicht testbar, ob es zur Erreichung von extern-vorgegebenen oder intern-entwickelten Zielgrößen korrekt gearbeitet hat. Zudem ist es (4) effektiv irreduzibel – also es ist nicht in einem analytisch überschaubaren Schema von Kausalketten rekonstruierbar und dadurch nicht mechanistisch erklärbar.

sind bei einem konkreten KI- oder Machine-Learning Verfahren und seinem konkreten Anwendungsfeld jeweils einzeln zu untersuchen, wie stark diese Limitierungen ausgeprägt sind. Grundsätzlich ist jedoch mit diesen Limitierungen zu rechnen.

Eine Konsequenz aus den o. g. Charakteristika ist die in einigen Anwendungsfeldern auftretende mangelnde Robustheit, die mit Begrenzungen der Reproduzierbarkeit oder Replikabilität (ad 2) in Verbindung steht. So zeigt sich, dass schon der Zusatz von wenig Rauschen bei wesentlichen Klassen von Machine Learning-Verfahren, die auf Neuronalen Netzen aufgebaut sind, zu einer strukturellen Instabilität oder Nicht-Robustheit führen kann (Szegedy et al. 2014; Goodfellow et al. 2015): Kleinste Störungen oder geringfügige Variationen der Inputgrößen, d. h. der Anfangs- und Randbedingungen,⁵⁵ wie sie in realen Daten aufgrund von Messfehlern, Messrauschen oder sonstigen Artefakten stets auftreten, induzieren deutliche Effekte bei den Resultaten und den daraus gewonnenen Aussagen. Bei der Mustererkennung wird ein Straßenhindernis (nur) als Gullydeckel klassifiziert, ein Tumor als Talgeinschluss oder ein Tiger als Katze. Die Problematik potenziert sich bei schneller Echtzeitanalyse mit kontinuierlich neuen Sensor-Inputdaten nochmals, wie sie im autonomen Fahren auftritt.⁵⁶

Die Fehler der Mustererkennung, die in Anwendungsfeldern als konkrete Risiken zu Tage treten, sind aufgrund der o. g. Opazität (*ex ante*) schwer zu antizipieren, aber auch sodann (*ex post*) schwer in den Griff zu kriegen. Umso relevanter ist es, grundlegender anzusetzen: Denn zunächst und allgemein stellen die vier Dimensionen der Opazität ein epistemisches Risiko⁵⁷ hinsichtlich der Qualitätssicherung des wissenschaftlichen Wissens (z. B. der o. g. Mustererkennung) dar. Wenn aufgrund der Opazität nicht mehr nachvollziehbar ist, wie ein Resultat zustande kommt, und Resultate aufgrund der Nicht-Robustheit oftmals nicht reproduzierbar sind, sind traditionelle Kriterien der Qualitätssicherung außer Kraft gesetzt.⁵⁸ Weitere prinzipielle epistemische Risiken, die vielfach aus grundsätzlichen Limitationen nicht

⁵⁵ Synonym verwendete Begriffe in der Informatik sind hier: Systemgrößen oder Parameter.

⁵⁶ Es schleichen sich dadurch auch in technischen Anwendungen der Nutzungspraxis Risiken ein, für die die Software-Problematik der Boeing 737 Max und ihrer Intransparenz womöglich nicht untypisch war.

⁵⁷ Allgemein zum Stichwort des epistemischen Risikos, siehe Biddle und Kukla (2017). Epistemische Risiken umfassen so genannte induktive Risiken. Ursprünglich war das Konzept der induktiven Risiken eng, nämlich falsche Schlussfolgerungen vorzunehmen, etwa die Annahme einer falschen Hypothese/Aussagen oder die Ablehnung einer richtigen Hypothese/Aussage. Heather Douglas (2000, S. 565) hat das Konzept erweitert und dafür argumentiert: „Significant inductive risk is present at each of three ‚internal‘ stages of science: choice of methodology, gathering and characterization of data, and interpretation of the data.“ Damit ist nicht mehr nur von Induktion, sondern allgemeiner von epistemischen Risiken die Rede, die mit Entscheidungen durch den Forscher in Verbindung stehen.

⁵⁸ Vor diesem Hintergrund bleibt auch oftmals unbestimmbar, ob in KI- oder Machine-Learning-Verfahren ein algorithmischer Programmfehler im Quellcode vorliegt. Die traditionelle Leitunterscheidung der Technikwissenschaften „funktioniert-kaputt“ kann nicht mehr verwendet werden (Kaminski 2014).

eliminiert werden können, betreffen die Verzerrungen („Bias“-Typen), wie sie in Kap. 2 diskutiert wurden.⁵⁹ Auch diese tragen zur Opazität bei.

Die Problematik der Opazität und der epistemischen Risiken ist der Informatik und den die informatischen Verfahren verwendenden Disziplinen, wie die (Teilchen-)Physik, seit langem bekannt. KI- und Machine-Learning-Verfahren werden in kritisch-konstruktiver Absicht als „Black Box“ bezeichnet. Von „Interpretationsproblemen“ ist die Rede, weshalb sich seit gut 20 Jahren Ansätze zu einer „erklärenden KI“ („explainable AI“) entwickelt haben (Murdoch et al. 2019; Chakraborty et al. 2017).⁶⁰ Denn, „[u]nderstanding the behaviour of artificial intelligence systems is essential to our ability to control their actions reap their benefits and minimize their harms.“ (Rhawan et al. 2019, S. 477) So entwickelt sich derzeit eine „Verhaltensforschung für KI-Systeme“: „Erklärende KI“ tritt an, die „Black Box“ (zumindest teilweise) zu öffnen („White-Boxing“). Damit soll sichergestellt werden, dass die jeweiligen informatischen Verfahren interpretierbar, d. h. transparent und nachvollziehbar sind. Nur so sei eine „algorithmic accountability“ möglich – was ethische und juristische Relevanz besitzt. Dabei setzt die „erklärende KI“ wiederum auf informatische Lösungen, die bis in die algorithmische Struktur der Programmcodes reinreichen. Allgemein weist der Zugang der „erklärenden KI“ eine Verwandtschaft zu den in den Technik- und Ingenieurwissenschaften üblichen Verfahren des „reverse engineering“ auf, welche (fertige) technische Produkte und Prozesse (*ex post*) untersucht und deren Verhalten zu beschreiben und zu erklären versucht (vgl. Rhawan et al. 2019, S. 480/484). Kontrovers ist, ob und wie vielversprechend die Ansätze der „erklärenden KI“ sind. Doch jenseits etwaiger Einschätzungen zukünftiger Erfolgsaussichten ist bemerkenswert, dass es diese Diskussion überhaupt gibt. Aus dieser ist auch ein Bedarf ableitbar, nämlich über *Ex-post*-Ansätze hinauszugehen und wissenschaftliche Anstrengungen in Richtung einer prospektiven, d. h. *Ex-ante*-Qualitätssicherung der jeweils in konkreten Kontexten verwendeten Algorithmen bzw. KI- und Machine Learning-Verfahren zu verstärken.

So gilt zusammenfassend, nicht nur die Objektsysteme, die durch die neuen informatischen Verfahren verbessert und vertieft zugänglich werden, sind komplex. Vielmehr sind auch die Verfahren selbst komplex. KI- und Machine-Learning-Verfahren sind in ihrer mathematisch-informatischen Struktur – in Aufbau und Ver-

⁵⁹ Diese treten bei KI- und Machine Learning-Verfahren stets auf und können nicht vollständig eliminiert werden. Allgemein umfassen die Verzerrungen: Activity-, Data-, Sampling-, Algorithmic- und Interaction-Bias. Den prominenten Typ der Data-Bias unterteilt man weiter in Label-, Feature-, und Sample Bias (s. o., Kap. 2 dieser Studie).

⁶⁰ Ein Beispiel: „Interpretability is a quickly growing field in machine learning, and there have been multiple works examining various aspects of interpretations (sometimes under the heading, explainable AI).“ (Murdoch et al. 2019) Terminologisch ist allerdings Vorsicht angebracht, weil der Begriff „erklärend“ (bei „erklärender KI“) nicht im engeren Sinne mit dem theoriebezogenen Erklärungs-begriff (z. B. deduktiv-nomologische Erklärungen) der traditionellen modernen Wissenschaft zusammenkommt, sondern primär auf eine gewisse Transparenz sowie eine gewisse Validität zielt. Insofern ist der Begriff „interpretierbar“, der auch in der Informatik vielfach verwendet wird, hier vorzuziehen.

halten – ebenfalls komplexe, nichtlineare, selbstorganisationsfähige (nicht-materielle) Objektsysteme. Mit anderen Worten: um mit (äußerer) Komplexität, also mit komplexen Gegenständen der Wissenschaft umgehen zu können, benötigt es auf Seiten der Verfahren, Instrumente und Mittel ebenfalls eine hinreichende (innere) Komplexität. Vermeintlich paradox formuliert: KI- und Machine-Learning-Verfahren zielen darauf ab, Komplexität durch Komplexität beherrschbar zu machen;⁶¹ sie treiben die Berechenbarkeit der Welt der Objektsysteme voran, ohne selbst berechenbar zu sein.⁶² Dabei bleiben nicht nur die Objektsysteme, auf die sie sich beziehen, sondern auch die Verfahren selbst in spezifischer Hinsicht opak, also eine „Black Box“. Ob mit etwaiger Opazität nicht voraufklärerische oder gar magische Elemente in die Wissenschaft einziehen, gegen die Wissenschaft einst eingetreten war und die als überwunden galten, etwa Praktiken im Umfeld des Orakels von Delphi, wie Kritiker befürchten, ist ein offener, in kritisch-gestalterischer bzw. regulatorischer Absicht durchaus diskutierbarer Punkt.

All das schmälert nicht den Erfolg von KI- und Machine-Learning-Verfahren, einen Zugang zu komplexen Objektsystemen zu gewinnen und deren Verhalten zu klassifizieren und zu prognostizieren. Was jedoch von Wissenschaftlerinnen und Wissenschaftlern aller Disziplinen aufgrund ihres Anspruchs auf Qualitätssicherung zuallererst zu leisten ist, ist die Wahrnehmung der informatisch induzierten Opazität. Dass sich ein opakes Instrument zu etablieren beginnt, ist allein nicht das Problem, sondern dass die damit verbundene Ambivalenz kaum hinreichend gesehen wird und die Herausforderung zum Umgang mit dem neuen Instrument noch nicht umfassend anerkannt wird. Die Opazität sollte beleuchtet und die durch diese erzeugten epistemischen Risiken sollten (wo immer möglich) minimiert werden – ansonsten werden sie zu Anwendungs-Risiken und eine hinreichende Qualitätssicherung ist für die technische Praxis nicht gewährleistet (Rhawan et al. 2019, S. 483).⁶³

⁶¹ Komplexe Objektsysteme werden nur durch (andere) komplexe Verfahren, Instrumente und Mittel zugänglich. Angesichts dieser informatisch generierten Komplexitätsproblematik gelangt man schnell in einen infiniten Regress – in ein Komplexitätsdilemma.

⁶² Auch in dieser Hinsicht deutet sich eine neue Sichtweise von Wissenschaft an. Traditionell wurde (in der modernen Wissenschaft seit Bacon im 16. Jahrhundert) davon ausgegangen, dass die Komplexität, wie sie sich in den Objektsystemen in Natur, Technik und Gesellschaft phänomenologisch zeigt, nur scheinbar ist. Eigentlich, so die selten explizierte Annahme, sind diese Objektsysteme in ihrem gesetzeshaften Kern einfach. Wissenschaft hat demnach die Aufgabe, die den (phänomenal komplex erscheinenden) Objektsystemen zugrunde liegenden einfachen Gesetze zu finden und offenzulegen. Der Fortschritt der Wissenschaften führt also komplexe Phänomene (der Objektsysteme) auf einfache Gesetze zurück. Dies ermöglichte dem Menschen, so die traditionelle Sichtweise, Prognose- und Kontrollmacht. Kurz gesagt: Komplexes war durch Einfaches beherrschbar. Nun bricht die aktuelle KI- und ML-getriebene Wissenschaft mit dieser Sichtweise. Sie geht nicht mehr davon aus, dass es (a) möglich und (b) hilfreich ist, Komplexität auf Einfaches zu reduzieren. Vielmehr ist, auf eine kurze Formel gebracht, Komplexes durch Komplexität beherrschbar.

⁶³ Damit verbunden wird von einigen kritischen Informatikern die Notwendigkeit der Entwicklung eines neuen, interdisziplinären Forschungsfeldes, was sich dieser Thematik annimmt: „Furthering the study of machine behaviour is critical to maximizing the potential benefits of AI for society.

4.6 Auf dem Wege zu einer nachmodernen Technik – verwendet auch in der Wissenschaft

KI- und Machine-Learning-Verfahren werden nicht selten als neue Instrumente bezeichnet. Wer die Instrumentenmetapher verwendet, sieht oftmals informatische (wie auch mathematische) Verfahren als neutrale Mittel und wertfreie Methoden an, welche die etablierte technische *Toolbox* der experimentellen Forschung (ein wenig) ergänzen und erweitern. Derartige technische Instrumente, so legt die Instrumentenmetapher nahe, mögen äußerst nützlich sein, doch bleiben sie Wissenschaft und Forschung äußerlich.

Dass diese Sichtweise zu kurz greift und zu einer Relativierung aktueller Entwicklungen führt, haben wir in den letzten Unterkapiteln dargelegt. Mit der hier vorgetragenen Skepsis gegenüber einer unterkomplexen Instrumentenmetapher wird aber nicht bestritten, dass die informatischen Verfahren nicht *auch* Instrumente sind: KI- und Machine-Learning-Verfahren sind *auch* – ähnlich wie traditionelle Apparate und Verfahren (z. B. Fernrohr oder Mikroskop) und wie herkömmliche Mess-, Experimentier- und Analysesysteme, die in der Forschung als Mittel Verwendung (z. B. Amperemeter, Teilchendetektor, Method-of-Least-Squares oder Maximum-Likelihood-Method) finden – als Instrumente und somit als technische (in diesem Fall: nicht-materielle) Objektsysteme anzusehen. Nimmt man eine solche begriffliche Zuweisung vor, dann ist aber zu fragen, um welchen Typ des Instruments oder, allgemeiner, um welche der Art von Technik es sich handelt. Damit verbunden ist die Frage, ob sich Technik in ihrem systemischen Kern verändert.

Um diese Fragen anzugehen, ist es notwendig, technische Systeme detaillierter zu betrachten. Technik transformiert, so kann man mit dem Kybernetiker Heinz von Foerster (1995) sagen, Input- in Output-Größen. Für klassische wie für moderne Technik ist eine hinreichend stabile, stark-kausale Input-Output-Transformationsregel konstitutiv, die auf einer definier- und erkennbaren Schnittstelle zwischen dem technischen System einerseits und seiner Umwelt andererseits basiert. Über diese Schnittstelle sind die Material-, Informations- und Energieflüsse geregelt und regelbar, die die Transformation kennzeichnen. Konstitutiv ist die Stabilität bzw. Robustheit der Transformationsregel, um die gewünschte Funktionalität zu gewährleisten. Stabilität meint, kleinste Störungen oder geringfügige Veränderungen der Randbedingungen oder der Inputgrößen sollen die Funktionalität nicht beeinträchtigen. Stabilität – und damit die jeweilige Funktionalität – wird durch die technische Konstruktion eingestellt, d. h. Reproduzierbarkeit wird sichergestellt: Der Nutzer kann das technische System aufgrund der Erwartbarkeit des Outputs oder Effekts verwenden, er kann mit diesem System als Mittel intentional handeln. Die Technikwissenschaften liefern über das technische System und deren Funktionalität ein

The consequential choices that we make regarding the integration of AI agents into human lives must be made with some understanding of the eventual societal implications of these choices. To provide this understanding and anticipation, we need a new interdisciplinary field of scientific study: machine behaviour.“ (ebd.)

grundlegendes Wissen, das auf modell- und gesetzesbezogenen Erklärungen basiert und eine hinreichende intersubjektive Nachvollziehbarkeit der internen Transformationsregeln ermöglicht.

Derartige klassisch-moderne Technik mit einer stabilen und statischen Transformationsregel wird von Heinz von Foerster unter dem Begriff der „trivialen Maschine“ näher ausgeleuchtet (Foerster 1995; vgl. auch Kaminski 2014). Klassisch-moderne Technik lässt sich mit von Foerster als „Trivialisierung einer komplexen Umwelt“ zur Regulation von Material-, Informations- und Energieflüssen verstehen. Von „funktionierender Simplifikation“ spricht Niklas Luhmann (2003) analog. Für von Foerster ist Technik im Sinne trivialer Maschine durch die statisch-stabile Input-Output-Transformationsregel bestimmt. Wesentliche Eigenschaften dieses Maschinen- bzw. Techniktyps sind nach von Foerster:⁶⁴ (1) *Synthetische Determiniertheit*: Input- und Output-Größen sind eindeutig festgelegt und starr bzw. strikt gekoppelt. Die dazu notwendige Transformationsregel ist statisch, d. h. unveränderlich. Sie hängt nicht vom weiteren Vergangenheitsgrößen ab. (2) *Analytische Determiniertheit*: Die Transformationsregel ist rekonstruierbar, insbesondere ist sie transparent, d. h. sie ist wahrnehmbar, interpretier- und erklärbar. (3) *Prognostizierbarkeit*: Die Output-Größen können aus den Inputgrößen prognostiziert werden; Effekte sind erwartbar.⁶⁵ – Die trivialen Maschinen, wie sie von Foerster im Blick hat, sind solche, die wir mit klassischer und moderner Technik in Verbindung bringen: Locher, Schere, Herd, Hausbeleuchtung, Taschenrechner, Fahrrad, Fön, Wecker, Ölheizung, bis hin zum traditionellen PKW, zu Werkzeugmaschine, Produktionsanlage und -prozessen. Nun müsste man, was hier nur angedeutet werden soll, in obiger Liste zumindest (4) eine weitere Eigenschaft ergänzen, nämlich dass bei trivialen Maschinen die Transformationsregel stabil, d. h. hinreichend robust ist und kleinere Störungen irrelevant sind. Das steckt indirekt in der analytischen Determiniertheit (ad 2) sowie der Prognostizierbarkeit (ad 3) drin,⁶⁶ sollte aber nochmals vor dem Hintergrund der Frage nach einer möglichen aktuellen Veränderung des Technik- und Instrumententyps hervorgehoben werden.

Unabhängig davon, ob und in welcher Hinsicht weitere Eigenschaften herkömmlicher, klassisch-moderner Technik zu ergänzen sind, wie von Kaminski (2014) gefordert, und ob nicht sowieso schon einige technische Systeme um uns herum nicht mehr der trivialen Technik entsprechen,⁶⁷ muss man sagen, dass KI- und Machine-Learning-Verfahren gewiss nicht von diesem klassisch-modernen Typ sind:⁶⁸ sie zielen nicht darauf ab, statische, stabile und nachvollziehbare Transformationsregeln

⁶⁴ Von Foerster führt weitere Eigenschaften an, die in unserem Zusammenhang nicht relevant sind.

⁶⁵ In wissenschaftsphilosophischer Terminologie würde man die soeben genannten drei Eigenschaften als ontologisch, epistemologisch und methodologisch bezeichnen. Zentrale Eigenschaft ist also die der synthetischen Determiniertheit der Transformationsregel, welche die notwendige Bedingung für die beiden anderen Eigenschaften darstellt.

⁶⁶ Das heißt dies kann im strengen Sinne als Folgerung angesehen werden.

⁶⁷ Wie beispielsweise die temperaturabhängige Heizungssteuerung oder Steuerungssysteme von Werkzeugmaschinen oder zur Verkehrsregelung, in Stellwerken der Bahn, u. a.

⁶⁸ Es soll hier nicht zwischen Mechanismus und Algorithmus unterschieden werden.

her- und sicherzustellen. Die vier oben genannten Eigenschaften gelten nicht. Vielmehr bilden die neuen informatischen Verfahren einen neuen, anders gearteten Techniktyp, den man als „nachmodern“ bezeichnen kann (Schmidt 2015; Schmidt 2016).⁶⁹ Das hat nichts mit postmodernen Modetrends zu tun, sondern soll andeuten, dass ein neuer Techniktyp, der technikhistorisch *nach der modernen* Technik entsteht und letztere heutzutage ergänzt, *nicht* aber ersetzt. Bei nachmoderner Technik ist die Input-Output-Transformationsregel nicht unveränderlich und nicht fest implementiert, sondern sie ist flexibel und dynamisch: sie wird in der zeitlichen Systementwicklung (selbstorganisierend) hervorgebracht. Sie ist nichtlinear (Rückkopplung) und sensitiv hinsichtlich kleiner Veränderungen (Instabilität). Für KI- und Machine-Learning-Verfahren ist dies eingängig: Im Lernvorgang wird die Transformationsregel über Trainingsdaten adaptiert bzw. erst gebildet. Maschinelles Lernen kann als Transformation der Transformationsregel angesehen werden. Dies gilt unabhängig davon, welche Lernstrategien verwendet werden, ob über neuronale Netze, evolutionäre/genetische Algorithmen oder Support Vector-Verfahren.

Mit den informatischen Lernprozessen, also mit der Transformation der Transformationsregel ist, wie ausgeführt, eine interne (Selbst-)Organisationsfähigkeit verbunden. Von evolutionären Prozessen ist die Rede, Rekurs genommen wird auf biozentrierte Begriffsfamilien⁷⁰ und Konzepte der Selbstorganisation.⁷¹ Diese für ein aktuelles Technikverständnis weitreichende Konzepte haben ihre eigene Geschichte. Wegbereiter für Selbstorganisationskonzepte sind die interdisziplinären System- und Strukturwissenschaften, die Strukturen und Dynamiken von Objekten unabhängig von ihren jeweiligen materiellen Manifestationen untersuchen. System- und Strukturwissenschaften wurden in den 1940er-Jahren im Rahmen der Kybernetik und Informationstheorie erstmals programmatisch entworfen (erste Phase: Ber-

⁶⁹ Auch andere Autoren vertreten Veränderungsthesen, die auf unterschiedliche Stichworte Bezug nehmen: „transklassische Technik“ (Hubig), „naturalized technology“ (Nordmann), „Biofakte“ (Karafyllis), „alternative design paradigm“ (Kroes) oder „informelle Technik“ (Kaminski). Allerdings wird in diesen Konzepten nur in Ansätzen der systemisch-mathematische bzw. informatisch-algorithmische Kern beleuchtet.

⁷⁰ Es ist von „evolutionären“ Prozessen die Rede, von „neuronalen“ Netzen, „genetischen“ Algorithmen und „zellulären“ Automaten.

⁷¹ Wenn von „Selbstorganisation“ die Rede ist, scheint eine begriffliche Präzisierung unabdingbar. Schließlich ist der Begriff äußerst schillernd. Als ein „Sich-selbst-Organisieren“ des Universums von Immanuel Kant im Jahre 1755 eingeführt, wurden Begriff und Grundgedanken von Friedrich W.J. Schelling aufgegriffen. Trotz einiger historischer Bedeutungsverschiebungen kann ein bleibender semantischer Kern von „Selbstorganisation“ identifiziert werden. Dieser bezieht sich auf den Prozess der Entstehung von etwas, meistens von etwas Neuem („Emergenz“). Kein Konstrukteur steuert diesen Prozess (direkt); er setzt lediglich Anfangs- und Randbedingungen, weshalb von „gerichteter“ Selbstorganisation gesprochen wird. Der Prozess selbst ist ihm sogar partiell unzugänglich – entweder aus prinzipiellen (ontologischen) oder aus pragmatischen (methodologischen) Gründen. So umfasst Selbstorganisation: (a) Entstehung von Neuem/Anderem, (b) Eigendynamik des Systems sowie (c) Entzogenheit relevanter Details, insbesondere die Unabschließbarkeit des Systems, d. h. kein klares Interface. In anderen Worten: Autonomie durch Produktivität, Prozessualität und Adaptivität.

talanny, Wiener, Shannon, von Neumann) und ab den 1960er-Jahren in den mathematischen geprägten Naturwissenschaften (zweite Phase: Prigogine, Haken, Maturana/Varela, von Foerster) detaillierter ausformuliert und ausgeweitet. Prominent sind die bereits genannten und diskutierten Theorien nichtlinearer dynamische Systeme, dissipative Strukturbildung, Synergetik, Autopoiesistheorie sowie Chaos- und Komplexitätstheorien. Während die *erste* Phase eng verbunden ist mit der Herausbildung der Informatik als eigene Wissenschaftsdisziplin und mit der theoretischen Informatik, ist die *zweite* Phase stark durch die konkrete Computerentwicklung geprägt. Und heute kann man von einer *dritten* Phase sprechen, die durch aktuelle KI- und Machine Learning-Verfahren hervorgerufen ist. Sie ist ermöglicht worden durch die allumfassende Digitalisierung: Rechner-, Sensor-, Speicher- und Netztechnologie, was eine Verarbeitung großer und heterogener Datenmengen möglich macht.⁷²

Nun muss man sagen, was sich im informatischen Feld der KI- und Machine Learning-Verfahren zeigt, scheint indes nur die Speerspitze eines allgemeinen Trends unterschiedlicher neuerer („emergenter“) Technologiebereiche zu sein: Selbstorganisationsprozesse spielen eine Rolle (a) in Synthetischer und System-Biotechnologie, (b) in den Nano- und Mikrosystemtechnologien sowie (c) in den Kognitions-, Neuro- und Pharmakotechnologien. Die jeweiligen technischen Systeme erscheinen fast selbst als handelnd, d. h. ein Handlungsstatus wird dieser Technik anthropomorphisierend zugeschrieben: Nachmoderne Technik erscheint demnach phänomenal als „autonom“; sie scheint schöpferisch tätig zu sein, Mittel zweckrational auszuwählen und Entscheidungsfähigkeit zu besitzen.⁷³

⁷² Man sieht, dass heutige KI- und Machine Learning-Verfahren in umfassende wissenschafts- und technikhistorische Kontexte eingeordnet werden können, was über die (engere) Begriffs- (als Wort-)Geschichte (Dartmouth Konferenz 1956, u. a.) hinausgeht.

⁷³ Nun kann man weiterfragen und darauf zurückkommen, was in den Unterkapiteln (s. o.) ausgeführt wurde: Konzepte der Selbstorganisation sind zentral. Was ist der Kern von Selbstorganisation? Konstitutiv für Selbstorganisation sind Instabilitäten bzw. Sensitivitäten. Selbstorganisation wird durch eine Instabilität der bisherigen Struktur gegenüber kleinen Schwankungen eingeleitet. Instabilität meint auch: kleine Ursache, große Wirkung. Sie induziert Rückkopplungs-, Wechselwirkungs- und Verstärkungsprozesse, verbunden mit einem veränderten (schwachen) Kausalitätstyp. Instabilität impliziert also nicht notwendigerweise, dass Systeme kollabieren, sondern vielmehr: Übergänge zu veränderten Zuständen. Wer nun Selbstorganisation nutzen will, wie in KI und Machine Learning, muss Instabilitäten provozieren und mit diesen umgehen; Selbstorganisation bedarf des Durchgangs durch Phasen der Instabilität. In Instabilitäten liegt die technologische Quelle jener Produktivität, die KI- und Machine Learning-Verfahren nutzen. Nun können Instabilitäten durchaus zweischneidig sein. Denn sie sind nicht nur zentrale Quelle von Produktivität. Sie führen auch zu Grenzen der Konstruierbarkeit sowie der Kontrollierbarkeit der Konstruktionen – wie gerade im unüberwachten Machine-Learning und Deep-Learning deutlich wird. Limits der Prognostizierbarkeit und (Re-)Produzierbarkeit treten auf. Schließlich kommt es bei Instabilität auf aller kleinste Details an: Zunächst geringfügig erscheinende Anstöße können große Wirkungen entfalten. Technisch können diese Details aus prinzipiellen wie aus pragmatischen Gründen niemals vollständig beherrscht werden; sie bleiben letztlich unzugänglich. Mithin ist es rückblickend durchaus plausibel, weshalb Instabilitäten in der klassischen und modernen Technik nichts verloren hatten. Wo sie auftraten, waren sie störend; es galt, sie zu eliminieren. Nur bei Stabilität waren die Bedingungen von Konstruierbarkeit und Kontrolle erfüllbar. Technik war

Dass indes grundlegende Probleme mit einem komplexen, nachmodernen Techniktyp verbunden sind, hat der Systemtheoretiker und Soziologe Niklas Luhmann gesehen. Traditionelle Technik ist für Luhmann, ähnlich wie für von Foerster, als „funktionierende Simplifikation“ und „simplifizierende Isolation“ eines Systems von seiner Umwelt zu verstehen. Bei herkömmlicher Technik könne die Isolation konstruktiv hergestellt werden, eine Schnittstelle könne definiert und ausgewiesen werden. Bei komplexer Hochtechnologie – Luhmann konnte zu seiner Zeit noch nicht ahnen, welche Entwicklung die informatischen Verfahren nehmen sollten – hingegen komme es zum „Anwachsen kausaler Komplexität“ (Luhmann 2003, S. 101). Diese basiert konstruktiv auf einer „immensen Komplexität von gleichzeitig [...] ablaufenden Kausalvorgängen“ (ebd.), wie es gerade auch für KI- und Machine-Learning-Systeme charakteristisch ist.

Aus dem Misslingen der Kausalisation und dem damit einhergehenden Verlust der Schnittstellen zwischen System und Umwelt ergeben sich Herausforderungen. Nach Luhmann hat man es „mit Chaosproblemen, mit Interferenzproblemen und mit jenen praktisch einmaligen Zufällen“ zu tun (ebd., 100). Die drei Problemtypen wurzeln für Luhmann in dem, was als Komplexität, Nichtlinearität und Sensitivität/Instabilität bezeichnet werden kann – und was (gleichzeitig) Quelle und Grundlage von Lern-, Adaptions- und Selbstorganisationsprozessen, d. h. für die Produktivität, darstellt. Ist das überhaupt noch Technik, die traditionell reproduzierbare und erwartbare Effekte her- und sicherstellt, fragt Luhmann, und er hat Zweifel. Liegen die Probleme im Kern dieser Systeme selbst, dann wird „die ‚Form‘ der Technik zum Problem. Sie markiert die Grenze zwischen eingeschlossenen und ausgeschlossenen (aber gleichwohl realen) Kausalitäten. Offenbar kommt es bei [nachmodernen] Technologien [wie bei KI- und Machine-Learning-Verfahren] aber laufend zu Überschreitungen dieser formbestimmenden Grenze, zur Einschließung des Ausgeschlossenen, zu unvorhergesehenen Querverbindungen. [...] Das führt zu der paradoxen Frage, ob Technik, auch wenn sie kausal funktioniert, technisch überhaupt möglich ist“ und noch (im traditionellen Sinne) Technik ist (ebd., 100). Damit scheinen sich, so Luhmann, „die Grenzen der technischen Regulation von Technik zu sprengen [...]: Die Probleme der Technik zeigen sich an den Versuchen, die Probleme der Technik mit technischen Mitteln zu lösen.“ (ebd., 99 f.) – Diese rekursive Beziehung wurde auch im letzten Unterkapitel offengelegt und als informatisch generiertes Komplexitätsdilemma bezeichnet: KI- und Machine-Learning-Verfahren sind als Ansatz zu verstehen, (äußere, d. h. objektseitige) Komplexität durch (innere, d. h. verfahrensseitige) Komplexität beherrschbar zu machen. Diese rekursive Beziehung ist höherstufig infinit fortsetzbar.

Eine solche Technik, das kann man mit Luhmann andeuten, scheint (wenn überhaupt) ein anderer Typ von Technik zu sein. Dieser Typ unterscheidet sich nicht nur graduell, sondern grundsätzlich von herkömmlichen technischen Instrumenten und

Technik, insofern sie stabil war. Man kann sagen, dass Technik einerseits und Instabilitäten andererseits als Antagonisten angesehen wurden, wie Feuer und Wasser. – In der nachmoderne Technik der KI- und Machine-Learning-Verfahren hat sich hier eine deutlich veränderte Sichtweise etabliert, was auf ein neues Technik-Paradigma hinweist.

Systemen, auch von jenen, die in Wissenschaft und Forschung eine Rolle spielen. Ob Fernrohr und Mikroskop, ob Amperemeter oder Thermometer, ob Zeitmesser oder Lichtschranke – diese Instrumente sind gekennzeichnet durch ein stabiles Input-Output-Schema. Berechenbarkeit, Reproduzierbarkeit, Prüfbarkeit und Erklärbarkeit war, weitgehend, durch die implementierte Transformationsregel (*ex ante*) festgelegt; Erwartbarkeit war sichergestellt. Das gilt für neuere Instrumente hingegen, die auf KI- und Machine-Learning-Verfahren aufbauen, nicht mehr. Hier wird die Transformationsregel selbst dynamisch transformiert und über Selbstorganisationsprozesse kontinuierlich adaptiert; Luhmann sieht selbstorganisierend-autopoetische Prozesse am Werke. Nachmoderne Technik hat einen Eigensinn, hält Überraschungen parat und ist schwer kontrollierbar, was *die andere Seite* ihrer enormen Leistungsfähigkeit darstellt: Ihr wird Autonomie, Produktivität und Adaptivität zugeschrieben, d. h. Eigenschaften, die auf hoher Sensitivität/Instabilität, auf großer Selbstorganisationsfähigkeit und auf einer weitreichenden Input-Offenheit gegenüber der Umwelt basieren.⁷⁴

Mit einer solchen qualitativen Veränderung des technischen Kerns wissenschaftlicher Instrumente verändert sich die Zusammenarbeit des Forschers mit seinen Experimental- und Analysesystemen und mithin der wissenschaftliche Erkenntnisprozess. So wird in Gesellschaft und Wissenschaft wohl nicht zu Unrecht bei diesem neuen Techniktyp ein Quasi-Akteursstatus erkannt, auch wenn das aus philosophischer Perspektive einen Kategorienfehler darstellen mag. Obwohl derzeit nichts dafür spricht, dass KI- und Machine-Learning-Systeme alleine den Nobel-Preis gewinnen könnten und der Forscher ersetzt werden könnte, wie Hiroaki Kitano (2016) nahelegt, so wird dennoch deutlich: die Verschiebung dieser Zuschreibungspraxis verweist auf grundlegende Verschiebungen in den Forschungspraktiken, d. h. auf neue und weitreichende Interaktionsformen zwischen menschlichem und technischem/nicht-menschlichem Akteur der wissenschaftlichen Wissensproduktion.

4.7 Technikrends: Ermöglichungstechnologie, Konvergenztechnologie, Technoscience

Während in den letzten Unterkapiteln der zweischneidige informatisch-systemische Kern von KI- und Machine-Learning-Verfahren untersucht wurde, soll im Folgenden eine Einordnung in allgemeine Technikrends vorgenommen werden. Drei Sondierungsbegriffe, die die aktuelle Technikphilosophie bereitstellt und die in den bisherigen Ausführungen (schon implizit) eine Rolle gespielt haben, sollen in Anschlag gebracht werden: Ermöglichungstechnologie, Konvergenztechnologie sowie Technowissenschaft („Technoscience“).

⁷⁴Die Offenheit ist sensorisch, d. h. durch kontinuierliche Sensordaten, notwendig, damit auf die Umwelt reagiert werden kann und an dieser gelernt werden kann.

KI- und Machine Learning-Verfahren können, *erstens*, als eine im Hinter- und Untergrund wirkende *Ermöglichungstechnologie* („enabling technology“; Schmidt 2004; Nordmann 2008) angesehen werden. Die neuen informatischen Verfahren sind in erster Linie keine Objekt- oder gegenstandsbezogene Technologie, wie etwa die Atom- oder Biotechnologie, auch wenn letztere durchaus neue technische Objektsysteme, hier Atomkraftwerke, dort Biosubstanzen, konstruktiv ermöglichen. Beispiele traditioneller Objekttechnologien sind auch Fertigungs-, Werkzeuganlagen- und Produktionstechnologien. Ermöglichungstechnologien wirken demgegenüber allgemeiner, grundlegender und hintergründiger. Sie stellen die Bedingung der Möglichkeit von anderen Technologien dar und sind i. w. S. als generische Infrastruktur- oder Systemtechnologien anzusehen. Das Internet ist eine der neueren Ermöglichungstechnologien; ältere Beispiele sind Verkehrsinfrastruktur- oder (Telefon-/Funk-/Morse-)Kommunikationstechnologien. Über den Status von Ermöglichungstechnologien wurde intensiv im Umfeld der Nanoforschung gesprochen, neuerdings auch bei Informations- und Kommunikationstechnologien. Noch ist der Sondierungsbegriff der Ermöglichungstechnologie semantisch nicht vollständig geklärt. Aber schon jetzt werden einige Charakteristika deutlich. Ermöglichungstechnologien sind stark wissenschaftsbasiert und vielfach verbunden mit formalen, mathematischen, theoretischen Entwicklungen. Sie bilden den wissenschaftlichen Hintergrund sowie die technische Grundlage für viele andere Technikentwicklungen, was über das von Innovationstheoretikern verwendete Begriffskonzeption der Basistechnologien hinausgeht.

KI- und Machine Learning-Verfahren fallen wie kaum ein anderer Techniktyp in die Kennzeichnungen einer Ermöglichungstechnologie. Zukünftig wird es kaum noch Objekt- oder gegenstandsbezogene Technologien mehr geben, in welchen diese informatischen Verfahren keine Rolle spielen. Wesentlich für Ermöglichungstechnologie ist, dass diese nicht eigenständig in Erscheinung treten, sondern in anwendungsbezogene technische Systeme integriert sind und für diese als konstitutiv anzusehen sind: Internetsuchmaschinen; Spracherkennung; Autocomplete-Funktionen; journalistische Texterstellung von Wetter- und Sportinformationen; Bots in sozialen Netzwerken; Recommender Systeme und Vorschläge bei internetbasierten Bestellungen; Online-Versicherungen; Spamfiltern; Schach- und Backgammon-Programmen; Fahrassistentensystemen; Motor- und Airbag-Steuerung; Wegeoptimierung in der Logistik; Parkplatzsuchsysteme in Städten; Umweltdatenüberwachung, inklusive Kenngrößenentwicklung; Serverfarmsteuerungen, usw.⁷⁵

⁷⁵Weite Beispiele ließen sich anfügen: Auswertung von medizinischen Daten und medizinische Diagnostik (von Covid-19 über Hautkrebsdiagnose bis hin zu Evaluation von Psychotherapien); Klassifikationen von Versicherungsdaten; Mustererkennung in der Kriminalistik; Prognose von Straffälligkeit; Klassifikation von Gruppen von Arbeitslosen zur prognostischen Bestimmung des Vermittlungserfolges; Prognose von Aktienkursen und Betriebsdaten; Zeitreihenanalyse in Kosmologie und Teilchenphysik; Auswertung von Klimadaten und Hilfestellung bei der Entwicklung von Klimamodellen; Rekonstruktion und Prognose von Kaufentscheidungen; bis hin zu Pflege- und Serviceroboter, zu Smart Home und Smart Grids (letztere sind für die deutsche Energiewende unumgebar) sowie zu Cyberphysical Systems, Ubiquitous Computing und Internet der Dinge.

Dem Endbenutzer ist zumeist nicht klar und transparent, dass KI- und Machine Learning-Verfahren zum Einsatz kommen. Die informatischen Ermöglichungstechnologien bleiben im Hinter- und Untergrund verborgen, ihre Spuren sind nicht sichtbar (Hubig 2006). Sie bilden damit das untergründige „Gestell“ für moderne oder spätmoderne Wissensgesellschaften, wie man mit Martin Heidegger sagen kann. Analog sieht der Soziologe Armin Nassehi (2019, S. 244), in Rekurs auf Niklas Luhmann, eine weithin „unsichtbare Technik“ heraufziehen, was zu einer veränderten Technikerfahrung führe und lebensweltliche Konsequenzen der Mensch-Technik-Beziehung nach sich ziehe. Das gesellschaftliche „Unbehagen“ hängt, so Nassehi, mit der Unsichtbarkeit und entsprechenden Un(an)greifbarkeit der hintergründig wirkenden KI- und Maschine-Learning-Verfahren als verborgene Ermöglichungstechnologien zusammen; das Unbehagen wird befördert durch den intransparenten Black-Box-Charakter dieser Verfahren. Die Black-Box sei es allerdings auch – das ist die andere (ebenfalls ambivalente) Seite der „unsichtbaren Technik“ –, dass dieser (ob gerechtfertigt oder nicht) Autonomie, Intelligenz und sogar Bewusstsein zugeschrieben werden könne. Nach Nassehi „ähnelt“ phänomenal dieser neue (nachmoderne) Techniktyp darin „eher dem menschlichen, für Alter Ego intransparenten Bewusstsein als einem sichtbaren Getriebe mit Zahnrädern, die ineinandergreifen und keine Variationsmöglichkeiten haben, außer dass sie funktionieren oder defekt sind.“ (Nassehi 2019, S. 244)

Ein verwandter Sondierungsbegriff, *zweitens*, ist der der technologischen Konvergenz oder, kurz, der *Konvergenztechnologie*. Dieser wurde in der Debatte über neuere Technikentwicklungen, die ihren Ausgang von einem Workshop der National Science Foundation der USA nahm (Roco und Bainbridge 2002), eingeführt. Programmatisch wird die Frage aufgeworfen, ob es nicht an der Zeit sei, die diversen und sich heterogen präsentierenden Technik- und Ingenieurwissenschaften – und ihre (bereichsspezifischen) Techniktypen – zusammenzuführen und zu vereinheitlichen: Denn Konvergenz erzeuge Synergie und fördere Innovation. Für das Konvergenzprogramm steht die konvergente Entwicklung der Physik der letzten 200 Jahre paradigmatisch Pate:⁷⁶ Das Ideal der Vereinheitlichung hat sich als Treiber wissenschaftlicher Innovation erwiesen. So wird gefragt, ob sich dieses produktive Ideal nicht auf die Technikwissenschaften, unter Einschluss der Informatik, der Lebenswissenschaften und der Medizin, übertragen lässt.

Die National Science Foundation der USA zielt dementsprechend auf „Convergent Technologies“ oder „NBIC-Convergence“ (Roco und Bainbridge 2002). Gemeint ist eine Zusammenführung von entstehenden („emergenten“) Zukunftstechnologien: Nano-, Bio-, Informations- und Kognitionstechnologien (NBIC-Konvergenz). Für eine Konvergenz wird zwar objektseitig der Nanolevel als zentral angesehen. Entscheidender ist aber, dass das methodische Instrumentarium der Konvergenz bei der Informatik sowie bei den Informationstechnologien liegt. Die Informatik gilt in dem Konvergenzprogramm als eine neue Leit- und Fundamentalwissenschaft, die die Nano-, Bio- und Kognitionstechnologien und deren forschungs-

⁷⁶Die Physik hat eine Vereinheitlichung von drei der vier fundamentalen Gesetze erreicht und weist eine hohe interne Kohärenz und Konsistenz auf.

basierte Technikentwicklung methodisch fundiert und diese zusammenführt. Damit verdrängt die Informatik in der wissenschaftlichen wie gesellschaftlichen Wertschätzung die exakten Naturwissenschaften, primär die Physik, sowie, in zweiter Linie, auch die Lebenswissenschaften. Eine neue Hierarchie der wissenschaftlichen Disziplinen scheint sich herauszubilden, wobei sich die Anwendungsorientierung von Forschung (im Dienste der Technikentwicklung) in den Vordergrund schiebt und die Grundlagenforschung eine reduzierte Anerkennung erfährt.

Vor dem Hintergrund der Doppeldiagnose von Ermöglichungs- und Konvergenztechnologie kann man, *drittens*, einen Schritt weitergehen. In den letzten 30 Jahren hat die Wissenschaftsphilosophie, im Verbund mit der sozialwissenschaftlichen Wissenschaftsforschung, Veränderungen im Gefüge der Wissenschaften diagnostiziert. Ein zentraler Begriff, mit dem dieser Wandel beschrieben wird, ist der der „Technowissenschaft“ (engl. „Technoscience“) (Latour 1987; Haraway 1995; Nordmann 2008; Nordmann et al. 2011). Technowissenschaft meint nicht einfach Technik- und Ingenieurwissenschaft, sondern beschreibt einen Wandel in der Forschungskultur aller Disziplinen in Richtung einer Orientierung am Nutzungs-, Verwertungs- und Anwendungskontext: Von Wissenschaft zur Technowissenschaft. Zwar gab es diese (Bacon'sche) Orientierung schon immer, entscheidend ist jedoch die Verschiebung der Schwerpunktsetzung des gesamten Wissenschaftssystems in diese Richtung. Neben traditionelle epistemische Werte, wie Erklärungsleistung, Reproduzierbarkeit, Prognosekraft, Testbarkeit, Theoriekonsistenz und -kohärenz, treten nicht-epistemische Werte, die gesellschaftlichen oder unternehmerischen Nutzen betreffen (Carrier 2011).⁷⁷ Der Philosoph Alfred Nordmann (2005, S. 215 f.) identifiziert einige „Symptome für einen Kulturwandel von der Wissenschaft hin zur Technowissenschaft“: „1. Statt darstellender Hypothesen über die Natur: eingreifende Gestaltung einer hybriden KulturNatur. 2. Statt quantitativer Voraussagen und hochgradiger Falsifizierbarkeit: Suche nach Strukturähnlichkeiten und qualitative Bestätigung. 3. Statt Artikulation von naturgesetzlichen Kausalbeziehungen oder Mechanismen: Erkundung interessanter, bzw. nützlicher Eigenschaften. 4. Statt Orientierung auf die Lösung theoretischer Probleme: Eroberung eines neuen Terrains für technisches Handeln. 5. Statt hierarchische Organisation von Natur und Wissenschaft: Orientierung auf transdisziplinäre Objekte und Modelle.“⁷⁸ Nord-

⁷⁷ Komplementär findet man hier und da auch in traditionellen angewandten Wissenschaften, wie den Technikwissenschaften, deutlich mehr Grundlagen- und Theoriearbeit.

⁷⁸ Ferner nennt Nordmann (2008, S. 215 f.): „6. Statt Trennung von (wissenschaftlicher) Gesetzmäßigkeit und (technischer) Machbarkeit: programmatische Gleichsetzung von natürlich, bzw. physikalisch Möglichem mit technisch Realisierbarem. [...] 7. Statt organisierter Skepsis: Konvergenz eklektischer Theorien auf verbindliche Artefakte hin. 8. Statt Universalismus und einer Wissenschaftsgemeinschaft aus Gleichgestellten: [...] ein Zusammenwirken vieler, ungleich situierter sozialer Akteure. 9. Statt gemeinsamem Eigentum: die Zirkulation von Produkten zwischen Instrumentenherstellern und Laboren, zwischen wissenschaftlich-technischen Einrichtungen, zwischen Labor, Industrie und Gesellschaft. 10. Statt ‚Interessenlosigkeit und Verpflichtung auf Wahrheit‘ als einzig geltendem Interesse: weder die wissenschaftliche Suche bloß nach besseren Theorien noch die technische Einwicklung bloß von besseren Geräten.“

mann adressiert den Wandel der Wissenschaftskultur unter Rekurs auf leitenden Werte und die von den Forschenden verfolgten Ziele in ihren jeweils konkreten Forschungsprojekten: Ihr Forschungshandeln orientiert sich verstärkt an Verwertungskontexten und weniger am aufklärerischen Ideal grundlegender Erkenntnis für das Weltverständnis. Entsprechende Verschiebungen zeigen sich auch in der Selbstbeschreibung des Forschungshandelns durch die Forschenden selbst.⁷⁹

KI- und Machine Learning-Verfahren kennzeichnen wie kaum ein anderes Feld das Regime der Technowissenschaft. Allerdings wäre es wohl treffender (anstatt von *Verfahren*) von KI- und Machine-Learning-*Technowissenschaft* zu sprechen. In einer solchen diagnostischen Zuspitzung fallen traditionelle Dichotomien, wie die von Grundlagenforschung vs. Anwendung weg. KI- und Machine-Learning-Technowissenschaft ist vor diesem Hintergrund als grundlagenbasierte Anwendungsforschung zu spezifizieren. Diese technowissenschaftliche Forschung *über, an* und *für* KI- und Machine Learning-Verfahren findet sich nicht nur im Rahmen der disziplinären Informatik, sondern gerade im Horizont all jener Disziplinen und Interdisziplinen, die mit vielen und heterogenen Datenreihen umgehen.⁸⁰

4.8 Fazit

Ein Regime *nachmoderner Wissenschaft* ist im Entstehen; KI- und Machine-Learning-Verfahren beginnen Wissenschaft in ihrem epistemischen Kern zu verändern. Derzeit ist damit eine Erweiterung und Ergänzung, keine Ersetzung des *Regimes der* herkömmlichen, *modernen Wissenschaft* impliziert.

Die Erweiterung bezieht sich auf zweierlei: Zunächst kann von einer *methodischen Erweiterung* gesprochen werden (Abschn. 4.2 und 4.3): KI- und Machine-Learning-Verfahren erweitern das Methoden- und Instrumentenspektrum der Wissenschaften grundlegend. Damit verbunden ist eine *Erweiterung der Klasse der Objektsysteme*, die sich bislang den Wissenschaften weithin entzogen haben, weil sie komplex, nichtlinear, selbstorganisationsfähig, vielfach instabil/sensitiv, räumlich verteilt und interaktiv sind (Abschn. 4.4). Aufgrund dieser Eigenschaften ist

⁷⁹ Verwandte Konzepte zur Beschreibung eines aktuellen wissenschaftshistorischen Wandels sind: *mode-2-science*, *post-normal science* oder *post academic science*.

⁸⁰ Wie z. B. die Teilchen- und Hochenergiephysik, die Kosmologie, die Bioinformatik, die System- und Synthetische Biologie, die Materialwissenschaften, der Maschinenbau, im Verkehrs- und Bauingenieurwesen, in der Klima-, Umwelt- und Nachhaltigkeitsforschung sowie viele Felder der Technikwissenschaften und der Medizin. Ferner sind die Sozial-, Human- und Wirtschaftswissenschaften zu nennen, etwa die Arbeitswissenschaften, die Psychologie oder die Wirtschaftsinformatik, aber auch Forschungsfelder wie die so genannten *Digital Humanities*. Dabei wurden die KI- und Machine-Learning-Technowissenschaft entwickelt von Informatikern, Mathematikern, Physikern, Biologen, Technikwissenschaftlern, Medizinerinnen, Psychologen, und das vielfach außerhalb von Universitäten und staatlichen Forschungsinstitutionen. Sie ist nicht primär aus der disziplinären Informatik entstanden, auch wenn es heute so scheinen mag und sie heute dort prominent beheimatet ist.

ihr Verhalten schwer berechenbar, kaum reproduzierbar, Aussagen sind schwer prüfbar und ihre Phänomene sind vielfach nicht deduktiv-nomologisch oder mechanistisch erklärbar.

Die methodische und objektseitige Erweiterung von Wissenschaft durch KI- und Machine Learning-Verfahren induziert einen Wandel. Dieser konnte offengelegt werden, indem vier Leistungsmerkmale von Wissenschaft – und damit verbundene (nicht-disjunkte) Wissenschaftsverständnisse – untersucht wurden (Abschn. 4.2 und 4.3): Prognostizierbarkeit, Reproduzierbarkeit, Testbarkeit und Erklärbarkeit. KI- und Machine Learning-Verfahren führen zu einer verstärkten Prognoseorientierung (ohne grundlegende Theorie- und Wissensbasis), zu einer reduzierten Notwendigkeit des reproduzierenden Experimentierhandelns (bei gleichzeitiger Nutzen- und Interventionsorientierung), zur Ersetzung traditioneller Testverfahren (bei gleichzeitigen neuen und erweiterten Prüfoptionen) sowie zu einem verringerten Interesse an Theorien und Erklärungen (bei gleichzeitig starkem Rekurs auf eine schwach-kausale Struktur der Wirklichkeit).

KI- und Machine Learning-Verfahren zielen darauf ab, so wurde gezeigt, versteckte Regelhaftigkeiten, Muster oder Strukturen, also schwache Kausalität, zu diagnostizieren und diese zu prognostischen Zwecken nutzbar zu machen, kurzum: es geht nicht (nur) um Korrelationen, sondern um Kausalität (Abschn. 4.4 sowie Teile von 4.3.1 und 4.3.4). Korrelationen sind zu schwach, um Prognosen vorzunehmen und Handlungen zu ermöglichen. Nur mit Kausalität sind effektive Handlungsstrategien möglich. Mit diesem Zugang schließen die neuen informatischen Verfahren an weitreichende Erkenntnisse der so genannten Struktur- und Systemwissenschaften an, die mit der Kybernetik und Informationstheorie zusammenhängen und zu den theoretischen Grundlagen der Etablierung der Informatik als wissenschaftliche Disziplin gehören. Die Struktur- und Systemwissenschaften haben sich seither weiterentwickelt und sind heutzutage reichhaltiger. Sie umfassen die Theorien nichtlinearer dynamischer Systeme mit Komplexitäts- und Selbstorganisationstheorien, d. h. Synergetik, Theorien dissipativer Strukturbildung und des Hyperzyklus, Chaos-, Katastrophen- und Autopoiesis-Theorien. Diese haben ab den 1970er- bis 1990er-Jahren mathematisch fundierte Verfahren entworfen und angewendet, mit deren Hilfe versteckte Regelhaftigkeit oder schwache Kausalität diagnostiziert und genutzt werden kann.

KI- und Machine-Learning-Verfahren nehmen komplexe, nichtlineare, selbstorganisationsfähige, partiell instabile Objektsysteme in den Blick und erweitern damit die wissenschaftlich zugänglichen Gegenstandsfelder (Abschn. 4.4). Allerdings hat diese Erweiterungen ihren Preis (Abschn. 4.5). Denn die Verfahren, mit denen der Zugang ermöglicht wird, nämlich KI- und Machine-Learning-Verfahren, weisen selbst diese Komplexität auf. Sie sind in ihrer mathematisch-informatischen Struktur ebenfalls komplexe, nichtlineare, selbstorganisationsfähige nicht-materielle Objektsysteme: (äußere) Komplexität wird nur durch (innere) Komplexität zugänglich und beherrschbar. Gleichzeitig zeigen sich Grenzen des Wissens. Werden diese Verfahren, Instrumente und Methoden in den Wissenschaften eingesetzt, entsteht eine neue nicht-eliminierbare Opazität bzw. prinzipielle Intransparenz, verbunden

mit epistemischen Risiken von Fehlschlüssen und -Diagnosen, die man bei herkömmlichen Verfahren, Instrumenten und Apparate so nicht kannte.

Der Wandel von Wissenschaft basiert also auf einem Wandel (der Entwicklung und dem Einsatz) der in der Wissenschaft verwendeten informatischen Technik(en) (Abschn. 4.6). Wenn die Anzeichen nicht trügen, beginnt sich – ergänzend zur bisherigen, d. h. zur *modernen Technik* – ein neuer Typ von Technik zu etablieren, welcher einen anderen (informatischen) Kern in sich trägt. Diese *nachmoderne Technik* weist eine Selbstorganisationsfähigkeit auf, sie erscheint als autonom, als selbst handelnd und als entscheidungsfähig. Sie ist komplex und dynamisch, basiert auf nichtlinearen und vielfach auch sensitiven/instabilen Prozessen, die ihr eine hohe Flexibilität und Adaptivität ermöglichen. Nachmoderne Technik ist nicht durch ein Input-Output-Schema mit einer analytisch rekonstruierbaren stabilen Transformationsregel zu beschreibbar.

Im Rahmen der Reflexion allgemeiner Technologietrends sind KI- und Machine-Learning-Verfahren charakterisierbar als Ermöglichungstechnologie, Konvergenztechnologie sowie Technowissenschaft (Abschn. 4.7). Die Informatik avanciert offenbar – in aller Ambivalenz – zu einer neuen Leit-, Fundamental- und Basiswissenschaft.

Dass KI- und Machine-Learning-Verfahren wissenschaftlich äußerst erfolgreich sind und gesellschaftliche und ökonomische Innovationen fördern, ist unstrittig. Allerdings bedarf es – um die Chancen dieses Zugangs zu komplexen, dynamischen, nichtlinearen, sich selbst organisierenden Objektsystemen in Natur, Technik, Wirtschaft und Gesellschaft nutzen und produktiv weiterentwickeln zu können – einer weitergehenden Reflexion, einem Monitoring und einer Minimierungsstrategie epistemischer Risiken der Opazität, die entstehen, wenn sich das von den neuen informatischen Verfahren getriebene Regime nachmoderner Wissenschaften ausweitet.

Gegen eine Erweiterung und Ergänzung des Regimes moderner Wissenschaft spricht nichts – im Gegenteil. Eine moderate Erweiterung und umsichtiger Ergänzung – institutionell und förderpolitisch abgesichert, begleitet von transparenten und diskursiv zugänglichen Prozeduren der Qualitätssicherung – kann zu neuen wissenschaftlichen Erkenntnissen sowie zu reichhaltigen gesellschaftlichen Gestaltungsoptionen, etwa in Richtung Nachhaltigkeit, führen. Allerdings ist einer möglichen Schwerpunktverlagerung in Richtung nachmoderner Wissenschaft, verbunden mit der sukzessiven Ersetzung und Eliminierung der modernen Wissenschaft vorzubeugen. Schließlich ist der Erfolg der bisherigen Wissenschafts- und Technikentwicklung eng verbunden mit der aufklärerischen und kritischen Tradition des Regimes der modernen Wissenschaft, die sich in der balancierten Koexistenz der vier (o. g.) Wissenschaftsverständnisse zeigt. Nicht nur ist eine Erkenntnis- und Theorieorientierung für moderne Wissenschaft – neben anderem – kennzeichnend; vielmehr ist sie grundlegend für mittel- und langfristige Anwendungs- und Verwertungserfolge. Nachmoderne Wissenschaft mag zwar vordergründig nützlich erscheinen und einem utilitären Zeitgeist entsprechen, doch sie allein greift zu kurz und ist nicht hinreichend grundlegend und hintergründig, auch für zukünftige Innovationen: Schließlich gehören Weltverständnis *und* Weltgestaltung zusammen –

wenn man das so traditionell formulieren mag. Moderne Wissenschaft ist als Orientierungsrahmen, Referenzsystem und Innovationsproduzent erhaltenswert.

Literatur

- Abarbanel HDI (1996) Analysis of observed chaotic data. Springer, New York
- Anderson C (2008). The end of theory: the data deluge makes the scientific method obsolete. WIRED Mag 16/07. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory. Zugegriffen am 10.12.2020
- Bacon F (1999) In: Krohn W (Hrsg) Neues Organon (1620), 2. Aufl. Meiner, Hamburg
- Bertels K, Neuberg L, Vassiliadis S et al (2001) On chaos and neural networks: the backpropagation paradigm. Artif Intell Rev 15:165–187
- Biddle JB, Kukla R (2017) The geography of epistemic risk. In: Elliott K, Richards T (Hrsg) Exploring inductive risk: case studies of values in science. Oxford University Press, Oxford, S 215–237
- Böhme G, van der Daele W (1977) Erfahrung als Programm. Über Strukturen vorparadigmatischer Wissenschaft. In: Böhme G, van der Daele W, Krohn W (Hrsg) Experimentelle Philosophie. Suhrkamp, Frankfurt, S 183–236
- Breiman L (2001) Statistical modeling: the two cultures. Stat Sci 16(3):199–231
- Bunge M (1987) Kausalität, Geschichte und Probleme. Mohr, Tübingen
- Carnap R (1936) Testability and meaning. Philos Sci 3(4):419–471
- Carrier M (2011) ‚Knowledge is power‘, or how to capture the relationship between science and technoscience. In: Nordmann A et al (Hrsg) Science transformed? Debating claims of an epochal break. University of Pittsburgh Press, Pittsburgh, S 43–53
- Cartwright N (1983) How the laws of physics lie. Clarendon Press, Oxford
- Chakraborty S et al (2017) Interpretability of deep learning models: a survey of results. In: El Baz D, Gao J, Grymes R (Hrsg) Interpretability of deep learning models: a survey of results. IEEE, San Francisco, S 1–6. <https://doi.org/10.1109/UIC-ATC.2017.8397411>
- Chandler D (2015) A world without causation: big data and the coming of age of posthumanism. Millenn J Int Stud 43(3):833–851
- Crutchfield JP (1994) The calculi of emergence: computation, dynamics and induction. Phys D 75:11–54
- Douglas HE (2000) Inductive risk and values in science. Philos Sci 67(4):559–579
- Dubois P et al (2020) Data-driven predictions of the Lorenz system. Physica D 408:132495. <https://doi.org/10.1016/j.physd.2020.132495>
- Fan H, Jiang J, Zhang C, Wang X, Lai Y (2020) Long-term prediction of chaotic systems with machine learning. Phys Rev Res 2(1):012080(R). <https://doi.org/10.1103/PhysRevResearch.2.012080>
- Foerster Hv (1995) Entdecken oder Erfinden. Wie läßt sich Verstehen verstehen? In: Foerster Hv, Glasersfeld E v et al (Hrsg) Einführung in den Konstruktivismus. Piper, München/Zürich, S 41–88
- Freeman WJ, Skarda CA (1985) Spatial EEG patterns, nonlinear dynamics and perception. Brain Res Rev 10:147–175
- Goodfellow IJ, Shlens J, Szegedy C (2015) Explaining and harnessing adversarial examples. CoRR abs/1412.6572 (2015) published as a conference paper at ICLR 201 (arXiv:1412.6572v3 [stat.ML] 20 Mar 2015)

- Gramelsberger G (2010) Computerexperimente. Zum Wandel der Wissenschaft im Zeitalter des Computers. Transcript, Bielefeld
- Hacking I (1983) Representing and intervening. Introductory topics in the philosophy of natural sciences. Cambridge University Press, Cambridge
- Haken H (1980) Dynamics of synergetic systems. Springer, Berlin
- Haken H, Schiepek G (2006) Synergetik in der Psychologie. Hogrefe, Göttingen
- Haraway D (1995) Die Neuerfindung der Natur. Primaten, Cyborgs und Frauen. Campus, Frankfurt
- Hertz H (1963) Die Prinzipien der Mechanik in neuem Zusammenhang dargestellt (1891). Wiss. Buchgesellschaft, Darmstadt
- Hubig C (2006) Die Kunst des Möglichen I. Technikphilosophie als Reflexion der Medialität. Transcript, Bielefeld
- Humphreys P (1991) Computer simulations. In: Fine A, Forbes M, Wessels L (Hrsg) PSA 1990, Bd 2. Publication of the Philosophy of Science Association, East Lansing, S 497–506
- Humphreys P (2004) Extending ourselves. Computational science, empiricism, and scientific method. Oxford University Press, Oxford/New York
- Jackson EA (1989) Perspectives of nonlinear dynamics, Bd 1. Cambridge University Press, Cambridge
- Janich P (1997) Kleine Philosophie der Naturwissenschaften. Beck, München
- Kaminski A (2014) Lernende Maschinen: naturalisiert, transklassisch, nichttrivial? In: Kaminski A, Gelhard A (Hrsg) Zur Philosophie informeller Technisierung. Wiss. Buchgesellschaft, Darmstadt, S 57–80
- Kaminski A, Resch M, Küster U (2018) Mathematische Opazität. Über Rechtfertigung und Reproduzierbarkeit in der Computersimulation. Jahrbuch Technikphilosophie 4:253–278
- Kant I (1989) Kritik der reinen Vernunft (1781/1787). Reclam, Stuttgart
- Karpatne A et al (2019) Machine learning for the geosciences: challenges and opportunities. IEEE Trans Knowl Data Eng 31(8):1544–1554
- Kitano H (2016) Artificial intelligence to win the Nobel Prize and beyond: creating the engine for scientific discovery. AI Mag 37(1):39–49
- Kitcher P, Salmon W (Hrsg) (1989) Scientific explanation (companion volume to Minnesota studies in the philosophy of science 14). University of Minnesota Press, Minnesota
- Küppers G, Lenard J (2005) Computersimulationen. Modellierungen 2. Ordnung. J Gen Philos Sci 36:305–329.
- Latour B (1987) Science in action. Harvard University Press, Cambridge
- Lengauer T (2019) Data Analysis in Medicine – from the bench to the bedside. Eingeladener Vortrag, 2. Juli 2019. Institut für qualifizierende Innovationsforschung und -beratung, Bad Neuenahr-Ahrweiler
- Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20:130–141
- Lorenz EN (1989) Computational chaos – a prelude to computational instability. Phys D 35:299–317
- Luhmann N (2003) Soziologie des Risikos (1991). de Gruyter, Berlin
- Mach E (1917) Erkenntnis und Irrtum. Skizzen zur Psychologie der Forschung, 3. Aufl. Barth, Leipzig
- Mainzer K (1996) Thinking in complexity. In: The complex dynamics of matter, mind, and mankind. Springer, Heidelberg
- Mainzer K (2014) Die Berechnung der Welt. Von der Weltformel zu Big Data. Beck, München
- Mantegna RN, Stanley HE (2000) An introduction to econophysics: correlations and complexity in finance. Cambridge University Press, Cambridge
- McCauley JL (2004) Dynamics of markets: econophysics and finance. Cambridge University Press, Cambridge

- Mittelstraß J (1998) *Die Häuser des Wissens*. Suhrkamp, Frankfurt
- Moon FC (Hrsg) (1998) *Dynamics and chaos in manufacturing processes*. Wiley, New York
- Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B (2019) Definitions, methods, and applications in interpretable machine learning. *PNAS* 116(44):22071–22080. (29 Oct 2019)
- Nassehi A (2019) *Muster. Theorie der digitalen Gesellschaft*. Beck, München
- Nordmann A (2008) *Philosophy of NanoTechnoScience*. In: Fuchs HH et al (Hrsg) *Nanotechnology*. Wiley, Weinheim, S 217–244
- Nordmann A, Radder H, Schiemann G (Hrsg) (2011) *Science transformed? Debating claims of an epochal break*. University of Pittsburgh Press, Pittsburg
- Pathak J, Hunt B, Girvan M, Lu Z, Ott E (2018a) Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir 570 computing approach. *Phys Rev Lett* 120(2):024 102
- Pathak J, Wikner A, Fussell R, Chandra S, Hunt BR, Girvan M, Ott E (2018b) Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos* 28:041101
- Pietsch W (2013) *Big data – the new science of complexity*. Unpublished
- Pietsch W (2021) *Big data (elements in the philosophy of science)*. Cambridge University Press, Cambridge
- Popper KR (1989) *Logik der Forschung* (1934). Mohr, Tübingen
- Prigogine I (1992) *Vom Sein zum Werden. Zeit und Komplexität in den Naturwissenschaften* (1979). Piper, München
- Rhawan I et al (2019) Machine behaviour. *Nature* 568:477–486
- Roco MC, Bainbridge WS (Hrsg) (2002) *Converging technologies for improving human performance. Nanotechnology, biotechnology, information technology, and cognitive science*. National Science Foundation, Arlington
- Sauer T, Yorke JA, Casdagli M (1991) Embedology. *J Stat Phys* 77(3/4):579–616
- Schmidt JC (2003) Zwischen Berechenbarkeit und Nichtberechenbarkeit. Die Thematisierung der Berechenbarkeit in der aktuellen Physik komplexer Systeme. *J Gen Philos Sci* 34:99–131
- Schmidt JC (2004) Unbounded technologies. Working through the technological reductionism of nanotechnology. In: Baird D et al (Hrsg) *Discovering the nanoscale*. IOS, Amsterdam, S 25–51
- Schmidt JC (2008) *Instabilität in Natur und Wissenschaft. Eine Wissenschaftsphilosophie der nachmodernen Physik*. de Gruyter, Berlin
- Schmidt JC (2015) *Das Andere der Natur. Neue Wege zur Naturphilosophie*. Hirzel, Stuttgart
- Schmidt JC (2016) Philosophy of late-modern technology. Towards a clarification and classification of synthetic biology. In: Boldt J (Hrsg) *Synthetic biology. Metaphors, worldviews, ethics, and law*. Springer VS, Wiesbaden/Heidelberg, S 13–30
- Smith G (2018) *The AI delusion*. Oxford University Press, Oxford
- Stegmüller W (1987) *Hauptströmungen der Gegenwartsphilosophie*, Bd 2. Kröner, Stuttgart
- Storch Hv (2019) Künstliche Intelligenz & Big Data in der Klimaforschung – Versuch einer Einschätzung. Eingeladener Vortrag, 02.07.2019. Institut für qualifizierende Innovationsforschung und -beratung, Bad Neuenahr-Ahrweiler
- Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow IJ, Fergus R (2014) Intriguing properties of neural networks. *ICLR*, abs/1312.6199. <http://arxiv.org/abs/1312.6199>. Zugegriffen am 07.09.2021
- Takens F (1985) Distinguishing deterministic and random systems. In: Barenblatt GI, Ioss G, Joseph D (Hrsg) *Nonlinear dynamics and turbulence*. Pitman, Boston, S 315–333
- Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD (1992) Testing for nonlinearity in time series: the method of surrogate data. *Phys D* 58:77–94
- Weinan E (2017) A proposal on machine learning via dynamical systems. *Commun Math Stat* 5:1–11

Weizsäcker CFv (1974) Die Einheit der Natur (1971). dtv, München

Winsberg E (2010) Science in the age of computer simulation. University of Chicago Press, Chicago/London

Wright GHv (1991) Erklären und Verstehen (1971). Athenäum, Meisenheim

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.