

Carl Friedrich Gethmann · Peter Buxmann
Julia Distelrath · Bernhard G. Humm
Stephan Lingner · Verena Nitsch
Jan C. Schmidt
Indra Spiecker genannt Döhmann

Künstliche Intelligenz in der Forschung

Neue Möglichkeiten und
Herausforderungen für die Wissenschaft

OPEN ACCESS

 Springer

Ethics of Science and Technology Assessment

Band 48

Reihe herausgegeben von

Carl Friedrich Gethmann, Universität Siegen, Siegen, Deutschland

Michael Quante, Philosophisches Seminar, Westfälische Wilhelms Universität
Münster, Deutschland

Bjoern Niehaves, Universitaet Siegen, Siegen, Deutschland

Holger Schönherr, Department of Chemistry and Biology, Universität Siegen
Siegen, Deutschland

The series *Ethics of Science and Technology Assessment* focuses on the impact that scientific and technological advances have on individuals, their social lives, and on the natural environment. Its goal is to cover the field of Science and Technologies Studies (STS), without being limited to it. The series welcomes scientific and philosophical reviews on questions, consequences and challenges entailed by the nature and practices of science and technology, as well as original essays on the impact and role of scientific advances, technological research and research ethics. Volumes published in the series include monographs and edited books based on the results of interdisciplinary research projects. Books that are devoted to supporting education at the graduate and post-graduate levels are especially welcome.

Weitere Bände in dieser Reihe: <http://www.springer.com/series/4094>

Carl Friedrich Gethmann • Peter Buxmann
Julia Distelrath • Bernhard G. Humm
Stephan Lingner • Verena Nitsch
Jan C. Schmidt • Indra Spiecker
genannt Döhmann

Künstliche Intelligenz in der Forschung

Neue Möglichkeiten und Herausforderungen
für die Wissenschaft



Springer

Carl Friedrich Gethmann
Universität Siegen
Siegen, Deutschland

Julia Distelrath
Schalkenbach, Deutschland

Stephan Lingner
Institut für qualifizierende
Innovationsforschung und -beratung
GmbH (IQIB)
Bad Neuenahr-Ahrweiler, Deutschland

Jan C. Schmidt
Department of Social Sciences
Hochschule Darmstadt
Darmstadt, Deutschland

Peter Buxmann
Fachgebiet Wirtschaftsinformatik
Technische Universität Darmstadt
Darmstadt, Deutschland

Bernhard G. Humm
Hochschule Darmstadt
Darmstadt, Deutschland

Verena Nitsch
Institut für Arbeitswissenschaft
RWTH Aachen
Aachen, Deutschland

Indra Spiecker genannt Döhmann
Goethe Universität Frankfurt a.M.
Frankfurt am Main, Deutschland



ISSN 1860-4803

ISSN 1860-4811 (electronic)

Ethics of Science and Technology Assessment

ISBN 978-3-662-63448-6

ISBN 978-3-662-63449-3 (eBook)

<https://doi.org/10.1007/978-3-662-63449-3>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer

© Der/die Herausgeber bzw. der/die Autor(en) 2022

Dieses Buch ist eine Open-Access-Publikation.

Open Access Dieses Buch 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 Buch 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.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung: Dieter Merkle

Springer ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

Vorwort

Neue Entwicklungen im Bereich der künstlichen Intelligenz (KI) und ihre zunehmende Verbreitung in zahlreichen Anwendungskontexten des täglichen Lebens lassen hier noch teilweise unbekannte Chancen aber auch Risiken für die Nutzer/ Nutzerinnen erwarten. Viele KI-Systeme versuchen dabei die Netzwerkarchitekturen des menschlichen Gehirns zu imitieren und dadurch das sog. „maschinelle Lernen“ zu ermöglichen. Die Komplexität, Lernfähigkeit und Autonomie dieser Systeme haben zur Folge, dass ihre internen Prozesse schwer nachvollziehbar und noch schwerer kontrollierbar sind. Die spezifische Intransparenz und Autonomie dieser KI-Systeme könnte – trotz ihres an sich vielversprechenden Einsatzes – insbesondere in der Forschung problematisch werden, da diese Eigenschaften mit der Transparenzforderung klassischer Peer review-Verfahren und der Selbstbestimmtheiterwartung der Forscher/Forscherinnen kollidieren könnten. Diese und weitere hiermit zusammenhängende Fragen waren im Rahmen eines Forschungsprojekts zu klären, über dessen Ergebnisse hier nun berichtet wird.

Zu diesem Zweck wurde am IQIB (Institut für qualifizierende Innovationsforschung und -beratung) die interdisziplinäre Projektgruppe „Digitale Arbeitswelten in Forschung und Entwicklung. Neue Möglichkeiten und Herausforderungen für die Wissenschaft (DiA)“ eingerichtet, die im Jahr 2018 ihre Arbeit aufnahm. Der Projektgruppe gehörten folgende Personen an: Professor Dr. Peter Buxmann (TU Darmstadt), Julia Distelrath (IQIB), Professor Dr. Dr. h.c. Carl Friedrich Gethmann (Universität Siegen), Professor Dr. Bernhard G. Humm (Hochschule Darmstadt), Dr. Stephan Lingner (IQIB), Professor Dr.-Ing. Verena Nitsch (RWTH Aachen), Professor Dr. Jan C. Schmidt (Hochschule Darmstadt) und Professor Dr. Indra Spiecker genannt Döhmann (Universität Frankfurt/M.). Zeitweise nahm auch Dr. Claudio Zettel (DLR-Projektträger, Bonn) an den gemeinsamen Diskussionen der Arbeitsgruppe teil. Unter dem Vorsitz von Carl Friedrich Gethmann brachten die Mitglieder dieser Projektgruppe ihre spezifischen Fachkompetenzen aus den Arbeitsgebieten Informatik, Wissenschafts- und Technikphilosophie, Ethik, Arbeitswissenschaften, Rechtswissenschaften, Technikfolgenabschätzung und Komplexitätstheorie in die Projektgruppe ein. Das Projekt DiA wurde von Stephan Lingner und Julia Distelrath koordiniert.

Die interdisziplinäre Projektgruppe traf sich über drei Jahre hin in ein- bis zweimonatigen Abständen zur Erörterung und Diskussion der relevanten Perspektiven zum Thema. Dieses Verfahren der iterativen Annäherung multidisziplinärer Perspektiven zu einem interdisziplinären Ganzen wurde am Vorläufer des IQIB, der Europäischen Akademie entwickelt und in diesem Vorhaben gleichfalls angewandt.¹ Die fortlaufende Reflexionsarbeit der Gruppe mündete schließlich in eine integrierte Beurteilung des vielschichtigen Forschungsgegenstands, die mit dieser Publikation vorliegt. Die Projektergebnisse profitierten maßgeblich von einer begleitenden, von der RWTH Aachen 2019 koordinierten Umfrage zu den Auswirkungen von KI und Digitalisierung auf die Arbeit und Arbeitsbedingungen von Forschern/Forscherinnen, die von Vera Rick (RWTH Aachen) sowie den Projektgruppenmitgliedern Julia Distelrath und Verena Nitsch durchgeführt wurde.

Über die Projektlaufzeit wurde der jeweilige Forschungsstand im Rahmen mehrerer Fachgespräche mit weiteren externen Experten aus der Informatik, den Technikwissenschaften, der Wissenschaftstheorie, der Medizin, den Biowissenschaften, der Physik, und der Klimaforschung diskutiert. Die Projektgruppe dankt Dr.-Ing. Mark Azzam (Deutsches Zentrum für Luft- und Raumfahrt, Köln), Dr. Nils Goerke (Universität Bonn), Professor Dr. Karl-Heinz Jöckel (Universitätsklinikum Essen), Professor Dr. Maik Kschischo (Hochschule Koblenz), Professor Dr. Thomas Lengauer (Max-Planck-Institut für Informatik, Saarbrücken und Universität Köln), Professor Dr. Peter Mättig (Universität Bonn), Professor Dr. Klaus Mainzer (Technische Universität München) und Professor Dr. Hans von Storch (Helmholtz-Zentrum Geesthacht) für ihre Diskussionsbeiträge, die der weiteren Arbeit im Projekt wichtige Impulse gaben.

An dieser Stelle möchte die Arbeitsgruppe auch Petra Prenzel (IQIB), Felicitas Schlitz (IQIB) und Oliver Vetter (TU Darmstadt) für ihre engagierte organisatorische, technische und wissenschaftliche Assistenz im Projekt danken.

Schließlich sei auch dem Deutschen Zentrum für Luft- und Raumfahrt (DLR) gedankt, das das Projekt DiA im Rahmen eines von Dr. Mark Azzam (DLR Think Tank) veranlassten Forschungsauftrags finanziell unterstützt hat. Die vorliegende

Der Text des vorliegenden Bandes wurde von der Arbeitsgruppe „DiA – Digitale Arbeitswelten in Forschung und Entwicklung“ erstellt, diskutiert und integriert. Die Ergebnisse dieser Studie geben die kollektiven Perspektiven und Positionen der Arbeitsgruppenmitglieder wieder. Aus praktischen Gründen wurden die Hauptverantwortlichkeiten für die einzelnen Kapitel wie folgt festgelegt:

- Empfehlungen: Alle Autoren/Autorinnen
- Kap. 1: Alle Autoren/Autorinnen
- Kap. 2: Bernhard G. Humm, Peter Buxmann und Jan C. Schmidt
- Kap. 3: Carl Friedrich Gethmann
- Kap. 4: Jan C. Schmidt
- Kap. 5: Verena Nitsch und Peter Buxmann
- Kap. 6: Indra Spiecker genannt Döhmann
- Kap. 7: Alle Autoren/Autorinnen

¹ Diese Arbeitsform ist genauer dargestellt in: Gethmann, C.F., Carrier, M., Hanekamp, G., Kaiser, M., Kamp, G., Lingner, S., Quante, M., Thiele, F. 2015. *Interdisciplinary Research and Transdisciplinary Validity Claims*, Heidelberg u. a.: Springer.

Publikation wurde dankenswerterweise aus dem Open Access-Publikationsfonds der Hochschule Darmstadt gefördert.

Der vorliegende Band spiegelt den fachlichen Wissensstand vom Januar 2021 wider.

Siegen, Deutschland
Bad Neuenahr-Ahrweiler, Deutschland
(Monat) 2021

Carl Friedrich Gethmann
Stephan Lingner

Empfehlungen für den Umgang mit KI in der wissenschaftlichen Forschung

Carl Friedrich Gethmann, Peter Buxmann, Julia Distelrath, Bernhard G. Humm, Stephan Lingner, Verena Nitsch, Jan C. Schmidt, Indra Spiecker genannt Döhmann

Die Digitalisierung von Arbeitswelten erreicht alle Bereiche von Forschung und Entwicklung. Insbesondere lassen neuartige Entwicklungen der künstlichen Intelligenz (KI) hier große Veränderungen im Forschungsalltag erwarten. Dennoch ist wissenschaftliche Forschung auch im Zeitalter von KI, Machine Learning (ML) und Big Data weiterhin auf den Menschen als Wissenschaftler/Wissenschaftlerin und Forscher/Forscherin angewiesen. Allerdings führt die Selbstorganisationsfähigkeit von KI-Systemen im Modus des Machine Learning zu einer Intransparenz von Forschungsprozessen, die ihre Kontrolle, Validierung und Anerkennung erschweren. Es ist daher damit zu rechnen, dass Vertrauensdefizite von entsprechenden KI-basierten Forschungsleistungen auftreten können, soweit ihnen nicht durch geeignete neuartige Validierungsverfahren entgegengewirkt werden kann. Auf der anderen Seite wird die besondere Leistungsfähigkeit von KI-Systemen neue, bislang kaum zugängliche Forschungsfelder erschließbar machen. Dadurch, dass sich KI-Systeme oft im Nachhinein als „multi-purpose tools“ erweisen, eröffnen sich nicht-vorhersehbare, auch disruptive Nutzungskontexte mit großer Tragweite für die Anwenderseite. Im Forschungsalltag wird dies zu Anpassungsanforderungen an die Forscher/Forscherinnen führen. Schließlich sind Forschungsevaluation und Datenschutz bei Anwendung datenintensiver KI-Methoden in besonderer Weise herausgefordert.

Die nachfolgenden Empfehlungen richten sich an die Forschung, die Arbeitswelt und die Regulierung von KI in der Wissenschaft. Sie wurden nach interdisziplinärer Reflexion der Beiträge zu den Kap. 2 bis 6 erarbeitet und sind im Schlusskapitel 7 ausführlicher begründet.

Vorschläge für die wissenschaftliche Forschung

Die Gestaltung des forschenden Wissenschaftssystems sollte sich auch im Zeitalter von KI weiterhin an bewährten Qualitätsstandards moderner Forschung orientieren.

Dabei soll KI-basierte Forschung unter Bedingungen erfolgen, unter denen sie ihre wissenschaftlichen Geltungsansprüche hinreichend begründet, ihre Prämissen offenlegt und ihre Resultate im wissenschaftlichen und ggf. auch im weiteren gesellschaftlichen Diskurs validiert. Nur so kann eine wissenschaftliche Robustheit (und ggf. gesellschaftliche Relevanz) von Ergebnissen KI-basierter Forschung erreicht und sichergestellt werden. Damit wissenschaftliche Ergebnisse nachvollziehbar sind, sollen die zugrunde liegenden Daten zugänglich sein. Des Weiteren sollen die eingesetzten Berechnungsverfahren benannt werden.

Hierzu sind insbesondere geeignete Vertrauen schaffende Regularien zu implementieren.

Sie sollten zunächst darauf gerichtet sein, die technisch-induzierte Intransparenz von KI-Systemen und ihrer Beiträge zu entsprechenden Forschungsleistungen weitgehend zu reduzieren. Weiterhin sollten die den in der Forschung eingesetzten KI-Systemen zugrunde liegenden wissenschaftlich-technischen Normen offenlegt werden, um z. B. mögliche technisch-bedingte Verzerrungen (Biases) von Forschungsergebnissen besser erkennbar zu machen. Schließlich sind auch flankierende Maßnahmen zur Offenlegung (und ggf. Legitimation) des jeweiligen Erkenntnisinteresses an KI-basierter Forschung zu ergreifen, um die Resultate in Wissenschaft und Gesellschaft besser einbetten zu können.

Maßnahmen zur Offenlegung von wissenschaftlich-technischen Normen eingesetzter KI-Systeme sollten möglichst frühzeitig implementiert werden.

Sie sind darauf zu richten, die Forscher/Forscherinnen für epistemische Risiken, die mit der partiellen Intransparenz, Eigendynamik und Unkontrollierbarkeit von KI, insbesondere im Modus des Machine Learning verbunden sind, rechtzeitig zu sensibilisieren und in ihre wissenschaftlichen Interpretationen der Gesamtergebnisse einfließen zu lassen. So wird vermieden, dass sich im Umgang mit KI in der Forschung eine „Sorglosigkeit“ herausbildet und verfestigt. Den Forschenden muss frühzeitig bewusst werden, dass KI im Modus des Machine Learning teilweise undurchschaubar ist und als Forschungsinstrument nur bedingt vertraut werden kann.

Vor diesem Hintergrund sollte das Forschungsfeld der so genannten „erklärbaren KI“ (engl. Explainable AI (XAI)) verstärkt gefördert werden, das darauf abzielt, die Intransparenz von KI-Systemen zumindest teilweise zu reduzieren und Nachvollziehbarkeit und Interpretierbarkeit zu ermöglichen.

Forschungsprojekte, in denen KI-Systeme eingesetzt werden, sollten daher auch stets personelle Informatik-Kompetenz vorhalten.

Diese Forderung ist darauf gerichtet, den wissenschaftlichen Risiken etwaiger naiver KI-Verwendung zu begegnen. Darüber hinaus sind in den jeweiligen Forschungskontexten „Gütemaße“ ggf. einzusetzender KI-Systeme festzulegen, die den Umfang ihres jeweiligen Einsatzes für projektspezifische Aufgaben sinnvollerweise bestimmen bzw. begrenzen. Auch hierfür wird informationstechnische Beratungskompetenz benötigt.

Forscher/Forscherinnen sollten sich auch im Zeitalter von KI weiterhin als Urheber des Forschungshandelns begreifen.

Sie bleiben trotz sog. „autonomer“ informationstechnischer Teilprozesse treibende Kraft in der Forschung – nicht zuletzt wegen der besonderen Reflexionsanforderungen zuweilen intransparenter KI-basierter Verfahren und im Interesse zustimmungsfähiger Wissenschaft.

Vorschläge für die wissenschaftliche Ausbildung und Weiterbildung

Heutigen und zukünftigen Forschern/Forscherinnen sollten verpflichtende Aus- und Weiterbildungsangebote mit KI-Bezug gemacht werden.

Sie sollen ihnen die notwendigen theoretischen und praktischen Grundkompetenzen KI-basierter Forschung inklusive entsprechender Erfahrungen und Fertigkeiten vermitteln. Hierfür sind frühzeitig entsprechende Curricula zu entwickeln oder zu ergänzen und ggf. um nicht mehr im Forschungsalltag benötigte Kompetenzen zu verschlanken.

Diese Bildungsmaßnahmen sollen nicht nur das technische und das praktische Verständnis über KI-Systeme in der Forschung stärken, sondern auch dazu beitragen, den „KI-Hype“ auf die Sachebene hin zu entmystifizieren.

Diese Maßnahmen sollen dabei die Reflexivität und Urteilsfähigkeit von Forschenden in neuartigen „kollaborativen“ KI-Umgebungen ermöglichen bzw. erhalten. Dabei sind neue wissenschaftliche Risiken aber auch neue erkenntnisbildende Chancen durch KI-basierte Forschung genauso zu vermitteln, wie auch herkömmliche Methoden, die im Zeitalter von KI mehr an Bedeutung gewinnen könnten (Statistik, Wahrscheinlichkeitstheorie etc.). Ferner sind die Forscher/Forscherinnen in diesem Zusammenhang weiterhin auf die bewährten wissenschaftskonstituierenden Normen, wie die Sicherung der Falsifizierbarkeit und Reproduzierbarkeit ihrer Forschungsergebnisse hin zu sensibilisieren und zu verpflichten.

Vorschläge an die Regulierung von KI-basierter Forschung

Die Verfahren der Forschungsevaluation sollten auch im Zeitalter von KI und Big Data einen fairen Wettbewerb zwischen den unterschiedlichen Fachkulturen erlauben.

Insbesondere sollte damit wenig datengenerierenden Disziplinen entgegengekommen werden, die andernfalls z. B. im Wettbewerb um Fördermittel benachteiligt sein könnten. Hierzu sind einschlägige wissenschaftliche Begutteilungskriterien ggf. anzupassen oder zu ergänzen.

Für öffentlich finanzierte, datenintensive Forschungsvorhaben sind klarere Regelungen der Europäischen Datenschutz-Grundverordnung wünschbar.

Insbesondere sollten in dem Zusammenhang das Minimierungsgebot und die Zweckbindung legitimer Datenverwendung eindeutiger und forschungsfreundlicher gestaltet werden, um auch dem explorativen Forschungsmodus von KI in Verbindung mit Big Data gerecht zu werden.

Recommendations for dealing with AI in scientific research

Carl Friedrich Gethmann, Peter Buxmann, Julia Distelrath, Bernhard G. Humm, Stephan Lingner, Verena Nitsch, Jan C. Schmidt, Indra Spiecker genannt Döhmann

The digitization of working environments is reaching all areas of research and development. In particular, new developments in artificial intelligence (AI) might lead to major changes in everyday research. Nevertheless, scientific research continues to rely on humans as scientists and researchers, even in the age of AI, machine learning (ML) and big data. However, the ability of AI systems to organize themselves in machine learning mode leads to a lack of transparency in research processes, which makes their control, validation and universal acceptance more difficult. It might be therefore expected that confidence deficits in corresponding AI-based research services will arise, unless they can be counteracted by suitable new validation procedures. On the other hand, the special performance of AI systems will enable to explore previously barely accessible research fields. Because AI systems often turn out to be “multi-purpose tools” in retrospect, unpredictable, even disruptive usage contexts open up with great importance for the users. In every-day research, AI will lead to significant adaptation requirements for researchers. Finally, research evaluation and data protection are particularly challenged when using data-intensive AI methods.

The following recommendations are aimed at research, the world of work and the regulation of AI in science. They were developed after interdisciplinary reflection on the contributions to Chaps. 2–6 and are explained in more detail in the final Chap. 7.

Suggestions for Scientific Research

Shaping the research-based science system in the age of AI should still continue to comply with the proven quality standards of modern research.

AI-based research should take place under those conditions under which it (a) adequately justifies its scientific validity claims, (b) reveals its premises and (c) validates its results in scientific and, if necessary, in further societal discourses. This is the only way to achieve and ensure a scientific robustness (and possibly social relevance) of the results of AI-based research. The underlying data should be made accessible so that scientific results are better comprehensible. Furthermore, the calculation methods used should be explicated.

In particular, suitable trust-building regulations must be implemented for this purpose.

These regulations should initially aim to largely reduce the technically-induced lack of transparency of AI systems and of corresponding research contributions. Furthermore, the underlying scientific and technical standards of the applied AI systems should be disclosed. This might – for example – enable to uncover any technically-related biases of research results. Finally, accompanying measures must also be taken to disclose and legitimize the respective cognitive interests in AI-based research in order to better embed the results in science and society.

Measures for the disclosure of scientific and technical standards of the AI systems used should be implemented as early as possible.

The measures should be directed towards sensitizing the researchers timely to those epistemic risks which are associated with the partial lack of transparency, inherent dynamics and uncontrollability of AI, especially in machine learning mode. This claim would allow for more sound scientific interpretations of overall AI-derived research results. It prevents also from the emergence and continuation of “careless” use of AI in research. Researchers must be timely aware about the fact that AI in machine learning mode is often not transparent and that it can only be trusted to a limited extent as a research tool. Against this background, the research field of the so-called “explainable AI” (XAI) should be increasingly supported,

which aims to at least partially reduce the inherent intransparency of corresponding AI systems while enabling better traceability and interpretability of results.

Research projects in which AI systems are used should therefore always include sufficient IT expertise at hand.

This requirement is aimed at countering the scientific risks of any naive AI use. In addition, “quality measures” of any AI systems to be used are to be specified in the respective research contexts, which sensibly determine or limit the scope of their respective use for project-specific tasks. Again, advisory skills in information technology are also required for this aim.

Even in the age of AI, researchers should continue to see themselves as the responsible originators of research.

Researchers remain the driving force in research, despite of any underlying “autonomous” IT sub-processes. This holds true not least because of the special reflection requirements of sometimes intransparent AI-based processes as well as in the interest of reliable and acceptable science.

Suggestions for scientific training and further education

Current and future researchers should be offered compulsory training and further education courses related to AI.

The courses should provide them with the necessary basic theoretical and practical competencies for AI-assisted research, while teaching them relevant experience and skills. For this purpose, appropriate curricula must be timely developed or supplemented. They also have to be streamlined with regard to skills that are no longer required in everyday research, if necessary.

These educational measures should not only strengthen the technical and practical understanding of AI systems in research, but also help to demystify the “AI hype” down to the factual level.

These measures are intended to enable or maintain the reflexivity and judgment of researchers in novel “collaborative” AI environments. New scientific risks but

also new knowledge-generating opportunities through AI-based research must be conveyed as well as conventional methods that could become more important in the age of AI (statistics, probability theory, etc.). Furthermore, in this context, the researchers must still be made aware of and committed to the tried and tested norms constituting science, such as ensuring the falsifiability and reproducibility of their research results.

Suggestions for the regulation of AI-based research

Research evaluation procedures should allow for fair competition between the different disciplinary cultures, even in the age of AI and big data.

In particular, barely data-generating disciplines should be met, which could otherwise be disadvantaged, e.g. in the competition for funding in the era of big data. For this purpose, relevant scientific assessment criteria might have to be adapted or supplemented.

The regulations of the European General Data Protection Regulation should be made clearer for publicly financed, data-intensive research projects.

In particular, the minimization requirement and the purpose limitation of legitimate data use should be made clearer and more research-friendly in order to meet the requirements of the exploratory research mode of AI in connection with big data.

Inhaltsverzeichnis

1	Einführung	1
	Carl Friedrich Gethmann, Peter Buxmann, Julia Distelrath, Bernhard G. Humm, Stephan Lingner, Verena Nitsch, Jan C. Schmidt und Indra Spiecker genannt Döhmann	
1.1	Fragestellung	2
1.2	Künstliche Intelligenz: eine Einordnung	5
1.2.1	Ein Thema von großer öffentlicher Aufmerksamkeit	5
1.2.2	KI im Kontext von Automatisierung und Digitalisierung	7
1.2.3	Definition von KI	8
1.3	Wissenschaftliches Arbeiten als Teilbereich der Wissensarbeit	9
1.4	KI in wissenschaftlicher Forschung	9
	Literatur	10
2	Grundlagen und Anwendungen von KI	13
	Bernhard G. Humm, Peter Buxmann und Jan C. Schmidt	
2.1	Historie von KI	13
2.2	Grundlagen der KI	15
2.2.1	Die Landkarte der KI	15
2.2.2	Machine Learning	18
2.2.3	Wissensbasierte KI	24
2.2.4	Zusammenspiel von wissensbasierter KI und Machine Learning	28
2.2.5	Die Rolle von Algorithmen in KI-Systemen	28
2.3	Anwendungen von KI	29
2.3.1	Teilchenphysik	29
2.3.2	Klimaforschung	33
2.3.3	Medizin	37
2.4	Fazit	39
	Literatur	40

3	Zur Frage der Ersetzbarkeit des Menschen durch KI in der Forschung	43
	Carl Friedrich Gethmann	
3.1	Einleitung	43
3.2	KI und menschliche Wissensbildung	44
3.2.1	Intelligenz	44
3.2.2	Denken können	46
3.2.3	Zusammenfassung	49
3.3	Künstliche Agenten und menschliche Handlungsurheberschaft	49
3.3.1	Handlungsurheberschaft und Zurechenbarkeit	50
3.3.2	Kausale Geschlossenheit	51
3.3.3	Merkmale wissenschaftlicher Intelligenz	52
3.3.4	Behaupten und Begründen	53
3.3.5	Wahrheit als Wohlbegründetheit	54
3.3.6	Zusammenfassung	58
3.4	Das wissenschaftliche Bacon-Projekt und die KI	58
3.4.1	Kontemplation versus Intervention	59
3.4.2	Poietisches (technisches) und praktisches Wissen	60
3.4.3	Zweck und Nutzen wissenschaftlichen Wissens	61
3.4.4	Kausalität und Korrelation	61
3.4.5	Die Bedeutung von Big Data für die Wissenschaft	64
3.4.6	Zusammenfassung	64
3.5	Deterministische und probabilistische Voraussagen	65
3.5.1	Erklären und Voraussagen	65
3.5.2	Fehlschlüsse: Stratifizierung	66
3.5.3	Beispiel: evidenzbasierte Medizin	67
3.5.4	Zusammenfassung	68
3.6	Methodische Verfahren und heuristische Urteilkraft	68
3.6.1	Falsifizierbarkeit und Falsifikation	69
3.6.2	Grenzen und Schwierigkeiten des Verfahrens der Falsifikation	70
3.6.3	Unterbestimmtheit wissenschaftlicher Theorien	71
3.6.4	Theoriendynamik und die epistemische Funktion von Wissenschaftlergemeinschaften	72
3.6.5	Zusammenfassung: Die Unverzichtbarkeit menschlicher Erkenntnissubjekte	74
3.7	Fazit	75
	Literatur	75
4	Wandel und Kontinuität von Wissenschaft durch KI. Zur aktuellen Veränderung des Wissenschafts- und Technikverständnisses	79
	Jan C. Schmidt	
4.1	Einleitung	79
4.2	Vier leitende Wissenschaftsverständnisse – das Referenzsystem	81
4.2.1	Wissenschaft als theoriebasiertes Prognoseverfahren	81

4.2.2	Wissenschaft als Experimentierhandeln und Interventionsform zur Herstellung von Reproduzierbarkeit.	83
4.2.3	Wissenschaft als stringente Prüfmethode und als kritischer Garant intersubjektiver Geltung.	84
4.2.4	Wissenschaft als Grundlage des Weltverständnisses und als Erklärungslieferant.	86
4.2.5	Zwischenfazit: Vielheit und Einheit wissenschaftlicher Disziplinen.	87
4.3	KI und Machine Learning als Katalysator eines Wandels der vier Wissenschaftsverständnisse?	88
4.3.1	Prognostizieren ohne Wissen: KI stärkt das prognoseorientierte Wissenschaftsverständnis – ohne Gegenstandswissen	88
4.3.2	Handlungsermöglichung ohne Experiment: KI folgt keinem experimentbasiertem Wissenschaftsverständnis – verstärkt dennoch Handlungsoptionen	92
4.3.3	Prüfen auf komplementären Wegen: KI relativiert das Ideal des testbasierten Wissenschaftsverständnisses – ermöglicht aber neue phänomenologische Prüfformen	94
4.3.4	Kausalität ohne Theorie: KI folgt keinem erklärungsbezogenem Wissenschaftsverständnis – verbleibt dennoch im Horizont von Kausalität	97
4.4	Von den Verfahren und Methoden zu den Objekten: Schwache Regelhaftigkeit nutzen – komplexe Objektsysteme werden zugänglich	99
4.5	Die Black Box und ihre Intransparenz.	104
4.6	Auf dem Wege zu einer nachmodernen Technik – verwendet auch in der Wissenschaft	110
4.7	Techniktrends: Ermöglichungstechnologie, Konvergenztechnologie, Technoscience	115
4.8	Fazit	119
	Literatur.	122
5	Auswirkungen von Digitalisierung und KI auf die wissenschaftliche Arbeit	127
	Verena Nitsch und Peter Buxmann	
5.1	Wissenschaftliche Arbeit als Teilbereich der Wissensarbeit	128
5.2	Auswirkungen der Digitalisierung auf die Arbeitsbedingungen von wissenschaftlich Arbeitenden	129
5.3	KI und Automation der wissenschaftlichen Arbeit	131
5.4	Empirische Befunde zu den Auswirkungen von KI und Digitalisierung auf die Arbeit und Arbeitsbedingungen von Forschenden	136

5.4.1	Auswirkungen der Digitalisierung auf den Forschungsalltag	136
5.4.2	Angaben zum Einsatz von KI-basierten Technologien	139
5.5	Fazit	143
	Literatur.	144
6	Die Regulierungsperspektive von KI/BigData in der Wissenschaft...	147
	Indra Spiecker genannt Döhmann	
6.1	Künstliche Intelligenz und BigData im Spannungsfeld von Wissenschaftsfreiheit und Forschungsevaluation	147
6.2	Regulierungsansätze von Künstlicher Intelligenz	149
6.3	Der Einsatz Künstlicher Intelligenz im allgemeinen Spannungsfeld von Wissenschaftsfreiheit und Datenschutz: Zwei Konstellationen	152
6.4	Der Einsatz Künstlicher Intelligenz als Mittel der Forschung.	154
6.4.1	Der grundsätzliche Konflikt zwischen Datenschutzrecht und Forschungsfreiheit.	154
6.4.2	Wissenschaftsfreiheit	155
6.4.3	Datenschutzrecht	155
6.4.4	Das Austarieren von Datenschutz und Wissenschaftsfreiheit nach dem Verständnis der DSGVO sowie des deutschen Rechts.	161
6.4.5	Die Lockerung der Zweckbindung in Art. 5 Abs. 1 lit. b 2. HS DSGVO	165
6.5	Datenschutzrecht und sonstige Anforderungen bei Kooperationen	167
6.6	Fazit und Ausblick	168
	Literatur.	170
7	Kurzfassung und Schlussfolgerungen	173
	Carl Friedrich Gethmann, Peter Buxmann, Julia Distelrath, Bernhard G. Humm, Stephan Lingner, Verena Nitsch, Jan C. Schmidt und Indra Spiecker genannt Döhmann	
7.1	Zentrale Ergebnisse	173
7.2	Wünschbare Forschungsbedingungen im Zeitalter von KI	176
7.2.1	Strategievorschlge an das Wissenschaftssystem	176
7.2.2	Empfehlungen fr den Ausbildungs- und Weiterbildungsbereich	178
7.2.3	Wnsche an die Regulierung von KI-basierter Forschung. . .	178
	Literatur.	178

Über die Autoren

Peter Buxmann ist Universitätsprofessor für Wirtschaftsinformatik | Software & Digital Business an der Technischen Universität Darmstadt. Er ist zudem Sprecher der Mission „Future Data Economy and Society“ im nationalen Forschungszentrum ATHENE und darüber hinaus Mitglied in zahlreichen Leitungs- und Aufsichtsgremien, u. a. im Beirat des Weizenbaum-Instituts für die vernetzte Gesellschaft – Das Deutsche Internet Institut in Berlin.

Seine Forschungsschwerpunkte sind die Digitalisierung von Wirtschaft und Gesellschaft, Methoden und Anwendungen der Künstlichen Intelligenz sowie das Spannungsfeld zwischen Datenökonomie und Privatsphäre. Er ist Autor von mehr als 300 Publikationen, die in internationalen Zeitschriften (z. B. Information Systems Research, Journal of Information Technology, European Journal on Information Systems, Information Systems Journal) und Konferenzbänden (z. B. International Conference on Information Systems sowie European Conference on Information Systems) erschienen sind. Die Arbeiten von Peter Buxmann und seinem Team wurden mehrfach ausgezeichnet, z. B. mit dem Paper of the Year der AIS im Jahr 2017 sowie dem Best Paper Award bei der Software Business Conference 2012 am Massachusetts Institute of Technology (MIT).

Peter Buxmann erhielt seine Promotion an der Universität Frankfurt. Nach einem Forschungs- und Lehraufenthalt an der Haas School of Business der University of California in Berkeley habilitierte er sich. Von 2000–2004 war er Professor für Wirtschaftsinformatik und Informationswirtschaft an der Technischen Universität Freiberg, bevor er an die Technische Universität Darmstadt wechselte.

Julia Distelrath, geb. Thelen, Master of Science Geographie, war vom 01.01.2018 bis zum 31.12.2020 Wissenschaftliche Mitarbeiterin und Projektkoordinatorin des Projekts „DiA – Digitale Arbeitswelten in Forschung und Entwicklung. Neue Möglichkeiten und Herausforderungen für die Wissenschaft“ am IQIB – Institut für qualifizierende Innovationsforschung & -beratung GmbH, in dessen Rahmen sie auch Beiträge zu einer Begleitstudie der RWTH Aachen zu Digitalisierung und KI in der Wissenschaft leistete. Julia Distelrath studierte zuvor Geographie in Tübingen (Bachelor of Science) und Bonn (Master of Science) und arbeitete zwischen

2013 und 2017 als studentische bzw. geprüfte wissenschaftliche Hilfskraft an der ehemaligen Europäischen Akademie Bad Neuenahr-Ahrweiler, später EA European Academy. Dort war sie u. a. an der Etablierung eines Projekts zur Technikfolgenabschätzung von „Industrie 4.0“ in Mittel- und Osteuropa beteiligt, in dem sie u. a. einen vergleichenden Länderreport als publizierten Abschlussbericht des Vorhabens koordinierte. Ihre Arbeitsgebiete umfassen die Themen Digitalisierung, Künstliche Intelligenz, Industrie 4.0 und Energiesystemforschung.

Carl Friedrich Gethmann Dr. phil. habil. Dr. phil. h.c., lic. Phil. ist Professor für „Wissenschaftsethik/Medizinethik“ in der Lebenswissenschaftlichen Fakultät der Universität Siegen.

Studium der Philosophie in Bonn, Innsbruck und Bochum; 1968 lic. phil. (Institutum Philosophicum Oenipontanum); 1971 Promotion zum Dr. phil. (Ruhr-Universität Bochum); 1978 Habilitation für „Philosophie“ (Universität Konstanz). 2003 Ehrenpromotion zum Dr. phil. h.c. an der Humboldt-Universität zu Berlin; seit 2010 Honorarprofessor an der Universität zu Köln.

Berufungen auf Ordentliche Professuren (C4) an die Universität Oldenburg (1990; abgelehnt), die Akademie für Technikfolgenabschätzung in Stuttgart (1991; abgelehnt), an die Universitäten Essen (1991; angenommen), Konstanz (1993; abgelehnt) und Bonn (1995; abgelehnt); 1996–2012 Direktor der Europäischen Akademie zur Erforschung von Folgen wissenschaftlich-technischer Entwicklungen Bad Neuenahr-Ahrweiler GmbH; 2012–2020 Professor für „Wissenschaftsethik“ am Forschungskolleg „Zukunft menschlich gestalten“ der Universität Siegen; 2021–2023

Mitglied der Academia Europaea (London); o. Mitglied der Berlin-Brandenburgischen Akademie der Wissenschaften; o. Mitglied der Deutschen Akademie der Naturforscher Leopoldina (Halle); o. Mitglied der Deutschen Akademie der Technikwissenschaften (acatech); Mitglied der Bio-Ethik Kommission des Landes Rheinland Pfalz (2000–2013); Präsident der Deutschen Gesellschaft für Philosophie e. V. (2006–2008); Mitglied des Deutschen Ethikrates (2013–2021); Mitglied des Direktoriums von ESYS (Energiesysteme der Zukunft 2015–2019); Mitglied des Ethikrates der Max Planck-Gesellschaft (ab 2015).

Verdienstkreuz am Bande des Verdienstordens der Bundesrepublik Deutschland (2006).

Forschungsschwerpunkte: Sprachphilosophie/Philosophie der Logik; Phänomenologie; Angewandte Philosophie (Medizinische Ethik/Umweltethik/Technikfolgenabschätzung).

Bernhard G. Humm ist seit 2005 Professor für Software Engineering und Projektmanagement am Fachbereich Informatik der Hochschule Darmstadt. Er ist geschäftsführender Direktor des Instituts für Angewandte Informatik Darmstadt (aIDA), sowie Promotionsausschuss-Vorsitzender des hochschulübergreifenden Promotionszentrums für Angewandte Informatik (PZAI).

Seine Forschungsschwerpunkte liegen in den Bereichen Angewandte Künstliche Intelligenz und Software Architektur. In diesen Bereichen führt er nationale und

internationale Forschungsprojekte im Industrie- und Hochschulumfeld durch und publiziert regelmäßig. Auszeichnungen für seine Forschungsarbeiten beinhalten den Hessischen Forschungspreis 2014 für HAW, den World Summit Award 2015 im Bereich Culture and Tourism, weitere Forschungs- und Lehrpreise auf Hochschulebene sowie Best Paper Awards in internationalen Konferenzen.

Prof. Humm arbeitete 11 Jahre beim Software- und Beratungshaus sd&m AG als technischer Seniorberater, Bereichsleiter und Leiter von sd&m Research. Dabei verantwortete er IT Großprojekte für namhafte Kunden aus der Finanzindustrie, Tourismus, Handel und der Luftfahrt. Seine Promotion (Ph.D. Computer Science) erhielt er 1994 an der University of Wollongong in Australien, sein Diplom mit Auszeichnung 1991 an der Universität Kaiserslautern.

Stephan Lingner ist Leiter des Forschungsbereichs „Technology Assessment“ am IQIB – Institut für qualifizierende Innovationsforschung und -beratung GmbH in Bad Neuenahr-Ahrweiler. Er studierte Geowissenschaften an den Universitäten Würzburg und Tübingen und wurde an der Universität Münster zum Dr. rer. nat. promoviert. Nach wissenschaftlichen Tätigkeiten an der Westfälischen Wilhelms-Universität Münster und am Deutschen Zentrum für Luft- und Raumfahrt in Köln wechselte er 1996 als Projektleiter zur Europäischen Akademie zur Erforschung wissenschaftlich-technischer Entwicklungen Bad Neuenahr-Ahrweiler GmbH. Dort war er seit 2005 stellvertretender Direktor bis zur Neuausrichtung und Umfirmierung der Akademie zum IQIB 2019. Stephan Lingner koordinierte zahlreiche interdisziplinäre Projekte und Studien zu Umwelt- und Digitalisierungsfragen sowie zu konzeptionellen Aspekten interdisziplinärer Forschung. Darüber hinaus war er mehrere Jahre Lehrbeauftragter der Hochschule Koblenz im Fachbereich Mathematik und Technik sowie Redaktionsleiter der Zeitschrift „Poiesis & Praxis. International Journal of Ethics of Science and Technology Assessment“. Stephan Lingner ist seit 2017 Mitherausgeber bzw. Mitglied im Wissenschaftlichen Beirat von „TATuP – Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis“ und seit 2005 koordinierendes Mitglied im „Netzwerk Technikfolgenabschätzung (NTA)“. Seine Arbeitsgebiete reichen von Technikfolgenbeurteilung über Umweltfragen bis hin zu Planetenforschung und Weltraumerkundung.

Verena Nitsch leitet das Institut für Arbeitswissenschaft an der RWTH Aachen University und die Abteilung Produkt- und Prozessergonomie am Fraunhofer Institut für Kommunikation, Informationsverarbeitung und Ergonomie. Nach ihrem Bachelorstudium der Angewandten Psychologie an der Charles Sturt University in Australien und der University of Central Lancashire in Großbritannien absolvierte sie ihr Masterstudium in der Arbeits- und Organisationspsychologie an der Manchester Business School. 2012 schloss sie ihre Promotion zum Doktor der Ingenieurwissenschaft im Fachgebiet Mensch-Technik-Interaktion ab und wurde 2013 auf die Juniorprofessur Kognitive Ergonomie an der Universität der Bundeswehr München berufen, wo sie 2016 bis 2018 das Institut für Arbeitswissenschaft leitete. Sie ist aktuell als Gutachterin und Beraterin in zahlreichen Ausschüssen und Beiräten, u. a. im Programmausschuss „Robotik und Automation“ des DLR Raum-

fahrtmanagements, im wissenschaftlichen Fachbeirat des Forschungsinstituts für Betriebliche Bildung, in der VDE Expertengruppe „Mensch-Maschine Schnittstelle“, im Steering Committee der DFG-Exzellenzinitiative „Internet of Production“ und als Mentorin des EXIST Gründerzentrums tätig. Zu ihren Forschungsschwerpunkten zählt die menschenzentrierte Arbeitsgestaltung und Mensch-Technik Interaktion in hochautomatisierten sozio-technischen Arbeitssystemen.

Jan C. Schmidt ist promovierter Physiker und habilitierter Philosoph. Seit 2008 ist er Professor für Wissenschafts- und Technikphilosophie an der Hochschule Darmstadt. Zuvor war er Professor für Technikphilosophie am Georgia Institute of Technology, Atlanta. Er hatte eingeladene Gast- und Vertretungsprofessuren an den Universitäten Jena, Wien und Klagenfurt inne. Zuvor war er Mitarbeiter am Institut für Physik der Universität Mainz sowie am Zentrum für Interdisziplinäre Technikforschung und am Institut für Philosophie der TU Darmstadt. Schmidt ist Mitglied verschiedener Beiräte und Kuratorien, etwa dem Transdisziplinaritäts-Beirat der Schweizerischen Akademie der Wissenschaften oder dem Konvent der ev. Akademie Frankfurt. Seine Arbeitsschwerpunkte umfassen: Wissenschafts- und Technikphilosophie, Konzepte und Methodologien inter- und transdisziplinärer Forschung, Geschichte der Natur- und Technikwissenschaften, Technik-, Umwelt- und Digitalethik, Nachhaltigkeitsforschung sowie Technikfolgenabschätzung.

Indra Spiecker genannt Döhmman, LL.M. (Georgetown Univ.) ist seit 2013 Professorin für Öffentliches Recht, Informations-, Umweltrecht und Verwaltungswissenschaften an der Goethe-Universität Frankfurt a. M. Dort leitet sie zudem die Forschungsstelle Datenschutz sowie das Institut für Europäische Gesundheitspolitik und Sozialrecht (ineges). Zuvor hielt sie nach ihrer Tätigkeit am MPI für Gemeinschaftsgüter, Bonn, den Lehrstuhl für Öffentliches Recht, Telekommunikationsrecht, Informationsrecht und Datenschutzrecht am Karlsruher Institut für Technologie (KIT).

Fr. Indra Spiecker genannt Döhmman forscht – vielfach gefördert auch in Projekten mit Unternehmen – im Staats- und Verwaltungsrecht unter Einbeziehung interdisziplinärer und rechtsvergleichender Erkenntnisse aus der Ökonomie, den Verhaltens- und den Technikwissenschaften; ihre Schwerpunkte liegen im Datenschutz-, Informations- und IT-Recht, ergänzt um Gesundheits- und Umweltrecht. Prof. Indra Spiecker genannt Döhmman wirkt als Mitglied in zahlreichen wissenschaftlichen Beiräten mit und berät Bund, Länder und Unternehmen zu vielfältigen Fragen insbesondere der Informationsgesellschaft. Sie ist Herausgeberin bzw. Schriftleiterin einer Reihe auch von internationalen wissenschaftlichen Zeitschriften sowie Herausgeberin des Großkommentars zur DSGVO (Simitis/Hornung/Spiecker gen. Döhmman). Prof. Indra Spiecker genannt Döhmman wurde 2016 als erste Juristin in die Acatech – Akademie der Technikwissenschaften – berufen.

1

Einführung



**Carl Friedrich Gethmann, Peter Buxmann, Julia Distelrath,
Bernhard G. Humm, Stephan Lingner, Verena Nitsch, Jan C. Schmidt
und Indra Spiecker genannt Döhmann**

„Künstliche Intelligenz“ (KI) ist ein Begriff, der zur Zeit große öffentliche Aufmerksamkeit erfährt. Die hohe Dynamik in diesem Bereich erfordert es, nicht nur bei der Klärung der Bedingungen, Wirkungen und Folgen hinsichtlich Gesellschaft und Wirtschaft stehenzubleiben: Es scheint vielmehr ein grundlegenderer Zugang

C. F. Gethmann
Universität Siegen, Siegen, Deutschland
E-Mail: carl.gethmann@uni-siegen.de

P. Buxmann
Fachgebiet Wirtschaftsinformatik, Technische Universität Darmstadt, Darmstadt, Deutschland
E-Mail: buxmann@is.tu-darmstadt.de

J. Distelrath
Schalkenbach, Deutschland
E-Mail: julia.distelrath@gmx.de

B. G. Humm
Hochschule Darmstadt, Darmstadt, Deutschland
E-Mail: bernhard.humm@h-da.de

S. Lingner (✉)
Institut für qualifizierende Innovationsforschung und -beratung GmbH (IQIB), Bad Neuenahr-Ahrweiler, Deutschland
E-Mail: Stephan.Lingner@iqib.de

V. Nitsch
Institut für Arbeitswissenschaft, RWTH Aachen, Aachen, Deutschland
E-Mail: v.nitsch@iaw.rwth-aachen.de

J. C. Schmidt
Department of Social Sciences, Hochschule Darmstadt, Darmstadt, Deutschland
E-Mail: Jan.Schmidt@h-da.de

I. Spiecker genannt Döhmann
Goethe Universität Frankfurt a.M., Frankfurt am Main, Deutschland
E-Mail: spiecker@jur.uni-frankfurt.de

zur Rolle von KI notwendig zu sein. Da in modernen Wissensgesellschaften wissenschaftliche Forschung wesentlicher Treiber gesellschaftlicher und wirtschaftlicher Entwicklungen ist, ist zu fragen, ob Forschungshandeln und Wissenserzeugung durch künstliche Intelligenz vor einem epochalen Wandel stehen und vor allem wie sich entsprechende Arbeitswelten verändern werden. Wird zum Beispiel der Mensch durch KI zu einer Randerscheinung in der Forschungspraxis oder wird er hier weiterhin zentraler Akteur bleiben?

In dieser Studie analysieren und beurteilen die Autoren mögliche Auswirkungen von KI auf die Arbeitswelten in Wissenschaft und Forschung, und sie formulieren auf dieser Grundlage Handlungsempfehlungen für eine verantwortungsbewusste Gestaltung von KI in der wissenschaftlichen Forschung. Die Studie ist das Ergebnis intensiver Beratungen einer interdisziplinären Arbeitsgruppe am Institut für qualifizierende Innovationsforschung und -beratung (IQIB). Die Arbeitsgruppe ist mit Fachleuten aus den Gebieten der Philosophie, Technikfolgenabschätzung, Informatik und Wirtschaftsinformatik mit Schwerpunkt KI, sowie aus den Arbeitswissenschaften und Rechtswissenschaften besetzt.

In diesem Einführungskapitel wird die Fragestellung des interdisziplinären Projekts erläutert und durch eine Einordnung in die Gebiete der KI sowie in das wissenschaftliche Arbeiten kontextualisiert. Kap. 2 beschreibt sodann die Grundlagen und Anwendungen von KI. In Kap. 3 wird die Frage behandelt, ob menschliche Forscher/Forscherinnen grundsätzlich durch Systeme künstlicher Intelligenz ersetzbar sind. Kap. 4 erörtert die durch KI herbeigeführte Veränderung des Wissenschafts- und Technikverständnisses. In Kap. 5 werden die Auswirkungen auf den wissenschaftlichen Beruf unter Heranziehung einer empirischen Studie untersucht. Kap. 6 richtet schließlich den Blick auf die Regulierungsperspektive von durch KI geprägte Forschung. Jedes dieser Kapitel endet mit einem kurzen Zwischenfazit. In Kap. 7 werden interdisziplinäre Schlussfolgerungen und Vorschläge für die Handlungsebene formuliert.

1.1 Fragestellung

Im Zuge der Digitalisierung von Arbeitswelten erreichen KI-Systeme bereits viele Bereiche öffentlich oder privat finanzierter Forschung und Entwicklung mit entsprechenden Konsequenzen für die wissenschaftliche Praxis. Aufgrund der besonderen gesellschaftlichen Rolle öffentlich geförderter Forschung konzentrieren sich die Untersuchungen wissenschaftlicher Arbeitswelten auf diesen Sektor. Dabei werden vornehmlich solche Wissenschaftsdisziplinen in den Blick genommen, deren Geltungsansprüche traditionell auf experimentellen Ansätzen fußen und eine hohe Affinität zu unterstützenden KI-Systemen versprechen. Es sind dies beispielsweise die Natur-, Lebens- und Technikwissenschaften, unter Einschluss der medizinischen Forschung. Diese Fokussierung ist eher praktischer Natur und soll mögliche Folgerungen für andere Disziplinen bzw. Förder-Regimes nicht ausschließen. Für die interdisziplinäre Beurteilung von KI in wissenschaftlichen Arbeitswelten stellen sich u. a. folgende Fragen aus computerwissenschaftlicher, wissenschaftsphilosophischer sowie arbeits- und rechtswissenschaftlicher Sicht:

Von *technischer Seite* aus ist zunächst zu fragen, welche Entwicklungsperspektiven, Meilensteine und Einsatzbereiche von KI in der Forschung „realistisch“ absehbar sind und wie sich das Zusammenspiel mit Big Data und anderen, auch klassischen Forschungsansätzen gestalten wird. Die Beantwortung dieser Fragen soll zugleich klären, welche Erwartungen an und Befürchtungen vor KI als überzogen entmystifiziert werden sollten. Auf der anderen Seite kann dadurch ermessen werden, worin der innovative Kern von KI-Systemen liegt und worin damit die Grenzen der KI liegen. So gilt es, den wissenschaftlich-technischen Kern von KI-Systemen für den öffentlichen Diskurs offenzulegen und reflektierbar werden zu lassen.

Zur Beantwortung solcher Fragen ist angesichts der begrifflichen Vielfalt im Bereich KI zunächst eine systematische Betrachtung dieses komplexen Feldes vor dem Hintergrund ihrer wissenschaftsgeschichtlichen Entwicklung vorzunehmen. Auf der Basis typischer KI-Anwendungen in Teilchenphysik, Klimaforschung und Medizin sind Schlussfolgerungen für die Qualität, Effizienz und Innovationskraft von KI in der Forschung zu entwickeln und mögliche Rückwirkungen auf das Wissenschaftssystem zu skizzieren.

Aus *wissenschaftsphilosophischer Sicht* fragt sich, ob, inwieweit und wo die Nutzung von künstlicher Intelligenz und Big Data den etablierten Forschungsmodus moderner Forschung verändert. Vor allem ist zu untersuchen, ob KI und Big-Data-basierte Forschung „mit der leichten Hand“ dazu führen, „Wahrheit“ allein in der großen Menge verfügbarer Daten zu vermuten. Ferner ist von Interesse, ob sich wissenschaftliche Prognosen künftig statt von Kausalitäten von Korrelationen bzw. automatisch erkannten Mustern ableiten lassen. Von der Beantwortung dieser Fragen hängt ab, welche Folgen dies für die künftige Rolle von Theorien und ihre Validierbarkeit sowie für die Erklärbarkeit von Forschungsergebnissen hätte.

In diesem Zusammenhang muss Wissenschaftler/Wissenschaftlerinnen vor allem die Sorge beschäftigen, ob Forschung als genuin menschliches Betätigungsfeld mittelfristig noch Bestand haben wird und welche Rolle den forschenden Akteuren in hybriden Mensch-Maschine-Ensembles bleibt. Verstehen sie sich noch als Urheber des jeweiligen Forschungshandelns oder füllen sie allenfalls noch kognitive oder sensorische Lücken in den jeweiligen Forschungsprojekten? Ein etwaiger Technikpaternalismus, der im Industriesektor bereits angedacht ist (Spath et al. 2012), würde bewährten Forschungsgrundsätzen und -traditionen widersprechen. Die vorliegende Studie greift insbesondere diese Punkte auf.

Spezieller ist auch der Einsatz von KI im Modus des Machine Learning in der Forschung zu hinterfragen, da hierbei die internen Lernprozesse in den künstlichen neuronalen Netzen opak bleiben und somit nicht nachvollziehbar sind (Black Box). Ohne Transparenz kritischer Schritte hin zum finalen Forschungsergebnis wäre aber die Nachvollziehbarkeit des Forschungsprozesses als immanente Kernforderung moderner Wissenschaft an sich selbst fraglich. Insoweit könnten sich zukünftig Forschungsergebnisse weitgehend der Kontrolle der Forschenden sowie ihrer gewohnten Validierung in der Wissenschaftsgemeinschaft entziehen. Zu untersuchen ist somit, ob für bestimmte KI-basierte Forschungsleistungen Abstriche an ihre Verlässlichkeit gemacht werden müssen. Die epistemische Robustheit von KI-unterstützter Forschung und Entwicklung ist daher zu klären. Das schließt die Frage ein, welche Aussichten bestehen, „neuronalen“ Prozesse von KI und ihre Resultate

künftig besser verstehbar zu machen. Für die Qualität und Validität wissenschaftlicher Arbeit ist somit zu klären, ob der Einsatz von KI und Big Data zu effizienterer und weiterhin exzellenter Forschung und Entwicklung führt oder ob man sich damit Defizite in ihrer Erklärbarkeit einhandelt. Im ungünstigen Fall stünden Vertrauen und gesellschaftliche Anerkennung von Wissenschaft auf dem Spiel.

Die wissenschaftsphilosophische Reflexion über KI-unterstützte Forschung zeigt, ob und wo bewährte und anerkannte Forschungspraxen auch künftig erhalten bleiben und wo ein technisch bedingter Wandel – möglicherweise auch disruptiver Art – zu erwarten sind. Beispielsweise könnte dies die Erschließung neuer, bislang nicht zugänglicher Gegenstandsfelder (vergl. Chou et al. 2017) und die künftige Rolle des Experimentierens in der Forschung betreffen, oder auch eine denkbare Verschmelzung von Wissenschaft und Technik zu „Techno-Science“.

Aus arbeitswissenschaftlicher Perspektive ist zunächst herauszustellen, in welchem Maße sich die allgemein vorausgesagten Konsequenzen der Digitalisierung für die Arbeitswelten (vergl. Frey und Osborne 2013; van Est und Kool 2015) auch auf die Beschäftigungslage auf dem akademischen Arbeitsmarkt übertragen lassen. So sind in diesem Zusammenhang neu entstehende und verlorengehende Berufsbilder und Arbeitsplätze in der Forschung von Interesse sowie deren mögliche Folgen für die sozialen Sicherungssysteme. Hierbei sind auch Fachkräfteangebot und demografische Indikatoren zu beachten.

Auf Mikroebene ist zu fragen, ob „intelligente“ Assistenzsysteme ihr grundsätzliches Potenzial, wissenschaftliches Arbeiten flexibler, effizienter und effektiver zu machen, auch in der Praxis ausspielen können. Auch sei auf den Arbeitsprozess im Forschungsalltag verwiesen, der sich sowohl als kognitiv reizvoll und abwechslungsreich als auch als monoton und einseitig belastend oder gar als überfordernd darstellen könnte. Explorative Umfragen unter Wissenschaftlern/Wissenschaftlerinnen sollen dieser Frage nachgehen und Hinweise geben, ob breite Digitalisierung und künstliche Intelligenz im Arbeitsalltag eine Flexibilisierung hin zu mobiler Arbeit, individuell angepassten Arbeitszeiten oder der Verlagerung von standardisierbaren Aufgaben an KI in der Praxis möglich erscheinen lassen (vergleiche auch Zettel et al. 2014). Auf der anderen Seite kann dieses Angebot gepaart mit hoher intrinsischer Motivation für die eigene Forschung auch in eine an Selbstausbeutung grenzende Bereitschaft für Einsatz und Erreichbarkeit münden. Die sich daraus ergebenden arbeitspsychologischen, und -medizinischen Konsequenzen der weiteren Digitalisierung im Forschungsalltag könnten auch wirtschaftlich bedeutsam werden (vergl. Storm 2017). Schließlich stellt sich die Frage, ob und wie KI-gestützte Arbeit menschengerecht gestaltet werden kann, um nicht nur psychische und physische Belastungen von Forschern/Forscherinnen zu reduzieren, sondern auch deren wissenschaftliche Autonomie im Forschungsprozess künftig zu erhalten.

Aus Regulierungsperspektive stellen sich Fragen durch den zunehmenden Einsatz von KI und Big Data in der wissenschaftlichen Forschung insbesondere hinsichtlich fairer Forschungsevaluation sowie möglicher Konflikte zwischen den Maximen der Wissenschaftsfreiheit einerseits und des Datenschutzes andererseits.

Da KI-basierte Forschung zumeist datenintensiv ist, fragt sich, (a) ob dies das Verhältnis und die Beurteilung von empirischen und reflektierenden Forschungskulturen in der Forschungsevaluation verändert, (b) ob diese Veränderung legitim wäre und ggf.

(c) welche Beurteilungsmethoden der Diversität von Fachkulturen im Zeitalter von KI gerecht werden. Parallel dazu sind ggf. bereits bestehende Regulierungsansätze von KI auf ihre Relevanz für forschungsnahe Bereiche hin zu untersuchen. Weiterhin ist das Konfliktfeld von Wissenschaftsfreiheit und Datenschutz zu erkunden, das insbesondere durch die Zweckbindungsnorm der Europäischen Datenschutz-Grundverordnung (DSGVO) für die wissenschaftliche Erhebung personenbezogener Daten gegeben ist. Hier ist im Interesse der Freiheit der Forschung nach legitimen Möglichkeiten des Austarierens und/oder der Lockerung der Zweckbindungsvorschrift zu suchen bzw. in Sonderfällen zu spezifizieren.

Schließlich ist im Zusammenhang mit der Regulierung von KI auch die Legitimität der KI-unterstützten Beurteilung von Wissenschaftlerprofilen im Arbeitsprozess und auf dem Arbeitsmarkt kritisch zu beurteilen.

Die interdisziplinäre Perspektive ist eine reflektierende, die die vorgenannten relevanten disziplinären Perspektiven zum Thema KI in der Forschung aufeinander bezieht und auf das Erkenntnis- und Beratungsinteresse hin weiterentwickelt und integriert. Das Ziel ist eine vorausschauende, prospektive Technikfolgenabschätzung von KI in der forschenden Wissenschaft, die begründbare und legitime Schlüsse für die Handlungsebene ermöglicht.

Entsprechende Vorschläge an die Adressaten in Forschung und Wissenschaft formulieren, welche qualitätssichernden Gestaltungsaufgaben für das Wissenschaftssystem durch die verbreitete Anwendung von KI in der Forschung erwachsen und welche Maßstäbe (z. B. Intersubjektivität) hierfür in Frage kommen und gerechtfertigt sind. Hierfür ist beispielsweise die Frage einschlägig, welche Motive und Normen das künftige Erkenntnisinteresse von Forschung leiten sollen und inwieweit KI-getriebene Forschung transparent sein sollte. Ferner ist zu klären, wie und wo der bewährte falsifikationistische Validierungsmodus moderner Forschung im Zeitalter von KI aufrechterhalten werden kann. Auf der Basis soll dann untersucht werden, welche individuellen Kompetenzen gefördert werden müssen, um „fit für die Zukunft“ in der Wissenschaft zu sein und welche Fähigkeiten dagegen nicht mehr im bisherigen Maß benötigt werden. Daraus lassen sich Hinweise für die strukturelle und inhaltliche Ausgestaltung von Aus- und Weiterbildungsinhalten ergeben. Die Beantwortung der vorgenannten Fragen werden damit auch einen konkreten Beitrag zum forschungspolitischen Anliegen der „Responsible Research and Innovation“ in Europa leisten (vgl. Owen et al. 2013).

1.2 Künstliche Intelligenz: eine Einordnung

1.2.1 Ein Thema von großer öffentlicher Aufmerksamkeit

Die hohe Öffentlichkeitswirksamkeit von KI ist mit überzogenen Erwartungen einerseits sowie mit übertriebenen Befürchtungen andererseits verknüpft. Dazu kommt, dass KI ein beliebtes Thema für Spielfilme ist und dadurch die Gefahr besteht, dass Fiktion und Wirklichkeit verschwimmen.

Hinsichtlich der Explikation des Begriffs „KI“ existiert eine große Bandbreite an Verständnisweisen und Definitionen. So meinte Herbert Simon, einer der Gründerväter von KI 1965: „*Machines will be capable [...] of doing any work that a man can do.*“ (Allen 2001). Und sein Kollege Marvin Minski prophezeite 1970: „*Within 10 years computers won't even keep us as pets.*“ (Allen 2001).

Solche überzogenen Erwartungen und Befürchtungen werden auch heute medienwirksam geäußert. Prominente Vertreter dafür sind Ray Kurzweil, Nick Bostrom, Yuval Noah Harai, Elon Musk, sowie der kürzlich verstorbene Stephen Hawking. So behauptet Ray Kurzweil in (Kurzweil 2015, Kap. 1): „Mit der Singularität werden wir die Grenzen unserer biologischen Körper und Gehirne überschreiten. Wir werden die Gewalt über unser Schicksal erlangen. Unsere Sterblichkeit wird in unseren Händen liegen. Wir werden so lange leben können wie wir wollen (...). Bis zum Ende des Jahrhunderts wird die nichtbiologische Komponente unserer Intelligenz Trillionen Mal mächtiger sein als bloße menschliche Intelligenz“. Die letzte Aussage hatte er übrigens 2005 noch für das Jahr 2045 prognostiziert (vgl. BrainyQuote.com 2021). Anhänger der Denkrichtung des Transhumanismus streben an, Grenzen menschlicher Möglichkeiten, sei es intellektuell, physisch oder psychisch, durch den Einsatz technologischer Verfahren zu erweitern. Von ihnen wird die Erwartung geäußert, dass alle Krankheiten und sogar der Tod mithilfe von Technologien besiegt werden können.

Die meisten professionellen Einschätzungen von KI sind demgegenüber deutlich nüchterner. Auch gibt es keine allgemein in der Wissenschaft akzeptierte Definition von KI. Stattdessen kursieren dutzende Definitionen in der einschlägigen Literatur. Beispielsweise führt die AGI Sentinel Initiative eine Liste von 17 Definitionen für Künstliche Intelligenz im Vergleich zu 17 weiteren Definitionen für menschliche Intelligenz.¹ So lautet beispielsweise die KI-Definition des United States Defense Science Board (DSB): „*Artificial intelligence, the capability of computer systems to perform tasks that normally require human intelligence (e.g., perception, conversation, decision-making).*“ (DSB 2016). Entsprechende Definitionen in Enzyklopädien und Wörterbüchern weisen in eine ähnliche Richtung, so z. B. in Encyclopaedia Britannica,² The English Oxford Living Dictionary³ oder Merriam-Webster.⁴

Für viele Unternehmen hat KI eine strategische Bedeutung. So stammt beispielsweise folgende Aussage von Amazon: „*Without ML, Amazon.com couldn't grow its business, improve its customer experience and selection, and optimize its logistic speed and quality.*“ (Marr 2018). Für Google hat Machine Learning hohe Priorität, um intelligentere und nützlichere Technologie zu entwickeln, sowie möglichst viele Menschen zu unterstützen. Ähnlich äußern sich Facebook, IBM und andere Technologie-Konzerne (Marr 2018). Aber auch die deutsche Bundesregierung hat 2018 eine KI-Strategie verabschiedet mit dem Ziel, die Erforschung, Entwicklung

¹ http://agisi.org/doc/AGISI_DefinitionsIntelligence.pdf.

² <https://www.britannica.com/technology/artificial-intelligence>.

³ https://en.oxforddictionaries.com/definition/artificial_intelligence.

⁴ <https://www.merriam-webster.com/dictionary/artificial%20intelligence>.

und Anwendung von künstlicher Intelligenz in Deutschland auf ein weltweit führendes Niveau zu bringen und zu halten.⁵

1.2.2 KI im Kontext von Automatisierung und Digitalisierung

KI ist ein Teil des umfassenderen Trends der Digitalisierung und der digitalen Transformation, der auf das Ausschöpfen von Potenzialen der Digitaltechnik gerichtet ist. Dies wiederum kann als ein Teil des umfassenderen Trends der Automatisierung gesehen werden, in der menschliche Arbeit von Maschinen übernommen wird. Besonders seit der industriellen Revolution verändert diese Automatisierung in rasanter Weise alle Lebens- und Arbeitsbereiche. Viele Berufsgruppen sind komplett verschwunden, neue sind entstanden, und Menschen sowie Gesellschaften mussten sich stets an veränderte Gegebenheiten anpassen, die durch technische Entwicklung verursacht wurden. Damit verbunden waren neue Gestaltungsnotwendigkeiten wie -möglichkeiten.

Während im Zuge der industriellen Revolution vorwiegend physische Tätigkeiten von Maschinen übernommen wurden, werden im Zuge der Digitalisierung, insbesondere mit KI-Techniken, zunehmend auch anspruchsvollere Tätigkeiten automatisiert. Beispielsweise können KI-Anwendungen auf Basis bildgebender Verfahren in ausgewählten Bereichen der Medizin menschliche Experten bei der Diagnose von Krankheiten nicht nur in der Geschwindigkeit, sondern auch in der Präzision übertreffen.

KI hat seit der Jahrtausendwende enorme Fortschritte gemacht und wird in Gesellschaft, Wirtschaft und Wissenschaft täglich genutzt. Beispiele dafür sind:

- Virtuelle Assistenten wie Apples Siri oder Amazons Alexa;
- Gesichtserkennung in modernen Kamerasystemen;
- Industrielle Bildverarbeitung;
- Business Intelligence;
- Robotik.

Diese großen Fortschritte von KI der letzten Jahre sind im Wesentlichen folgenden vier großen Entwicklungstrends zu verdanken:

1. Immer *leistungsfähigere Hardware*;
2. Stetig *verbesserte KI-Software* (Algorithmen, Bibliotheken, Entwicklungsumgebungen, Programmiersprachen);
3. *Zunehmende Verfügbarkeit großer Datenmengen*, vielfach auf der Basis umfassender Sensorik gewonnen (Big Data);
4. *Verbesserter Ausbildungsstand* von Software-Entwicklern im Bereich KI.

⁵<https://www.ki-strategie-deutschland.de>.

1.2.3 Definition von KI

In Ermangelung einer allgemein anerkannten Definition von KI, geben wir hier eine Definition an, welche wir für das gemeinsame Verständnis dieser Studie verwenden:

Künstliche Intelligenz bezeichnet die Fähigkeit von Computersystemen, auf sie zugeschnittene Aufgaben selbsttätig zu lösen, die aufgrund ihrer Komplexität bislang menschliche Fähigkeiten erforderten.

Diese Definition lehnt sich an die Definition des Defence Science Board (DSB 2016), sowie an zahlreiche ähnliche Definitionen an. Computersysteme können dabei Anwendungen auf PCs, Tablets oder Smartphones sein, aber auch Roboter, eingebettete Systeme, Wearables, etc. Es ist in dem Zusammenhang wichtig darauf hinzuweisen, dass Computersysteme mit KI-Fähigkeiten gemäß dieser Definition nicht intelligent *sind*, sondern lediglich in der Lage sind, bestimmte anspruchsvolle Aufgaben zu lösen. Diese Aufgaben setzen bislang exklusiv menschliche Fähigkeiten des *Wahrnehmens* (*Sehen, Hören, Fühlen etc.*), *Lernens, Wissens, Denkens*, *Kommunizierens* (*Sprechen, Schreiben etc.*) und *Handelns* voraus. KI-Systeme ahmen diese Fähigkeiten nach und führen so – im Rahmen ihrer bestimmungsgemäßen Möglichkeiten – komplexe Problemlösungen automatisch durch. Siehe zur Illustration die folgende Abb. 1.1.

Beispiele für solche KI-Systeme sind selbstfahrende Autos oder virtuelle Assistenten wie Apples Siri oder Amazons Alexa. Virtuelle Assistenten simulieren dabei die Fähigkeiten des Hörens (Spracheingabe, z. B. „Alexa, mach den Fernseher an“), des Sprechens (Sprachausgabe), sowie des Handelns (z. B. Fernseher einschalten). Selbstfahrende Autos ahmen die Fähigkeiten des Sehens (Verarbeitung von Kamera-

Abb. 1.1 Fähigkeiten intelligenten Verhaltens.
(Quelle: Humm 2020)
(Abb. 1.1 verändert nach „Artaxerxes“ gemäß der geltenden Creative-Commons-Lizenz.
Original: https://commons.wikimedia.org/wiki/File:Our_Community_Place_Sandbox.jpg)



bildern sowie weiteren Sensordaten), des Lernens (z. B. Erkennung von Stoppschildern), des Wissens (z. B. über Verkehrsregeln), sowie des Handelns (z. B. Ansteuerung von Gas, Bremse und Lenkrad) nach.

1.3 Wissenschaftliches Arbeiten als Teilbereich der Wissensarbeit

Die Auswirkungen der Digitalisierung und des Einsatzes von KI-Methoden auf wissenschaftliche Arbeit lassen sich erst abschätzen, wenn ein grundlegendes Verständnis der charakteristischen Merkmale dieser Arbeit vorliegt. Die wissenschaftliche Arbeit zeigt viele Parallelen zur Wissensarbeit auf, zu der bereits viele arbeitswissenschaftliche Untersuchungen vorliegen (vgl. Arlinghaus 2017).

Wissensarbeit beinhaltet zunächst einmal allgemein die Verarbeitung von Informationen und die Anwendung von Wissen. Darüber hinaus zeichnet sie sich jedoch auch durch eine hohe Selbstständigkeit einerseits und eine gewisse Unsicherheit andererseits aus. Sie ist ferner komplex, kommunikationsintensiv und umfasst routinierte ebenso wie kreative Tätigkeiten. Insofern kommt hier – zumindest partiell – auch die Erzeugung von neuer Information hinzu.

Es soll in dieser Studie argumentiert werden, dass wissenschaftliche Forschung eine spezielle Form der Wissensarbeit darstellt. Forschende sind in diesem Sinne als hoch qualifizierte Fachkräfte zu bezeichnen, die ihr Wissen anwenden, um neues Wissen oder ambitionierte wissenschaftliche Methoden zu entwickeln. Es wird ferner davon ausgegangen, dass sich wissenschaftliche Arbeit im Bereich der Forschung durch ein hohes Maß an Innovationsarbeit auszeichnet und dass wissenschaftliches Arbeiten von ausgesprochen hoher Selbstständigkeit, ergebnis- und prozessbezogener Unsicherheit und Kommunikationsintensität geprägt ist, das letztendlich auf die Bildung neuer wissenschaftlich begründeter Erkenntnis gerichtet ist.

1.4 KI in wissenschaftlicher Forschung

Wie im Privat- und Geschäftsleben sind KI-Anwendungen auch in der Wissenschaft und Forschung mittlerweile vielfältig im Einsatz. Beispiele hierfür sind:

- Theorieentwicklung in der Teilchenphysik auf der Basis von Data Mining Technologien und maschinellem Lernen;
- Mustererkennung in der Klimaforschung;
- Prognose von Epidemien in der Medizin.

Der zunehmende Einsatz von KI lässt folgende positive Effekte für die Forschung erwarten:

1. Forschungsaktivitäten können nun mittels KI effizient durchgeführt werden, die bislang *nicht praktikabel* waren. Beispiel hierfür ist die Extraktion und Korrelation von Daten aus unterschiedlichen Quellen bzw. Forschungsdatenbanken.
2. KI erlaubt solche Analysen, die *bislang unmöglich* waren. Beispielsweise kann hierfür die individualisierte Genomanalyse für bestimmte medizinisch-therapeutische Zwecke ins Feld geführt werden.
3. Forschende werden durch KI bei der *Hypothesenfindung und Modellbildung* unterstützt. So kann Maschinelles Lernen und Data Mining z. B. zur Aufdeckung bislang verborgener Korrelationsmuster beitragen. Dabei kann von den Forschenden bereits das Aufdecken verborgener Muster prognostisch genutzt werden – auch abseits der Erkenntnisbildung über zugrunde liegende kausale Erklärungen.

Die vorliegende Studie wird klären, inwieweit der Mensch hierbei handelnder Akteur in der wissenschaftlichen Forschung bleibt und ob und inwieweit KI-Anwendungen dabei die Rolle von neuartigen *Werkzeugen* einnehmen werden. An dieser Stelle sei vermerkt, dass bereits die mächtigen Werkzeuge in der Wissenschaftsgeschichte fundamentale Änderungen von neuen Methoden und Erkenntnissen hervorgerufen und ganze Wissenschaftszweige begründet haben. Beispielsweise hat die Erfindung des Mikroskops die Biologie und Medizin revolutioniert und das Gebiet der Molekularbiologie erst ermöglicht. Dennoch könnte sich der Werkzeugcharakter von bestimmten KI-Instrumenten grundlegend anders darstellen. Waren alle bisherigen Werkzeuge in der Geschichte der Wissenschaften kausale, stabile Maschinen, die im Prinzip transparent und kausal erklärbar waren, wäre hier nun zu untersuchen, ob und inwieweit sich dies bei KI-Werkzeugen aufgrund ihrer Komplexität, ihrer Dynamik, und Autonomie, sowie der Unklarheit der Systemgrenzen der Untersuchungsobjekte verändern wird.

In den folgenden Kapiteln analysieren die Autoren/Autorinnen aus Perspektive verschiedener einschlägiger Disziplinen mögliche Chancen und Herausforderungen von KI für die wissenschaftliche Forschung und machen Gestaltungsvorschläge für ihre wünschbare Einbettung.

Literatur

- Allen FE (2001) The myth of artificial intelligence. *American Heritage* 52(1), Feb./March 2001. <https://www.americanheritage.com/myth-artificial-intelligence>. Zugriffen am 04.09.2020
- Arlinghaus A (2017) Wissensarbeit. Aktuelle arbeitswissenschaftliche Erkenntnisse. Mitbestimmungsförderung Report, Nr. 35. Hans-Böckler-Stiftung, Düsseldorf.
- BrainyQuote.com (2021) Ray Kurzweil Quotes. BrainyMedia Inc. https://www.brainyquote.com/quotes/ray_kurzweil_591137. Zugriffen am 26.01.2021
- Chou F, Hawkes A, Northon K (2017) Artificial Intelligence NASA Data Used to Discover Eighth Planet Circling Distant Star. Release 17-098. Dec. 14, 2017. NASA. <https://www.nasa.gov/press-release/artificial-intelligence-nasa-data-used-to-discover-eighth-planet-circling-distant-star>. Zugriffen am 20.11.2020

- DSB (2016) Summer study on autonomy. Washington, DC. 20301-3140: Defense Science Board. <https://goo.gl/mm6N6y>. Zugegriffen am 08.10.2018
- van Est R, Kool L (Hrsg) (2015) Working on the robot society. Visions and insights from science concerning the relationship between technology and employment. Rathenau Instituut, The Hague
- Frey CB, Osborne MA (2013) The future of employment: how susceptible are jobs to computerization? Working Paper. Oxford Martin School
- Humm BG (2020) Applied artificial intelligence – an engineering approach, 2. Aufl. Eigenverlag. ISBN: 979-8635591154
- Kurzweil R (2015) Menschheit 2.0: Die Singularität naht, 2. Ausg. Lola Books, Juli 2015. ISBN: 9783944203133
- Marr B (2018) The key definitions of artificial intelligence (AI) that explain its importance. Forbes Feb 14, 2018. <https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/#54d0f59a4f5d>. Zugegriffen am 04.09.2020
- Owen R, Stilgoe J, Macnaghten P, Gorman M, Fisher E, Guston D (2013) Responsible innovation: managing the responsible emergence of science and innovation in society. J. Wiley, Oxford
- Spath D, Schlund S, Gerlach S, Hämmerle M, Krause T (2012) Produktionsprozesse im Jahr 2030. IM – Fachzeitschrift für Information Management und Consulting 27(3):50–55
- Storm A (Hrsg) (2017) Gesundheitsreport 2017. Beiträge zur Gesundheitsökonomie und Versorgungsforschung (16). DAK, Hamburg
- Zettl C, Lehmann C, Bouncken RB (2014) Vertrauensarbeitszeit im Wissenschaftsbetrieb. In: Bornewasser M, Zülch G (Hrsg) Arbeitszeitmanagement. AWV, Eschborn, S 293–306

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.





Bernhard G. Humm, Peter Buxmann und Jan C. Schmidt

2.1 Historie von KI

Der Begriff „Artificial Intelligence“ wurde auf dem Dartmouth Workshop 1956 geprägt. Teilnehmer dieses Workshops waren Pioniere der KI, u. a. John McCarthy, Marvin Minski, Allen Newell und Herbert Simon. Im Workshop Proposal wurde folgende Vermutung formuliert: „[E]very aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it“ (McCarthy et al. 1955).

Die Grundlagen der KI wurden jedoch schon deutlich früher gelegt, vor allem durch Alan Turing (1912–1954) mit seinen grundlegenden Arbeiten über Berechenbarkeit und dem so genannten „Turing-Test“ zur Überprüfung des intelligenten Verhaltens einer KI-Anwendung.

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.

B. G. Humm (✉)

Hochschule Darmstadt, Darmstadt, Deutschland

E-Mail: bernhard.humm@h-da.de

P. Buxmann

Fachgebiet Wirtschaftsinformatik, Technische Universität Darmstadt, Darmstadt, Deutschland

E-Mail: buxmann@is.tu-darmstadt.de

J. C. Schmidt

Department of Social Sciences, Hochschule Darmstadt, Darmstadt, Deutschland

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

In den 1960–1980er-Jahren folgte ein KI-Hype ohnegleichen mit enthusiastischen Versprechungen von schnellen und fantastischen KI-Erfolgen. Dieser Hype führte zu enormen finanziellen Förderungen der KI-Forschung, vor allem in den USA. Es passierte, was passieren muss, wenn unrealistische Erwartungen aufgeblasen und Versprechungen maßlos überzogen werden: die Blase platzte. Bildverarbeitungssysteme für autonomes Fahren konnten nicht zwischen geöffneten Kanaldeckeln und Schatten unterscheiden; Logik-Anwendungen konnten in Punkto Allgemeinwissen nicht mit kleinen Kindern mithalten. Es folgte eine Desillusionierung, verbunden mit einem Kollaps des KI-Markts. Diese Zeit zwischen den späten 1980er- und frühen 2000er-Jahren wird manchmal auch als „KI-Winter“ bezeichnet.

Allerdings wurden in dieser Zeit KI-Techniken kontinuierlich weiterentwickelt, häufig ohne Nennung des damals diskreditierten Begriffs KI. Vor allem die führenden IT-Unternehmen waren hier wegbereitend. Beispielsweise sagte der Google-Mitbegründer Larry Page 2006: „We want to create the ultimate search engine that can understand anything. Some people could call that artificial intelligence“ (Sarathi 2014). KI gewann dadurch implizit enorm an Bedeutung und wurde allgegenwärtiger Bestandteil des täglichen Lebens: Navigationsgeräte mit Sprach-Ein- und Ausgabe, digitale Assistenten auf Smartphones, Bilderkennung in Kameras, KI-Module in Computerspielen etc. Erst in den letzten Jahren wurde dafür auch wieder der Begriff KI verwendet.

Heute sehen wir einen ähnlichen Hype wie in den 1970er-Jahren. Wieder gibt es überzogene Erwartungen bzw. Befürchtungen. Beispielsweise prognostiziert Nick Bostrom in seinem Buch „Superintelligenz: Szenarien einer kommenden Revolution“ (Bostrom 2014), dass innerhalb eines Menschenlebens so genannte Superintelligenzen, KI-Systeme, die der menschlichen Intelligenz weit überlegen sind, entstehen, was das Ende der Menschheit bedeuten könnte. Erwartungen von KI-Systemen, die sich selbst, unabhängig vom Menschen außerhalb ihres ursprünglichen Anwendungszwecks weiterentwickeln (wie z. B. Bostrom es postuliert) werden auch als *starke KI* bezeichnet.

Ob es jemals eine solch starke KI geben kann, ist eine spekulative Frage, die niemand mit Sicherheit beantworten kann. Bis heute sind keine Ansätze für starke KI bekannt. Stand heute gibt es also keine Evidenz dafür, dass starke KI überhaupt jemals möglich sein wird. Die allerwenigsten KI-Forscher/-Forscherinnen und -Praktiker/-Praktikerinnen rechnen auf absehbare Zeit mit einer solchen. Allerdings nehmen Vorstellungen von einer starken KI in der öffentlichen Debatte einen großen Raum ein – ähnlich zum früheren KI-Hype. Es ist durchaus möglich, dass sich hier die Geschichte wiederholt und eine folgende Phase der Desillusionierung zu einem neuen KI-Winter führt. Wir empfehlen die Lektüre der fundierten Diskussion des Robotik-Pioniers Rodney Brooks zu Prognosen über die Zukunft der KI (Brooks 2017), sowie den Blogeintrag von Bernhard Humm zu diesem Thema (Humm 2017).

In dieser Studie verwenden wir nicht den Ausdruck „Künstliche Intelligenzen“, da er eine Personifizierung der Maschine suggeriert. Stattdessen sprechen wir von *KI-Anwendungen* („Apps“), besonders wenn sie auf Standard-Hardware wie PCs oder Smartphones laufen. Wir verwenden auch den Ausdruck *KI-Systeme*, wenn die Hardware inhärenter Systembestandteil ist, wie z. B. bei Robotern.

Alle KI-Anwendungen, die heute existieren, werden der *schwachen KI* zugeordnet. Solche KI-Anwendungen sind von Ingenieuren speziell auf einen Anwendungszweck hin entwickelt worden, wie z. B. Gesichter erkennen, Schach spie-

len, autonom fahren, etc. Auch wenn Verfahren des Maschinellen Lernens eingesetzt werden, ändert sich der Anwendungszweck nicht, solange Ingenieure nicht die Software anpassen. Ein selbstfahrendes Auto wird niemals von sich aus das Schachspiel lernen und ein Schachprogramm wird niemals von sich aus Gesichter erkennen.

Solche maßgeschneiderten KI-Anwendungen haben mittlerweile eine hohe Verbreitung im Privat- und Geschäftsleben. Nachfolgend einige Beispiele:

- Automatische Sprach-Eingabe und -Ausgabe in Autos, Mobiltelefonen, Help-desks etc.
- Gesichtserkennung in Kameras, Fotoarchiven, Flughäfen, etc.
- Spam-Klassifikation mittels Maschinellern in E-Mail-Programmen
- KI-Komponenten in Computerspielen
- Semantische Suche im Internet
- Maschinelle Übersetzung zwischen Sprachen wie Deutsch und Englisch als Web-Service
- Business-Intelligence-Auswertungen von Geschäftsdaten
- Sentiment-Analyse von Einträgen in sozialen Medien, z. B. für bestimmte Produkte
- Roboter-Anwendungen, z. B. in der industriellen Fertigung
- Anwendungen für Militär, Luft- und Raumfahrt, z. B. Drohnen und autonome Fahrzeuge wie Mars Rover.

Diese Anwendungen haben oft eine erstaunliche Leistungsfähigkeit, die manchmal sogar die Fähigkeit menschlicher Experten übertrifft. Das gilt nicht nur für Spiele wie Schach oder Go. Auch in der Medizin übertreffen in eingeschränkten Bereichen bildgebende KI-Verfahren zur Tumorerkennung die Vorhersagegenauigkeit von Medizinern.

Allerdings gibt es auch Negativbeispiele. So können neuronale Netze zur Bildverarbeitung durch leicht modifizierte Bilder fehlgeleitet werden. Solche Modifikationen, für den Menschen kaum zu erkennen, provozieren Falschklassifikationen. Wird beispielsweise ein Stoppschild so manipuliert, dass ein Vorfahrtsschild erkannt wird, dann kann dies fatale Auswirkungen bei selbstfahrenden Autos haben. Ebenso kann eine Verzerrung (Bias) in Trainingsdaten, z. B. aufgrund von Diskriminierungen in der Vergangenheit, eine solche Diskriminierung für die Zukunft fortschreiben. Solche Beobachtungen wurden in Machine-Learning-Anwendungen für Arbeitsämter oder im Rechtswesen gemacht. Die scheinbare Objektivität von Daten kann in diesem Zusammenhang trügen.

2.2 Grundlagen der KI

2.2.1 *Die Landkarte der KI*

In Kap. 1 haben wir Künstliche Intelligenz definiert als die Fähigkeit von Computersystemen, Aufgaben auszuführen, die normalerweise menschliche Intelligenz erfordern. Als Beispiele haben wir genannt: Wahrnehmen (Sehen, Hören, Fühlen

etc.), Lernen, Wissen, Denken, Kommunizieren (Sprechen, Schreiben etc.) und Handeln. Wir verwenden nun diese Aufgaben und Fähigkeiten, um die Teilgebiete der KI strukturieren. Siehe hierzu eine Landkarte der KI in der folgenden Abb. 2.1.

Bewusst haben wir die Fähigkeiten in Anführungszeichen gesetzt, um nicht zu suggerieren, dass KI-Anwendungen wirklich lernen, denken, wissen und handeln können wie Menschen. Dennoch kann man gewisse funktionale Äquivalente von KI-Funktionen und menschlichen Fähigkeiten sehen, die die sechs Begriffe rechtfertigen. Dargestellte Überlappungen zwischen Kreisen sollen illustrieren, dass diese Fähigkeiten miteinander in Beziehung stehen.

- **„Wahrnehmen“:** Das KI-Teilgebiet *Computer Vision* beschäftigt sich mit der Verarbeitung von Bildern. Ein Anwendungsbeispiel ist die automatische Erkennung von Krankheiten wie Krebs mittels bildgebender Verfahren wie MRT, CTG, Ultraschall oder Röntgen.
- **„Kommunizieren“:** *Natural Language Processing* (NLP) behandelt die Verarbeitung natürlicher Sprache. Beispiel ist die Spracherkennung und Sprachgenerierung in digitalen Assistenten wie Apple Siri oder Amazon Alexa.

Information Retrieval (Informationsgewinnung) umfasst die Suche nach Dokumenten basierend auf einem Informationsbedarf. Anwendungen sind Suchmaschinen wie Google, Yahoo und Bing.

- **„Lernen“:** *Machine Learning* (Maschinelles Lernen) umfasst mathematische Verfahren zur Verallgemeinerung von beobachteten Beispielen. Es wird verwendet, um Vorhersagen zu treffen. Ein Beispiel ist die Klassifikation von Emails als Spam.

Data Mining umfasst mathematische Verfahren zum Erkennen von Mustern in (großen) Datenmengen. Ein Beispiel ist die Gruppierung von Kundendaten.

- **„Wissen“:** *Knowledge Representation* ist das KI-Teilgebiet, welches sich mit der expliziten Modellierung von Wissen beschäftigt. Ein Beispiel ist der Google Knowledge Graph, die Informationsquelle für Infoboxen, welche passend zu Google Suchanfragen angezeigt werden.
- **„Denken“:** *Logische Programmierung* umfasst spezielle Programmiersprachen, mit denen komplexe Regelwerke implementiert werden können, beispielsweise für die Versicherungsbranche.

Probabilistic Reasoning befasst sich mit Schlussfolgern bei unsicherem Wissen, wie dies beispielsweise in der Robotik nötig ist.

Complex Event Processing befasst sich mit der Verarbeitung von fortlaufenden Ereignissen, z. B. Kreditkartenbuchungen. Dabei sollen komplexe Muster in Ereignisfolgen erkannt werden, z. B. zur Betrugserkennung.

- **„Handeln“:** Das KI-Teilgebiet *Planen* beschäftigt sich mit der Optimierung von Handlungsschritten zur Erfüllung von Zielen, z. B. in der industriellen Fertigung.

Agenten-Technologie wird eingesetzt bei komplexen KI-Anwendungen wie autonomem Fahren, die viele KI-Gebiete umfassen. Intelligente Agenten verbinden Komponenten für Computer Vision, Natural Language Processing, Machine Learning, Knowledge Representation etc.

Bei der *Robotik* geht es um die Entwicklung von stationären oder mobilen Robotern sowie deren Programmierung.

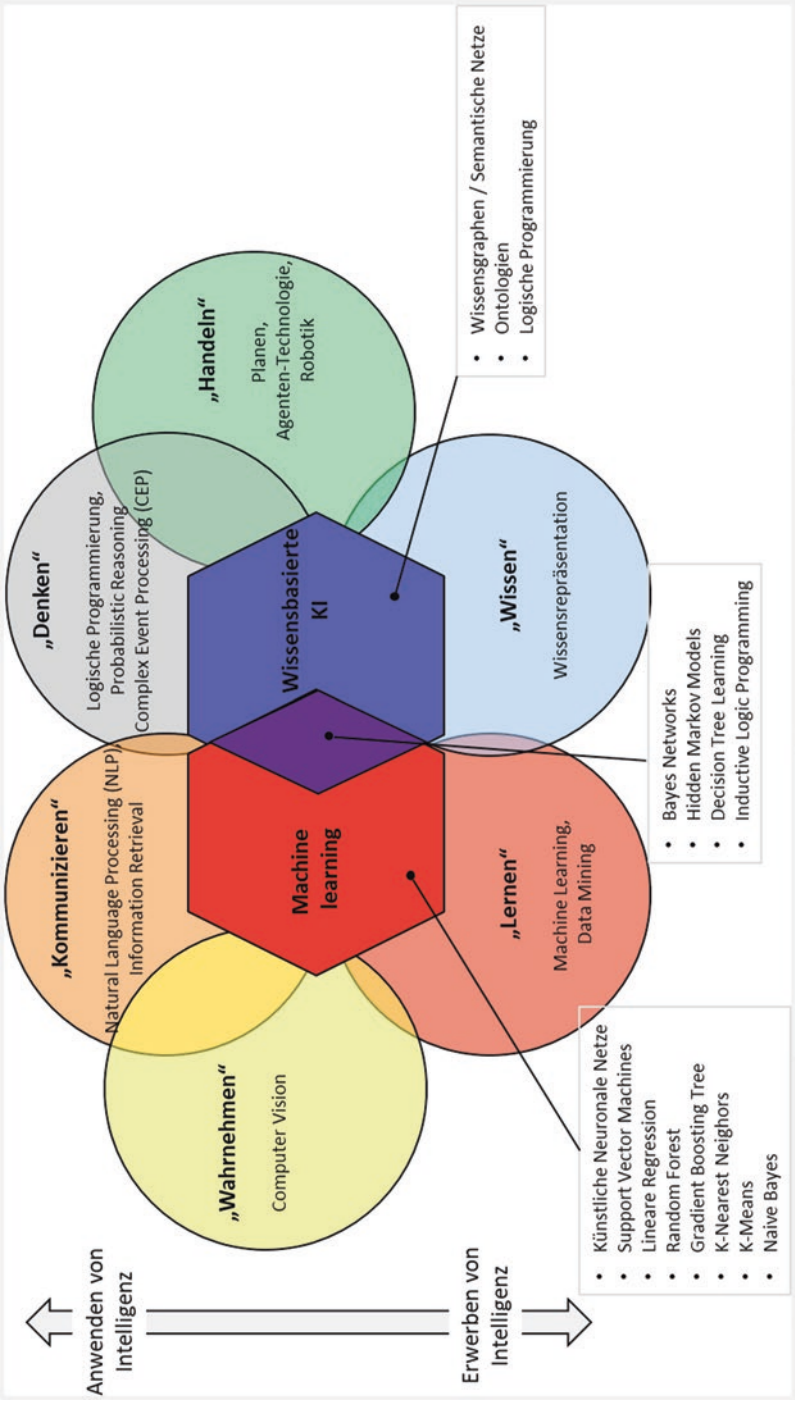


Abb. 2.1 Landkarte der KI (nach Humm 2020)

Die von uns als Teilgebiete der KI dargestellten Aspekte werden in der Fachliteratur nicht einheitlich bezeichnet. Was KI ist, ist – wie in jedem jungen und dynamischen Forschungsfeld – nicht ad hoc uneindeutig definierbar.

Im Laufe der KI-Geschichte haben sich zwei Hauptströmungen von *KI-Verfahren* herausgebildet, welche in der Grafik als Sechsecke dargestellt sind:

- **Wissensbasierte KI:** Hier wird Wissen von menschlichen Experten explizit modelliert in einer Form, die sowohl von Menschen gelesen und verstanden als auch von Computern interpretiert werden kann. Bekannte Verfahren sind *Wissensgraphen / Semantische Netze*, *Ontologien* und *Logische Programmiersprachen* (in der Grafik als Notizen in Rechtecken dargestellt).
- **Machine Learning:** Hier werden Aussagen anhand von numerischen, z. B. statistischen Verfahren berechnet. Häufig kann eine hohe Vorhersagegenauigkeit erzielt werden, aber die Ergebnisse sind für Menschen in der Regel nicht erklärbar.¹ Verfahren sind *Künstliche Neuronale Netze*, *Support Vector Machines*, *Lineare Regression* etc.

Hybride Verfahren vereinen die Vorteile von wissensbasierter KI und Machine Learning. Sie sind in der Grafik durch die Schnittmenge zwischen den beiden Sechsecken dargestellt (violette Raute). Sie umfassen Verfahren wie *Bayes Networks*, *Hidden Markov Models*, *Decision Tree Learning* etc.

In allen Teilgebieten der KI kommt wissensbasierte KI und Machine Learning zum Einsatz. Allerdings überwiegen in den KI-Teilgebieten für die Fähigkeiten Wahrnehmen, Kommunizieren und Lernen die Machine Learning Verfahren, während in den KI-Teilgebieten für Wissen, Denken und Handeln die wissensbasierten Verfahren meist bevorzugt werden.

In den nächsten beiden Abschnitten stellen wir die beiden KI-Hauptströmungen, Machine Learning und wissensbasierte KI, kurz vor.

2.2.2 Machine Learning

Dass das Interesse an der Entwicklung und Nutzung von KI in den letzten Jahren stark zugenommen hat, liegt insbesondere an den Fortschritten, die im Bereich des Machine Learning (ML) erzielt wurden. Im Allgemeinen umfasst Machine Learning Methoden, die Zusammenhänge in Datensätzen erkennen, um darauf basierend Vorhersagen zu treffen (Murphy 2012). Die fast universelle Anwendbarkeit von ML-Algorithmen macht die Nutzung interessant für Wirtschaft und Wissenschaft. Ebenfalls kann KI einen Beitrag zur Lösung gesellschaftlicher Herausforderungen leisten. So kann ML die Diagnose von Krankheiten unterstützen,

¹ Gängige ML-Verfahren wie künstliche neuronale Netze arbeiten assoziativ, nicht kausal. Das bedeutet, dass Korrelationen erkannt, aber nicht begründet werden können. Derzeit gibt es intensive Bestrebungen, auch ML-Verfahren mit Erklärungskomponenten anzureichern. Dies wird auch als erklärbare KI (explainable AI) bezeichnet.

Therapievorschläge entwickeln oder bei der Entwicklung von Medikamenten helfen. Daher sprechen Erik Brynjolfsson und Andrew McAfee vom MIT auch von der wichtigsten „general purpose technology“ unseres Zeitalters (Brynjolfsson und McAfee 2017).

Im Laufe der Aktivitäten der Arbeitsgruppe wurde von eingeladenen Gästen häufig das Argument vorgebracht, dass Künstliche Intelligenz bzw. Machine Learning nichts Neues seien, da viele der auch heute genutzten Methoden, wie z. B. Künstliche Neuronale Netze oder Support Vector Machines, schon seit Jahrzehnten existieren. Aus diesen Gründen den Schluss zu ziehen, dass es sich nicht lohnt, sich mit KI oder ML zu beschäftigen, wäre aber ein Fehler. Es gibt eine Vielzahl von Beispielen, die zeigen, dass es häufig an den Rahmenbedingungen liegt, ob sich eine Technologie durchsetzt oder nicht – und dass es nicht die Technologie „an sich“ ist, die zu betrachten ist. Das prominenteste Beispiel ist vielleicht das Internet.

In den letzten Jahren haben sich die Rahmenbedingungen für die Nutzung von KI- bzw. ML-Algorithmen stark geändert. Diese Rahmenbedingungen unterstützen dabei die Durchsetzung und Verbreitung dieser Typen von Algorithmen. Insbesondere sind die folgenden vier Punkte zu nennen:

- Erstens sind Daten heute in einer nie gekannten Menge verfügbar – sowohl im Internet als auch in Unternehmen. Diese Daten sind die Grundlage für den Einsatz von ML-Verfahren. Darüber hinaus existieren Datenplattformen, wie z. B. Kaggle, die Daten für eine Vielzahl von möglichen KI-Anwendungen bereitstellen.
- Zweitens sind Rechenleistung und Speicherplatz so kostengünstig wie noch nie zuvor und können von Cloud-Anbietern problemlos bezogen werden. Möglicherweise werden Entwicklungen im Bereich des Quanten-Computing in der Zukunft weiteren rechenintensiven KI-Anwendungen zum Durchbruch verhelfen.
- Drittens hat sich die Performance von ML-Algorithmen in den letzten Jahren verbessert. Insbesondere gilt das für die Technik des Deep Learning, welche mehr und mehr Anwendung findet.
- Viertens existieren viele kostenlos verfügbare Toolkits und Bibliotheken zur Entwicklung von KI-Anwendungen. Beispiele sind Scikit-learn, Apache Spark MLlib, Keras, CNTK, PyTorch oder TensorFlow (siehe hierzu auch Buxmann und Schmidt 2019). Die meisten dieser Werkzeuge stehen unter einer Open-Source-Lizenz. Diese Werkzeuge erleichtern die Entwicklung von ML-Anwendungen enorm. So können mit Hilfe von Tools wie Tensorflow oder Scikit-learn beispielsweise Methoden wie Künstliche Neuronale Netze vergleichsweise einfach in eine Software eingebunden werden. Darüber hinaus existieren Frameworks, wie RapidMiner, die den gesamten Entwicklungsprozess, einschließlich Aufgaben wie die Modellierung sowie die Aufarbeitung, Bereinigung und Visualisierung von Daten, unterstützen.

Zusätzlich haben sich die Möglichkeiten zur Nutzung von ML-Algorithmen auch dadurch vereinfacht, dass Anbieter wie beispielsweise Google, IBM, Microsoft oder Amazon mittlerweile KI-Services auf Basis eines Pay-per-Use-Zahlungsmodells anbieten. Das bedeutet, Anwender können Dienste, wie z. B. zur Umwandlung von Stimme in Text oder zur Erkennung von Objekten über ein

Software-as-a-Service-Modell beziehen. Es entwickeln sich also Geschäftsmodelle rund um den ML-Einsatz, was die Nutzung und Verbreitung zukünftig weiter befördern wird.

Methoden zur Messung der Qualität von ML-Algorithmen

Natürlich führt die Nutzung von ML-Algorithmen nicht immer zu guten Ergebnissen. Die Qualität ist abhängig von dem gewählten Verfahren sowie der Datenbasis, d. h. der Trainingsdaten. Vor diesem Hintergrund stellt sich die Frage nach der Anwendung von Verfahren zur Messung der Qualität. Das Grundprinzip dieser Messung soll am Beispiel von Klassifikationsproblemen erläutert werden. Hierzu zählen beispielsweise Objekterkennung auf Bildern (Dalal und Triggs 2005) oder medizinische Diagnosen (Kononenko 2001). Um zu prüfen, wie gut ML-Lösungen für solche Probleme funktionieren, kann die so genannte *Confusion Matrix* aufgestellt werden.

Dies lässt sich am einfachsten anhand eines Beispiels erklären (Buxmann und Schmidt 2021): Es soll eine trainierte ML-basierte Anwendung evaluiert werden, die auf Grundlage von Gesundheitsdaten ausgibt, ob ein Patient an Krebs erkrankt ist oder nicht. Die Anwendung gibt folglich an, in welche der Klassen – „krebskrank“ versus „nicht krebskrank“ – ein Patient eingeordnet wird. Um zu verstehen, wie gut diese Klassifizierung funktioniert, werden echte Diagnosen als Vergleichswerte herangezogen, in denen vermerkt ist, ob ein Patient wirklich an Krebs erkrankt ist oder nicht. Es stehen folglich für jeden Patienten des Testdatensatzes jeweils die tatsächliche Klasse und die von der Anwendung vorhergesagte Klasse zur Verfügung. Werden anschließend die beiden Klassen eines Patienten verglichen, liegt stets einer der folgenden vier Fälle vor (Fawcett 2006):

- Richtig-Positiv (RP): Der Patient ist tatsächlich krebserkrankt und wird auch so von der Anwendung klassifiziert.
- Richtig-Negativ (RN): Der Patient ist tatsächlich nicht krebserkrankt und wird auch so von der Anwendung klassifiziert.
- Falsch-Positiv (FP): Der Patient ist tatsächlich nicht krebserkrankt, aber die Anwendung klassifiziert den Patienten fälschlicherweise als krebserkrankt.
- Falsch-Negativ (FN): Der Patient ist tatsächlich krebserkrankt, aber die Anwendung klassifiziert den Patienten fälschlicherweise als nicht krebserkrankt.

Anschließend wird gezählt, wie oft diese vier Fälle in den klassifizierten Testdaten vorkommen, um die Confusion Matrix zu erstellen. Abb. 2.2 a) zeigt den allgemeinen Aufbau einer Confusion Matrix. Sie besteht aus den Dimensionen „tatsächlicher Wert“ und „vorhergesagter Wert“, die die Daten in die obigen vier Fälle unterteilen. Jede der vier resultierenden Zellen spiegelt die Anzahl des jeweiligen Falls wider. Durch das Ablesen der vier Werte kann grundlegend eingeschätzt werden, wie gut die Anwendung Klassen richtig zuordnen kann (Fawcett 2006). Aus Abb. 2.2 b) wird beispielsweise ersichtlich, dass das entsprechende Modell relativ viele Patienten als krebskrank klassifiziert hat, die tatsächlich nicht krebskrank sind (150 Patienten wurden richtigerweise als krebskrank klassifiziert, während 60 fälschlicherweise als krebskrank klassifiziert wurden). Dafür wird aber auch er-

		Tatsächlich	
		Positive Klasse	Negative Klasse
Vorhergesagt	Positive Klasse	#RP	#FP
	Negative Klasse	#FN	#RN

a)

		Tatsächlich	
		krebsskrank	nicht krebsskrank
Vorhergesagt	krebsskrank	150	60
	nicht krebsskrank	11	300

b)

Abb. 2.2 (a) Allgemeiner Aufbau einer Confusion Matrix; (b) beispielhafte Confusion Matrix. (Quelle: Buxmann und Schmidt 2021)

sichtlich, dass das Modell nicht krebssranke Patienten relativ gut klassifizieren kann (300 Patienten wurden richtigerweise als nicht krebsskrank klassifiziert während nur 11 fälschlicherweise als nicht krebsskrank klassifiziert wurden).

Die Confusion Matrix kann also eingesetzt werden, um einen Einblick in die Verteilung von tatsächlichen und vorhergesagten Werten zu erhalten. Unternehmen können die Informationen nicht nur nutzen, um zu verstehen, wie viele Fehler eine Anwendung insgesamt produziert, sondern vor allem, um potenzielle Fehlerarten steuern zu können. So können je nach Relevanz ML-Algorithmen während der Entwicklung so parametrisiert werden, dass eine Art der potenziellen Fehler reduziert wird, was jedoch gewöhnlich zu einer Steigerung der anderen Fehlerart führen wird. Wenn dem Unternehmen beispielsweise wichtiger ist, dass wenige FP erzeugt werden, dann kann der Algorithmus stärker auf diese Fehlerart fokussiert werden, was jedoch zu einem erhöhten Aufkommen von FN führen kann.

Eine Confusion Matrix repräsentiert allerdings noch keinen konkreten Zielwert, auf dessen Grundlage eine ML-Lösung optimiert werden kann. Basierend auf den Werten der Confusion Matrix können konkretere Qualitätsmaße aufgestellt werden. Hierzu zählen vor allem Accuracy, Precision, Recall, F1 Score sowie die ROC-Curve, welche unterschiedliche Vor- und Nachteile mit sich bringen (Fawcett 2006; Powers 2011).

Bias im Machine Learning

Bei der Nutzung von ML-Anwendungen ist zu berücksichtigen, dass die Ergebnisse verzerrt sein können. Diese Verzerrungen sind in der Regel auf die verwendeten Trainingsdaten zurückzuführen und können in Label-, Feature-, und Sample Bias unterschieden werden (Buxmann und Peters 2020).

Diese Verzerrungen können beispielhaft anhand eines Systems für die Vorhersage von Rückfallkriminalität veranschaulicht werden. Um die Wahrscheinlichkeit vorherzusagen, dass ein Strafgefangener rückfällig wird, werden z. B. in den USA intelligente Systeme eingesetzt, deren Einschätzungen auf historischen Verhaftungsstatistiken beruhen (Corbett-Davies und Goel 2018). Das System analysiert die Merkmale des Strafgefangenen (eingehende Variablen), z. B. demografische Daten

oder seine kriminelle Vergangenheit (z. B. historische Verhaftungsdaten) und trifft dann eine Vorhersage darüber, wie hoch das Risiko eines kriminellen Rückfalls ist (Zielvariable). Auf Basis dieser Vorhersage kann dann beispielsweise die Entscheidung getroffen werden, ob der Gefangene auf Bewährung freigelassen wird.

Beim Label Bias treten Verzerrungen in der vorherzusagenden Zielgröße auf. Oftmals ist die gewählte, beobachtbare Zielgröße eine fehlerhafte Näherungsvariable für die tatsächlich gewünschte Zielgröße. Im oben genannten Beispiel ist die tatsächlich gewünschte Zielgröße eine Risikoeinschätzung über die Durchführung einer erneuten Straftat im Falle einer Freilassung. Da die Vorhersage auf historischen Verhaftungsdaten beruht, wird jedoch implizit die Wahrscheinlichkeit einer erneuten Verhaftung im Falle der Freilassung prognostiziert statt dem Risiko über die Durchführung einer neuen Straftat. In Studien wurde jedoch gezeigt, dass Verhaftungsstatistiken für Drogenkriminalität verzerrt sein können, beispielsweise zum Nachteil von Minderheiten wie Hispano- und Afroamerikanern (Lum und Isaac 2016), sodass die tatsächliche Kriminalitätsrate nicht ausreichend genau abgebildet wird. Die Verzerrungen beruhen beispielsweise darauf, dass Stadtviertel, in denen bestimmte Minderheiten stärker vertreten sind, von der Polizei stärker überwacht werden und somit die Wahrscheinlichkeit der Verhaftung für diese Minderheiten steigt.

Neben dem Label Bias kann im gewählten Beispiel auch ein Feature Bias auftreten. Dieser liegt beispielsweise dann vor, wenn eine im Trainingsdatensatz enthaltene Variable unterschiedliche Vorhersagekraft für bestimmte Bevölkerungsgruppen hat. Die Entwickler des Systems aus dem obigen Beispiel könnten auf Basis allgemeiner Kriminalitätsstatistiken beispielsweise annehmen, dass Langzeitarbeitslosigkeit die Wahrscheinlichkeit einer zukünftigen Straftat erhöht. Wird der Beschäftigungsstatus nun als Variable in das Modell aufgenommen, wird der Algorithmus diese scheinbar allgemeingültige Korrelation verwenden, um eine Vorhersage darüber zu treffen, wie hoch die Wahrscheinlichkeit einer erneuten Verhaftung von Strafgefangenen ist. Nehme man jetzt an, dass die getroffene Annahme für eine bestimmte Bevölkerungsminderheit, z. B. Hispanoamerikaner, in der Wirklichkeit nicht zutrifft. Das gelernte Modell wird die gelernte Korrelation jedoch trotzdem anwenden, arbeitslosen Hispanoamerikanern fälschlicherweise ein höheres Rückfallrisiko zuweisen und sie dadurch benachteiligen.

Der Sample Bias wird dadurch hervorgerufen, dass ungenügende Daten für das Training ausgewählt werden. Dieser Bias entsteht, wenn die Population in den Trainings-Daten die tatsächliche Population im Einsatzbereich nicht adäquat repräsentiert, sondern mindestens eine relevante Gruppe über- oder unterrepräsentiert ist (Heckman 1979). Im Falle des vorangegangenen Beispiels würde dies bedeuten, dass ein System zur Vorhersage möglicher Kriminalitätsrückfälligkeit von Strafgefangenen in einem spezifischen US-Bundesstaat trainiert wurde, jedoch in der gesamten USA eingesetzt wird. Hier liegt das Problem darin, dass sich die Demografie der einzelnen Bundesstaaten grundlegend voneinander unterscheiden kann. Letztendlich kann ein intelligentes System nur dann richtig funktionieren, wenn es zuvor genügend Daten zu allen Fällen bekommen hat, die es später klassifizieren soll.

Label-, Feature-, und Sample Bias werden zusammenfassend als Data Bias bezeichnet. Data Bias ist aber nur eine von mehreren Formen von Bias im Machine Learning (siehe Abb. 2.3).

Data Bias ist also durch Verzerrungen in den Daten verursacht. Diese können beispielsweise aus der unterschiedlichen Aktivität von Benutzergruppen resultieren (Activity Bias). So stammen die meisten verfügbaren Medizindaten aus westlichen Industrienationen, die jedoch nur einen kleinen Prozentsatz der Weltbevölkerung ausmachen. Damit verwandt ist der Sampling Bias, nämlich die Auswahl an existierenden Daten für das Trainieren von ML Anwendungen. Algorithmic Bias repräsentiert Verzerrungen, die durch ML-Algorithmen hervorgerufen werden. Da alle ML-Verfahren Näherungsverfahren sind, kann es verzerrungsfreie ML-Algorithmen prinzipiell nicht geben. In der Interaktion zwischen ML-Anwendung und Anwender kann es Bias geben, besonders wenn die inhärente Unsicherheit von ML-Verfahren dem Anwender nicht deutlich gemacht wird und Prognosen der ML-Anwendung als Tatsachen suggeriert werden. Schließlich kann die Nutzung von ML-Anwendungen neue Daten generieren, welche möglicherweise wiederum verzerrt sind. Im schlimmsten Fall kann dies zu einem Teufelskreis führen (Second Order Bias).

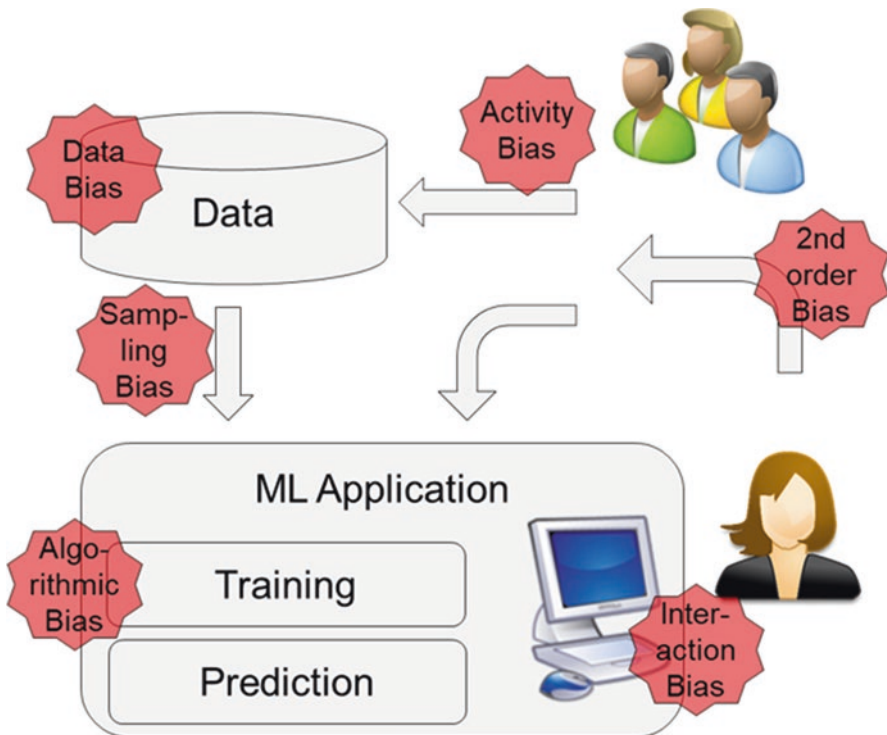


Abb. 2.3 Bias im Machine Learning (aus Humm 2020)

2.2.3 Wissensbasierte KI

In der Populärliteratur wird KI heute manchmal mit Machine Learning gleichgesetzt. Das ist eine stark verkürzte Sicht, denn wissensbasierte KI hat eine ebenso große Bedeutung in Forschung und Praxis.

Expertensysteme

Erste praktische Relevanz für wissensbasierte KI hatten die so genannten Expertensysteme der 1970er- und 1980er-Jahre, noch lange bevor Machine Learning eine Bedeutung im Alltag erlangte. Noch heute unterstützen die Expertensysteme beispielsweise bei medizinischen Diagnosen oder der Fehlersuche in IT-Systemen. Ein Expertensystem löst für Menschen komplexe Probleme in einem begrenzten Fachgebiet. Es agiert quasi als Experte und unterstützt durch Handlungsempfehlungen. Die Problemlösungen und Handlungsempfehlungen leitet es aus einer Wissensbasis ab (Luber und Litzel 2019).

Expertensysteme kommen in vielen verschiedenen Bereichen und Fachgebieten zum Einsatz. Typische Anwendungen finden sich dort, wo es an Experten fehlt oder die vorhandenen Experten von der Auswertung großer Datenmengen entlastet werden sollen. Darüber hinaus erhöhen Expertensysteme durch die unmittelbare Bereitstellung von Lösungen die Sicherheit in kritischen Situationen oder verbessern die Qualität eines Produktes. In der Medizin unterstützen Expertensysteme bei Diagnosen oder der Auswertung von Röntgenaufnahmen. Ein weiterer Anwendungsbereich ist in der Chemie zu finden. Dort analysieren Expertensysteme die Struktur chemischer Verbindungen oder unterstützen bei organischen Synthesen. Ebenfalls Anwendungen für Expertensysteme sind geologische Erkundungen, die militärische Aufklärung, Erdölbohrungen, Erdbebenvorhersagen, die Umweltentwicklung oder die Überwachung und Steuerung von Kernreaktoren (Luber und Litzel 2019).

Die Wissensbasis repräsentiert das formalisierte Expertenwissen meist in Form von Wenn-dann-Regeln. Expertensysteme sind in der Lage, das aus Fakten und Regeln bestehende Wissen zu interpretieren und Schlussfolgerungen abzuleiten. Hierfür sind sie mit einer Inferenzmaschine ausgestattet, die entscheidet, wie und in welcher Reihenfolge oder Form die Regeln zur Lösung eines Problems herangezogen werden. Über eine sogenannte Erklärungsmaschine macht das Expertensystem das Zustandekommen der Problemlösungen und Handlungsempfehlungen dem Menschen gegenüber verständlich. Moderne Expertensysteme bestimmter Fachgebiete sind in ihrem Problemlösungsverhalten mit den Leistungen menschlicher Experten vergleichbar oder übertreffen sie in Teilbereichen sogar. Typische Aufgaben für Expertensysteme sind:

1. Interpretation von Daten durch den Vergleich von Soll- und Ist-Werten
2. Klassifikation von Ereignissen
3. Konfiguration komplexer Systeme unter Berücksichtigung verschiedener Bedingungen
4. Erkennen von Fehlerursachen und Reduzierung von Arbeitsfehlern
5. Beseitigung kritischer Zustände durch das Einleiten von Aktionen

6. Planung einer Folge von Aktionen zur Erreichung eines bestimmten Ziels
7. Dialogorientierte, fachspezifische Beratung von Menschen
8. Vorhersage von Ereignissen auf Basis bestimmter Geschehnisse.

Ein Expertensystem besteht u. a. aus folgenden Komponenten: eine Wissenserwerbskomponente, die Wissensbasis, eine Inferenzmaschine, eine Erklärungskomponente, sowie die Schnittstelle für den Benutzerdialog. Eine zentrale Rolle übernehmen die Wissensbasis und die Inferenzmaschine. In der Wissensbasis ist das komplette problembezogene Wissen gespeichert. Auf Basis dieses Wissens generiert die Inferenzmaschine mittels Schlussfolgerungen Problemlösungen. Sie verknüpft hierfür Fakten und Regeln und generiert daraus neues Wissen. Diese präsentiert das System über den Benutzerdialog dem Anwender und erklärt ihr Zustandekommen mithilfe der Erklärungskomponente. Die Wissenserwerbskomponente gestattet die Erweiterung der Wissensbasis. Neue Fakten und Regeln lassen sich zum vorhandenen Wissen hinzufügen. Die Wissenserwerbskomponente kann darüber hinaus die Vollständigkeit und Konsistenz des gespeicherten Wissens prüfen (Luber und Litzel 2019).

Wissensnetze

Die Wissensbasen von wissensbasierten KI-Systemen werden häufig als sog. *Wissensnetze* (Knowledge Graph)² dargestellt. Wissensnetze spezifizieren formal die wesentlichen Konzepte eines Fachgebiets, sowie deren Beziehungen zueinander. Abb. 2.4 zeigt beispielhaft einen Ausschnitt aus einem Wissensnetz für die Medizin.

In diesem Wissensnetz-Ausschnitt wird dargestellt, dass Krebs und Allergien Krankheiten sind, sowie Chemotherapien und Immuntherapien Behandlungen sind. Des Weiteren wird dargestellt, dass Chemotherapien für Krebserkrankungen eingesetzt werden können, während Immuntherapien sowohl für Allergien als auch für Krebserkrankungen eingesetzt werden.

Neben solchen Wissensnetzen für spezielle Fachgebiete gibt es auch umfassende Wissensnetze. Mit ca. 70 Mrd. Fakten (Vincent 2016) ist eines der umfangreichsten

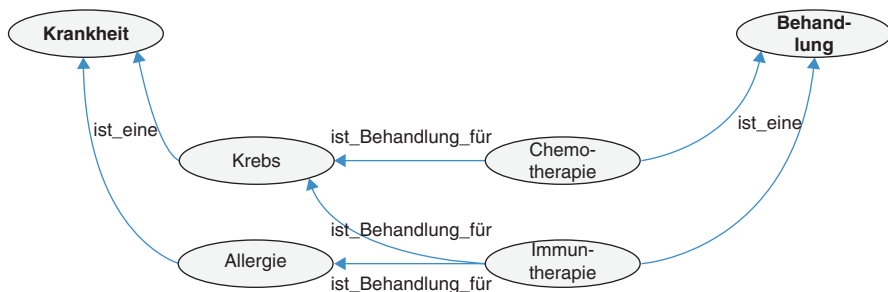


Abb. 2.4 Ausschnitt eines Wissensnetzes für die Medizin (aus Hoppe 2020)

²Für die genaue Abgrenzung von Knowledge Graph, Ontologie, Thesaurus, WordNet und kontrolliertem Vokabular siehe Hoppe 2020, Kap. 4.

und am meisten genutzten Wissensnetze der Google Knowledge Graph. Er ist die Grundlage für alle Info-Boxen, die bei Google-Suchen angezeigt werden. Er wird auch verwendet, um die Suche zu verbessern, sowie allgemeine Fragen zu beantworten. Versuchen Sie mal, in die Google-Suche eine Frage einzugeben wie z. B. „Wann wurde die Mauer gebaut?“; siehe Abb. 2.5 für einen Screenshot der Antwort.

Die Google-Suche assoziiert das Wort „Mauer“ mit der Berliner Mauer und gibt als Ergebnis der Anfrage ein konkretes Datum, den 13. August 1961 zurück. In einer Infobox auf der rechten Seite ist eine kurze Beschreibung der Berliner Mauer, inklusive wichtiger Daten wie der Baubeginn, sowie ein Link auf den Wikipedia-Eintrag angegeben. Die Wissensbasis für all diese Informationen ist der Google Knowledge Graph.

Linked Open Data

Neben kommerziellen Wissensnetzen wie dem Google Knowledge Graph, dem Facebook Graph oder dem Microsoft Graph, existiert eine große Menge freier Wissensnetze. Linked Open Data (LOD) ist eine Initiative des World Wide Web Consortiums (W3C), welche Organisationen Wissensnetze im Web zur freien Verfügung bereitstellt. Durch gegenseitige Verweise zwischen Wissensnetzen soll deren Nutzen noch gesteigert werden. Abb. 2.6 zeigt eine Übersicht über öffentlich verfügbare Wissensnetze, die sogenannte LOD Cloud.

Jeder Knoten in der LOD Cloud (repräsentiert durch einen Kreis in Abb. 2.6) beschreibt ein Wissensnetz, oft mit hunderttausenden Fakten. Verbindungen zwischen Knoten zeigen an, dass die entsprechenden Wissensnetze über Hyperlinks miteinander verbunden sind. Die Wissensnetze stammen aus so unterschiedlichen

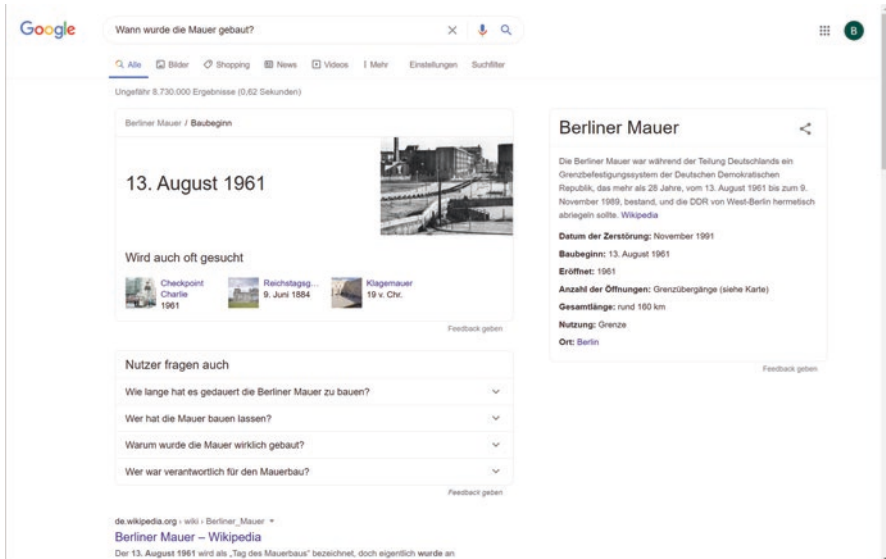


Abb. 2.5 Google Resultat auf die Frage „Wann wurde die Mauer gebaut?“ (Google 2020)

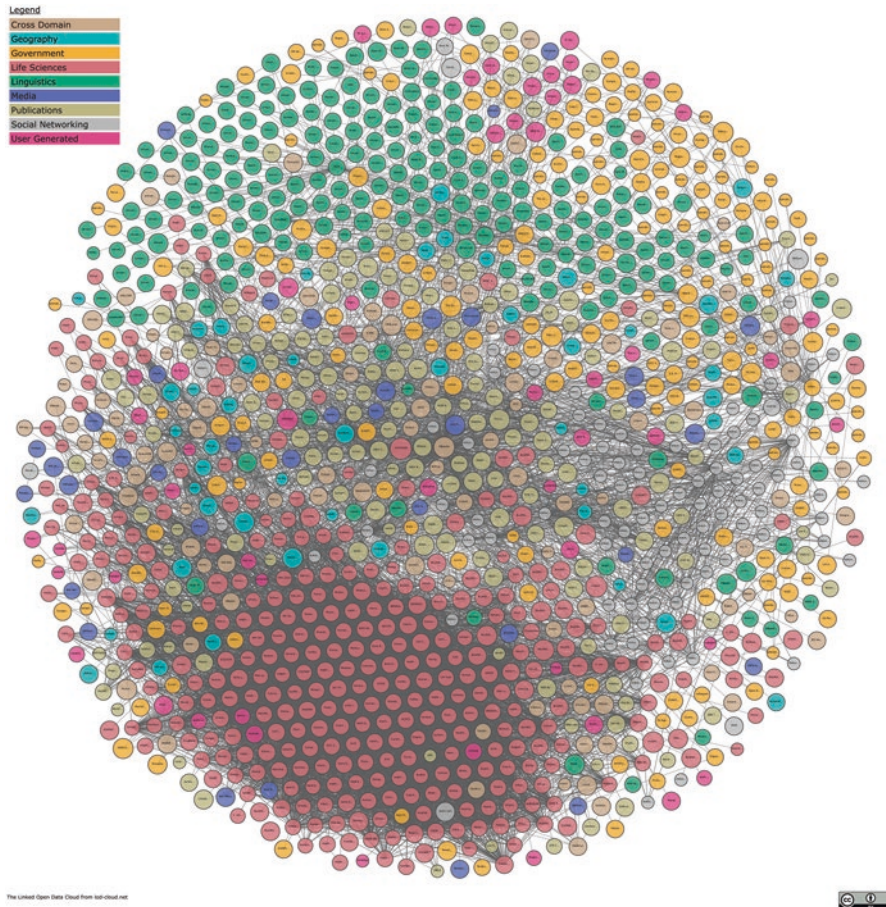


Abb. 2.6 Linked Open Data Cloud (Insight Centre for Data Analytics 2020)

Domänen wie Lebenswissenschaften, Geografie, öffentliche Verwaltung, Medien etc. oder sind domänen-übergreifend.

Besonders im Bereich der Lebenswissenschaften existieren viele, umfangreiche Ontologien zu Krankheiten, Symptomen, Körperteilen, Behandlungen, Medikamenten, Genen, u. v. m. Beispiele sind der NCIt (NCI Thesaurus), MeSH (Medical Subject Headings) und die GO (Gene Ontology, Ashburner et al. 2000; The Gene Ontology Consortium 2019). Zahlreiche umfangreiche domänen-übergreifende Ontologien existieren ebenfalls, z. B. Wikidata, DBpedia oder YAGO. In diesen Ontologien sind Konzepte aus Wikipedia formalisiert, die Verwendung in domänen-übergreifenden KI-Anwendungen finden (Siehe auch Hoppe 2020, Kap. 4).

2.2.4 Zusammenspiel von wissensbasierter KI und Machine Learning

Wissensbasierte KI und Machine Learning ergänzen sich gegenseitig. Faktenwissen wird mit Wissensnetzen repräsentiert. Für Problemfelder, in denen keine gesicherte Faktenlage verfügbar ist, wo Unsicherheit und Rauschen vorliegt, wo Entscheidungen dynamisch an Nutzerverhalten angepasst werden soll, da spielt Machine Learning seine Stärken aus. Der Nachteil der meisten Machine-Learning-Verfahren ist jedoch, dass sie wie eine Black Box funktionieren und Vorhersagen meist nicht begründet werden können. Hier liegt wiederum die Stärke von wissensbasierten Verfahren, da Aussagen aufgrund der Wissensbasis begründet werden können. Jedoch ist die Erstellung und Pflege einer Wissensbasis durch Experten aufwändig.

Auch die KI unterliegt Moden und Strömungen. Während in den 1970–1990er-Jahren die wissensbasierte KI dominant war und Machine Learning eher eine Nischenrolle einnahm, hat sich das in den letzten beiden Jahrzehnten eher umgedreht. In der Populärliteratur wird sogar manchmal KI mit Machine Learning gleichgesetzt. Aber in letzter Zeit werden zunehmend auch Probleme mit Machine Learning in den Blick genommen, wie z. B. Bias (s. o.) oder die fehlende Erklärbarkeit von Entscheidungen.

So bildet sich aktuell ein neuer Forschungsschwerpunkt „Explainable AI“ (Erklärbare KI) heraus, in dem die Vorteile beider KI-Ansätze kombiniert werden sollen. Wir sehen solche hybriden Ansätze als zukunftsweisend an, weil sie die jeweiligen Stärken der beiden Ansätze verbindet, sowie deren Schwächen adressiert.

2.2.5 Die Rolle von Algorithmen in KI-Systemen

Neben der unpräzisen Gleichsetzung von KI und Machine Learning stellt auch die populäre Verwendung des Begriffs „Algorithmen“ für KI-Systeme eine verkürzte Sicht dar, die klärungsbedürftig ist: Ein Algorithmus ist eine Handlungsvorschrift zur Lösung eines Problems oder einer Klasse von Problemen, z. B. das Sortieren von Elementen einer Liste. Ein Algorithmus ist vergleichbar mit einem Kochrezept, in dem Schritt für Schritt beschrieben wird, wie ein Gericht hergestellt werden soll. Aber genauso wenig, wie ein Kochrezept das Gleiche wie ein Abendessen ist, ist ein Algorithmus das Gleiche wie ein KI-System. Ein Kochrezept macht nicht satt. Für ein Abendessen benötigt man Köche, welche viele Zutaten gemäß der einer Vielzahl von Kochrezepten mit Hilfe von Werkzeugen wie Messern, Töpfen, Backöfen und Herden und mit Einsatz von Energie zubereiten und anrichten. Erst dann kann gegessen werden.

Um ein KI-System zu entwickeln, verwenden Software-Entwickler sogenannte Software-Bibliotheken: konkrete Software-Implementierungen von hunderten und tausenden von Algorithmen mittels konkreter Programmiersprachen. Für eine konkrete KI-Anwendung wird eine Software-Architektur entwickelt: ein Diagramm,

welches beschreibt, wie verschiedene Implementierungen aus Software-Bibliotheken für einen bestimmten Anwendungszweck zusammengeschaltet werden sollen. Basierend auf der Software-Architektur wird die KI-Anwendung implementiert und anschließend auf Computer-Hardware ausgeführt. Danach muss die KI-Anwendung mit Daten trainiert oder um eine Wissensbasis angereichert werden. Erst dann kann sie auf einer Hardware (z. B. einem Roboter) ausgeführt werden und kann Nutzen für Anwender stiften. Ein Algorithmus stiftet hingegen für einen Endanwender keinerlei Nutzen – genauso wie ein Kochrezept nicht satt macht.

Wir vermeiden daher in dieser Studie die Verwendung des Begriffs „Algorithmen“ (außer im Sinne der Definition) und sprechen von KI-Anwendungen bzw. -Systemen, wie oben bereits eingeführt.

2.3 Anwendungen von KI

Nachfolgend stellen wir beispielhaft KI-Anwendungen im Umfeld der Forschung von drei Wissenschaftsdomänen vor: Physik, Klimaforschung und Medizin.

2.3.1 Teilchenphysik³

KI, Machine Learning und Big Data sind in der Physik und in der physikalischen Grundlagenforschung seit Jahrzehnten wohletabliert. Die Teilchen- bzw. Hochenergiephysik bildet ein weitreichendes Anwendungsfeld, in dem diese Methoden und Instrumente in Kooperation mit Informatikern nicht nur angewendet, sondern auch weiterentwickelt werden. Dieses Feld der Physik ist deshalb so beispielhaft für Forschung allgemein, weil in Experimenten große Datenmengen anfallen, deren Informationsgehalt zunächst unklar ist. Hier unterstützen KI- und Machine Learning-Verfahren seit über drei Dekaden die Grundlagenforscher/-forscherinnen dabei, aus diesen Daten Informationen zu gewinnen, Theorien zu entwickeln und Wissen zu generieren. Insofern zukünftig auch in anderen Wissenschaftsdisziplinen und Forschungsfeldern große Datenmengen erzeugt und zugänglich werden, die zur Erkenntnisgewinnung genutzt werden können, kann die Teilchenphysik als paradigmatisch angesehen werden: allgemeine Entwicklungstrajektorien kristallisieren sich vermutlich in kaum einem anderen Feld so prägnant heraus wie hier.⁴

³Dieser Abschnitt basiert in Teilen auf dem Vortrag von P. Mättig, Universität Bonn, zum Thema „Big Data und maschinelles Lernen am LHC“ beim Workshop Digitale Arbeitswelten in Forschung und Entwicklung, Bad Neuenahr-Ahrweiler, 02.10.2019.

⁴Auch aus historischer Perspektive gilt dies für die Physik, aber auch insbesondere für das CERN: Am CERN sind von Physikern grundlegende Aspekte des World Wide Web entwickelt worden (Gruppe um Tim Berners-Lee u. a.): Sprache HTML, Transferprotokoll http, die URL Signaturen, der erste robuste Browser, u. a.

Die weltweit größten und aufwendigsten Experimente der Naturwissenschaften werden am Large Hadron Collider (LHC) am Europäischen Kernforschungszentrum CERN bei Genf durchgeführt. In dem Teilchenbeschleuniger werden sehr hohe Energien (mit fast Lichtgeschwindigkeit) erzeugt, um in die grundlegende Struktur der Materie einzudringen und so fundamentale Erkenntnisse über den Aufbau, Struktur und Entwicklung der Materie zu erhalten. Derart hohe Energien waren auch kurz nach dem Urknall präsent; der Teilchenbeschleuniger simuliert mithin experimentell Aspekte des Urknalls. Damit ist die Teilchenphysik auch für die Wissenserzeugung der Kosmologie und für ein grundlegendes Verständnis des Universums relevant. Im Zentrum der Teilchenphysik steht seit einigen Jahrzehnten das so genannte Standardmodell der Elementarteilchen. Dieses Modell kann als Theoriekern der Teilchenphysik angesehen werden. Es bündelt alle wesentlichen Erkenntnisse der Physik in einem kompakten Modellrahmen, insbesondere die Existenz und Eigenschaften der Elementarteilchen. Diese umfassen (a) die Fermionen, also „Materieteilchen“ wie Leptonen, z. B. das bekannte Elektron, sowie die Quarks, aus denen z. B. Protonen aufgebaut sind, und (b) die Bosonen, also „Feldteilchen“ wie die Eichbosonen, d. h. „Fundamentalkräfteteilchen“, sowie das Higgs-Boson. Die Teilchenphysik rückte vor 10 Jahren erneut in die Schlagzeilen, als am Large Hadron Collider im Jahre 2012 das schon in den 1960er-Jahren vorhergesagte Higgs-Boson experimentell nachgewiesen werden konnte. Durch diesen Fund – genauer: durch Rekonstruktion zentraler Eigenschaften des Higgs-Bosons als spezielle Signaturen in den großen Datenmengen der Detektorsignale – gilt das Standardmodell im Prinzip als experimentell hinreichend bestätigte, vollständige Beschreibung der Grundlagen der Materie und der subatomaren Welt. Das Standardmodell beschreibt fast alle beobachteten subatomaren Phänomene und kann als eine sehr genau getestete Theorie angesehen werden.

Doch es gibt offene Fragen, die die Physik beschäftigen – und die, unter Anerkennung der Leistung des Standardmodells, gleichwohl zu einer Suche nach einer „Physics beyond the Standard Model“ geführt haben. Eine der offenen Fragen betrifft die so genannte Dunkle Materie, die vom Standardmodell nicht erfasst wird, obwohl sie mehr als 20 % des Kosmos ausmacht. Bisher ist sie allerdings nur durch Gravitationseffekte, also durch ihre Gravitations-Wechselwirkung, nachgewiesen. Nicht ausgeschlossen ist, dass Dunkle Materie auch nicht-gravitativ wirkt und etwa im Rahmen schwacher Wechselwirkung charakterisierbar ist. Fragen nach dem Anteil und den Eigenschaften Dunkler Energie schließen sich hier an.⁵ Zudem ist mit dem Standardmodell noch keine hinreichende Vereinheitlichung der vier fundamentalen Kräfte, d. h. der vier Wechselwirkungen erreicht. Dies betrifft teilweise die Zusammenführung von starker und elektroschwacher Kraft, aber insbesondere die Integration der Gravitationskraft in einen umfassenden und widerspruchsfreien

⁵ Auch wenn man beides als Falsifikation des Standardmodells ansehen könnte, wird das Standardmodell von Physikern keinesfalls verworfen, sondern es wird daraus ausgegangen, dass dieses in einen neuen und größeren Theorierahmen, d. h. in eine umfassendere Theorie, eingebettet werden kann, ähnlich wie Newtonsche Gravitationstheorie in die Allgemeine Relativitätstheorie.

Theorierahmen. Dies zeigt, dass die fast als klassisch zu bezeichnende Problematik der Physik, nämlich die Zusammenführung von Quantenmechanik bzw. Quantenfeldtheorien mit der Allgemeinen Relativitätstheorie (als zentraler Gravitationstheorie) noch nicht erreicht ist. Im Prinzip war die Vereinheitlichung im Sinne der Entwicklung eines kohärenten und konsistenten Theoriekorpus stets ein grundlegendes Erkenntnisideal der Physik, das das physikalische Forschungshandeln geprägt hat. Heute sind indes auch kritische Stimmen zu hören, die das Erkenntnisideal einer vereinheitlichten Theorie hinterfragen und eine neue Ära einer „Prediction without Explanation“ heraufziehen sehen.

Um eine „Physik jenseits des Standardmodells“ zu entwickeln, ist es – jenseits der Existenz und einiger Eigenschaften der dunklen Materie – notwendig, weitere empirische Abweichungen vom Standardmodell zu finden und so den ersten Schritt über das Standardmodell hinauszugehen. Auf theoretischer Seite gibt es unterschiedliche Ansätze, eine Vielzahl von Modellen und Theoriefragmenten, wie eine solch neue Physik aussehen könnte. Doch bis dato sind zwischen den Ableitungen bzw. Prognosen des Standardmodells einerseits sowie jenen der Konkurrenten der neuen Physik experimentell kaum Differenzen zu beobachten. Das könnte sich natürlich zukünftig ändern. Auf dem weiteren Weg zu einer „Physik jenseits des Standardmodells“, darin scheinen sich die meisten Physiker einig zu sein, wird KI, Machine Learning und Big Data eine kaum zu überschätzende Rolle spielen. Das betrifft in erster Linie die Entdeckung und Rekonstruktion von Signaturen in sehr großen Datenmengen, die auf neue Teilchen und ihre Eigenschaften hindeuten könnten – und weniger das Ziel, theoriebasierte Erklärungen zu liefern.

Nun wurden KI- und Machine Learning-Verfahren von Teilchenphysikern auch bisher vielfach eingesetzt, etwa zur Entdeckung des Higgs-Bosons. Derartige Elementarteilchen liegen nämlich nicht an sich vor und müssen nur noch vermessen werden. Vielmehr müssen sie durch hohe Energien zugänglich gemacht bzw., besser gesagt, erst erzeugt werden. Dazu werden Kollision- und Streuexperimente durchgeführt: bei sehr hohen Geschwindigkeiten werden Teilchen zur Kollision gebracht und Streuungen induziert, wobei Elementarteilchen hervortreten oder generiert werden. Die Kollisionen werden von hochtechnischen Detektoren vermessen, die entsprechende Signale auffangen. Dabei entstehen große Datenmengen, die in der Größe von Millionen Terabytes, d. h. Exabytes, liegen. Entscheidend ist die anschließende Datenanalyse mit dem Ziel, Informationen zu erhalten, was bei der jeweiligen Kollision passierte und welche Elementarteilchen in welcher Form beteiligt waren. Es geht dabei um die Rekonstruktion der Existenz und des Verhaltens der Elementarteilchen durch Identifikation ihrer Eigenschaften, d. h. die Elementarteilchen sind operational durch Ereignis-Signaturen definiert, die sich als Muster in den Datenreihen finden (vgl. Aaboud et al. 2018). Diese Muster zu erkennen und mit Bedeutung zu versehen, ist herausfordernd. In praktischer Hinsicht ist eines der zentralen Probleme die Trennung von gesuchtem Elementarteilchen-Signal einerseits und anderen, unspezifischen Signalen sowie dem Hinter- und Untergrund(rauschen) andererseits. Machine Learning-Verfahren können diese Mustererkennung unterstützen und helfen, die relevanten Ereignis-Signaturen zu identifizieren. Dieses Klassifikationsproblem ist ein klassisches Problem, welches ähnlich in vielen empirischen Wissenschaftsdisziplinen auftritt.

Dazu werden im Rahmen von Machine Learning-Verfahren Algorithmen, die in diesem Fall auf künstlichen Neuronalen Netzen basieren, anhand von modellbasierten, simulierten Teilchenkollisionen trainiert. In diesem Verfahren des überwachten Lernens („supervised Machine Learning“) wird gelernt, irrelevante von relevanten Ereignis-Signaturen zu unterscheiden. Damit verbunden ist ein Klassifikationsschema. Erst im Anschluss werden die mit Theorie-Modellen trainierten Machine Learning Algorithmen eingesetzt, um in experimentell erzeugten Datenreihen Muster, wie etwa die Ereignis-Signatur des Higgs-Teilchens, zu identifizieren.

Ein verwandter Fall von Machine Learning, in dem sich die Relevanz des überwachten Lernens in der Teilchenphysik zeigt, bezieht sich auf mathematische Modellierungen und modellbasierte Simulationen im Zusammenhang des Standardmodells. Um konkrete, vor dem Theorie-Hintergrund des Standardmodells entwickelte teilchenphysikalische Modelle zu optimieren, wird Machine Learning zum Auffinden von Parametern oder von Raum- und Zeitdimensionen sowie zum Test von wissenschaftlicher Robustheit verwendet. Ferner unterstützt Machine Learning das Auffinden von raumzeitlichen Skalierungseigenschaften und die Auflösung von Substrukturen, was für die Analyse von experimentellen Datenreihen von Interesse ist.

Es gibt noch ein weiteres, deutlich avancierteres Feld der Teilchenphysik, in dem Machine Learning – und zwar in der Variante des „unsupervised Machine Learning“ (unüberwachten Lernens) – eine Rolle spielt. Insofern sich Grenzen des Standardmodells, wie oben beschrieben, zeigen, gibt es Bestrebungen zu einer „Physik jenseits des Standardmodells“. Dabei weiß niemand, wie diese aussehen könnte. Unklar ist, wonach man in Datenreihen suchen soll. Machine Learning ermöglicht hier im Prinzip, modell- und theorieunabhängig Datenreihen auf Signaturen zu untersuchen – auch solche, die im Standardmodell nicht vorkommen. Vieles ist hier noch Zukunftsmusik, aber es zeigt sich ein zunehmendes Interesse von Teilchenphysikern, diese Verfahren zumindest heuristisch anzuwenden, um neue Entdeckungen zu machen und auf weitere Ideen zu kommen.

Zusammengenommen stellen KI und Machine Learning für die Teilchenphysik methodische Verfahren bereit, die in der Erzeugung von neuem Wissen eine erkenntnisgenerierende Funktion besitzen („Context of Discovery“). Diese Verfahren fördern die Effizienz und Effektivität des physikalischen Erkenntnishandelns. (1) Sie bieten die Möglichkeit, neue physikalische Objekte zu identifizieren, zu klassifizieren bzw. gar diese zu definieren sowie Substrukturen oder Eigenschaften von bekannten und neuen Objekten besser auflösen zu können. (2) Entscheidend ist, dass die informatischen Verfahren eine bessere Trennbarkeit der Signale vom Untergrund ermöglichen und die Signifikanz von Observablen, d. h. der Signalmessgrößen erhöhen. (3) Zudem werden die Verfahren auch im Rahmen von Parametrisierungen und Optimierungen von Simulationen eingesetzt, also zu Bestimmung und Verbesserung der Größen, die in die Modellbildung und anschließende numerische Simulation eingehen. So bilden sie auch eine entscheidende Grundlage für Computerexperimente der Teilchenphysik.

Trotz der breiten Anwendung und des weitreichenden Erfolgs der informatischen Verfahren bleiben Teilchenphysiker insgesamt kritisch gegenüber KI und Machine Learning: Sie verwenden daher einen Teil ihrer Forschungszeit auf die kritische Validierung dieser Verfahren. So bleibt es dabei, dass Korrelationen, Klassifikationen oder Prognosen, wie man sie über KI- und Machine Learning-Verfahren generieren kann, für die grundlagenforschende Physik zwar hilfreich, aber am Ende (als Resultate) zu wenig sind. Es geht (nach wie vor) um kausale Erklärungen und um Einordnung neuer Ereignisse und Phänomene in einen gegebenen Theorierahmen, d. h. um deduktiv-nomologische Erklärungen. Die Physik möchte, ganz traditionell, die Ordnung, Struktur und Entwicklung der Materie und des Universums verstehen.

2.3.2 Klimaforschung⁶

Die Klimaforschung ist traditionell ein Teilgebiet der Physik und der Angewandten Mathematik: Ihr Ursprung liegt in der Kontinuumsmechanik, d. h. in der Fluid- oder Hydrodynamik, die Mitte des 19. Jahrhunderts entstand, und sie verwendet zudem Erkenntnisse der Thermodynamik. Doch auch wenn ihr Ursprung in der Physik und Mathematik liegt, ist sie heute interdisziplinär ausgerichtet. Sie umfasst nicht nur zentrale Fragestellungen der Chemie und den Geowissenschaften, sondern integriert insbesondere die Informatik und einige Technikwissenschaften (u. a. wegen Sensor-, Mess- und Detektortechnik). Und insofern aktuelle Klimaforschung im Horizont des so genannten Anthropozäns betrieben wird und anthropogene Ursachen und Folgen des Klimawandels untersucht, reicht ihre Forschung bis in die Biologie und die Gesellschaftswissenschaften hinein.

Die Klimaforschung ist wissenschaftlich äußerst herausfordernd. Das globale Klimasystem ist nicht – wie üblicherweise in den exakten Naturwissenschaften – abgrenzbar und unter Laborbedingungen kontrollierbar. Phänomene sind selten reproduzierbar. Ein traditionelles Experimentieren ist nicht möglich, ähnlich wie in der Kosmologie und in Teilbereichen von Biologie und Ökologie. Zudem ist das Klimasystem, wie die allgemeine Grundgleichungen der Klimaforschung (Navier-Stokes-Gleichungen) nahelegen, hochgradig nichtlinear und damit vielfach auch sensitiv und instabil. Kleinste Veränderungen von Variablen und Parameter, wie beispielsweise lokale Drücke, Temperaturen oder Strömungsgeschwindigkeiten sowie ihre jeweiligen Gradienten, können große Folgen nach sich ziehen. Systemdynamiken kippen oder bifurkieren: der Golfstrom kann zum Erliegen kommen, Hitze- und Dürreperioden können entstehen. Damit verbunden sind mathematische und numerische Probleme, aber auch Herausforderungen adäquater Messung und Datenerhebung. Allgemein ist für die Klimaforschung herausfordernd, die relevanten Variablen oder

⁶ Dieser Abschnitt basiert teilweise auf dem Vortrag von Hans von Storch zum Thema „Künstliche Intelligenz & Big Data in der Klimaforschung“ auf dem Workshop Digitale Arbeitswelten in Forschung und Entwicklung, Bad Neuenahr-Ahrweiler, 2. Juli 2019.

Zustandsgrößen zu bestimmen, die in einem bestimmten Raumzeitbereich die Klimadynamik bestimmen. Mit anderen Worten, es ist jeweils offen, wie hochdimensional der Zustandsraum ist, in dem die Dynamik beschrieben und sodann, auf Basis der Numerik, durch Computersimulation dargestellt werden kann. Ferner sieht sich die Klimaforschung angesichts voranschreitender Sensorik und der Implementierung sehr vieler neuer Messstationen mit sehr großen und sehr unterschiedlichen Zeitreihen an Daten und ihren jeweiligen Typen konfrontiert.

Angesichts der grundlegenden Nichtlinearität und Instabilität des Klimasystems einerseits und der großen Datenmengen andererseits stellen KI-, Machine Learning- und Big Data-Verfahren ein zukunftsweisendes Instrumentarium zur Verfügung, die wissenschaftlichen Herausforderungen anzugehen. Dies kann anhand von drei Fällen beispielhaft dargestellt werden:

Rekonstruktion des Zustandsraumes, Identifikation lokaler Unterräume und Entwicklung von Prognosen

Wie oben angedeutet, weist das Klimasystem eine große Anzahl an Freiheitsgraden auf. Für die Klimaforschung sind damit eine Vielzahl möglicher Zustandsgrößen – also Variablen, die das Klimasystem beschreiben – verbunden, die schwer zu überblicken sind. Zudem besitzen diese Zustandsgrößen zumeist komplizierte Skalierungseigenschaften: die raumzeitliche Skalenabhängigkeit ist empirisch und mathematisch schwer in den Griff zu kriegen. So ändern sich beispielsweise die für die Klimaforschung relevanten Differenzialgleichungen mit fortschreitender Gitterauflösung, d. h. mit Skalierung strukturell. Das heißt die Differenzialgleichungen, die jeweils partikuläre Modelle für Teile des Klimasystems darstellen, sind nicht immer strukturell stabil. Approximationsmöglichkeiten sind dadurch effektiv reduziert, verbunden mit wesentlichen Prognoseunsicherheiten. Die durch Nichtlinearität und Instabilität/Sensitivität des Klimasystems sowie der dieses repräsentierende mathematische Modell (Differenzialgleichungen) erzeugten Probleme der Approximation zeigen sich beispielsweise bei atmosphärischen Konvektionen oder bei Wolkenbildungsphänomenen. Aufgrund der großen Anzahl der Freiheitsgrade ist, wie angedeutet, der Zustandsraum des Klimasystems grundsätzlich als hochdimensional anzusehen. Im Detail ist das wissenschaftliche Wissen effektiv begrenzt; detaillierte Modelle sind schwer zu entwickeln und zu validieren, auch wenn die Klimaforschung in den beiden letzten Jahrzehnten deutliche Fortschritte verzeichnen kann. So ist es bis dato schwierig, quantitative Zusammenhänge für viele der Zustandsgrößen aufzufinden und zu rechtfertigen. Eine gewisse Abhilfe leistet hier das Konzept der „Principal Interaction Patterns“, das dynamisch weitgehend abgeschlossene niederdimensionale Unterräume bzw. lokale Zustandsräume des Klimasystems zu definieren ermöglicht, die an ihren „Rändern“ mit dem Rest des Zustandsraums stochastisch oder ggf. auch deterministisch-konditioniert in Wechselwirkung treten (Hasselmann 1988). Die Modellparameter bzw. Zustandsgrößen, inklusive der Skalierungseigenschaften dieser Unterräume sind empirischen Verfahren der Quantifizierung ihrer internen Dynamiken zugänglich, mit denen man z. B. bereits die langfristigen stochastischen Oszillationen im südlichen Ozean durch kontinuierliche Frischwasserzufuhr erklären kann (Zhang et al. 2019). Siehe Abb. 2.7.

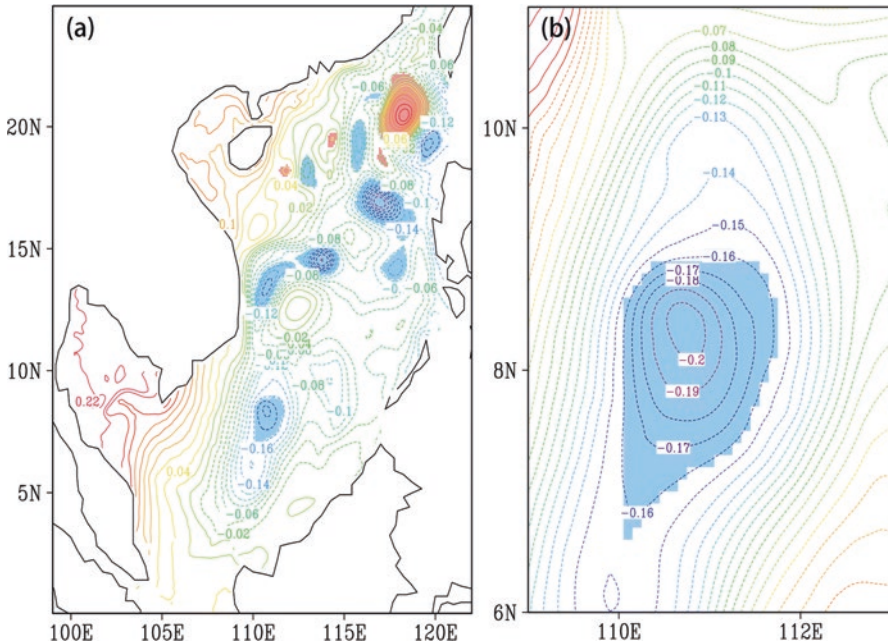


Abb. 2.7 Mustererkennung für Strömungswirbel im Südchinesischen Meer (aus Zhang et al. 2019)

In diesem Beispiel geht es um Mustererkennung für Wirbelbildungen im Südchinesischen Meer. Gesucht wird nach einer Statistik der jährlichen Häufigkeit, Intensität und Größe von wandernden Wirbeln in diesem Meeresabschnitt. Eine entsprechende KI-Anwendung detektiert und charakterisiert anhand des täglich schwankenden Meeresspiegels im Mittel pro Jahr 28 antizyklonische Wirbel und 54 zyklonische Wirbel in dieser Region. Diese Erkenntnisse sind für das Verständnis der regionalen und überregionalen Klimadynamik von zentraler Bedeutung.

Mit der zunehmenden Verfügbarkeit und Verarbeitung großer Datenmengen in Verbindung mit neuen informatischen Verfahren lassen sich nun weitere Unterräume des Klimasystems besser beschreiben und verstehen. Das trifft insbesondere für das Verständnis solcher Unterräume zu, welche auf die Interpretation großer Datenmengen mit hoher technischer Rechen- und Speicherkapazität angewiesen sind. Ein Beispiel hierfür ist das noch ungenaue Verständnis der nordatlantischen Oszillation und der ihr zugrunde liegenden nichtlinear-instabilen Dynamik von dominanter Westwinddrift und ihrer zeitweisen Blockade. Hier könnten vertiefende niederdimensionale Analysen, d. h. die Analyse der Unterräume von der Anwendung geeigneter Big-Data- und Machine-Learning-Verfahren profitieren. Ein anderes Beispiel ist die Modellierung und Simulation des Golfstroms, die bislang nur durch Kopplung mit einem stark vereinfachten Atmosphärenmodell in grober Form gelingt. Erst die heute verfügbaren Daten und Computer-Leistungen ermöglichen Kopplungen mit wesentlich realitätsnäheren Atmosphärenmodellen, in denen auch wetterbedingte Fluktuationen und ihre Stochastik einfließen, um die Dynamik des

Golfstroms auch im Detail verstehen zu können. Entsprechend abgeleitete und abgesicherte Vorhersagen der Dynamik von Meeresströmungen und Atmosphäre in bzw. über dem (Nord-)Atlantik sind neben dem wissenschaftlichen Erkenntnisgewinn auch für die mittel- bis langfristige Klimavorsorge in Europa von signifikanter Bedeutung.

Entwicklung von Modellen bzw. Zerlegung in Teilmodelle

Wie angedeutet, zielt die Klimaforschung darauf ab, mathematische Modelle, basierend auf partiellen Differenzialgleichungen, zu entwickeln. Dies sind so genannte Zustandsraummodelle, mit denen die klimatischen Zustandsgrößen oder Zustandsvariablen in einem gewählten raumzeitlichen Ausschnitt des (als Ganzes unzugänglichen) Klimasystems auf Basis von zwei oder mehreren (interagierenden) Teilmodellen gegeben sind. Dabei berechnet ein dynamisches Teilmodell – von einem bekannten bzw. definierten Anfangszustand aus – die weitere Entwicklung der Zustandsgrößen über gewünschte Zeitskalen hinweg. Diese Vorhersagen werden schrittweise mit beobachteten Daten und dem so genannten Beobachtungsmodell abgeglichen. Bei diesem werden die modellierten mit den bekannten Zustandsgrößen desselben Zustandsraums in Beziehung gesetzt und ggf. bei zu hoher Abweichung über geeignet Korrekturterme jeweils angepasst. So erhält man konsistente Zeitreihenverläufe in den gewählten Zustandsräumen.

Entsprechende algorithmischen Verfahren werden beispielsweise – jeweils von den aktuellen atmosphärischen Zuständen ausgehend – mit großem Erfolg für die Wettervorhersage verwendet. Darüber hinaus eignen sie sich auch für widerspruchsfreie Re-Analysen von langfristigen Klimaentwicklungen sowie für das konsistente „Down-Scaling“ großskaliger Strukturen im Klimasystem. In diesen Bereichen der Zustandsraumanalyse und -vorhersage weisen KI-, Machine Learning- und Big Data-Verfahren bis dato das größte Potenzial für die Klimaforschung auf, das noch weiter zu nutzen ist.

Signal-Rauschen-Unterscheidung zur Identifikation exzeptioneller Ereignisse

Eine Machine-Learning- und Big-Data-gestützte Kernaufgabe der Klimaforschung ist die Unterscheidung von bedeutsamen Signalen einerseits und interner Variabilität („Rauschen“) andererseits im hochdimensionalen Zustandsraum eines lokal betrachteten Klimateilsystems. Erst wenn sich dort Strukturen, d. h. bedeutsame Signale detektieren lassen, die nicht durch das interne Rauschen dieses Systems generiert werden, schließt man auf die Existenz eines außerhalb dieses lokalen Klimasystems liegenden Signals. Hieraus ergibt sich dann die Frage nach der externen Ursache dieses Signals an. Dieser Ansatz wird heute standardmäßig auf den Nachweis von Klimawirkungen erhöhter Treibhausgaskonzentrationen angewandt sowie auf deren Unterscheidung von natürlichen Schwankungen, etwa wie die der wechselnden Sturmtätigkeit in Nordeuropa oder der süd pazifischen „El-Nino“-Oszillation und ihrer Wirkungen auf die Umwelt.

Ein ambitioniertes Ziel ist die sogenannte „Event Attribution“, bei welcher einzelne Ereignisse wie Stürme oder Hitzewellen auf charakteristische Eigenschaften hin untersucht werden, um zu prüfen, ob diese noch unter natürlichen Klimaverhältnissen plausibel sind oder ob sie bereits den Einfluss des anthropo-

genen Klimawandels widerspiegeln. Für derart kurzfristige Analysen und Interpretationen entsprechender Extremereignisse werden geeignete Algorithmen und neue informatische Verfahren eingesetzt.

Zwischenfazit

In der Erforschung des Erdklimas wurde schon lange und mit wechselndem Erfolg nach wirksamen Methoden und Verfahren gesucht, die eine Beschreibung von Klimasystemkomponenten und -prozessen (Zustandsgrößen, relevante Unterräume und lokale Modelle) ermöglichen. Nur auf dieser Basis können vereinfachte, effiziente und automatisierbare Klimavorhersagen entwickelt werden. Inzwischen sind KI- und Big-Data-Verfahren so weit, dieses Ziel zu erreichen: Sie werden bereits routinemäßig zur konsistenten raumzeitlichen Modellierung von Klima und Wetter, ihrer Dynamik und daraus abgeleiteter Prognosen eingesetzt. Der Einsatz neuronaler Netze eignet sich darüber hinaus auch für die automatisierte Analyse von dynamisch relevanten Teilen des Phasenraums im Klimasystem. Unter anderem kann KI dabei auch die Trennung von signifikanten Signalen von Hintergrundrauschen erleichtern, was eine notwendige Bedingung für entsprechend belastbare wissenschaftliche Aussagen sowie deren Erklärung darstellt.

Auch wenn KI-, Machine Learning- und Big Data-Verfahren damit zum besseren Verständnis der Komponenten und Prozesse des Klimasystems beitragen können, betonen Klimaforscher, dass entsprechende Erkenntnisbildung durch künstliche Intelligenz alleine – also automatisiert und ohne Zutun von Forschenden – prinzipiell nicht erreichbar ist. Der Klimaforschung geht es nicht nur um Prognosen und Klassifikationen allein, sondern um ein Verständnis des Klimasystems und von wissenschaftlichen Erklärungen. Eine gute Prognostik ist in der Klimaforschung – ganz traditionell – abhängig von wissenschaftlichem Wissen, welches freilich in Verfahren von KI und Big Data Eingang findet.

2.3.3 Medizin⁷

KI und insbesondere Machine Learning haben das Potenzial, Ärzte bei der Diagnose von Krankheiten und mit Vorschlägen zu Therapien zu unterstützen. Beispielsweise können Symptome, Röntgenbilder und Krankenakten in wenigen Sekunden überprüft und mit anderen Datenbeständen, beispielsweise aus verschiedenen Krankenhäusern, abgeglichen werden (Jäkel 2018). Auf diese Weise können deutlich größere medizinische Wissensbestände bei den Analysen berücksichtigt werden, als dies ein einzelner Arzt könnte. So kann Spezialwissen aus den verschiedenen medizinischen Bereichen (zum Beispiel Anatomie, Chirurgie) auch für Hausärzte

⁷ Dieser Abschnitt basiert teilweise auf dem Vortrag von Thomas Lengauer zum Thema „Data Analysis in Medicine“ auf dem Workshop Digitale Arbeitswelten in Forschung und Entwicklung, Bad Neuenahr-Ahrweiler, 02.07.2019, sowie auf dem Vortrag von Karl-Heinz Jöckel zum Thema „Künstliche Intelligenz in den Medizinischen Wissenschaften“ auf dem Workshop Digitale Arbeitswelten in Forschung und Entwicklung, Bad Neuenahr-Ahrweiler, 06.09.2019.

zugänglich gemacht werden. Seltene Krankheiten können schneller und zu geringen Kosten identifiziert werden (Kuhn et al. 2018). Vor diesem Hintergrund haben KI-Methoden bereits viele Anwendungszwecke in der Medizin, so zum Beispiel:

- Erkennung von pathogenen Mustern, z. B. in der Bildgebung von Organen und Geweben
- Erkennung von Gesetzmäßigkeiten, z. B. in der biomedizinischen Forschung
- Übersetzung zwischen unterschiedlichen Sprachen oder Terminologien, z. B. Übersetzung pathologischer Befunde von natürlicher Sprache in die medizinische Klassifikation ICD-O
- Prädiktion für Individuen, z. B. in der präzisionsmedizinischen Differenzialdiagnostik
- Prädiktion für Gruppen/Populationen, z. B. zur Einschätzung von Epidemien
- Unterstützung menschlichen Handelns in konkreten Entscheidungssituationen, z. B. für Therapieempfehlungen
- Unterstützung bei abstrakten Planungsprozessen, z. B. bei der Studienplanung
- „Intelligente“ Interaktion mit Menschen, z. B. bei automatisierter Durchführung von Befragungen
- Entwicklung von Medikamenten

Ein Beispiel ist die *personalisierte Therapie* von HIV. So existieren auf der einen Seite Millionen von HIV-Varianten. Auf der anderen Seite gibt es mehrere Dutzend Medikamente für HIV mit hunderten von Therapie-Kombinationen. Für eine beim Patienten beobachtete Population von Virenvarianten eine möglichst effektive Therapie aus zwei bis drei Medikamenten zu finden, stellt eine kombinatorische Herausforderung dar (Deutsches Ärzteblatt 2018, vgl. Shah et al. 2019). Diese Herausforderung kann mittels KI-Methoden auf der Basis von konkreten Therapiebeispielen angegangen werden. Das Ergebnis sind maßgeschneiderte Behandlungen für individuelle Patienten mit verbesserten Erfolgchancen.

Die Chancen der Nutzung von KI in der Medizin liegen auch im Fortschritt von Algorithmen in der Bilderkennung begründet. Die folgende Abb. 2.8 zeigt, wie sich die Fähigkeiten von KI-Anwendungen im Vergleich zu Menschen in den letzten Jahren verbessert hat.

Damit sind sie grundsätzlich gut geeignet, auch eine Analyse von CT- oder MRT-Bildern durchzuführen und Ärzte bei der Diagnose von Krankheiten zu unterstützen. Das Potenzial wird auch deutlich aus einem Gespräch, das einer der Autoren mit Prof. Dr. Jürgen Schäfer (Mediziner am Universitätsklinikum Marburg für seltene Krankheiten) geführt hat: „Zum Beispiel gibt es jetzt schon so gute Daten in der Pathologie, dass eine KI-Anwendung sehr präzise Gewebeschnitte analysieren und tumorsuspekte Zellen erkennen kann. Ein Pathologe, der 10.000 Gewebeschnitte befunden hat, der hat natürlich mehr Erfahrung als jemand, der nur 500 Befunde erstellt hat. Aber wenn man dann sieht, dass eine KI-Anwendung nicht nur 10.000, sondern Millionen Befunde im Speicher hat, dann bekommt man eine Vorstellung, wie viel profunder solch eine Befundung in Zukunft sein mag. Allerdings wird KI nicht so schnell einem Patienten die Hand halten und mit menschlichem

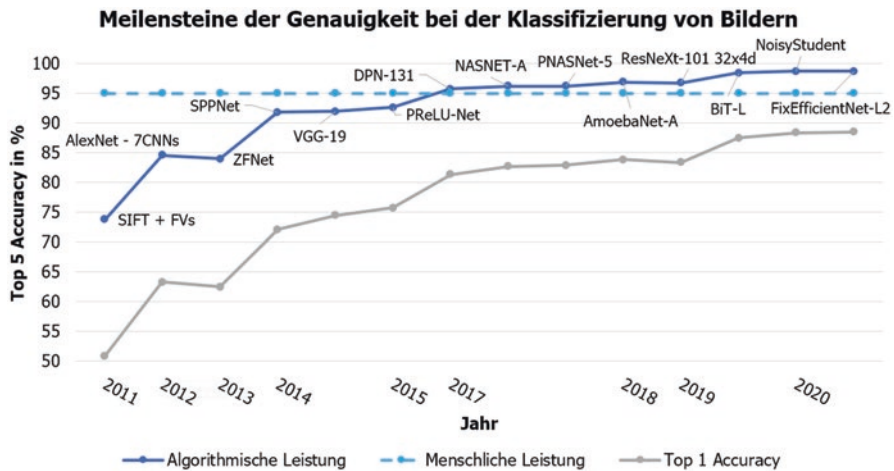


Abb. 2.8 Meilensteine der Genauigkeitsverbesserung bei der Klassifizierung von Bildern. (Quelle: ImageNet 2020)

Einfühlungsvermögen begegnen“ (Buxmann 2020). Solange dies der Fall ist, brauchen sich Ärzte keine Sorgen machen, dass sie vom Computer verdrängt werden.

Eine wichtige zukünftige Herausforderung besteht darin, dass KI-Anwendungen in die Lage versetzt werden, ihre Entscheidungen zu erklären. Kaum ein Arzt oder kaum ein Patient würde beispielsweise Therapievorschlge einer KI-Anwendung akzeptieren, bei denen nicht nachvollziehbar sind, wie sie entstanden sind. Hierbei gibt es weltweit gerade eine Vielzahl von Forschungsarbeiten, die auch unter dem Stichwort „Explainable AI“ (Erklrbare KI) zusammengefasst werden.

2.4 Fazit

KI hat nicht nur eine hohe Bedeutung fr Wirtschaft und Gesellschaft. Auch in Forschung und Wissenschaft wird KI erfolgreich eingesetzt, wie beispielhaft in der Physik, der Klimaforschung und der Medizin beleuchtet. KI zeigt sich damit als ein mchtiges Werkzeug, welches die Wissensarbeit verndert:

1. *Hhere Effizienz:* Durch die Automatisierung von Routine-Ttigkeiten knnen (Forschungs-)Arbeiten beschleunigt werden. Allerdings ndern sich dadurch auch Ttigkeitsbereiche und Stellenprofile. Beispiele hierfür sind KI-gesttzte Laborauswertungen in den Lebenswissenschaften.

Die Erhhung des Automatisierungsgrades kann auch indirekt bewirken, dass Forschende einen immer weniger tiefen Einblick in die zu untersuchende Materie bekommen; Forschende wrden dann Ergebnissen von KI-Anwendungen vertrauen, ohne sie nachvollziehen und kritisch hinterfragen zu knnen.

2. *Größere Genauigkeit:* Vor allem durch die Verfügbarkeit von großen Datenmengen können Forschungsfragen in der Regel präziser beantwortet werden. Beispiele hierfür sind KI-gestützte Auswertungen von Versuchen in der Physik. Falsch eingesetzt können Ergebnisse jedoch auch falsch interpretiert werden. Garantien gegen Falschaussagen gibt es nicht.
3. *Neue Möglichkeiten:* KI und verwandte Technologien erlauben Dinge, die bislang unmöglich waren, z. B. Genomanalyse und darauf aufbauende Forschungen. Besonders die Hypothesenbildung aufgrund von KI-Techniken eröffnet in vielen Wissenschaftsdisziplinen ganz neue Möglichkeiten.

Eine mögliche Folge von datengetriebenen KI-Methoden in der Wissenschaft könnte ein Paradigmen-Shift sein, bei welchem statistisch valide Korrelationen an Bedeutung zunehmen und erklärbare Kausalzusammenhänge an Bedeutung abnehmen (Näheres hierzu wird in Kap. 4 dieses Bandes beschrieben). Ob und welche Folgen das für die künftige Rolle der menschlichen Forscher/Forscherinnen hat, erläutert das Kap. 3.

Literatur

- Aaboud M et al (2018) Measurements of Higgs boson properties in the diphoton decay channel with 36 fb⁻¹ of *pp* collision data at $s\sqrt{=}$ 13 TeV with the ATLAS detector. Phys Rev D 98(5):052005. <https://doi.org/10.1103/PhysRevD.98.052005>
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25(1):25–29. <https://doi.org/10.1038/75556>
- Bostrom N (2014) Superintelligenz: Szenarien einer kommenden Revolution. Suhrkamp, Berlin
- Brooks R (2017) The seven deadly sins of predicting the future of AI. <https://rodneymbrooks.com/the-seven-deadly-sins-of-predicting-the-future-of-ai/>. Zugegriffen am 13.10.2020. Deutsche Übersetzung: Algorithmenethik. 2017. Die sieben Todsünden der Prognosen über die Zukunft der KI. <https://algorithmenethik.de/2017/11/14/die-sieben-todsunden-der-prognosen-ueber-die-zukunft-der-ki>. Zugegriffen am 13.10.2020
- Brynjolfsson E, McAfee A (2017) The business of artificial intelligence. Harv Bus Rev 1–20. <https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence>. Zugegriffen am 13.10.2020
- Buxmann P, Peters F (2020) Ethik der Künstlichen Intelligenz. In: Canibol H-P, Hossenfelder J, Lünendonk J (Hrsg) *Handbuch Consulting 2020*. Lünendonk und Hossenfelder GmbH, Mindelheim, S 92–103
- Buxmann P, Schmidt H (2019) Open source tools Für Künstliche Intelligenz & Maschinelles Lernen. <https://www.ki-business.de/tools>. Zugegriffen am 25.11.2020
- Buxmann P (2020) Künstliche Intelligenz – Wirtschaft und Innovation. Studie im Auftrag des Ministeriums für Digitales für das Land Hessen. Wiesbaden (im Erscheinen)
- Buxmann P, Schmidt H (2021) Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens. In: Buxmann P, Schmidt H (Hrsg) Künstliche Intelligenz: Mit Algorithmen zum wirtschaftlichen Erfolg, 2. Aufl. Springer Gabler, Berlin
- Corbett-Davies S, Goel S (2018) The measure and mismeasure of fairness: a critical review of fair machine learning. arXiv preprint arXiv:1808.00023
- Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Bd 1, San Diego, S 886–893

- Deutsches Ärzteblatt (2018) Die Forderung nach transparenten Algorithmen kann ihre Vorhersagekraft schwächen. <https://www.aerzteblatt.de/nachrichten/95778/Die-Forderung-nach-transparenten-Algorithmen-kann-ihre-Vorhersagekraft-schwaechen>. Zugriffen am 26.11.2020
- Fawcett T (2006) An introduction to ROC analysis. Pattern Recogn Lett 27(8):861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Google (2020) Wann wurde die Mauer gebaut? – Google Suche. <https://www.google.de/search?source=hp&q=Wann+wurde+die+Mauer+gebaut%3F&oq=Wann+wurde+die+Mauer+gebaut%3F>. Zugriffen am 10.09.2020
- Hasselmann K (1988) PIPs and POPs: the reduction of complex dynamical systems using principal interaction and oscillation patterns. J Geophys Res 93(D9):11015. <https://doi.org/10.1029/JD093iD09p11015>
- Heckman JJ (1979) Sample selection bias as a specification error. Econometrica 47(1):153–161
- Hoppe T (2020) Semantische Suche. Grundlagen und Methoden semantischer Suche von Textdokumenten. Mit Beiträgen von B.G. Humm. Springer Vieweg, Wiesbaden
- Humm BG (2017) „Technische Superintelligenzen werden die Menschheit bedrohen“ – Wie kann man eine so absurde Behauptung aufstellen? <https://www.cognitiveagent.org/2017/05/05/technische-superintelligenzen-werden-die-menschheit-bedrohen-wie-kann-man-eine-so-absurde-behauptung-aufstellen/>. Zugriffen am 13.10.2020
- Humm BG (2020) Applied Artificial Intelligence. An Engineering Approach, 2. Aufl. Unabhängig veröffentlicht bei Amazon KDP, Seattle, USA 2020. ISBN: 979-8635591154. Online <https://www.amazon.de/Applied-Artificial-Intelligence-Engineering-Approach/dp/B086Y6H7ZL>
- ImageNet (2020) Image Classification on ImageNet. <https://paperswithcode.com/sota/image-classification-on-imagenet>. Zugriffen am 01.10.2020
- Insight Centre for Data Analytics (2020) The linked open data cloud. <https://lod-cloud.net/>. Zugriffen am 12.03.2020
- Jäkel L (2018) Wie Künstliche Intelligenz unser Leben berührt. <https://orange.handelsblatt.com/artikel/53316>. Zugriffen am 13.02.2019
- Kononenko I (2001) Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med 23(1):89–109. [https://doi.org/10.1016/S0933-3657\(01\)00077-X](https://doi.org/10.1016/S0933-3657(01)00077-X)
- Kuhn S, Jungmann SM, Jungmann F (2018) Künstliche Intelligenz für Ärzte und Patienten: „Googeln“ war gestern. <https://www.aerzteblatt.de/archiv/198854/Kuenstliche-Intelligenz-fuer-Aerzte-und-Patienten-Googeln-war-gestern>. Zugriffen am 25.11.2020
- Luber S, Litzel N (2019) Was ist ein Expertensystem? Big Data Insider. <https://www.bigdata-insider.de/was-ist-ein-expertensystem-a-819539>. Zugriffen am 09.09.2020
- Lum K, Isaac W (2016) To predict and serve? Significance 13(5):14–19
- McCarthy J, Minsky ML, Rochester M, Shannon CE (1955) A proposal for the Dartmouth summer research project on artificial intelligence. <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>. Zugriffen am 02.02.2021
- Medical Subject Headings. U. S. National Library of Medicine. <https://www.nlm.nih.gov/mesh/meshhome.html>. Zugriffen am 13.10.2020
- Murphy KP (2012) Machine learning. A probabilistic perspective. MIT Press, Cambridge
- NCI Thesaurus. U. S. National Cancer Institute. <https://ncithesaurus.nci.nih.gov>. Zugriffen am 13.10.2020
- Powers DMW (2011) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. J Mach Learn Technol 2(1):37–63
- Sarathi P (2014) 12 famous quotes on Artificial Intelligence (AI) by Google founders. Big Data Made Simple. <http://bigdata-madesimple.com/12-famous-quotes-on-artificial-intelligence-by-google-founders>. Zugriffen am 30.09.2019
- Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, Ringel M, Schork N (2019) Artificial intelligence and machine learning in clinical development: a translational perspective. npj Digit Med 2(1):69. <https://doi.org/10.1038/s41746-019-0148-3>
- The Gene Ontology Consortium (2019) The gene ontology resource: 20 years and still GOing strong. Nucleic Acids Res 47(D1):D330–D338. <https://doi.org/10.1093/nar/gky1055>

- Vincent J (2016) Apple boasts about sales; Google boasts about how good its AI is. The Verge. <https://www.theverge.com/2016/10/4/13122406/google-phone-event-stats>. Zugegriffen am 12.03.2020
- Zhang M, von Storch H, Chen X, Wang D, Li D (2019) Temporal and spatial statistics of travelling eddy variability in the South China Sea. *Ocean Dyn* 69(8):879–898. <https://doi.org/10.1007/s10236-019-01282-2>

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.



Zur Frage der Ersetzbarkeit des Menschen durch KI in der Forschung



Carl Friedrich Gethmann

3.1 Einleitung

Spricht man KI als Instrument wissenschaftlicher Arbeit an, wird zugleich unterstellt, dass ein menschlicher Akteur dieses Instrument (vermeintlich) zweckgerichtet einsetzt. Mit welchem Erfolg und unter welchen Erfolgskriterien der instrumentelle Einsatz auch immer erfolgt, grundsätzlich ist der menschliche Akteur im epistemischen Prozess als solcher nicht grundsätzlich gefährdet, wenn sich seine Rolle auch ändern kann. Die Metapher von digitalen Agenten verlegt dagegen die Akteursrolle in das Instrument selbst, das begrifflich genau in diesem Moment aufhört, eines zu sein. Aus diesem Grunde ist auch der Begriff der „Autonomie“ in Bezug auf technische Artefakte allenfalls metaphorisch zu verwenden. Nur ein Akteur kann genau genommen sich selbst die Regeln des Handelns auferlegen. Ein technisches Gerät dagegen, dem die Regeln seiner Operationen vom Konstrukteur, Software-Entwickler u. a. vorgegeben werden, ist grundsätzlich „heteronom“ bestimmt, auch wenn es im Rahmen der vorgegebenen Regeln weiter „lernt“, seinen zweckgerichteten Einsatz zu optimieren. Ein digitaler Agent wäre demgegenüber der Souverän, der dank seiner künstlichen Intelligenz die Mittel bestimmte. Dieses Bild nimmt die Vorstellung in Anspruch, dass der Prozess des Generierens von Wissen prinzipiell ein regelbestimmtes Verfahren ist, also einer Methodologie folgt, dass ein solches Verfahren somit auch von einem technischen Artefakt übernommen und ohne die menschlichen Schwächen (Müdigkeit, Lustlosigkeit, Ehrgeiz, Geldgier u. a.) sogar mit größerer Perfektion ausgeführt werden kann. Sollte dieses Bild von einem regelgeleiteten Verfahren wissenschaftlicher Erkenntnisproduktion jedoch unzutreffend sein, wäre auch der Gedanke der Übernahme eines solchen Verfahrens durch einen „Erkenntnisroboter“ irreführend.

C. F. Gethmann (✉)
 Universität Siegen, Siegen, Deutschland
 E-Mail: Cfg-ea@t-online.de

Eine kohärente Rekonstruktion des Wissens- und Handlungsbegriffs kommt zu dem Ergebnis, dass der Mensch – trotz Entwicklungen von KI, Big Data und Machine Learning – der zentrale Akteur der forschenden Wissensarbeit bleibt. Als Handelnder ist der Mensch in der Genese wissenschaftlichen Wissens sowie in der method(olog)ischen Geltungsausweisung nicht (vollständig) substituierbar.

Zur Fundierung dieser These wird zunächst das Verhältnis von KI und menschlicher Wissensbildung untersucht (2). Der entscheidende Unterschied wird darin gesehen, dass künstlichen Agenten im Unterschied zum Menschen keine Handlungsurheberschaft zugeschrieben werden kann (3). Diese Unterscheidung ist Anlass, die weiterhin gültige Bedeutung des Bacon-Projekts der neuzeitlichen Wissenschaft angesichts künstlicher Intelligenz zu untersuchen. (4). Die Unterscheidung zwischen deterministischen und probabilistischen Voraussagen (5) sowie zwischen methodischen Verfahren und heuristischer Urteilskraft (6) sollen die Nicht-Ersetzbarkeit menschlicher Intelligenz weiter illustrieren.

3.2 KI und menschliche Wissensbildung

Von „künstlicher“ Intelligenz wird gesprochen, wenn Computer-Anwendungen in technischen Artefakten (Geräten, devices) Merkmale menschlicher Intelligenz aufweisen. In diesem Zusammenhang wird seit den siebziger Jahren des 20. Jahrhunderts kontrovers über die Frage diskutiert – in Anlehnung an einen bekannten Buchtitel von H.L. Dreyfus (1979, 1992) – was Computer können und nicht können bzw. demnächst können und nicht können werden. Um diese Frage zu beantworten, muss zunächst geklärt werden, von welchen impliziten Annahmen hinsichtlich der menschlichen Intelligenz dabei Gebrauch gemacht wird. Dazu wird jedoch in den Wissenschaften keine einfache Antwort angeboten.

3.2.1 Intelligenz

In der Psychometrie wird „Intelligenz“ zumeist als Fähigkeit definiert, zu lernen, sich an neue Umstände anzupassen oder komplexe Informationen zu verarbeiten. In der Konstruktion von Intelligenztests, die sich auf diese Definitionen stützen, werden Aufgaben gestellt, die verschiedene mentale Operationen erfordern. Der Wechsler-Intelligenztest (1958), auf dem die bekanntesten modernen Intelligenztests aufbauen, besteht aus zwei Teilbereichen: einem verbalen und einem sprachfreien Test. Der verbale Test enthält Unteraufgaben für Information, Verständnis, Wortschatz, Gedächtnisspanne für Zahlen, arithmetisches Denkvermögen und Erkennen verbaler Ähnlichkeiten. Im sprachfreien Teil werden verschiedene Aufgaben zur Mustererkennung und -verarbeitung gestellt, z. B. sind auseinandergeschnittene Teile zu einer vollständigen Figur zusammenzufügen, Bilder zu ergänzen

oder zu einer sinnvollen Sequenz zusammenzufügen oder Muster aus Klötzchen sowie Zahlen mit Symbolen zu kombinieren. In der Validierung von Intelligenztests steht zumeist deren prognostische Leistung im Zentrum des Interesses: Eine hohe Validität wird hierbei durch eine starke Korrelation von Testwerten und dem pragmatischen Erfolg (z. B. dem beruflichen Erfolg nach einem Eignungstest) angezeigt.

Insbesondere die Frage zur Struktur der Intelligenz, speziell danach, ob Intelligenz als eine einheitliche Fähigkeit betrachtet wird oder viele voneinander unabhängige Fähigkeiten umfasst, ist bis heute umstritten. Mit der Entwicklung der Kognitionspsychologie in den 1970er-Jahren rückte zunehmend die Analyse der Prozesse, die nötig sind, um die Aufgaben der Intelligenztests zu lösen, in den Vordergrund der wissenschaftlichen Untersuchung. Umfangreiche Forschungen zur Bearbeitung informationsverarbeitender Aufgaben, wie Informationskodierung und geteiltes Hören führten zu der Annahme, dass (verbale) Intelligenz durch die Fähigkeit zur Auswahl und Benutzung von Informationsverarbeitungsmethoden bestimmt wird (Hunt et al. 1975). Die Kernthemen der kognitionspsychologischen Forschung jedoch, nämlich Denken, Problemlösung und Entscheidungsfindung, die außerhalb der Disziplin der Psychologie häufig mit dem Begriff Intelligenz assoziiert werden und auf die häufig in der Entwicklung künstlicher Intelligenz Bezug genommen wird, fallen in der Psychologie nicht unter diesen Begriff.

In Intelligenztests wird auf ein Ensemble menschlicher Leistungen Bezug genommen, nämlich sprachliche Kompetenz, elementare Rechenfähigkeiten, räumliches Vorstellungsvermögen, logisches Schlussfolgern oder Gedächtnisleistungen. In jüngerer Zeit werden solche Leistungen durch soziale und emotive Beziehungen, kreative Gestaltung und komplexe Problemlösungen ergänzt. Spätestens diese Ergänzungsversuche werfen die grundsätzliche Frage auf, wie die Übertragung auf technische Artefakte zu verstehen ist. Bei einigen Merkmalen wie beispielsweise der elementaren Rechenfähigkeit, dem Schlussfolgern oder Gedächtnisleistungen scheint naheliegend, dass menschliche Fähigkeiten univok auf technische Artefakte übertragen werden können. Auch diesbezüglich sind schon kritische Fragen zu stellen, wie beispielsweise, ob die menschliche Erinnerungsfähigkeit in gleicher Weise eine Gedächtnisleistung ist wie die Aktivierung eines technischen Speichers. Einerseits sind die quantitativen Leistungen technischer Speicher der menschlichen Erinnerungsfähigkeit um Größenordnungen überlegen. Andererseits sortiert der Mensch seine Gedächtnisleistungen beispielsweise nach der jeweils kontextuell bestimmten Bedeutung, während ein technischer Speicher unterschiedslos je nach den technischen Vorgaben Daten aufnimmt oder nicht. Bei Intelligenzleistungen mit emotiven und kreativen Qualitäten verstärkt sich der Verdacht, dass es sich hierbei um anthropomorphe und somit äquivalente Übertragungen handelt. Die Klärung solcher Vergleichbarkeitsprobleme hängt somit wesentlich von den Kriterien ab, durch die man eine spezifisch menschliche Intelligenzleistung bestimmt sieht. Grosso modo wird man die Verwendung des Ausdrucks „Intelligenz“ in der Wortverbindung „künstliche Intelligenz“ eher als eine Metapher einordnen müssen, deren Beschreibungs- und Erklärungsfunktion genauerer Aufklärung bedarf.

3.2.2 Denken können

Auch ohne eine genauere Klärung des Begriffs der Intelligenz ist offenkundig, dass er das Verhältnis zwischen menschlichen und maschinellen Leistungsmerkmalen vor allem bezüglich kognitiver Fähigkeiten hervorhebt. Entsprechend war in der ersten Phase der Diskussion um KI ab den siebziger Jahren des 20. Jahrhunderts die Frage, „ob Computer denken können“ leitend. Für die Fragerichtung ist die kontroverse Diskussion um die Interpretation des von A.M. Turing 1950 (Turing 1950) vorgeschlagenen Turing-Test paradigmatisch.¹ Nach Turing kann die Frage, ob technische Artefakte „denken“ können, dadurch entschieden werden, dass ein Fragesteller (für ihn verdeckten) Menschen und Geräten beliebige Fragen stellt. Wenn in einer größeren Zahl von Durchgängen mit wechselnden Fragestellern und wechselnden Menschen bzw. Geräten die Antworten zu einem hinreichend großen Anteil (z. B. 50 %) nicht eindeutig Mensch bzw. Gerät zugeordnet werden können, gibt es nach Turing keinen Grund, technischen Artefakten weniger Denkvermögen zuzuschreiben als Menschen. Für die Beurteilung der Angemessenheit dieses Tests hängt offenkundig alles davon ab, nach welchen Kriterien in überprüfbarer Weise einer Entität das Vermögen „Denken“ zugeschrieben wird. Computerwissenschaftler/-wissenschaftlerinnen unterstellen dabei im Anschluss an Turing ein Verständnis von „menschlicher“ Intelligenz, relativ zu dem sie die Leistungsfähigkeit von „künstlichen“ Computern interpretieren. Beispielsweise wird das menschliche Gedächtnis nach dem Operationsraum von technischen Speichern „modelliert“. Für dieses von den Naturwissenschaften vom Menschen transferierte Verständnis von menschlicher Intelligenz sind drei *anthropologisch folgenreiche Annahmen* kennzeichnend:

1. *Messbarkeit*: Unter Intelligenz wird eine Eigenschaft verstanden, die vielen Sorten von Entitäten graduell zugeschrieben wird. Menschen wie technische Artefakte können somit mehr oder weniger intelligent sein, je nach ihrer messbaren Leistung. Qualitative Gewichtungen lassen sich auf quantitative zurückführen, oder aber sie sind bedeutungslos.
2. *Leiblosigkeit*: Die Eigenschaft der Intelligenz ist nicht an eine leibliche bzw. organische Realisierung gebunden. Die physische (organische oder anorganische) Beschaffenheit intelligenter Entitäten ist für die Frage, ob sie denken können, irrelevant. Damit wird das Zuspreehen von Intelligenz von der physischen Realisierung unabhängig gemacht.
3. *Nicht-Individualität*: Zustände, auf die mentale Termini referieren, sind durch äußere Reize und die Reaktionen auf sie zu erklären. Gleiche mentale Zustände müssen durch gleiche Ursachen festgelegt sein. Wenn Kognitionen intelligenter Entitäten affektive und emotive Varianz bei gleichen Reiz-Reaktions-Bedingungen aufweisen, sind diese als zu behebende technische Störungen einzuordnen.

¹Vgl. die kritische Darstellung bei K. Mainzer (1995), S. 113 f. und passim.

Diese Unterstellungen stammen aus identitätstheoretischen und funktionalistischen Konzeptionen der Philosophie des Geistes, gemäß der jede menschliche kognitive Leistung im Prinzip funktional äquivalent durch ein technisches Artefakt darzustellen und nachzustellen ist. Viele KI-Forscher/-Forscherinnen übernehmen somit von den Naturwissenschaften vom Menschen ein funktionalistisches Verständnis menschlicher Fähigkeiten, für die (ohne weitere Rechtfertigung) das „Denken“ für symptomatisch gehalten wird (vgl. die kritische Untersuchung bei Carrier und Mittelstraß 1989). Die Interpretation und Identifikation der Leistung des „Denkens“ übernehmen viele Neurowissenschaftler/-wissenschaftlerinnen von Beschreibungen der „künstlich“ hergestellten Funktionsweisen eines Computers, anscheinend ohne, dass ihnen der damit verbundene Explikationszirkel Probleme bereitet.

Die philosophische Kritik an der Vorstellung des „denkenden“ Computers setzt an den Präsuppositionen des Funktionalismus an. Sie bezieht sich dabei auf eine lange Geschichte eines oft mehr oder weniger explizit vertretenen anthropologischen Naturalismus. Der Ausdruck „Naturalismus“ kennzeichnet nicht die kognitiven Bemühungen von Naturwissenschaftlern/-wissenschaftlerinnen als solche, sondern aus den Naturwissenschaften vom Menschen abgeleitete, aber über sie hinausgehende Deutungsansprüche hinsichtlich des Selbstverständnisses des Menschen und seiner gesellschaftlichen Selbstorganisation (exemplarisch: im Zusammenhang mit dem Strafrecht und bezüglich Erziehungsmaßnahmen). Mit der Erfolgsgeschichte der Naturwissenschaften generell und der Naturwissenschaften vom Menschen speziell ist der Naturalismus in der Neuzeit – abhängig von der Entwicklung der Naturwissenschaften selbst – in immer neuen Varianten entwickelt worden. Als besonders einflussreich sind zu nennen:

- Mechanismus (La Mettrie, d’Holbach);
- Physikalismus (Laplace);
- Evolutionismus (Darwin, Haeckel);
- Empirismus (Schlick, Carnap)
- Soziobiologismus (Wilson);
- Neurowissenschaften (Singer, Roth).

Dabei geht es in einer auffälligen Unschärfe der Begriffe mal um die „natürliche“ Erklärung kognitiver, mal operativer Fähigkeiten des Menschen. In allen Fällen hatte der Naturalismus eine „wissenschaftliche“ Weltanschauung vom Menschen zum Ziel, die eine weitgehende *Transformation philosophischer Fragestellungen* in solche der Wissenschaften ermöglichen soll (vgl. ausführlicher Gethmann 2016).

Die Kritik am anthropologischen Funktionalismus, die besonders prominent von H.L. Dreyfus² und J. Searle (Searle 1984) entwickelt worden ist, versucht zu zeigen, dass die anthropologischen Präsuppositionen des Funktionalismus grundsätzlich

²Die von Hubert M. Dreyfus im Anschluß an M. Heideggers Analyse des In-der-Welt-seins formulierte Kritik an den anthropologischen Präsuppositionen der Vertreter der KI bleibt unbeachtlich der Fortschritte in der KI-Forschung gültig (Dreyfus 1979).

unangemessen sind, um als hinreichende Beschreibungsinstrumente für das Handeln menschlicher Akteure eingesetzt werden zu können:

1. *Messbarkeit*: Im Unterschied zu technischen Artefakten bestimmen menschliche Akteure ihre Lebensumstände in einem mehr oder weniger großen Umfang „selbst“, und zwar nach Maßgabe subjektiver qualitativer Kriterien. Diese lassen sich nicht (restlos) adäquat quantitativ darstellen. Die Wichtigkeit einer Freundschaft, die Verzweiflung über eine Erkrankung oder die Freude an einer Mahlzeit lassen sich in der mentalen Binnensphäre nicht quantitativ messen. Emotionen können zwar andere „anstecken“, aber nicht ohne weiteres mit Anspruch auf Geltung verbindlich gemacht werden.³
2. *Leiblosigkeit*: Die Eigenschaft der Intelligenz menschlicher Akteure ist immer an eine leibliche Realisierung gebunden. Einer der Hauptmängel des Funktionalismus ist die mangelnde Unterscheidung von Leib und Körper. Der Mensch *hat* nicht einen Leib, sondern er *ist* Leib. Deswegen sind beispielsweise Angriffe gegen seinen Leib (etwa durch Folter) nicht Angriffe gegen Gegenstände seines Eigentums, sondern Angriffe auf ihn selbst. Wesentliches Merkmal der Leiblichkeit ist die organische Struktur: Leib ist Leben und Leben lässt sich nicht (vollständig) mit physikalischen Beschreibungsinstrumenten erfassen.⁴
3. *Nicht-Individualität*: Wenn ein technisches Artefakt auf eine Ursache reagiert, ist zu erwarten, dass jedes baugleiche ungestörte technische Artefakt in gleicher Weise reagiert. Baugleichheit ist in Bezug auf lebende Organismen jedoch eine uneinschlägige Kategorie. Selbst genetische Zwillinge sind nicht in dem Sinne gleich, dass sie auf gleiche Stimuli auf gleiche Weise reagieren. Erst recht determiniert die genetische Gleichheit nicht die Urheberschaft gleicher Handlungen. Das menschliche Individuum reagiert nicht einfach auf Symbole, sondern verwendet und kontrolliert sie. Verhält sich ein Individuum relativ stabil gegenüber seiner Umwelt, entwickelt es auf diese Weise seine Identität. Identität als eine Handlung prägende Bedingung ist somit etwas völlig anderes als eine Handlung konditionierende Baugleichheit.⁵

³Dazu ist die in der Tradition der philosophischen Anthropologie entwickelte Konzeption von Emotionen und Affekten heranzuziehen, vgl. Scheler 1923.

⁴Auf den begrifflichen Unterschied von „Leib“ und „Körper“ hat wohl zuerst Scheler 1916, S. 397–402 hingewiesen. Vgl. ferner Plessner 1928, S. 367; Plessner 1941, S. 238 u. ö.; Hengstenberg 1957, S. 88–101; Scherer 1976, S. 157–173. Zur Bedeutung der Leiblichkeit im Zusammenhang mit der Debatte um das Gehirn vgl. auch Fuchs (2013) bes. S. 33–40, 95–110. An die Philosophische Anthropologie schließt sich Dreyfus 1979, S. 235–255 an.

⁵Der Unterschied zwischen physischer Baugleichheit und sozialer Identität ist vor allem im symbolischen Interaktionismus herausgearbeitet worden (Mead 1934).

3.2.3 Zusammenfassung

Die Vorstellung der vollständigen Substituierbarkeit menschlicher kognitiver Leistungen macht von anthropologischen Präsuppositionen Gebrauch, die einer kritischen Betrachtung nicht standhalten. Dabei ist nicht ausgeschlossen, dass kognitive Teilfunktionen wie beispielsweise Gedächtnisleistungen von technischen Artefakten nicht nur ersetzt, sondern sogar hinsichtlich geringerer Störanfälligkeit, höherer Leistungsfähigkeit, tieferer Vernetzung, u. a. übertroffen werden können. Der technische Erfolg, der zweifellos besteht, erzeugt dabei die problematische Suggestion, dass die anthropomorphe Interpretation der technischen Artefakte unproblematisch ist.

3.3 Künstliche Agenten und menschliche Handlungsurheberschaft

Die Diskussion um die KI hat sich etwa seit der Jahrtausendwende unmerklich von der Konzentration auf die möglicherweise kognitiven Fähigkeiten abgewandt und zunehmend auf die operativen Fähigkeiten konzentriert. Nicht die Frage, in welcher Weise, wenn überhaupt, Computer „denken“ können, sondern in welchem Sinne Computer (z. B. als humanoide Roboter) „handeln“ können, steht inzwischen im Vordergrund. Diese Verschiebung der Aufmerksamkeit geht Hand in Hand mit der technischen Entwicklung, die unter intelligenten Systemen nicht nur solche mit hoher Informationsverarbeitungs-Kapazität versteht, sondern Quasi-Akteure, die menschliches Handeln nicht nur nachahmen und übertreffen, sondern zunehmend „autonom“ ersetzen. Damit stellt sich die Frage, in welchem Sinne solche „Agenten“ überhaupt handeln und in welchem Umfang und in welcher Weise solche „autonomen“ Systeme in der Folge für ihr Handeln verantwortlich gemacht werden können. Entsprechend fordert diese Frage nicht nur – wie früher – die Epistemologie, sondern auch die Ethik heraus. Im Rahmen der Jurisprudenz hat sich eine breite Debatte um die Zuordnung der Handlungsurheberschaft zu „autonomen“ Systemen als e-persons und beispielsweise Fragen der Haftung ergeben (z. B. Hilgendorf 2015). Allerdings wäre die Diskussionslage zu einfach beschrieben, wenn man sie als Übergang vom Denken (Wissen) zum Handeln beschriebe; kognitive und operative Fähigkeiten und Leistungen lassen sich technisch nicht trennen, so dass die Frage, ob Computer denken können, mit der Frage, ob Computer handeln können, in aufzuklärender Weise zusammenhängt; dieser enge Zusammenhang wird sprachlich schon dadurch suggeriert, dass die Ausdrücke, mit denen kognitive Prozesse beschrieben werden wie beispielsweise „denken“ grammatisch „Tu-Wörter“ sind. Die pragmatische Wende in der Diskussion muss also so rekonstruiert werden, dass die epistemischen Kategorien in sie eingebettet werden können.

3.3.1 *Handlungsurheberschaft und Zurechenbarkeit*

Die in Abschn. 3.2.2 angesprochenen funktionalistischen Frage-Transformationen unterstellen eine Entität, deren Handlungen und die mit ihnen verbundenen Formen des Wissens und Sprechens grundsätzlich als Wirkungen von Ursachen zu interpretieren sind (kausalistische [ätiologische] Handlungstheorien). Dem steht die Selbst-Erfahrung des Akteurs entgegen, (gegebenenfalls bedingter, aber nicht verursachter) Urheber seiner eigenen Handlungen zu sein; Handlungen sind dementsprechend nicht Wirkungen von Ursachen, sondern Zweck-Realisierungsversuche (finalistische [teleologische] Handlungstheorien). Die Pointe finalistischer Handlungstheorien liegt darin, die Handlungserfahrung des Akteurs bezüglich seiner eigenen Handlungen, die „Vollzugsperspektive“, von der Handlungsbeschreibung eines Außenstehenden, der „Berichtsperspektive“, zu unterscheiden. Diese Unterscheidung ist grundlegend, weil die Naturwissenschaften vom Menschen ihre Gegenstände in der Berichtsperspektive (mehr oder weniger angemessen) beschreiben, methodologisch dagegen für die Vollzugsperspektive nicht zuständig sind (Gethmann 2010).

Die Selbstzuschreibungen in der Vollzugsperspektive (Askriptionen) lassen sich nicht restlos auf die Fremdzuschreibungen in der Berichtsperspektive (Deskriptionen) reduzieren. Das zeigt sich vor allem darin, dass der Ich-Autor einer (Rede-) Handlung nicht ohne semantischen Verlust auf den Akteur im Rahmen einer Handlungsbeschreibung reduziert werden kann. Während die Äußerung „Ich verspreche dir, dich morgen zu besuchen“ unter sprachlichen Normalbedingungen als Vollzug eines Versprechens bestimmbar ist, ist die Äußerung „Jemand verspricht dir, dich morgen zu besuchen“ jedenfalls kein Versprechensvollzug. Unter naheliegenden sprachlichen Randannahmen könnte es der Bericht über einen Versprechensvollzug sein (oder aber eine Beschreibung, Erzählung, ...). Die für das Abgeben eines Versprechens konstitutive Selbstverpflichtung ist für einen Bericht über ein Versprechen nicht konstitutiv. Während der Vollzug des Versprechens nur dann gelingt, wenn der Autor des Versprechens sich daran macht, zu gegebener Zeit die versprochene Handlung auszuführen, hängt das Gelingen des Vollzugs des Berichts über ein Versprechen davon ab, dass der Bericht z. B. verlässlich u. a. ist. Die Verwechslung des Vollzugs-Ich mit einem Bericht über einen Berichtsgegenstand in der Berichtsperspektive ist als Fehler des Vollzugswiderspruchs (*contradictio exercita*) anzusprechen.

Die Nicht-Reduzierbarkeit von Askriptionen auf Deskriptionen ist die anthropologisch-apriorische Basis dafür, dass Nicht-Menschen (das gilt für Tiere wie für technische Geräte in gleichem Maße) nicht durch Handlungsurheberschaft ausgezeichnet sind und ihnen somit raumzeitliche Episoden, die sie ursächlich auslösen, nicht als Handlungen zuzurechnen sind. Daher wird man niemals einem technischen Artefakt wie einem Roboter moralische Vorwürfe beispielsweise wegen eines gebrochenen Versprechens machen und deswegen auch keine moralischen Sanktionen in Form von Tadel, Mißachtung oder Strafe im Sinne sozialer Desintegration verhängen. Ein Mensch, der eine falsche Wegbeschreibung abgibt, wird

getadelt, bestraft und/oder aufgeklärt, ein Navigationsgerät bei gleicher Fehlleistung wird reklamiert, repariert oder ersetzt. Entsprechend ist auch unvorstellbar, von technischen Artefakten zu verlangen, in Rechtfertigungsdiskurse als Proponenten einzutreten. Entsprechend wird man niemals noch so „intelligente“ Roboter vor Gericht stellen und strafrechtlich sanktionieren.

Diese Rekonstruktionen lassen sich zu der Definition zusammenfassen, dass eine Intelligenzleistung genau dann als *menschliche* anzusprechen ist, wenn die Zuordnung zu einem Wesen mit Handlungsurheberschaft pragmatisch angemessen ist. KI-Leistungen können unter Gesichtspunkten angemessener Beschreibung menschlichen Intelligenzleistungen ähnlich sein, die technischen Artefakte, von denen KI-Leistungen ausgehen, können jedoch nicht in angemessener Weise als Handlungsurheber angesprochen werden. Der von John. R. Searle (1984, bes. Ch. 2) am Beispiels des Chinesischen Zimmers illustrierte Unterschied zwischen einer syntaktischen Prozedur und dem Beherrschen einer Sprache folgt aus dem Kriterium der Handlungsurheberschaft: nur ein Wesen, das (Rede-)Handlungen hervorbringt, kann diesen eine Bedeutung verleihen bzw. eine Bedeutung verstehen (vgl. Mainzer 1995, S. 653 ff.).

3.3.2 Kausale Geschlossenheit

Der Begriff der Handlungsurheberschaft und die damit verbundene finalistische (intentionalistische) Handlungserklärung folgt der Analyse von Kant, wonach die Vorstellung der Verursachung primär in der Erfahrung des Akteurs als Verursacher der Handlungsfolgen liegt („Kausalität aus Freiheit“; vgl. Kant 1956, S. A538/B566 ff.). Im Anschluss an Kant wurde dieser Gedanke durch die Agent-Causality-Konzeption des menschlichen Handelns weiterentwickelt (Chisholm 1964). Gelegentlich wird eingewandt, dass die Vorstellung eines Handlungsurhebers als unverursachter Verursacher das *Grundpostulat von der kausalen Geschlossenheit* des physikalischen Weltbildes verletze. In diesem Zusammenhang muß man sich allerdings vor Augen halten, dass das Kausalitätsprinzip („Jedes Ereignis ist als Wirkung einer Ursache zu interpretieren.“) – im Unterschied zu den Kausalgesetzen – nicht das Ergebnis physikalischer Beobachtung und Theoriebildung ist, sondern ein „regulative(r) Grundsatz“ (Falkenburg 2012, S. 27), eine grundlegende methodologische Präsupposition der Naturwissenschaften. Das Kausalitätsprinzip hat einen ähnlichen apriorischen Status wie das Prinzip der Homogenität des Raumes oder der Isotropie der Zeit. Somit steht die Vorstellung des „unverursachten Verursachers“ möglicherweise nicht in Einklang mit den methodologischen Grundlagen der Physik, was noch zu prüfen wäre, sie widerspricht jedoch nicht naturwissenschaftlichen Erkenntnissen (Falkenburg 2012, bes. S. 267–326). Dass menschliche Handlungen etwas „echt Neues“ in die Welt bringen, schließt im übrigen logisch nicht aus, dass sie in anderer Hinsicht verursacht sind. So wird schon von Aristoteles das Phänomen des Zufalls erklärt: dass ein Ereignis zufällig ist, bezieht sich auf eine bestimmte Kausalreihe, nicht auf alle Kausalreihen (Aristoteles 1936; *Phys* II 4–9 (197 b)).

3.3.3 Merkmale wissenschaftlicher Intelligenz

Die Frage nach den Folgen der Einführung künstlicher Intelligenz für Wissenschaftler-Arbeitsplätze spezialisiert die Frage nach dem Unterschied zwischen künstlicher und menschlicher Intelligenz auf die Sub-Spezies (menschlicher) Wissenschaftler/Wissenschaftlerinnen. Prima facie dürfte unbestritten sein, dass sich die Tätigkeit von Wissenschaftlern/Wissenschaftlerinnen von anderen Tätigkeiten unterscheidet, nämlich dadurch, dass Wissenschaftler/Wissenschaftlerinnen als spontane Handlungsurheber von Wissen verstanden werden. Ein Gerät der künstlichen Intelligenz kann aber a priori kein Handlungsurheber sein, sondern immer nur ein „Anwender“ von etwas, was durch menschliche Intelligenz vorgegeben wurde.⁶ Die Explikation des Begriffes der künstlichen Intelligenz muss sich somit auf ein nicht-naturalistisches (finalistisches, intentionalistisches) Verständnis kognitiver Handlungen stützen.

Für die Beurteilung von Systemen Künstlicher Intelligenz in der Forschung bedeutet das näherhin, eine nicht-naturalistische Rekonstruktion der Tätigkeit des Wissenschaftlers zu entwickeln. Als Definitions-Skizze für die Explikation des Wissensbegriffs wird hier ein begründungstheoretischer Ansatz gewählt:

X weiß, dass p := Für alle Y , es gibt ein p : X kann p gegenüber Y begründen

Dieser Ansatz weicht von der oft verwendeten Definition ab, die Wissen durch Überzeugt-sein von einem Sachverhalts p und dem der-Fall-sein von p charakterisiert (Lenzen 1980, S. 53 ff.):

X weiß, dass p := Üp und p .*

Dieser „nicht-epistemische“ Ansatz ist durch zwei kaum auflösbare Probleme belastet. Einmal ist „Überzeugt-sein“ in wohl keinem Kontext semantisch „einfacher“ als „Wissen“; die Definition steht unter einem Obscurum per obscurius-Verdacht. Ferner ist unklar, wie ein performativ oder modal uneingebundenes „ p “ zu verstehen ist. Die Kommentare legen nahe, implizit den Modaloperator der Möglichkeit oder Wahrscheinlichkeit unterstellen (s. die Diskussion Lenzen 1980, S. 61 ff.).⁷ Vor allem aber besteht ein Explikationszirkel dadurch, dass man die alethischen Modaloperatoren nicht ohne direkten Rekurs auf „Wissen“ explizieren kann. Schließlich besteht das Problem, dass ein modal oder performativ uneingebettetes „ p “ zwar formal ein Satzradikal ist, das sich jeder Einbettung „andienen“ könnte, aber konnotativ einen präsupponierten epistemologischen Realismus suggeriert. Es wäre jedoch ein Verstoß gegen pragmatische Definitionsprinzipien,

⁶Vergl. Humm et al. in diesem Band, Kap. 2.

⁷Eine solche Deutung würde aber die Probleme der Semantik der Modallogik wie die De re- de dicto-Problematik in die Explikation des Wissensbegriffs einschleusen, was jedenfalls das Obscurum per obscurius-Problem verschärft.

durch eine Wortgebrauchsregel bereits eine Position in einer philosophischen Großdebatte (Realismus vs. Anti-Realismus) sozusagen unter der Hand zu entscheiden.

Auf der Basis des begründungspragmatischen Ansatzes lassen sich definieren:

X erkennt p := X ist dabei, Wissen zu erwerben / herzustellen

X meint, dass p := p ist „Kandidat“ des Wissens, dass p

Y zweifelt, dass p := Y fordert X auf, p zu begründen

Die gemeinsprachliche Verwendung des Ausdrucks „Wissen“ erweckt in Verbindung mit einem weitverbreiteten Vulgär-Cartesianismus den Eindruck, als sei Wissen ein privater innerer Vorgang, der gelegentlich von seinem „Besitzer“ „ausgedrückt“ wird.⁸ Demgegenüber verwendet die obige Definitionsskizze den Ausdruck resultativ, d. h. als ein Ergebnis eines sozialen Prozesses, nämlich des Begründens. Begründen ist eine regelgeleitete Sequenz von Redehandlungen, die mit einem konstativen performativen Modus beginnt, für den das „Behaupten“ hier exemplarisch eingesetzt wird. Eine regelgeleitete Sequenz von Redehandlungen heie „Diskurs“. Diskurse lassen sich wie alle sprachlichen Phänomene empirisch beschreiben und erklären und fallen somit in den Gegenstandsbereich der empirischen Sprachwissenschaften. Um *interlingual* die Regeln korrekter diskursiver Verfahren zu rekonstruieren (Logik, Topik), muss man demgegenüber auf die Instrumente einer formalen Pragmatik von Redehandlung und Redehandlungssequenzen zurückgreifen (Gethmann 1979).

3.3.4 Behaupten und Begründen

Eine Redehandlungssequenz soll „Diskurs“⁹ heißen, wenn gilt:

- (i) Die vorkommenden Redehandlungen lassen sich nach ihrem spezifischen „Ort im Diskurs“ klassifizieren, und zwar können sie sein:
 1. Eröffnungshandlungen (initiative Redehandlungen), wie beispielsweise das Behaupten oder Auffordern;
 2. Fortsetzungshandlungen (re-aktive Redehandlungen), wie beispielsweise das Zweifeln;
 3. Beendigungshandlungen (resultative Redehandlungen), wie beispielsweise das Zustimmung oder Abweisen.

⁸Zur generellen Kritik am nach-cartesischen Mentalismus s. Gethmann und Sander (2002).

⁹Mit „Diskurs“ wird hier der in der Erlanger Schule rekonstruierte „Dialog“ bezeichnet. Der Ausdruck „Dialog“ erscheint aus zwei Gründen verbal verfehlt: Einmal heißt $\delta\iota\alpha$ nicht „zwei“ (vielmehr $\delta\upsilon\omicron$), sondern „durch“ (per); ferner wird mit „Dialog“ oft eine eher emphatische Form von Verständigung gemeint, worauf es hier nicht ankommen kann.

- (ii) Eröffnungshandlungen sind solche, deren Zweck die Diskursbeendigungshandlung der Zustimmung ist. Sie können auch „geltungsbeanspruchende“ Redehandlungen heißen, sofern ihr Anspruch die Zustimmungsfähigkeit („Zustimmbarkeit“, „Akzeptabilität“) ist.
- (iii) Die Teilnehmer an Diskursen („Parteien“) nehmen genau zwei Rollen wahr:
 - erste Rolle: Diskurseröffnung und diese stützende („fundierende“): Proponent;
 - zweite Rolle: Diskursfortsetzung oder -beendigung: Opponent.
 Dabei können die Parteien Individuen oder Kollektive sein. Ferner können Parteien auch beide Rollen in Personalidentität wahrnehmen.
- (iv) Die Abfolge von Redehandlungen zwischen den Rollenträgern erfolgt gemäß impliziten Korrektheitsregeln, die konstitutiv in die Gelingensbedingungen der Einzelhandlungen eingehen.
 - „Ich behaupte, dass p“
 - kann dann paraphrasiert werden durch:
 - „Ich bin gegenüber einem Adressaten bereit, im Zweifelsfall eine Begründungshandlung zu vollziehen.“

Diskurse, deren Eröffnungshandlungen zum Typ der Konstativa gehören, sollen „Begründungsdiskurse“ heißen.

3.3.5 *Wahrheit als Wohlbegründetheit*

Die vorstehende Explikation des Begriffs des wissenschaftlichen Wissens bindet diesen an die Fähigkeit eines zweckorientierten Akteurs, Behauptungen im Zweifelsfall begründen zu können. Wissen kann somit nur ein Akteur ausbilden, der wenigstens über die Handlungskompetenzen des Behauptens, Bezweifeln, Aufforderns, Zustimmens bzw. Abweisens verfügt. Ein solcher Akteur verfügt über nicht messbare propositionaler Einstellungen, Leiblichkeit und Individualität und kann daher den Turingtest nicht bestehen. Daraus ergeben sich folgende Konsequenzen:¹⁰

(a) *Wahrheit*

Das Zusprechen des Beurteilungsprädikators „wahr“ zu einem propositionalen Gehalt ist nur zulässig, wenn ein diskursives, für jedermann nachvollziehbares Verfahren angegeben werden kann, das diesen propositionalen Gehalt als zustimmungsfähig auszeichnet. Das Hauptargument für einen solchen, wesentlich auf *Verfahren* bezogenen Wahrheitsbegriff („Prozeduralismus“) besagt in Anlehnung an M. Dummett (1978), dass das Wort „Wahrheit“ auf nichts referieren kann, was das Erkenntnisvermögen seiner Benutzer prinzipiell transzendiert; deshalb muss das Zu- und Absprechen von Wahrheit an Verfahren der Verifikation bzw. Falsifikation zurückgebunden werden. Solange eine Aussage weder als begründbar noch unbegründbar ausgewiesen ist, kann man deshalb nicht davon ausgehen, dass sie kraft ihrer Dar-

¹⁰Die folgenden Abschnitte sind eine Zusammenfassung von: Gethmann (2014).

stellungsfunktion von einer erkenntnisunabhängigen Realität „wahr gemacht“ wird. Die Bedeutung einer Aussage kann folglich nicht durch Angabe ihrer *Wahrheitsbedingungen* im Sinne der formalen Semantik erklärt werden, sondern nur durch Angabe ihrer *Behauptbarkeitsbedingungen*, d. h. der Umstände, unter denen ein Sprecher eine Aussage (korrekt) behaupten darf. Wäre eine Aussage allein deshalb wahr, weil der durch sie dargestellte Sachverhalt besteht, dann könnte die Feststellung ihrer Wahrheit das Erkenntnisvermögen des Sprechers unter Umständen prinzipiell übersteigen. Dies hätte zur Folge, dass sich die Kenntnis der Bedeutung dieser Aussage nicht im Handeln manifestieren kann, weil es keine Möglichkeit gibt, über das Zutreffen der Wahrheitsbedingungen zu befinden. Unentscheidbare Aussagen sind weder wahr noch falsch, weil die Wahrheit einer Aussage nicht über ihre prinzipielle Verifizierbarkeit hinausgehen kann. In dieser Fassung ist die Problemstellung nicht spezifisch auf mathematische oder andere analytische Wahrheiten beschränkt, sondern sie gilt für beliebige Kontexte; sie betrifft in gleichem Maße lebensweltliche Erfahrung, empirische Laborwissenschaften, exakte Formalwissenschaften und historisch-hermeneutische Kulturwissenschaften. In allen diesen Kontexten ist die epistemologische Grundfrage zu beantworten, wie das Verhältnis zwischen Wahrheit (als Eigenschaft einer mentalen oder lingualen Entität wie Urteil oder Behauptung) und dem Verfahren zu denken ist, das zur Feststellung dieser Eigenschaft führt.¹¹

Die Bindung des Wahrheitsbegriffs an menschliche Verfahrenssouveränität gibt Anlass, auf den grundlegenden Unterschied von *Algorithmik* und *Heuristik* hinzuweisen. Mit der Entwicklung der KI ist die Vorstellung verbunden, dass die Wahrheitsfindung grundsätzlich durch Algorithmen übernommen werden kann und somit wenigstens langfristig (nach weiterer technischer Perfektionierung) Forscher/Forscherinnen „aus Fleisch und Blut“ durch KI-Maschinen abgelöst werden können. Dieser Vorstellung steht allerdings schon das von Gödel in den 30.er Jahren des 20. Jahrhunderts aufgestellte Unvollständigkeitstheorem im Wege, demgemäß jedes formale System bei der algorithmischen Erzeugung von ableitbaren Sätzen wahre Gesetze auslöst; dazu gehören insbesondere meta-sprachliche Sätze, beispielsweise solche über die Widerspruchsfreiheit des Systems. Aus philosophischer Sicht zeigen Gödels Theoreme die Grenzen des algorithmisch Erzeugbaren (Mainzer 1995, S. 100 f., 789 f.; Nida-Rümelin und Weidenfeld 2018, S. 111 f.) Spezifisch menschliche Intelligenz übersteigt deswegen jedes algorithmische Verfahren, weil heuristische Verfahren die algorithmischen nicht nur bei Bestimmung der Eigenschaften des formalen Systems übersteigen, sondern ihr wesentlich vorausgehen muss, indem im vorhinein der Zweck des Algorithmus festgelegt werden muss,

¹¹ Gegen diese Explikation des Wahrheitsbegriffs wird eingewandt, sie unterscheide nicht zwischen der Wahrheit einer mentalen oder lingualen Entität und ihrer epistemischen Qualifikation; das Wahrschein sei vom Fürwahrgehaltenwerden strikt zu unterscheiden. Diesem Einwand ist entgegenzuhalten, dass die Abstraktion der Wahrheit vom Verfahren des Bewahrheitens zu einem sinnlosen Ergebnis führt: es kann nicht mehr erklärt werden, was es bedeutet, eine Proposition p für wahr zu halten, wenn man keine Vorstellung hat, wie man die Behauptung von p diskursiv begründen kann. – Zu dem hier vertretenen „Anti-realismus“ vgl. weiterführend: Wright 1992; Tennant 1997; Siegart 1997; s. a. den Diskussionsband Forum für Philosophie 1992.

damit auch seine Gelingensbedingungen und damit seine Qualitätsstandards. Jeder Anwendung von Algorithmen geht eine Heuristik voraus.

Am Ablaufschema eines Begründungsverfahrens lässt sich die Unvermeidbarkeit heuristischer, d. h. regelüberwachter Suchverfahren leicht illustrieren. Angenommen der Opponent äußert einen Zweifel an dem vom Proponenten behaupteten Sachverhalt. ein solcher Zweifel „folgt“ aus der Behauptung keineswegs. Der Opponent könnte auch sofort zustimmen, zum Beispiel, weil er bereits eine Begründung kennt oder an einer solchen kein Interesse hat. Die Behauptung könnte ihm relativ zu seinem bisherigen Überzeugungs-Ensemble als schwer integrierbar erscheinen. Er könnte auch überzeugt von der Behauptung sein, aber neugierig, worauf der Proponent sich im nächsten „Zug“ stützt. Der Proponent auf der anderen Seite muss „suchen“, welche propositionalen Gehalte zur Überzeugungswelt des Opponenten gehören *und* durch welche Regel er den Übergang zwischen den Prämissen und seiner Anfangsbehauptung herstellt. Insgesamt ist leicht zu sehen, dass das Begründen ein Suchprozess, ist, der sich grundsätzlich vom Ablauf eines algorithmischen Verfahrens unterscheidet. Was für Begründungen generell gilt, gilt auch für den Sonderfall des *Beweisens*, d. h. des Begründens im Rahmen einer formal konstituierten Sprache. Von formalen Sonderfällen abgesehen muss der Beweis für ein Theorem gesucht/gefunden werden, selbst dann, wenn die Beweisbarkeit des Theorems aus formalen Gründen a priori (im Rahmen der konstituierten Sprache) feststeht.

(b) Realität

Die Behauptung der Existenz ist an den effektiven Existenznachweis gebunden. Verallgemeinert heißt das: *Realität* ist immer durch Verfahren vollzogene Realität. Es gibt keine schlechthin verfahrenstranszendente Realität. Versteht man unter „Realität“ das Ensemble der Tatsachen, d. h. *wahrer* Sachverhalte, dann hängt die Frage, was „real“ ist, von den Verfahren ab, die eine sprachliche Entität als „wahr“ auszeichnen. Eine solche Position ist erfahrungsgemäß dem Missverständnis ausgesetzt, die Realität werde als bloßes Produkt subjektiver Vollzüge angesehen. Es wird jedoch nicht behauptet, Tatsachen seien *Resultate* des Wahrmachens (dies wäre klassischer „Idealismus“ oder „Irrealismus“), sondern nur, dass es Tatsachen *nicht ohne* Verfahren des Wahrmachens „gibt“: Verfahren sind notwendige, nicht hinreichende Bedingung für das Zusprechen von Realität. Die Bindung des Wahrheitsbegriffs an den Begriff der Begründbarkeit (oder einen ähnlichen) setzt ein vorgängiges Verständnis der Fähigkeiten eines handelnden Wesens voraus. *Irrelative Wahrheit wird nicht durch Transzendenz (einer „Realität“), sondern durch Invarianz (eines wahr machenden Verfahrens) garantiert.* Die Qualifikationen der Begründbarkeit oder Unbegründbarkeit hängen von Verfahrensstrukturen („Strategien“), nicht vom jeweiligen raumzeitlichen Vollzug („Partie“) ab.¹²

¹² Die Unterscheidung von Strategie und Partie entspricht derjenigen von Handlungsschema und -vorkommnis (bzw. type und token); vgl. Lorenz 1976.

(c) Subjektivität

Damit sind hinsichtlich des Verfahrensverständnisses sowohl ein Moment der Faktizität als auch ein solches der Invarianz zu unterscheiden. Die Faktizität liegt darin, dass es einen Handlungsrahmen geben muss, relativ zu dem Verfahrensvorschriften formuliert werden, deren Ausführungen Handlungen sind. Solche faktischen Rahmenbedingungen sind z. B. Materialeigenschaften derjenigen Körper, aus denen Messgeräte hergestellt werden, mit deren Hilfe wiederum Messverfahren durchgeführt werden; oder die Kompetenzen derjenigen Akteure, die Verfahren durchführen, z. B. sprachliche elementare Fähigkeiten und Fähigkeiten der Körperbearbeitung. Diese setzen einen leiblich verfassten Akteur voraus, der in eine faktische Lebenswelt mit sozialen und natürlichen Momenten eingebettet ist, die (kontingent) sind.

Der Bezug auf einen Akteur, der die Kompetenz besitzt, Verfahren durchzuführen, traditionell gesprochen: das erkennende *Subjekt*, ist für das Zusprechen von Wahrheit und das Erkennen von Realität somit wesentlich. Das Verfügenkönnen über ein Verfahren ist vor allem deswegen als Wahrheitskriterium geeignet, weil es nicht nur um das private und raum-zeitlich einmalige Für-wahr-Halten, sondern um die prinzipielle Nach-Vollziehbarkeit für jedermann geht. Der Verfahrensbegriff ordnet die Wahrheitsqualifikation nicht nur einer faktisch-allgemeinen, sondern einer prinzipiell öffentlichen Sphäre zu. Dies bedeutet aber, dass das Verfügenkönnen über ein Verfahren nicht adäquat allein mit mentalen Termini wie Vorstellen, Denken usw. erfolgen kann. Ein Proponent, der äußert: „Ich kann mir ein Verfahren denken, aber ich kann es nicht mitteilen“, verstößt gegen eine wesentliche Gelin- gensbedingung des Wahrheitsdiskurses. Der Wahrheitsdiskurs verläuft nicht auf einer Hinterbühne, deren Spiel gelegentlich auch auf der Vorderbühne „ausgedrückt“ wird, sondern das Wahrheitsverfahren ist wesentlich ein öffentlicher Vorgang (vgl. Gethmann und Sander 2002).

Dieses „Manifestationsprinzip“ (Dummett 1993, S. 37) besagt also, dass eine Verdoppelung der Wirklichkeit in eine Denk- und eine Sprachsphäre überflüssig ist, weil mentale Prozeduren wiederum mit lingualen, also öffentlichen Kategorien rekonstruiert werden müssten. Um nämlich sichergehen zu können, dass zwei Sprecher Dasselbe „meinen“, muss sich die Bedeutung der verwendeten Ausdrücke im öffentlichen Sprachgebrauch niederschlagen können. Der über öffentlich zugängliche Verfahren verfügende bzw. diese kontrollierende und nachvollziehende menschliche Akteur muss als einer angenommen werden, der Handlungen gemäß Verfahren als Zweckrealisierungsversuche vollziehen kann.

Eine Verständigung über Handlungen ist generell nur möglich, wenn Akteure nicht nur über Handlungsvorkommnisse (tokens), sondern auch über Handlungsschemata (types) kontrolliert reden können. Über ein Handlungsschema reden heißt, wissen, wann *dasselbe noch einmal* getan wird (Kamlah und Lorenzen 1973, Kap. II.2; Lorenz 1976, S. 258). Je nach Handlungstyp gibt es Kennzeichen dafür, ob jemand diesen Handlungstyp beherrscht oder nicht. Ein Verfahren des Begründens muss wenigstens folgende Anforderungen erfüllen, was auf Seiten des Akteurs Fähigkeiten nicht-trivialer Art unterstellt:

- (i) Das Verfahren muss *regelmäßig* sein, d. h. es müssen wenigstens implizite Regeln rekonstruierbar sein, die für jedermann die Wiederholbarkeit sichern. Damit wird für jeden Akteur die Fähigkeit unterstellt, eine kanonische Handlungsfolge einzuhalten (regulative Kompetenz).
- (ii) Das Verfahren muss *lückenlos* sein, d. h. es darf beim schrittweisen Durchlaufen keine Schritte geben, die nicht durch die Regeln erlaubt sind. Dies unterstellt für jeden Akteur die Fähigkeit, Handlungsfolgen als Folgen zu bilden und wahrzunehmen (konsekutive Kompetenz).
- (iii) Das Verfahren muss *zirkelfrei* sein. Dies unterstellt die Fähigkeit, Wiederholungen in einer Handlungsfolge zu erkennen und zu vermeiden (repetitive Kompetenz).

Das Subjekt, auf das der Verfahrensbegriff verweist, muss also mit spezifischen, durchaus anspruchsvollen Fähigkeiten wie regulativer, konsekutiver und repetitiver Kompetenz ausgestattet gedacht werden.

Der Gedanke der Bindung des „Wahrseins“ an das Vorführenkönnen eines Verfahrens ist als Versuch zu verstehen, das Problem der Unkontrollierbarkeit von prä-tendierten Wahrheitsintuitionen zu vermeiden. Ein endlicher Akteur muss sich die Wahrheit einer sprachlichen Entität gemäß dem regelgerechten, lückenlosen und zirkelfreien Durchlaufen eines Verfahrens – diskursiv, nicht-intuitiv – im Prinzip jederzeit reproduzierbar vor Augen führen können.

3.3.6 Zusammenfassung

Technische Artefakte, die mit sog. künstlicher „Intelligenz“ ausgestattet sind, können grundsätzlich nicht als Akteure in Begründungsdiskursen auftreten, weil ihnen die pragmatischen Merkmale der Handlungsurheberschaft und Zurechenbarkeit nicht zukommen. Wie immer in unterschiedlichen kognitiven Kontexten Handlungen in Begründungsdiskursen wie Behaupten, Bezweifeln, Zustimmung u. a. propositional bestimmt sind, technische Artefakte können diese Handlungen zwar unterstützen, aber nicht selbst ausführen.

3.4 Das wissenschaftliche Bacon-Projekt und die KI

Der schärfste Epochenbruch, den die Wissenschaftshistoriografie bisher beschrieben hat, ist der Übergang von der antik-mittelalterlichen Auffassung der Wissenschaft zur früh-neuzeitlichen. Dieser Übergang ist besonderer Prägnanz in der programmatischen Schrift *Novum Organon* des Francis Bacon beschrieben worden, die diese Transformation zu methodologischen Forderungen zusammenfasst; deshalb spricht man auch vom Bacon-Projekt (Schäfer 1993; Gethmann 2003). Als markante Schwellenphänomene beschreibt Bacon eine zweifache Transformation der

neuen Wissenschaft gegenüber der alten. Es geht *einmal* (a) um eine Transformation der Erkenntnisstile, die man mit der Wendung „Kontemplation versus Intervention“ zusammenfassen kann: Die Kontemplation der Natur wird abgelöst durch die Intervention in die Natur. Es geht *zum anderen* (b) um eine Transformation der Wissensformen des poetischen und praktischen Wissens, die zueinander in Beziehung gesetzt werden. Beide Unterscheidungen sind im übrigen nicht neu, neu ist aber das Verhältnis zwischen ihnen, wie Bacon es bestimmt.

3.4.1 *Kontemplation versus Intervention*

Nach antiker und mittelalterlicher Auffassung kann sich Wissen nur aus einer kontemplativen, also in einer der inneren Schau gewidmeten Einstellung gegenüber der Natur ergeben. Die Begründung hat Platon prägnant formuliert: Wissen kann es nur vom Allgemeinen und Unveränderlichen geben, die Welt um uns herum ist aber partikulär und veränderlich, also können wir nur durch Kontemplation des Allgemeinen zum Wissen gelangen. Demgegenüber verdankt sich neuzeitliche Wissenschaft, deren Wurzeln freilich schon im späten Mittelalter liegen, der Überzeugung, dass nur durch Intervention in Naturabläufe kausale oder konditionale Beziehungen erkennbar werden. Das Paradigma dieser Transformation des Erkenntnisstils ist das Experiment, das – entgegen einem verbreiteten Ondit über die *Naturwissenschaften* – nicht in der gegebenen Natur, sondern an einem kulturell hergestellten Ort menschlicher Arbeit – dem *Labor* – stattfindet. Dabei transformiert sich durch Bacon das Natur- wie das Technikverständnis. Natur ist nicht mehr nur die gegebene oder phänomenale Natur, sondern das entscheidende Merkmal der Natur ist das Naturgesetz, dessen Struktur mathematisch darstellbar ist; Technik ist nicht mehr, wie bei Aristoteles, das Wider-Natürliche, sondern Technik ist das, was naturgesetzlich möglich ist. Zwischen einem kontemplativ erworbenen Wissen und der Verfügungsmacht (*potestas*) über die Natur gibt es in der vor-neuzeitlichen Wissenschaft keinen Konnex, wie Bacon kritisiert. Und somit sieht man bis ins späte Mittelalter auch so gut wie keinen systematischen Zusammenhang zwischen Wissen und Naturbeherrschung.

Erst durch ein aufgrund von Intervention in Naturabläufe gebildetes Wissen wird die Allianz von Wissen und Ingenieurskunst, die ein Markenzeichen der Neuzeit ist, möglich. Verfügungsmacht tritt im Verhältnis zum Wissen an zwei Orten auf, *einmal* als Folge des Wissens. Wer der Natur durch geschicktes Isolieren von Kausalfaktoren auf die Schliche kommt, kann die Abläufe im Prinzip auch verändern, er gewinnt Macht über die Natur. *Zum anderen* aber erreicht der Mensch Verfügungsmacht über die Natur durch Intervention in sie nur dadurch, dass er bereits über ein bestimmtes Know-how der Materialbearbeitung und der Herstellung von Messgeräten verfügt (Janich 2015).

Dass dieser Umstand häufig übersehen wird, zeigt sich zum Beispiel in der üblichen Bezugsetzung zwischen Ingenieurskunst und Wissenschaft. Dass wissenschaftlich fundierte Ingenieurskunst eine Folge von Wissen ist, ist bekannt, und dass

Ingenieurskunst insoweit angewandte Wissenschaft ist, mag dahinstehen. Aber die Ingenieurskunst geht auch dem Wissen voraus. Ohne eine lebensweltlich verfügbare elementare Körperbearbeitung und die Fertigkeit im Umgang mit elementaren Geräten beispielsweise der Körperbearbeitung (bis hin zu Messgeräten) gibt es keine Labore, keine Experimente, also keine Interventionen. In diesem Sinne ist Wissenschaft auch angewandte Ingenieurskunst, und deswegen schätzen die Wissenschaftler/Wissenschaftlerinnen die Erfinder. Bacon ist wohl der erste, der die Rolle der Erfinder für die Wissenschaft mit besonderer Hochschätzung würdigt.

Wenn das Know-how eine Folge bestimmten Wissens ist, aber auch möglicherweise Bedingung eines nächsten Wissens, dann kann man sich vorstellen, dass eine Art Zusammenhang, eine Kette von Wissensformen entsteht, jeweils verbunden über bestimmte Fähigkeiten im Geräteumgang und in der Naturintervention, ein kumulativer Zusammenhang, der sich zu einem „Fort-schritt“ verkettet. Und in diesem Sinn ist Fortschritt ein spezifisches Merkmal neuzeitlicher Wissenschaft. Nicht, dass es in Antike und Mittelalter nicht gelegentlich etwas Neues gegeben hätte, aber auf dem Hintergrund des interventionistischen Erkenntnisstils läßt sich ein Handlungswissenskontinuum vorstellen, das den Wissenserwerb planbar macht. „Wissen ist Macht“ heißt auch: Fortschritt ist planbar.

3.4.2 *Poietisches (technisches) und praktisches Wissen*

Die zweite Unterscheidung, die wesentlich in das Diktum von Bacon eingeht, ist die Unterscheidung zwischen dem *poietischen*¹³ und dem *praktischen* Wissen. Neu ist auch hier ihre Verhältnisbestimmung. Aristoteles bestimmt definitorisch, dass das poietische Wissen, seinen Zweck außerhalb des erkennenden Akteurs hat, etwa im Zusammenhang mit der Herstellung eines Artefakts, während das praktische Wissen seinen Zweck im Akteur selbst oder in seinen Ko-Akteuren hat. Praktisches Wissen ist das Wissen, das man braucht, um einen Freund zu beraten, um eine Gesellschaft zu organisieren, etwa durch Ausbildung von Jurisprudenz, oder das man braucht, um die Heilkunst auszuüben. Bacons Diktum verbindet nun diese beiden Aspekte auf eine eigentümliche Weise. Das poietische, das technisch-naturwissenschaftliche Wissen, das durch Intervention in Naturabläufe gewonnen wird, dient einem praktischen Zweck, der im erkennenden Akteur liegt. Etwas verallgemeinert, Wissenschaft dient der Befreiung des Menschen von natürlichen und sozialen Zwängen. Eine solche Welt – befreit von natürlichen und sozialen Zwängen – hat Bacon in seinem utopischen Reiseroman „Atlantis“, einer der früh-neuzeitlichen Utopien, vorgestellt. Neuzeitliche Wissenschaft ist nach Bacon gerade nicht zweckfrei, sondern an einen allgemeinen humanen Zweck gebunden. Sie ermöglicht im gelingen-

¹³*Gr. ποίησις (herstellendes Handeln) gegenüber πράξις (zwischenmenschliches Handeln). Statt „poietisch“ sagt man heute „technisch“, aber τέχνη (ars, Lehre/Wissenschaft) ist bei Aristoteles beides; die Unterscheidung zwischen Wissenschaft und Technik ist also eigentlich windschief (at odds).

den und günstigen Falle, das Verfügungswissen zu erlangen, das gebraucht wird, um die Befreiung von natürlichen und sozialen Zwängen zu bewirken. Zusammenfassend kann man vom *praktischen Zweck* neuzeitlicher Wissenschaft sprechen.

3.4.3 *Zweck und Nutzen wissenschaftlichen Wissens*

Die Rede von der *Zweckorientierung* der Wissenschaft muss allerdings gegen Missverständnisse gesichert werden, da damit nicht selten Vorstellungen von einer Indienstnahme der Wissenschaft für partikuläre ökonomische Interessen (wessen auch immer) verbunden werden. Vor allem diese Partikularisierung des Zweckbegriffs und die damit verbundene Partikularisierung des Interessenbegriffs sind maßgeblich dafür, dass viele Wissenschaftler/Wissenschaftlerinnen für die Beschreibung ihres Selbstverständnisses auf der „Zweckfreiheit“ bestehen. Zwar ist die Debatte über einer Gesellschaftstheorie und Geschichtsmetaphysik hinweggegangen, die die Wissenschaften für partikuläre Zwecke (durch Einnahme des sog. Klassenstandpunkts) instrumentalisiert. An die Stelle dieses Theoriesyndroms ist jedoch eine verbreitete Ökonomisierung der Wissenschaft im Rahmen der Arena des Marktes getreten. Demgegenüber ist daran festzuhalten, dass Wissenschaft zwar nicht utilitären und partikulären Nutzenerwägungen unterworfen ist, gleichwohl aber einem trans-utilitären Zweck dient; „Zweck“ – im Sinne einer universellen und humanen Zwecksetzung – ist von (ökonomischem) „Nutzen“ zu unterscheiden.

3.4.4 *Kausalität und Korrelation*

„Big Data“ und ihre Verarbeitung durch KI setzt nach Meinung einiger Wissenschaftsphilosophen die Wissenschaften in die Lage, auf das schwierige Geschäft der Kausalerklärung zu verzichten und sie durch Korrelationsanalysen zu ersetzen. In diesem Sinne wird von einem „Ende der Theorie“ (Anderson 2008) gesprochen, nicht in dem Sinne, dass man mit dem Untergang der Theorie etwas Wertvolles verlöre, sondern in dem Sinne, dass eine lästige, eigentlich undurchführbare Aufgabenstellung, nämlich die Explikation eines adäquaten Kausalitätsverständnisses, endlich aufgegeben werden kann. Kausale Hypothesen werden unter den Rahmenbedingungen des Bacon-Projekts durch Experimente, d. h. durch Intervention in die Natur gefunden. Somit bedeutet das Ende der Theorie auch das Ende der Notwendigkeit des Experiments. An die Stelle des Experiments sollen durch Algorithmen gesteuerte Simulationen und Modellkonstruktionen treten. Damit würde die neuzeitliche Vorstellung von Wissenschaft durch KI abgelöst und somit ein neuer Epochenbruch herbeigeführt.

Mit der Losung vom „Ende der Theorie“ wird somit in erster Linie auf die Ablösung der regulativen Idee der deterministischen Kausalerklärung durch Korrelati-

onsanalysen in Form von Mustererkennung in großen Datenmengen abgehoben. Der interventionistische Erkenntnisstil zielt dagegen auf die Beherrschung der Natur durch die Aufdeckung von Ursache-Wirkungs-Verhältnissen ab, deren Erkenntnis die notwendige Bedingung dafür ist, dass man durch Handlungen planvoll in sie eingreifen kann. Somit hängt der Erfolg dieses Wissenschaftsparadigmas von der korrekten Explikation von Ursache-Wirkungs-Verhältnissen ab. Ausgangspunkt der wissenschaftsphilosophischen Bemühungen um die begriffliche Rekonstruktion von Ursache-Wirkungs-Verhältnissen ist die bahnbrechende und folgenreiche Feststellung Humes, dass man zwar die scheinende Sonne und den etwas später erwärmten Stein optisch und haptisch wahrnehmen kann, nicht aber deren zeitliche Abfolge und schon gar nicht die Ursache-Wirkungs-Beziehung zwischen den Ereignissen. Behauptungen über zeitliche Sukzessivität und Kausalität zwischen zwei oder mehreren Ereignissen haben keine Referenzbeziehung zu wahrnehmbaren Gegenständen bzw. Ereignissen, sondern verbinden diese vielmehr durch bereits investierte Kategorien der zeitlichen Abfolge bzw. der Verursachung. Zeitliche Abfolge und Ursache-Wirkungs-Verhältnisse gehören damit neben anderen Kategorien zu den semantischen Netzwerken, über die das erkennende Subjekt bereits verfügen muss, wenn es Ereignisse zueinander in Beziehung setzt.

Hinsichtlich des Verhältnisses von Korrelationsanalyse und Kausalerklärung interagiert die begriffliche Organisation und Interpretation interventionistischer („experimenteller“) Forschung mit der grundsätzlichen Interpretation des Wahrscheinlichkeitsbegriffs. Wer der Überzeugung ist, dass die Welt „an sich“ deterministisch organisiert ist, Wahrscheinlichkeitsaussagen sich daher lediglich (vorläufigen) Erkenntnisdefiziten verdanken („Gott würfelt nicht.“), wird die Relationen zwischen unterschiedlichen Ereignissen vollständig und disjunkt in kausale und korrelative einteilen; es handelt sich dabei um eine *kontradiktorische* Unterscheidung. Wer demgegenüber unterstellt, dass Relationen zwischen Ereignissen grundsätzlich probabilistisch organisiert sind, wird diese Relationen auf einem Kontinuum (zum Beispiel zwischen 0 und 1) anordnen, dessen Endpunkte sich *polar-konträr* gegenüberstehen. Die damit angedeutete Kontroverse ist eine innerhalb des interventionistischen Paradigmas.

Sowohl Deterministen wie Probabilisten unterstellen ein Verständnis von Kausalität, das ihre Feststellung bezüglich des Zusammenhangs (Korrelation) bestimmter Ereignisse begründet. Der Determinist will kausale Zusammenhänge von bloß zufälligen, der Probabilist starke (signifikante) von schwachen (kontingenten) Korrelationen unterscheiden. Das Kausalitätsprinzip spielt somit eine bezüglich der hier betrachteten Positionen invariante kriteriale Rolle, die grundsätzlich unvermeidbar ist. Für das wissenschaftliche Geschäft wäre daher ein klares Verständnis von Kausalität unverzichtbar. Bedauerlicherweise ist die Rekonstruktion dieses Verständnisses jedoch eine der wissenschaftstheoretischen Baustellen, deren prekärer Zustand sogar zu tiefer Skepsis gegenüber dem interventionistischen Erkenntnisstil führen könnte.

Die wichtigste Funktion des Kausalitätsprinzips ist die Abwehr sogenannter ursachenloser Ereignisse oder Ereignisse mit obskuren Ursachen (zu denen nach Hume auch wundersame göttliche Eingriffe gehören; s. o. Abschn. 3.3.2). Das Kau-

salitätsprinzip ist das grundlegende Rationalitätsprinzip der neuzeitlichen interventionistischen Wissenschaften; würde man für bestimmte Ereignisse das Kausalitätsprinzip außer Kraft setzen, würde man auch kausal unerklärte oder durch obskure Ursachen erklärte Ereignisse in den Wissenschaften zulassen. Unter der Bedingung der Gültigkeit des Kausalitätsprinzips sind Situationen, in denen man eine Ursache nicht benennen kann, grundsätzlich als Antrieb zu verstehen, die Ursachenforschung fortzusetzen, da ausgeschlossen ist, dass es keine oder nur obskure Ursachen gibt. Es ist offenkundig, dass die Ausdehnung des Kausalitätsprinzips auf im klinischen Kontext diagnostizierte Symptome von großer Bedeutung für das Verständnis der medizinischen Praxis (gegenüber manchen sogenannten alternativen Medizinverständnissen) ist.

Während das Kausalitätsprinzip als Grundlage neuzeitlicher interventionistischer Wissenschaft weithin unumstritten ist, beginnen die Probleme mit der Explikation von *Kausalgesetzen*. Kausalgesetze sind allgemeine Behauptungen über Ursache-Wirkungs-Verhältnisse zwischen Wirkungsereignissen und Ursacheereignissen. Nachdem Hume konstatiert hat, dass Ursache-Wirkungs-Verhältnisse nicht Referenzobjekte der menschlichen Wahrnehmung oder menschlicher Wahrnehmungs-Behauptungen sein können, wird in der Wissenschaftsphilosophie kontrovers diskutiert, welche Mindestbedingungen erfüllt sein müssen, um zu gültigen Kausal-Behauptungen zu gelangen.

Der Ansatz von G.H. von Wright schließt sich unmittelbar an den interventionistischen Erkenntnisstil an und versucht, das Verhältnis von Ursache und Wirkung unter Einbeziehung *handlungstheoretischer Kategorien* zu erklären (von Wright 1971). Von Wright weist auf die alltägliche Handlungserfahrung hin, zu der es gehört, dass der Akteur sicher ist, durch bestimmte Handlungen bestimmte Wirkungen herbeiführen zu können. Dabei wird „Handeln“ grundsätzlich nicht als physikalisch beschreibbaren Episode betrachtet (dies würde ersichtlich zu Zirkelproblemen führen), sondern als Zweck-Realisierungsversuch. Betrachtet man eine Handlung als absichtliche (geplante) Herbeiführung einer Wirkung, wird dagegen ein Kausalitätsgesetz noch nicht vorausgesetzt, da die Wahrheit von Aussagen über den Zusammenhang von verursachender Handlung und bewirkter Zweck-Erreichung keine methodische Vorbedingung für die Möglichkeit des Handelns ist. Die Allgemeinheit von Naturgesetzen im Sinne regelmäßig wiederkehrender (kausaler) Verbindungen von Ereignissen wird durch ein aus Handlungskontexten verallgemeinertes Verfügungswissen gewonnen, das sich in experimentellen Zusammenhängen bewähren muß. Dabei ist zu berücksichtigen, dass Experimente ihrerseits technisch hergestellte Zweckkontexte sind, die letztlich auf erfolgreich ausgeführte Herstellungshandlungen zurückzuführen sind (Janich 1997). Falsche Kausalitätsbehauptungen beruhen dann auf unrichtig ausgeführten Handlungssequenzen und sind durch Fehler in diesen erklärbar. Der handlungstheoretische Ansatz gibt somit grundsätzlich Kriterien für die Unterscheidung von wahren und falschen Kausalbehauptungen an die Hand. Der handlungstheoretische Ansatz steht jedoch mit dem Kausalitätsprinzip als grundlegendem Rationalitätsprinzip interventionistischer Wissenschaft nicht in Einklang, wenn man das Kausalitätsprinzip als generelle Bedingung für wissenschaftliches Wissen betrachtet. Handlungen im Sinne eines „finalistischen“ (inten-

tionalistischen) Handlungsverständnisses sind danach selbst keine möglichen Gegenstände interventionistischer Wissenschaft. Entsprechend kommt man bei einem so verstandenen Instrumentalismus nicht darum herum, die Existenz unterschiedlicher Typen wissenschaftlicher Wissensbildung zu unterscheiden.

3.4.5 Die Bedeutung von Big Data für die Wissenschaft

Am Anfang des Prozesses der Genese von Informationen stehen elektrotechnisch durch Folgen von 0 und 1 hergestellte Signale, die zu Symbolen codiert werden. Diese werden durch syntaktische Regeln zu Daten zusammengeführt. Solche Daten werden in einen Kontext gesetzt und bekommen dadurch Bedeutung. Erst durch diese Bedeutung erhält man eine Information. So bedeutet das Datum 39,8 mit der Maßeinheit Celsius im Kontext der Medizin Fieber. Informationen werden dann mit anderen Informationen zu Wissen verbunden, beispielsweise die Diagnose einer Krankheit, um auf dieser Grundlage Probleme zu lösen, wie zum Beispiel eine Therapie einzuleiten. Durch die moderne Entwicklung u. a. im Bereich der Speichertechnik geht es heute allerdings nicht mehr nur um strukturierte Daten, sondern vor allem um unstrukturierte Daten. Dadurch entsteht eine gewaltige, amorphe Datenmasse, die durch die herkömmlichen Datenbanktechnologien und Algorithmen nicht bewältigt werden kann. Neuartige Such-Algorithmen, die diese Datenmassen nach Datenkorrelationen und Datenmustern durchsuchen, führen schließlich die Ergebnisse zusammen, um daraus beispielsweise Trends, Profile von Produkten und Personen abzuleiten. Diese Datenmassen führen zu einer tiefgehenden Veränderung in der Wissenschaft.¹⁴ Zu den herkömmlichen Datenkorrelationen und daraus entwickelten Hypothesen kommen Machine Learning Algorithmen; durch das dadurch mögliche In-silico-Experiment mit Computersimulation und der Verbindung mit dem herkömmlichen In-vitro-Experiment gibt es die Möglichkeit, zu neuem Wissen zu kommen. Beispielsweise lässt sich anhand der Genomsequenz eines patientenspezifischen HIV-Erregers die Resistenzwahrscheinlichkeit für bestimmte Wirkstoffe berechnen und daraus das spezifische Profil eines Patienten berechnen. Zusammen mit dem „Internet der Dinge“ können so in Großzentren der Medizin Daten der Patienten und der Ärzte mit der technischen Infrastruktur zusammenwachsen.

3.4.6 Zusammenfassung

Algorithmen sind auf dem Hintergrund von Theorien und Gesetzen zu interpretieren und auch kritisch zu beurteilen. Korrelationen und Datenmuster ersetzen keine Erklärungen und Begründungen von Ursachen. Das heißt, diese Daten müssen nicht nur qualitativ evaluiert, sondern auch normativ beurteilt werden. Die Algorithmen

¹⁴Vergl. Schmidt in diesem Band, Kap. 4.

müssen auf ihre Zwecke und Ziele hin überprüft werden. Diese Überprüfung kann nur von einem Wesen vorgenommen werden, das in der Lage ist, Handlungen als Zweckrealisierungsversuche auszuführen, zu verstehen und zu überprüfen. In diesem Sinne bleibt auch durch die KI-induzierten Innovationen das Bacon-Projekt in seinen Grundstrukturen erhalten.

3.5 Deterministische und probabilistische Voraussagen

Kausalerklärungen haben den großen pragmatischen Vorteil, dass sie sichere Voraussagen (Prognosen) ermöglichen. Viele Wissenschaftsphilosophen sehen in der Prognosefähigkeit – und nicht in einer repräsentativen Abbildung der „realen“ Wirklichkeit – sogar die zentrale operative Bedeutung von interventionistischen „Theorien“ (van Fraassen 1980). Korrelationsanalysen auf der Basis großer Datenmengen gelangen demgegenüber grundsätzlich nicht zu derartigen deterministischen Prognosen. Ihre Zukunftsantizipationen, die unter dem Begriff der *predictive analytics* entwickelt werden, hängen von probabilistischen Präsuppositionen ab.

3.5.1 Erklären und Voraussagen

Deterministische Voraussagen in der Physik unterstellen ein deterministisch organisiertes Referenzobjekt, wie es beispielsweise für das Planetensystem angenommen wird. Das Referenzobjekt von Big Data ist jedoch grundsätzlich probabilistisch organisiert. Der Ersetzung von Kausalerklärungen durch Korrelationsanalysen entspricht – aufgrund der Struktur-Isomorphie von Erklärungen und Voraussagen – die Ersetzung von unkonditionierten Voraussagen durch konditionierte Voraussagen, wie sie bei Wettervorhersagen verwendet werden. Die Voraussage der nächsten Sonnenfinsternis ist danach strukturell etwas völlig anderes als die Voraussage des nächsten Regenschauers. Wettervorhersagen beruhen nicht auf Kausalerklärungen, sondern auf mehr oder weniger großen in der Vergangenheit gewonnenen Datenmengen, die ein zukünftiges Ereignis mehr oder weniger nahelegen. Durch Mustererkennung bezogen auf Big Data und neue Techniken der Datenanalyse wächst die Sicherheit der Vorhersage, nähert sich in günstigen Fällen der Sicherheit der deterministischen Voraussage an, fällt aber grundsätzlich nicht mit ihr zusammen. Wenn es um lapidare pragmatische Probleme geht, im Beispiel: die Frage, ob man den Regenschirm mitnimmt, mag es dann zwischen diesen Fällen keinen pragmatischen Unterschied geben. In anderen pragmatischen Kontexten kommt es aber gerade darauf an, ob man von einer Restunsicherheit ausgehen muß oder nicht. Ob man einen Regenschirm unnötigerweise mitgenommen hat, mag in den meisten Fällen belanglos sein. In medizinischen Kontexten, in denen es nicht selten um Leben und Tod geht, versucht man daher, Restunsicherheiten falsch-positiver oder falsch-negativer Voraussagen möglichst einzugrenzen.

3.5.2 *Fehlschlüsse: Stratifizierung*

Wird der Unterschied zwischen die deterministischen Voraussagen und probabilistisch konditionierten Voraussagen nicht beachtet, stellen sich typische Fehlschlüsse ein. Beispielsweise werden durch auf der Grundlage von Big Data erstellten Korrelationen – durch Einteilung einer statistischen Stichprobe in kleinere Gruppen – sog. „Strata“ – zugeordnet (etwa regional definierten Kohorten wie die Anwohner einer Straße, Alterskohorten, Kohorten, die durch diagnostische Merkmale definiert sind), die ohne jede soziale Erlebnisqualität sein können. Die Bildung solcher Kohorten und die auf ihrer Basis durch Algorithmen produzierten Voraussagen können für den Verwender durchaus nützlich und ihre durch die Rest-Unsicherheit erzeugte mögliche Fehlermarge mit Bezug auf seine Zwecke erträglich sein. Für das Individuum mag der Effekt einer solchen statistischen Stratifizierung (etwa durch Verweigerung eines Kredits) ein erheblicher Nachteil sein, für den Verwender der Daten, (beispielsweise eine Bank) in der Summe aufgrund der summierten Risikovermeidung ein Vorteil.

Besonders prekär ist diese strukturelle Situation allerdings im Gesundheitsbereich, wenn ein Individuum aufgrund diagnostizierter Merkmale einem „Stratum“ zugeordnet und aufgrund von Algorithmen eine therapeutische Strategie festgelegt wird. Man kann von einem statistischen Kollektivismus sprechen, gegenüber dem das Individuum geltend machen können muß, als solches betrachtet und behandelt zu werden. Das bedeutet, dass im Rahmen der Medizin als praktischer Wissenschaft eine korrelative Zuordnung eines Individuums zwar ein gutes diagnostisches Indiz sein kann, das wiederum eine gute praktische Heuristik rechtfertigt, aber grundsätzlich nicht die einzige Evidenzgrundlage für eine Behandlung sein darf. Epidemiologische Studien grenzen entsprechend denkbare prioritäre Diagnosen ein und legen auf dieser Grundlage gewisse prioritäre Therapien nahe, dürfen jedoch nie allein die Grundlage für eine Behandlung sein. Der individuelle Patient kann immer ein statistischer Ausreißer sein, mag die Wahrscheinlichkeit auch noch so klein sein. Daher ist die Gefahr des „epidemiologischen Fehlschlusses“ immer zu beachten. Eine genetische Disposition beispielsweise kann durchaus zu unterschiedlichen Wahrscheinlichkeitsverteilungen bei großen Kollektiven führen, aber diese determinieren nicht das Individuum. Sollte es eine Wahrscheinlichkeitskorrelationen zwischen der Disposition „Musikalität“ und der Inzidenz von Magenkarzinomen geben, besagt das für das vor dem Arzt stehende Individuum so gut wie nichts. Big-Data Vorhersagen tragen entsprechend wesentlich einen Rest an Ungewissheit in sich, die nicht auf Unzulänglichkeit der eingesetzten Daten zurückgeführt werden kann. Auch auf der Basis von Gesetzeswissen gibt es Ungewissheiten; diese sind jedoch jedenfalls prinzipiell zu überwinden. Ein Akteur, der sich Big-Data-Vorhersagen zunutze machen will, muss daher entscheiden, welche Kosten (im monetären oder nicht-monetären Sinn) er zu übernehmen bereit ist.

3.5.3 *Beispiel: evidenzbasierte Medizin*

Diese pragmatischen Probleme bezüglich aus Big Data gewonnener Voraussagen werden noch einmal um eine Dimension verschärft, wenn das Referenzobjekt, auf das sich Wahrscheinlichkeitsaussagen beziehen, durch diese Aussagen selbst beeinflussbar ist, wie das für alle Kontexte menschlichen Handelns, z. B. die Volkswirtschaft oder das Gesundheitssystem gilt.

Die aus epidemiologischen Studien gewonnenen konditionierten Voraussagen können bei passender öffentlicher Verbreitung kollektive Verhaltensweisen erheblich verändern, derart, dass die Datengrundlage nicht mehr verlässlich ist. Von weitreichender praktischer Bedeutung ist die Heranziehung großer Datenmengen für das Konzept der evidenzbasierten Medizin.¹⁵ Diese stützt sich vorzugsweise auf Erfahrungen zur Indikation, Wirksamkeit und Sicherheit von Therapieverfahren, die in Form von Expertenbewertungen oder systematischen Reviews („Studien“) zur Verfügung gestellt werden. Evidenzbasierte Medizin in diesem Sinne ist ein mittels nachvollziehbarer Kriterien standardisierter Prozeß der methodisch-rationalen Informationsreduktion und der Bewertung des jeweiligen, in Studien dokumentierten Standes der Wissenschaft. Die Anwendung des gesicherten und aktuellen Wissens setzt Einvernehmen darüber voraus, was als aktuelles und gesichertes Wissen zu verstehen ist. Die traditionell übliche Formel ‚Stand des Wissens‘ ist offenbar als unzureichend anzusehen. Die evidenzbasierte Medizin präzisiert deshalb diesen Stand des Wissens durch die Forderung, dass dieser durch die wissenschaftliche Literatur dargestellt wird, wobei nur diejenigen Publikationen als wissenschaftlich gewertet werden, die definierte wissenschaftliche Kriterien der Korrelationsanalyse erfüllen. In diesem Zusammenhang ist auch die Forderung nach direkt-vergleichenden randomisierten Studien als Mittel zur Verzerrungsvermeidung bei korrelativen Analysen zu verstehen. Aus diesen Studien ergeben sich die Kriterien der Indikation für die Intervention. Durch sie wird der Anwendungsbereich für therapeutische Interventionen bestimmt, wie er beispielsweise in der amtlichen Zulassung von Arzneimitteln und Medizinprodukten festgelegt wird. Wesentliche Dimensionen für die Beurteilung der Wirkungen von Interventionen sind Eintrittswahrscheinlichkeit, Erheblichkeit und Behandelbarkeit, wobei die Eintrittswahrscheinlichkeit in diesem Zusammenhang von besonderem Interesse ist. Die Eintrittswahrscheinlichkeit betrifft alle Arten von gesundheitsbezogenen Ereignissen wie Symptome, Testergebnisse, Behandlungserfolge, Krankheitsresiduen, Behinderungen (Funktions- und Leistungseinschränkungen). Grundsätzlich unterstellt also die evidenzbasierte Medizin ein probabilistisches Kausalitätsverständnis, und sie stellt somit probabilistisch konditionierte Voraussagen zur Verfügung. Dieses bietet für das ärztliche Handeln zweifellos eine sicherere Grundlage als bloß narrativ präsentierte okkasionelle Erfahrung. Dennoch ist der Ansatz der evidenzbasierten Medizin pragmatisch schwächer als eine deterministische Kausalanalyse, sofern man unterstellt, dass ein Kausalgesetz epistemisch mehr bietet als jede Korrelationsbe-

¹⁵Vergl. Humm et al. in diesem Band, Abschn. 2.3.3.

ziehung. Die Gefahr von Stratifizierung und epidemiologischen Fehlschlüssen ist grundsätzlich gegeben. Diese Gefahren lassen es als systematisch unmöglich erscheinen, Diagnose und Therapie generell durch Big Data-gestützte predictive analysis zu ersetzen.

3.5.4 Zusammenfassung

Was für die Diagnose und Therapie durch den den Einzelfall würdigenden Arzt gilt, gilt analog für das Judiz des Richters, das pädagogische Ermessen des Lehrers und die kreative Einschätzung eines wissenschaftlichen Problems durch den Forscher/die Forscherin. In allen Fällen ist Big Data unter günstigen Bedingungen ein Hilfsmittel, aber kein geeignetes Instrument der definitiven Lagebeurteilung und Entscheidung. Pragmatische und heuristische Faktoren wie Prüfung der Kohärenz mit anderen Evidenzquellen, Erfolgseinschätzungen u. a. spielen eine nicht zu vernachlässigende Rolle.

3.6 Methodische Verfahren und heuristische Urteilstkraft

Bis in das 20. Jahrhundert folgte die methodologische Grundvorstellung des Bacon-Projekts dem Verifikationismus, d. h. dem Versuch, wissenschaftliche Hypothesen durch Verfahren des „Wahrmachens“ auszuweisen. Allerdings war spätestens seit Hume („Humesches Problem“) geläufig, dass endlich viele Bestätigungsinstanzen, beispielsweise Experimente, aus logischen Gründen nicht einen generalisierenden Satz vom Typ „Alles Kupfer leitet Elektrizität.“ wahr machen können. Aus dieser methodologischen Verlegenheit hat Karl Raimund Popper (1934) durch den Vorschlag geführt, den Verifikationismus durch den Falsifikationismus zu ersetzen. Danach soll man nicht versuchen, Hypothesen durch Erfahrung zu beweisen, sondern sie versuchsweise durch Erfahrung zu widerlegen. Wenn eine Hypothese dem Widerlegungsversuch widersteht, dann hat sie sich vorläufig bewährt. Aufgrund der Anzahl der (unabhängigen) Widerlegungsversuche läßt sich der Grad der Bewährung einer Hypothese angeben. Poppers Falsifikationismus weist eine starke Affinität mit dem von C.S. Peirce eingeführten Fallibilismus auf; danach geht es bei den menschlichen Erkenntnisbemühungen nicht darum, „Gewißheit“ zu erlangen („Certismus“), sondern Irrtümer zu vermeiden, soweit das einem endlichen Wesen überhaupt möglich ist (Spinner 1974). Der Grundgedanke des Falsifikationismus hat sich in den empirischen (sich auf Experimente, Befunde oder Befragungen stützenden) Wissenschaften weitgehend faktisch etabliert. Man kann von einem falsifikationistischen Ethos im Sinne eines meistens nicht explizit in Frage gestellten common sense der Wissenschaftler/Wissenschaftlerinnen sprechen.

3.6.1 *Falsifizierbarkeit und Falsifikation*

Der Falsifikationismus bietet eine Antwort-Strategie für zwei unterschiedliche wissenschaftsphilosophische Probleme an. Einmal geht es um die Abgrenzung von wissenschaftlich sinnvollen gegenüber wissenschaftlich sinnlosen Aussagen. Zu den sinnlosen Aussagen gehören nicht die Aussagen, die aufgrund einfacher syntaktischer („Aua!“) oder semantischer („Alle Bulgaren sind Primzahlen.“) Defizite nicht falsifizierbar sind, sondern vor allem solche, die wegen grundsätzlich empirisch nicht überprüfbarer Unterstellungen der Falsifikation entzogen sind. Dies gilt insbesondere für in Anspruch genommene wissenschaftliche „Theorien“, die inhaltliche Falsifikationsverfahren durch eine Immunisierungsstrategie verbieten.¹⁶ Falsifizierbarkeit ist also zum einen ein Abgrenzungskriterium. Darüber hinaus bildet das Verfahren der Falsifikation ein Wahrheitskriterium im Sinne eines Bewährungskriteriums. Um eine Theorie gleich welcher epistemischen Herkunft (von der bloßen Fantasie bis zu Annahmen, die sich aus anderen Theorien nahelegen) zu überprüfen, muss folgende Schrittfolge vollzogen werden:

- (i) Aus Theorie-Kandidaten werden mit Hilfe der Logik Hypothesen deduziert.
- (ii) Aus diesen Hypothesen werden Prognosen deduziert; diese Prognosen können empirisch (durch Experiment, ...) widerlegt werden. Gelingt diese Widerlegung nicht, gilt die Theorie in Abhängigkeit von Häufigkeit und Härte des Testverfahrens als mehr oder weniger gut bestätigt. Dabei wird die Reproduzierbarkeit unterstellt, denn ein empirischer Test, der nur ein einziges Mal durchgeführt werden kann, ist kein Verfahren. Falsifizierbarkeit setzt Reproduzierbarkeit voraus. Bei der Falsifikation geht es nicht um eine generelle Erkenntnissicherung (wie dies im Rahmen des Fallibilismus unterstellt wird). Primäre Erfahrungen, wie sie Menschen durch ihre fünf Sinne im Alltag „widerfahren“ (nicht: „machen“), sind ausdrücklich nicht angesprochen. Vielmehr geht es um sekundäre Erfahrungen, die wesentlich auf Verallgemeinerung im Sinne von Hypothesen, Modellen, Theorien beruhen. Bereits der Einsatz von Messgeräten im Rahmen von Experimenten, die der Widerlegung von Prognosen dienen sollen, beruht auf Verallgemeinerungen und führt immer zu sekundärer Erfahrung. Vom einfachen Thermometer bis zum komplexen Geiger-Zähler sind empirische Testverfahren von vornherein theorie-imprägniert. Modelle. Aus ihnen lassen sich mathematisch formulierte Szenarien bilden, die viele Einzelbeobachtungen aus ganz unterschiedlichen Erkenntnisquellen zusammenfassen, um zu Verallgemeinerungen zu kommen. Das Verfahren der Falsifikation muß daher bei einer sehr schnell erreichten Stufe von Komplexität in zunehmendem Maße Vereinfachungen und Idealisierungen im Rahmen von Modellierungen vornehmen. Quantität und Qualität von Modellierungen werden nicht durch das

¹⁶Popper nennt insbesondere die Marxistische Gesellschaftstheorie (Wer diese in Frage stellt, gilt als Klassenfeind.), die Freudsche Psychoanalyse (Wer diese in Frage stellt, gilt als besonders schwer erkrankt) und die Christliche Metaphysik (wer diese in Frage stellt, gilt als Ungläubiger.) (Popper 1962). Zu Immunisierungsstrategien vgl. Albert (1968).

Referenzobjekt der Theorien festgelegt, sondern hängen wesentlich vom Zweck ab, der mit dem Geltungsanspruch der Theorie verbunden ist.

3.6.2 Grenzen und Schwierigkeiten des Verfahrens der Falsifikation

Ein großes Feld der wissenschaftsphilosophischen Diskussion des 20. Jahrhunderts besteht in der Auseinandersetzung mit Grenzen und Schwierigkeiten des Verfahrens der Falsifikation und Versuchen ihrer Überwindung. Eine bedeutende Grenze liegt darin, dass es sich ausschließlich auf empirisches Wissen bezieht, wie es beispielsweise durch Experimente bestätigt wird. Die für diese Form von Wissen konstitutiven Formalwissenschaften (vor allem Logik und Mathematik) fallen dabei völlig heraus. Zwar gibt es für diese Disziplinen kein Humesches Problem. Abgesehen davon, dass bei weitem nicht für alle interessanten Kalküle schon Widerspruchsfreiheitsbeweise geführt worden sind, gibt es bezüglich des Einsatzes formaler Instrumente aber durchaus ein Wahlproblem. Ob beispielsweise für die Modellierung von Wolkenbildung die Monte Carlo Methode der Wahrscheinlichkeitstheorie angemessen ist oder nicht, ist eine Frage der pragmatischen Adäquatheit, für deren Entscheidung es keine durch eine Methodologie vorgegebene Prozedur gibt. Ferner kann das Verfahren der Falsifikation nicht auf solche Wissensformen bezogen werden, die konstitutiv dafür sind, dass das Verfahren überhaupt sinnvoll ist. Es geht um grundlegende methodologische Postulate wie die der Konsistenz. Bezüglich dieser Wissensformen gibt es keine empirische Kontrolle, vielmehr bedarf es begrifflicher Überlegungen bezüglich der Angemessenheit und Reichweite dieser Postulate. Dabei spielt auch der Anwendungserfolg dieser Postulate eine Rolle, wobei die Festlegung der Maßstäbe des Erfolgs durchaus problematisch ist. Aber auch im Raum des empirischen Wissens gibt es erhebliche Stolpersteine. In den Naturwissenschaften gibt es keineswegs nur generelle Aussagen vom Typ der Gesetzesaussagen. Es gibt durchaus singuläre Phänomene und entsprechende Existenzbehauptungen (vom Typ „Es gibt einen 80. Mond des Jupiter.“ „Dies ist das Schlüsselbein des Australopithecus.“), die nicht nach dem Standardverfahren der Falsifikation auf ihre Bewährtheit hin überprüft werden können. Auch einmalige Groß-Phänomene („Das Universum nach dem Urknall.“, „Das Weltklima.“) haben die ontologische Eigenschaft, dass es gewissermaßen keinen Außenraum gibt, von dem aus falsifizierende Phänomene vorgebracht werden können. Erhebliche methodologische Komplikationen werfen Hypothesen über ferne Zeiten und Räume auf. Da die einfache Falsifikation durch Abwarten („Das Klima im Jahr 3000 ...“) oder durch Fernreisen („Der Exoplanet in 1000 Lichtjahren Entfernung ...“) als Teststrategie entfällt, müssen indirekte Ersatzstrategien herangezogen werden, bei denen Hypothesen eingesetzt werden müssen, die ihrerseits einem Falsifikationsverfahren zu unterwerfen wären. In anderer Weise sind komplizierte indirekte Testverfahren einzusetzen, wenn es sich um Wahrscheinlichkeitsaussagen handelt, die sich auf ein empirisches Referen-

zobjekt beziehen. Wahrscheinlichkeitsaussagen enthalten gewissermaßen die Falsifikationsinstanzen schon in sich. Während der Satz „Alle Schwäne sind weiß.“ leicht zu falsifizieren wäre, wirft der Satz „Wahrscheinlich sind alle Schwäne weiß“ erhebliche Probleme auf.

3.6.3 *Unterbestimmtheit wissenschaftlicher Theorien*

Sowohl verifikationistische als auch falsifikationistische Methodologien unterstellen zunächst, daß zwischen Instanzen der Bestätigung und der mehr oder weniger bestätigten Theorie eine Relation besteht der Art, daß die Theorie um so mehr bestätigt ist, je mehr Instanzen für sie bzw. je weniger Instanzen gegen sie sprechen. Diese einfache Bestätigungsrelation ist schon dadurch in Frage gestellt, dass die Qualifizierung der Bestätigung einer Theorie auch von der absoluten Zahl der Verifikations- bzw. Falsifikationsverfahren abhängt, mehr aber noch von der Nicht-Trivialität der Bestätigungsinstanzen, beispielsweise der Experimente. Beide Gesichtspunkte stehen der intuitiv naheliegenden Vermutung im Wege, den Bestätigungsgrad einer Theorie einfach durch die Zahl der Verifikations- bzw. Falsifikationsinstanzen bestimmt sein zu lassen. Noch gravierender ist der Umstand, dass die logische Beziehung zwischen Bestätigungsinstanzen und Theorien in beiden Richtungen uneindeutig ist. Ein Experiment kann viele Theorien bestätigen und eine Theorie kann von unterschiedlichen Bestätigungsinstanzen bestätigt werden. Die Uneindeutigkeit wird noch einmal durch die zuerst von C.G. Hempel (1946) formulierte Paradoxie der Bestätigung (Rabenparadoxie) verschärft. Demgemäß bestätigt die Beobachtung eines weißen Schwans mit der Standardlogik¹⁷ den Satz „Alle Raben sind schwarz.“ bestätigt.

Eine Folgerung aus der Hempelschen Paradoxie ist das Unbestimmtstheorem (Duhem-Quine These). Danach kann eine Theorie durch sehr verschiedene (sogar widersprüchliche oder disparate) Mengen von Evidenzen bestätigt werden. Vice versa gilt, dass eine konsistente Menge von Evidenzen viele Theorien bestätigen kann; man erhält damit ein Nachfolgeproblem des Humeschen Induktionsproblems, das man als Problem des semantischen Überschusses einer Theorie bezeichnen kann: Endlich viele Evidenzen reichen nicht aus, um einen generellen Satz (z. B. ein Naturgesetz) zu bestätigen. Anders formuliert: Eine Theorie weist immer einen (nicht bestätigten) Überschuß über das empirische Wissen aus. Aus beiden Problemen folgt zusammenfassend das Mehrdeutigkeitsproblem: Das Verhältnis der Mengen primärer Erfahrung zu wissenschaftlichen Theorien ist mehr-mehr-deutig. Primäre Erfahrung (z. B. durch ein Experiment erzeugt) bestätigt viele Theorien; eine Theorie wird durch verschiedene (sogar widersprüchliche) Erfahrungen bestätigt. Es bedarf also materieller Kriterien, die beispielsweise – mit Blick auf die Raben-Paradoxie – Schwäne grundsätzlich nicht als Bestätigungsinstanzen zulassen. Da-

¹⁷ Wegen $\Lambda x (Fx \rightarrow Gx) \leftrightarrow \Lambda x (\neg Gx \rightarrow \neg Fx)$ [kl. Kontraposition] gilt Best $[(\neg Fa \wedge \neg Ga), \Lambda x (Fx \rightarrow Gx)]$. Vgl. zum Logikwahlproblem in diesem Zusammenhang Gethmann (1980).

mit entsteht jedoch das gegenläufige Problem, dass man jede Theorie durch die Auswahl von Bestätigungsinstanzen verifizieren bzw. falsifizieren kann, indem man die falsifizierenden Instanzen als inhaltlich nicht einschlägig ausschließt oder als einschlägig einschließt. Die Unterbestimmtheit läßt sich auch schon am einfachen Verfahrensschema der Falsifikation erkennen: auch wenn die Auswahl der möglicherweise falsifizierenden Instanzen aus den Hypothesen logisch deduziert wird, wird dadurch keine inhaltliche Auswahl dieser Instanzen vorgegeben. Erst recht besteht keine eindeutige logische Beziehung zwischen der angeblich falsifizierenden Instanz und der Hypothese, die falsifiziert wird. In dieser Richtung handelt es sich um ein Schlußverfahren vom Typ der „Abduktion“. Der Wissenschaftler, der für die zur Debatte stehende Hypothese das Verfahren der Falsifikation durchlaufen will, muß also im Vorhinein unter Inkaufnahme möglicher Zirkel im Interesse der epistemischen Kontrolle pragmatische Kriterien heranziehen. Zu diesen gehören: wechselseitige Kohärenz mit anderen Theorien, Analogien zu anderen Wissensbereichen, der Blick auf andere Meinungen der Wissenschaftlergemeinschaft und andere Verfahren kluger Abwägung. Auf die Rolle der Heuristik im Sinne regelüberwachter Suchverfahren wurde bereits hingewiesen (siehe Abschn. 3.3.5 dieses Kapitels).

3.6.4 Theoriendynamik und die epistemische Funktion von Wissenschaftlergemeinschaften

Die dargestellten Probleme, das Verfahren der Falsifikation eindeutig zu rekonstruieren, haben seit der Mitte des 20. Jahrhunderts zu erheblichen Zweifeln an der Vorstellung geführt, der faktische Forschungsprozess entwickle sich entlang der Befolgung eines normativen Sets von methodischen Regeln. Statt dessen wurde der Blick auf historisch-kontingente Prozesse der Wissensbildung in den Wissenschaften gelenkt, insbesondere die sozialen Interaktions- und Kommunikationsverfahren innerhalb der Wissenschaftlergemeinschaften in Wechselwirkung mit wissenschaftsexternen Einflussfaktoren. Besonders einflussreich war dabei zunächst die Konzeption der diachronen Wissenschaftsphilosophie, die der amerikanischen Wissenschaftshistoriker und -philosoph Th. S. Kuhn in mehreren Anläufen entwickelt hat (Kuhn 1970, 2000). In Auseinandersetzung mit der Vorstellung, dass die Wissenschaftsentwicklung kumulativ-kontinuierlich, nach wissenschaftsinternen (methodologischen) Rationalitätskriterien verlaufe, beschreibt Kuhn einen diskontinuierlichen, auch von wissenschaftsexternen (sozialen) Faktoren bestimmten Wandel in der Wissenschaftsgeschichte. Der zentrale Begriff für die Erklärung des Phänomens des diskontinuierlichen Wandels ist der des Paradigmas, womit der Sachverhalt bezeichnet werden soll, dass einige anerkannte Vorbilder konkreter wissenschaftlicher Praxis die Muster liefern, gemäß denen die Wissenschaftlergemeinschaften ihre methodischen und sozialen Entscheidungen fällen. Was in einer Disziplin ‚rational‘ heißt, hängt danach jeweils von einem solchen Paradigma ab. Kuhn unter-

scheidet näherhin drei Phasen der Wissenschaftsentwicklung: (i) In der ‚vor-normalen (vor-paradigmatischen) Phase‘ hat sich noch kein verbindliches Paradigma durchgesetzt, vielmehr ist ein Pluralismus von konkurrierenden Ansätzen festzustellen; die so genannten Tatsachen werden im Rahmen der verschiedenen Ansätze unterschiedlich gewichtet. Es gibt keine einheitlichen Methodologien und keinen festumrissenen Wissenskanon. Demgegenüber ist (ii) in der ‚normalen Phase‘ ein wissenschaftliches Paradigma für die Definition einer Disziplin nach innen und außen leitend. Es besteht eine einheitliche Methodologie und eine allgemein von den Mitgliedern der Wissenschaftlergemeinschaft anerkannte Kommunikationsstruktur (z. B. wissenschaftliche Zeitschriften); das bereits errungene Wissen ist kanonisiert, z. B. in anerkannten Lehrbüchern niedergelegt. (iii) Die ‚revolutionäre Phase‘ besteht in einem sich mehr oder weniger schnell vollziehenden Paradigmenwechsel.

Für den Übergang von einer Phase zur nächsten gibt Kuhn soziale Indikatoren an, die diesen Prozess häufig begleiten. Zum Beispiel, ist der Übergang von der vor-normalen (vor-paradigmatischen) zur normalen Phase (i) an der Auflösung der Schulverbände, dem Ausschluss der das Paradigma nicht akzeptierenden Wissenschaftler/Wissenschaftlerinnen erkennbar (diese werden etwa als ‚Philosophen‘ aus der Disziplin fortgelobt), (ii) an der Entwicklung von Fachzeitschriften und Fachvereinigungen, (iii) an der Aufnahme der Disziplin in die akademischen Lehrpläne, (iv) an der Änderung der Publikationsformen von der Monografie (die alles von Anfang an darstellt) zum Fachartikel (der den Stand der Forschung als bekannt und anerkannt unterstellt), wodurch die wissenschaftlichen Publikationen einen esoterischen Charakter bekommen (das ‚gebildete Publikum‘ wird von der Diskussion ausgeschlossen). Die Tätigkeit des normalen Wissenschaftlers gleicht nach Kuhn (im Gegensatz zum Ideal des kritischen Wissenschaftlers bei K. R. Popper) der eines routinierten Aufräumens und Verdeutlichens; der paradigmatische Bestand des Wissens wird nicht angetastet, Erfolg hängt von der Anerkennung des Paradigmas ab. So leistet der normale Wissenschaftler z. B. im empirischen Bereich Präzisionsarbeit bezüglich der vom Paradigma als aufschlussreich bestimmten Fakten, bezüglich der gemäß der Theorie prognostizierten Fakten, bezüglich der Größen (Konstanten, Gesetze) der paradigmatischen Theorie selbst; er weitet die Anwendungsbereiche der anerkannten Theorien aus und bemüht sich um mathematische Formulierung und Verbesserung bestehender Formulierungen der Theorie (puzzle solving).

Im Anschluss an Kuhns Kritik an Popper hat es in der Wissenschaftsphilosophie des 20. Jahrhunderts eine breite Diskussion über das Verhältnis methodologischer Regeln zu kontingenten sozialen Einflüssen mit zahlreichen Lösungsversuchen gegeben (Radnitzky und Andersson 1980). Vor allem Imre Lakatos hat einen einflussreichen Vermittlungsversuch zwischen Popper und Kuhn entwickelt (Lakatos 1970). Im Gegensatz zu Kuhns Diskontinuitätsthese ist der Prozess der Wissenschaftsentwicklung nach Lakatos durch eine Kontinuität sich ablösender Theorien gekennzeichnet. Das Kontinuum entwickelt sich aus einer Menge methodologischer Regeln (‚Forschungsprogramm‘), die teilweise Verbote (‚negative Heuristik‘), teilweise Gebote (‚positive Heuristik‘) sind. Forschungsprogramme sind durch ei-

nen methodologischen ‚harten Kern‘ zu charakterisieren, der durch Hilfhypothesen gegen Überprüfung geschützt werden muss.

Die Diskussion über die Rolle einer wissenschaftlichen Methodologie und die Regeln interner Wissenschaftskommunikation in Wechselwirkung mit externen Einflussfaktoren, wie sie durch gesellschaftliche Erwartungen gebildet werden, die sich wiederum zu Instrumenten der Forschungsförderung verdichten können, spielt offenkundig eine epistemische und nicht nur eine pragmatische Rolle bei der Rekonstruktion wissenschaftlicher Geltungsansprüche und ihrer Einlösung. Das Gegenstandsfeld der Wissenschaftsphilosophie konvergiert demzufolge offenkundig mit dem der Wissenschaftshistoriografie und der Wissenschaftssoziologie; dies macht Versuche plausibel, die Forschungsperspektiven dieser drei Disziplinen zu einem einheitlichen Konzept der Wissenschaftsforschung/Wissenschaftswissenschaft zu vereinen. Im Rahmen eines solchen Programms sind die Beziehungen zwischen sozialen und kognitiven Strukturen und Vorgängen zu problematisieren. Die schematischen Deutungen der Wissenschaftsentwicklung, wonach diese entweder durch die Geltung ahistorischer Standards nur als intern induzierter Prozeß oder als ein Reflex nur externer sozialer Faktoren verstanden wird, müssen nach P. Weingart (1972) durch eine „intervenierende Variable“ des organisatorischen Aufbaus der Wissenschaft, ihrer Strukturen, Regeln und prozessualen Mechanismen miteinander verknüpft werden. Durch diese findet einerseits eine Vermittlung von sozialen Bedingungen und Einflüssen in wissenschaftliche, d. h. kognitive, Prozesse statt, andererseits werden ihre Resultate über die spezifische Organisation der Wissenschaft in die Gesellschaft vermittelt und wirken auf diese. Der Bezug der sozialen und kognitiven Strukturen aufeinander im Rahmen einer einheitlichen Konzeption ergibt sich nur, wenn die Institution der Wissenschaft (als intervenierende Variable) auf ihre normativen Grundlagen hin untersucht und mit den normativen Ansprüchen, die Wissenschaftler/Wissenschaftlerinnen kognitiv als Gründe erfahren, verglichen wird (Gethmann 1981).

3.6.5 Zusammenfassung: Die Unverzichtbarkeit menschlicher Erkenntnissubjekte

Der Falsifikationismus ist eine interne methodologische Wende innerhalb des Bacon-Projekts. Schon Popper hatte zunächst versucht, das Falsifikationsverfahren durch eine kanonische Prozedur anzugeben, die es sogar erlauben sollte, ein numerisches Maß für den Grad der Bestätigung einer Hypothese zu ermitteln. Ein solches Verfahren ließe sich grundsätzlich durch einen Bestätigungs-Algorithmus auf der Basis von Big Data verfeinern. Die weitere wissenschaftsphilosophische Debatte über den Begriff der Bestätigung, insbesondere das Problem der Unbestimmtheit, hat jedoch gezeigt, dass Bestätigungsversuche von materiellen Hintergrundannahme, Risiko-Abwägungen und heuristischen Fähigkeiten abhängen, die sich nicht in ein kanonisches Verfahren überführen lassen, sondern von faktisch-kontingenten

sozialen Faktoren abhängen. Die Vorstellung, die Wahrheitsprüfung von wissenschaftlichen Hypothesen letztlich Computern zu übertragen, hat sich daher als unhaltbar erwiesen. Erfahrung, Einfallsreichtum, Fingerspitzengefühl, Übersicht und kommunikative Kompetenz sind konstitutive Fähigkeiten, die Falsifikationsverfahren anleiten.

3.7 Fazit

Die Vorstellung der vollständigen Substituierbarkeit menschlicher kognitiver Leistungen macht von anthropologischen Präsuppositionen Gebrauch, die einer kritischen Betrachtung nicht standhalten. Dabei ist nicht ausgeschlossen, dass kognitive Teilfunktionen wie beispielsweise Gedächtnisleistungen von technischen Artefakten nicht nur ersetzt, sondern sogar hinsichtlich geringerer Störanfälligkeit, höherer Leistungsfähigkeit, tieferer Vernetzung, u. a. übertroffen werden können. Technische Artefakte, die mit sog. künstlicher „Intelligenz“ ausgestattet sind, können jedoch grundsätzlich nicht als Akteure in wissenschaftlichen Begründungsdiskursen auftreten, weil ihnen die pragmatischen Merkmale der Handlungsurheberschaft und Zurechenbarkeit nicht zukommen. Hinsichtlich praktischer Aufgabenstellungen ist Big Data unter günstigen Bedingungen ein Hilfsmittel, aber kein geeignetes Instrument der definitiven epistemischen Leistungsbeurteilung und Entscheidung. Pragmatische und heuristische Faktoren wie Prüfung der Kohärenz mit anderen Evidenzquellen, Erfolgseinschätzungen u. a. spielen eine nicht zu vernachlässigende Rolle. Die wissenschaftsphilosophische Debatte über den Begriff der Bestätigung, insbesondere das Problem der Unbestimmtheit, hat gezeigt, dass Bestätigungsversuche von materiellen Hintergrundannahme, Risiko-Abwägungen und heuristischen Fähigkeiten abhängen, die sich nicht in ein kanonisches Verfahren überführen lassen, sondern von faktisch-kontingenten sozialen Faktoren abhängen. Erfahrung, Einfallsreichtum, Fingerspitzengefühl, Übersicht und kommunikative Kompetenz sind konstitutive Fähigkeiten, die Falsifikationsverfahren anleiten müssen, aber nicht durch sie ersetzt werden können.

Literatur

- Albert H (1968) Traktat über kritische Vernunft, 1. Aufl. J.C.B. Mohr, Tübingen
- Anderson Ch (2008) The end of theory: the data deluge makes the scientific method obsolete. WIRED. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory. Zugriffen am 11.12.2020
- Aristoteles (1936) In: Ross WD (Hrsg) Aristotle's physics. A revised text. Clarendon, Oxford
- Carrier M, Mittelstraß J (1989) Geist – Gehirn – Verhalten. de Gruyter, Berlin
- Chisholm RM (1964) Human freedom and the self. The Lindley lecture. University of Kansas Press, Lawrence

- Dreyfus HL (1979) What computers can't do. A critique of artificial reason. Cambridge University Press, Cambridge
- Dreyfus HL (1992) What computers still can't do: a critique of artificial reason. New Edition. MIT Press, Cambridge.
- Dummett M (1978) Truth. In: Dummett M (Hrsg) Truth and other enigmas. Harvard University Press, Cambridge, S 1–24
- Dummett M (1993) What is a theory of meaning (II). In: Dummett M (Hrsg) The seas of language. Clarendon Press, Oxford, S 34–93
- Falkenburg B (2012) Mythos Determinismus. Springer, Heidelberg
- Forum für Philosophie (1992) Ed. Realismus und Antirealismus. Suhrkamp, Frankfurt am Main
- Fuchs T (2013) Das Gehirn – ein Beziehungsorgan, 4E. Aufl. Kohlhammer, Stuttgart
- Gethmann CF (1979) Protologik. Untersuchungen zur formalen Pragmatik von Begründungsdiskursen. Suhrkamp, Frankfurt am Main
- Gethmann CF (1980) Die Logik der Wissenschaftstheorie. In: Gethmann CF (Hrsg) Theorie des wissenschaftlichen Argumentierens. Suhrkamp, Frankfurt am Main, S 15–42
- Gethmann CF (1981) Wissenschaftsforschung? Zur philosophischen Kritik der nachkuhnischen Reflexionswissenschaften. In: Janich P (Hrsg) Wissenschaftstheorie und Wissenschaftsforschung. Beck, München, S 9–38
- Gethmann CF (2003) Wissen als Macht. Wissenschaftsphilosophische Überlegungen. In: Emmermann R et al (Hrsg) An den Fronten der Forschung: Kosmos – Erde – Leben. Hirzel, Stuttgart, S 238–245
- Gethmann CF (2010) Menschsein – Menschbleiben. Zur Grammatik askriptiver Äußerungsmodi. In: Rüsen J (Hrsg) Perspektiven der Humanität. Menschsein im Diskurs der Disziplinen. Transcript, Bielefeld, S 41–58
- Gethmann CF (2014) Wahrheit und Beweisbarkeit. Die philosophische Bedeutung von Intuitionismus und Konstruktivismus in der Logik. In: Mittelstraß J, von Bülow C (Hrsg) Dialogische Logik. mentis, Münster, S 13–28
- Gethmann CF (2016) What remains of the fundamentum inconcussum in light of the modern sciences of humans? J Gen Philos Sci 47:1–20
- Gethmann CF, Sander T (2002) Anti-Mentalismus. In: Gutmann M et al (Hrsg) Kultur – Handlung – Wissenschaft. Velbrück, Weilerswist, S 91–108
- Hempel CG (1946) A note on paradoxes of confirmation. Mind 55:79–82
- Hengstenberg HE (1957) Philosophische Anthropologie. Kohlhammer, Stuttgart
- Hilgendorf E (2015) Recht und autonome Maschinen – ein Problemaufriß. In: Hilgendorf E, Hötitzsch S (Hrsg) Das Recht vor den Herausforderungen der modernen Technik (4). Nomos, Baden-Baden, S 11–40
- Hunt E, Lunneborg C, Lewis J (1975) What does it mean to be high verbal? Cogn Psychol 7:194–227
- Janich P (1997) Das Maß der Dinge. Protophysik von Raum, Zeit und Materia. Suhrkamp, Frankfurt am Main
- Janich P (2015) Handwerk und Mundwerk. Über das Herstellen von Wissen. Beck, München
- Kamlah W, Lorenzen P (1973) Logische Propädeutik. Vorschule des vernünftigen Redens. Bibliographisches Institut, Mannheim
- Kant I (1956) Kritik der reinen Vernunft. Meiner, Hamburg
- Kuhn TS (1970) The structure of scientific revolutions, 2. Aufl. University of Chicago Press, Chicago
- Kuhn TS (2000) The road since „structure“. In: Conant J, Haugeland J (Hrsg) Philosophical essays. University of Chicago Press, Chicago
- Lakatos I (1970) Falsification and the methodology of scientific research programs. In: Lakatos I, Musgrave A (Hrsg) Criticism and the growth of knowledge. Cambridge University Press, Cambridge, S 91–196
- Lenzen W (1980) Glauben, Wissen und Wahrscheinlichkeit. Systeme der epistemischen Logik. Springer, Wien
- Lorenz K (1976) Sprachtheorie als Teil einer Handlungstheorie. In: Wunderlich D (Hrsg) Wissenschaftstheorie der Linguistik. Athenäum, Kronberg, S 250–280

- Mainzer K (1995) Computer – Neue Flügel des Geistes? de Gruyter, Berlin
- Mead GH (1934) In: Morris CW (Hrsg) *Mind, self and society from the standpoint of a social behaviorist*. Chicago University Press, Chicago
- Nida-Rümelin J, Weidenfeld N (2018) *Digitaler Humanismus: Eine Ethik für das Zeitalter der Künstlichen Intelligenz*. Piper, München
- Plessner H (1928) *Die Stufen des Organischen und der Mensch*. de Gruyter, Berlin. (Wiederabdr. *Gesammelte Schriften IV* (2003). Suhrkamp, Frankfurt am Main)
- Plessner H (1941) *Lachen und Weinen. Eine Untersuchung nach den Grenzen menschlichen Verhaltens*. Van Loghum Slaterus, Arnheim. (Wiederabdr. *Gesammelte Schriften VII* (2003). Suhrkamp, Frankfurt am Main)
- Popper KR (1934) *Logik der Forschung*. Julius Springer, Wien
- Popper KR (1962) *The open society and its enemies*. Routledge, London
- Radnitzky G, Andersson G (Hrsg) (1980) *Fortschritt und Rationalität der Wissenschaft*. J.C.B. Mohr (Paul Siebeck), Tübingen
- Schäfer L (1993) *Das Bacon-Projekt. Von der Erkenntnis, Nutzung und Schonung der Natur*. Suhrkamp, Frankfurt am Main
- Scheler M (1916) *Der Formalismus in der Ethik und die materiale Wertethik*. Niemeyer, Halle an der Saale
- Scheler M (1923) *Wesen und Formen der Sympathie*, 2. Aufl. F. Cohen, Bonn
- Scherer G (1976) *Strukturen des Menschen*. Ludgerus, Essen
- Searle J (1984) *Minds, brains and science*. Cambridge University Press, Cambridge
- Siegwart G (1997) *Vorfragen zur Wahrheit. Ein Traktat über kognitive Sprachen*. R. Oldenbourg, München
- Spinner HF (1974) *Pluralismus als Erkenntnismodell*. Suhrkamp, Frankfurt
- Tennant N (1997) *The taming of the true*. Clarendon, Oxford
- Turing AM (1950) *Computing machinery and intelligence*. *Mind* NS 59:433–460
- Van Fraassen BC (1980) *The scientific image*. Clarendon, Oxford
- Von Wright GH (1971) *Erklären und Verstehen*. Fischer-Athenäum, Frankfurt am Main
- Weingart P (1972) *Einführung. Wissenschaftsforschung und wissenschaftssoziologische Analyse*. In: Weingart P (Hrsg) *Wissenschaftssoziologie I*. Fischer Athenäum, Frankfurt am Main, S 11–42
- Wright C (1992) *Truth and Objectivity*. Harvard University Press, Cambridge

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.



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 Wissenskultur 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 Wissenskulturskultur.

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

²²Nun 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.

²³Dieses Argument wird in Abschn. 4.4 ausgeführt.

²⁴Allerdings 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 Storchenpopulation 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 Storchenpopulation, 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 machenenden 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. Zusammengefasst 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). Speziell 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 Quellecode 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 zuallerst 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 Veränderungsthese, 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-, Adaption- 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, raumzeitlich 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.



Auswirkungen von Digitalisierung und KI auf die wissenschaftliche Arbeit



Verena Nitsch und Peter Buxmann

Seit Jahrzehnten untersucht die arbeitswissenschaftliche Forschung die Auswirkungen des Digitalisierungstrends auf Beschäftigte und Arbeitsprozesse (siehe auch Mütze-Niewöhner und Nitsch 2020). Empirische Untersuchungen existieren bereits zu zahlreichen Funktionsbereichen (z. B. Produktion und Service) und Berufsgruppen (z. B. Elektro- und Metallberufe). Bislang liegen jedoch kaum empirische Befunde zu den Effekten der Digitalisierung auf die Arbeit von Forschenden vor. Entsprechend schwierig gestaltet sich die Aufgabe, mögliche gegenwärtige und zukünftige Auswirkungen des Digitalisierungstrends und dem Einsatz von KI auf die wissenschaftliche Arbeit und damit Beschäftigte abzuschätzen. Jedoch existieren zahlreiche Befunde zur thematisch verwandten Wissensarbeit. Es lässt sich daher davon ausgehen, dass einige dieser Befunde auf die wissenschaftliche Arbeit übertragbar sind.

Das folgende Kapitel widmet sich daher zunächst der Charakterisierung von wissenschaftlicher Arbeit als spezielle Form der Wissensarbeit. Vor diesem Hintergrund wird anschließend die Evidenzlage zu Auswirkungen der Digitalisierung auf Wissensarbeitende skizziert und eine Abschätzung der Übertragbarkeit der Befunde auf den wissenschaftlichen Beruf vorgenommen. Es wird ferner postuliert, dass der

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.

V. Nitsch (✉)

Institut für Arbeitswissenschaft, RWTH Aachen, Aachen, Deutschland

E-Mail: v.nitsch@iaw.rwth-aachen.de

P. Buxmann

Fachgebiet Wirtschaftsinformatik, Technische Universität Darmstadt, Darmstadt, Deutschland

E-Mail: buxmann@is.tu-darmstadt.de

Einsatz von KI-Technologien in erster Linie eine zunehmende Automatisierung von informatorischen/kognitiven Routinetätigkeiten in der wissenschaftlichen Arbeit bewirken wird. Dementsprechend werden einige Effekte der Automatisierung solcher Tätigkeiten vorgestellt und mögliche Auswirkungen auf den Arbeitsmarkt und Arbeitstätigkeiten diskutiert. Um ein genaueres Bild zu Digitalisierungseffekten, dem Einsatz von KI-Methoden und der Einstellung von wissenschaftlich Beschäftigten zu KI-Methoden zeichnen zu können, wurde eine Befragung durchgeführt, deren wesentliche Ergebnisse in diesem Kapitel vorgestellt werden. Im weiteren Verlauf des Kapitels werden die Auswirkungen von KI aus einer ökonomischen Perspektive beleuchtet. Das Kapitel schließt ab mit einem kurzen Resümee und einigen Empfehlungen für die Gestaltung von wissenschaftlicher Arbeit im Zeitalter der Digitalisierung und KI.

5.1 Wissenschaftliche Arbeit als Teilbereich der Wissensarbeit

Um die Auswirkungen der Digitalisierung und des Einsatzes von KI-Methoden auf die wissenschaftliche Arbeit abschätzen zu können, bedarf es zunächst eines grundlegenden Verständnisses dieser Arbeit. Obgleich in der Fachdisziplin der Arbeitswissenschaft und verwandten Fachbereichen wie der Arbeits- und Organisationspsychologie sowie der Arbeitssoziologie bislang wenig Literatur zur Charakterisierung der wissenschaftlichen Arbeit existiert, beschäftigt sich eine Vielzahl von Expertinnen und Experten bereits seit Jahrzehnten mit der näheren Beschreibung und Erforschung mit der verwandten Thematik der Wissensarbeit.

Mit der seit Mitte des 20. Jahrhunderts beobachteten gesellschaftlichen Entwicklung von einer Industrie- zu einer Wissensgesellschaft und der damit verbundenen verstärkten Ausrichtung auf professionalisierte, technisierte und dienstleistungsorientierte Tätigkeiten, gewann die Wissensarbeit zunehmend an Bedeutung und Beachtung (Nerdinger et al. 2014). Wissensarbeit ist heutzutage in vielen Berufen zu finden, z. B. bei Journalisten, Medizinern, Informatikern und Beratungsdienstleistenden.

Die Verwendung des Begriffs der Wissensarbeit in der deutschen Fachliteratur lässt sich auf die Übersetzung des englischen Begriffs „knowledge work“ zurückführen, wo er jedoch zumeist breiter ausgelegt wird. Drucker (1991) prägte maßgeblich den Begriff des „knowledge workers“ und beschrieb diesen als „an employee whose major contribution depends on his employing his knowledge rather than his muscle power and coordination, frequently contrasted with production workers who employ muscle power and coordination to operate machines“ (Drucker 1991, S. 564). Weiterhin definierte Drucker „knowledge workers“ als hoch qualifizierte Fachkräfte, die durch formale Ausbildung erworbenes theoretisches und analytisches Wissen im Beruf anwenden, um neue Produkte oder Dienstleistungen zu entwickeln. Drucker war ferner der Auffassung, dass kontinuierliches Lernen und ständige Innovationsbereitschaft wichtige Voraussetzungen für eine erfolgreiche Wissensarbeit seien.

Eine häufig zitierte Definition der Wissensarbeit aus dem deutschsprachigen Raum stammt hingegen von Willke (1998). Willke unterscheidet zwischen wissensbasierter Arbeit, die auf spezialisierter Expertise von Personen gründet und in den meisten Formen der Facharbeit vorliegt, und Wissensarbeit, die Tätigkeiten wie Kommunikation, Transaktionen und Interaktionen umfasst und erfordert, dass relevantes Wissen „(1) kontinuierlich revidiert, (2) permanent als verbesserungsfähig angesehen, (3) prinzipiell nicht als Wahrheit sondern als Ressource betrachtet wird und (4) untrennbar mit Nichtwissen gekoppelt ist, sodass mit Wissensarbeit spezifische Risiken verbunden sind.“ (Willke 1998, S. 161).

In Anlehnung an die Definitionen von Drucker und Willke charakterisieren May et al. (2002) die Wissensarbeit etwas näher. Sie spezifizieren unter anderem, dass die Aufgaben- und Problemstellungen bei der Wissensarbeit meist komplex und problemhaltig sind. Für die Problemlösung müssen oftmals neue Vorgehensweisen unter Einbezug aktuellen theoretischen, aber auch kontextuellen Wissens erarbeitet werden, was Kreativität erfordern kann. Letzteres wird gelegentlich auch als Innovationsarbeit bezeichnet. So beschreibt z. B. Hacker (2018) Innovationsarbeit als „Suchen von Lösungen und Wegen zur Lösung bei gut sowie schlecht definierten Problemen“ (Hacker 2018, S. 108).

Während die hier genannten und weitere Definitionen zur Wissensarbeit sich in einigen Aspekten unterscheiden, sind viele sich darin einig, dass die Wissensarbeit mehr umfasst als die Verarbeitung von Informationen und Anwendung von Wissen. Vielmehr ist sie spezifisch gekennzeichnet durch eine hohe Selbstständigkeit und eine gewisse (ergebnis- und prozessbezogene) Unsicherheit; sie ist komplex, kommunikationsintensiv und umfasst routinierte ebenso wie kreative Tätigkeiten, wodurch sich die Wissensarbeit von der klassischen Facharbeit unterscheidet. Die bei anderen Arbeitsformen üblicherweise eingesetzten innerbetrieblichen Kontroll- und Steuerungsformen finden somit bei der Wissensarbeit keine Anwendung (Wilkesmann 2005).

Im Sinne der Definition von Drucker (1991) sind Forschende als hoch qualifizierte Fachkräfte zu bezeichnen, die ihr Wissen anwenden, um neues Wissen oder wissenschaftliche Methoden zu entwickeln. Die Gruppe der Forschenden wäre dementsprechend der Kategorie der Wissensarbeitenden zuzuordnen. Es ist davon auszugehen, dass sich wissenschaftliche Arbeit im Bereich Forschung und Entwicklung durch ein besonders hohes Maß an Wissensarbeit (inkl. Innovationsarbeit) auszeichnet und damit ebenso von hoher Selbstständigkeit, (Prozess-)Unsicherheit und Kommunikationsintensität geprägt ist.

5.2 Auswirkungen der Digitalisierung auf die Arbeitsbedingungen von wissenschaftlich Arbeitenden

Die Digitalisierung zeigt in vielen Arbeitsbereichen sowohl positive als auch negative Auswirkungen auf Arbeitsbedingungen und die dadurch beeinflusste Gesundheit und Arbeitszufriedenheit von Beschäftigten. Es ist zu vermuten, dass Digitalisierungsmaßnahmen bei wissenschaftlich Arbeitenden ähnliche Effekte zeigen wie sie bei anderen wissensintensiven Berufsgruppen bereits ermittelt wurden.

Auf die physische und psychische Gesundheit von Beschäftigten wirkt sich der Digitalisierungstrend insgesamt eher nachteilig aus. Ein Teil der häufig festgestellten negativen gesundheitlichen Auswirkungen von Wissensarbeit lässt sich auf den damit verbundenen hohen Anteil von Büroarbeit an Computerarbeitsplätzen zurückführen. So zeigen arbeitswissenschaftliche Untersuchungen zu Bildschirmarbeit bereits ab einer Arbeitszeit von 2 Stunden diverse Beeinträchtigungen der Augen, darunter Sehschärfeminderungen, Farbsinnstörungen sowie Akkomodations- und Adaptionstörungen (Köchling 1985). Eine Umfrage von über 7000 Beschäftigten zeigte weiterhin stark verbreitete körperliche Fehlbeanspruchungen und eine deutliche Fehlbeanspruchung in Bezug auf Arbeitsintensität und Länge der Arbeitszeit im Zusammenhang mit Büroarbeit. Insbesondere die Gefahr der informationellen Überforderung ist hinreichend belegt. So zeigten z. B. Untersuchungen zur E-Mail-Kommunikation einen Anstieg erlebter Überforderung in Abhängigkeit von der Anzahl eingehender Meldungen je Zeiteinheit, vom Inhalt, von der Darstellung sowie von der beim Eingang der Meldung vorliegenden Arbeitssituation (vgl. Moser et al. 2002; Simpson und Prusak 1995). Unübersichtliche Informationsdarstellungen oder unerwartetes Systemverhalten können den „Technikstress“ weiter ansteigen lassen (Hoppe 2010). Mit zunehmendem Einsatz digitaler Technologien in der wissenschaftlichen Arbeit und Kommunikation lässt sich auch eine Zunahme solcher Fehlbeanspruchungen bei wissenschaftlich Beschäftigten annehmen.

Arlinghaus (2017) stellte eine Übersicht gesicherter arbeitswissenschaftlicher Erkenntnisse zur Belastung und Beanspruchung von Wissensarbeitenden zusammen. Unter anderem werden hier negative gesundheitliche und soziale Auswirkungen der durch die Digitalisierung geförderten ständigen Erreichbarkeit aufgeführt. Weiterhin ist die Arbeit im Home-Office häufig mit einer sozialen Isolation verbunden und dem Wegfall von Pausen, jedoch zeigen Studien auch eine subjektiv erhöhte Leistungsfähigkeit und eine bessere Vereinbarkeit von Aufgaben des Berufs- und Privatlebens. Unregelmäßige und lange tägliche und wöchentliche Arbeitszeiten sowie zu kurze Ruhezeiten zeigen wiederum zahlreiche gesundheitliche Auswirkungen, ein erhöhtes Unfallrisiko und verringerte Leistung sowie eine Verschlechterung der Work-Life-Balance.

In ihrer Arbeit sind Wissensarbeitende zwar häufig konfrontiert mit hohen (meist kognitiven) Arbeitsanforderungen, gleichzeitig sind sie jedoch auch zumeist in der Lage, über Art und Inhalt ihrer Arbeitstätigkeiten maßgeblich mitzubestimmen. Ihre Arbeit gewährt ihnen folglich einen hohen Handlungsspielraum, was häufig mit einer größeren Arbeitszufriedenheit einhergeht. Selbstbestimmte Arbeitszeiten wurden auch mit positiven Effekten auf die Gesundheit, Work-Life Balance und die soziale Teilhabe assoziiert (Arlinghaus 2017).

Es ist anzunehmen, dass über die o. g. Aspekte der Wissensarbeit auch wissenschaftliche Arbeit von einer erhöhten zeitlichen und örtlichen Flexibilisierung durch die Digitalisierung profitieren könnte. Die mit der Digitalisierung oftmals einhergehende Automatisierung reduziert in vielen Bereichen (physische und kognitive) Routinetätigkeiten, während der Anteil an nicht-routinierten, insbesondere komplexen, Tätigkeiten zunimmt. Dies schafft Bedingungen, die der wissenschaftlichen Arbeit, welche sich in der Regel mit komplexen Sachverhalten auseinandersetzt und

die häufig nach kreativen und innovativen Einfällen verlangt, welche oftmals auch außerhalb regulärer Büroarbeitszeiten entstehen, entgegenkommen könnten. Es liegt daher nahe, dass Forschende vom Einsatz digitaler Technologien mehr profitieren könnten als andere Berufsgruppen. Forschende haben außerdem in der Regel hohe Freiheitsgrade in Bezug auf Arbeitsinhalte, was mit einer erhöhten intrinsischen Arbeitsmotivation einhergeht. Es ist daher zu vermuten, dass Forschende, aufgrund eines höheren Maßes an Selbstbestimmtheit und intrinsischer Arbeitsmotivation resilienter gegenüber digitalen Stressoren sind als andere Berufsgruppen.

Die durch die Digitalisierung vorangetriebene Entgrenzung der Arbeit geht häufig mit einer Verlagerung in risikoreiche Arbeitsrechtsverhältnisse einher, wie Hacker (2018) in seiner Diskussion zur innovativen Erwerbsarbeit darstellt. Durch den Mangel an Zeit- oder Leistungsnormen in diesem Bereich erhöht sich außerdem die Gefahr einer kontinuierlichen Überforderung: Eine hohe Flexibilisierung der Arbeit in Kombination mit einer bei wissenschaftlich Beschäftigten häufig vorliegenden hohen intrinsischen Arbeitsmotivation erhöht damit die Wahrscheinlichkeit, dass Arbeitspersonen Arbeits- und Pausenzeiten wählen, die nicht nach ergonomischen Maßstäben und damit wenig menschenverträglich gestaltet sind. Da Wissenschaftler/Wissenschaftlerinnen an universitären und außeruniversitären Forschungseinrichtungen in der Regel heute bereits in zeitlich befristeten Arbeitsverhältnissen mit unklarer Berufsperspektive arbeiten, in Befragungen zumeist eine erhebliche Anzahl geleisteter (unbezahlter) Überstunden sowie eine deutliche Entgrenzung von Berufs- und Privatleben angeben (z. B. Müller 2015; Eichhorst 2015), ist noch nicht abzusehen, ob sich die Arbeitsbedingungen für Beschäftigte in der Wissenschaft mit der Digitalisierung maßgeblich verändern würden. Wahrscheinlicher erscheint es, dass das in der Politik bereits erkannte und häufig im gesellschaftlichen Diskurs thematisierte Problem prekärer Beschäftigungsverhältnisse von wissenschaftlich Arbeitenden zu einer positiven Entwicklung für diese Berufsgruppe führt. Um den gesundheitlichen Nebenwirkungen der Digitalisierung entgegen zu wirken, bedarf es hingegen noch weiterführender Maßnahmen zur gesellschaftlichen Sensibilisierung und Aufklärung in Bezug auf eine menschenverträgliche Arbeitsgestaltung.

5.3 KI und Automation der wissenschaftlichen Arbeit

Betrachtet man einige der in vielen Bereichen als „intelligent“ bezeichneten Technologien genauer, lässt sich feststellen, dass diese heutzutage in erster Linie für die Automatisierung von Arbeit eingesetzt werden, also für die technikbetriebene Durchführung von Arbeitsprozessen, die zuvor vollständig oder teilweise von menschlichen Arbeitskräften durchgeführt wurden (z. B. Hirsch-Kreinsen und Karačić 2019). Das Phänomen der Automatisierung von Arbeit wurde in den vergangenen Jahrzehnten in der Arbeitsforschung, Ingenieurpsychologie und weiteren verwandten Fachgebieten ausgiebig erforscht. Während in der Arbeitsforschung insbesondere die Erforschung und Ermittlung des Substituierbarkeitspotenzials von

Automationstechnologien und dessen Auswirkungen auf den Arbeitsmarkt von Interesse ist, steht in der Ingenieurpsychologie und Human Factors Forschung die Untersuchung von Wirkmechanismen auf der Ebene des Individuums im Fokus. Es ist davon auszugehen, dass der Einsatz von KI-Technologien auch die Automatisierung der wissenschaftlichen Arbeit weiter vorantreibt. Im folgenden Abschnitt sollen daher näher auf absehbare und mögliche Folgen der Automation für die wissenschaftliche Arbeit eingegangen werden.

Studien zur Schätzung des Substituierbarkeitspotenzials von Automationstechnologien stützen sich in der Regel auf dem in Tätigkeitsprofilen oder durch Selbstauskünfte beschriebenen Anteil von Routinetätigkeiten, also regelmäßig wiederkehrenden Vorgängen eines Berufs. Je größer der Anteil dieser Tätigkeiten, desto größer wird die Wahrscheinlichkeit eingeschätzt, dass die Arbeit durch Technologien ausgeführt werden kann. Die grundlegende Annahme ist hier, dass Arbeitsaufgaben automatisiert werden können, wenn sich diese in Gänze durch eine Computer-Anwendung beschreiben und ausführen lassen. So kam z. B. eine Untersuchung des Zentrums für Europäische Wirtschaftsforschung (ZEW) zu dem Schluss, dass neue Technologien nicht nur fähig seien, manuelle Routinearbeit zu ersetzen, sondern vor allem auch Arbeit mit einem hohen Anteil geistiger Routinearbeit zunehmend substituiert wird (vgl. Arntz et al. (2019)).

Diese Verschiebung der Anteile, die Menschen und Software-Lösungen bei der Erfüllung einer Aufgabe haben, wurde von der SAE International für das Thema „Autonomes Fahren“ beschrieben (SAE International 2016). Das entsprechende Phasenmodell ist in Abb. 5.1 dargestellt.

Das Beispiel des autonomen Fahrens ist gut geeignet, um diese Veränderungen darzustellen. Während ursprünglich die Fahrer alle Aktionen, wie Lenken, Bremsen, Gas geben etc., selbstständig ausgeführt haben, werden zukünftig immer mehr Aufgaben von Software-Lösungen übernommen, wobei nicht zwingend KI-Algorithmen eingesetzt werden müssen (Buxmann 2020). Das hier dargestellt Grundprinzip lässt sich auf eine Vielzahl anderer Bereiche aus Wissenschaft und Wirtschaft übertragen. Mit dem Einsatz von KI-Anwendungen können Effizienz-

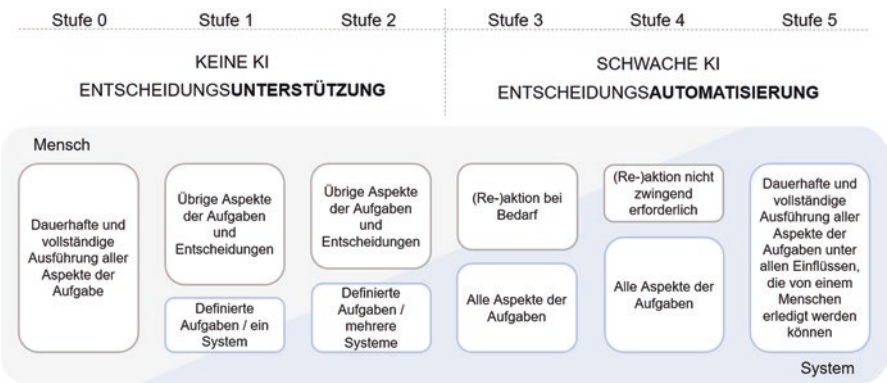


Abb. 5.1 Arbeitsteilung zwischen Mensch und Maschine in Anlehnung an das Phasenmodell der SAE International (verändert nach SAE International 2016)

gewinne verbunden sein, darüber hinaus ist es aber auch denkbar, dass KI-Anwendungen bessere Entscheidungen als Menschen treffen. So könnten KI-Anwendungen einen immer größeren Teil der Literaturrecherche übernehmen oder bei Formulierung von Standardtexten helfen. Neben den möglichen Effizienz- und Effektivitätseffekten ist aber auch zu berücksichtigen, dass ein höherer Anteil von Arbeiten, die von KI-Anwendungen übernommen werden, Verantwortlichkeiten für Entscheidungen vom Menschen weg hin zu KI-Anwendungen verschieben. Vor diesem Hintergrund ist eine breite gesellschaftliche Debatte notwendig wieviel Verantwortung zukünftig bei KI-Anwendungen liegen sollte.

Auf makroökonomischer Ebene wurden in einer 2018 vom World Economic Forum durchgeführten Studie „The Future of Jobs 2018“ über 300 Unternehmen nach den Auswirkungen der Digitalisierung auf die Zukunft der Arbeit gefragt. In Abb. 5.2 ist dargestellt, welche Verschiebungen zwischen Mensch und Maschine bis 2022 stattfinden könnten.

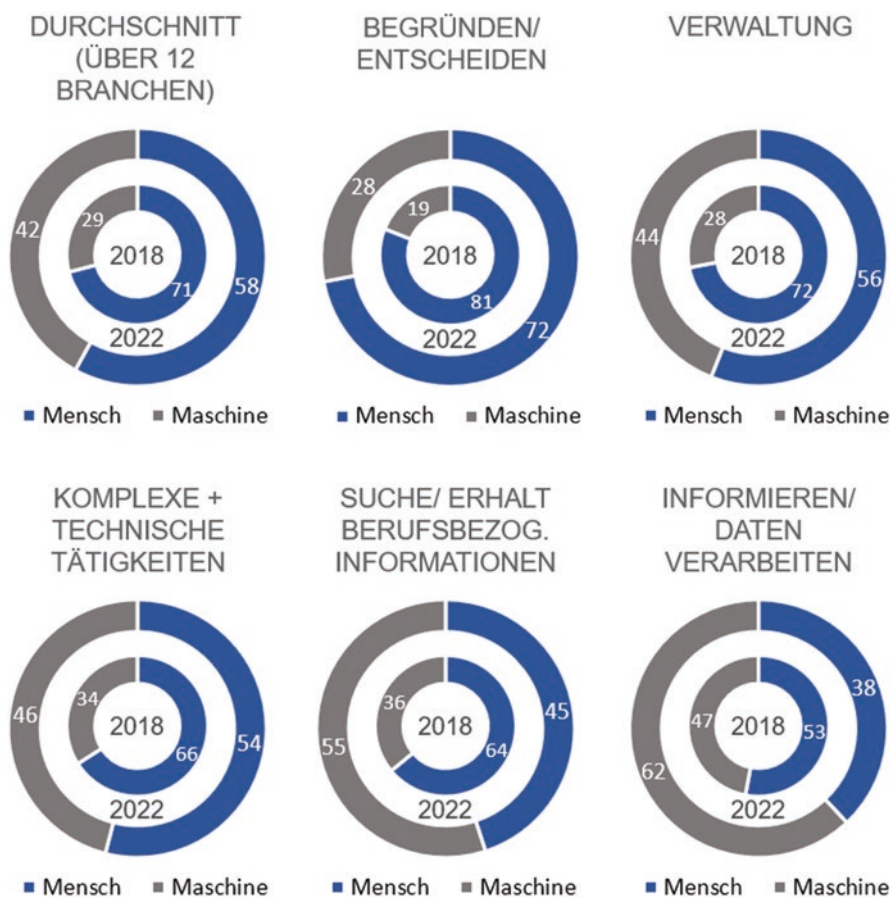


Abb. 5.2 Maschinen übernehmen die Arbeit – Aufteilung der Arbeitsstunden zwischen 2018 und 2022/Anteile in Prozent (verändert nach World Economic Forum 2018 und Schmidt 2018)

Ein zentrales Resultat der Studie ist, dass KI-Anwendungen zukünftig viele Aufgaben übernehmen werden, die bislang von Menschen durchgeführt wurden. Ein weiteres wichtiges Ergebnis ist, dass KI immer stärker auch Aufgaben übernehmen könnte, für die kognitive Fähigkeiten erforderlich sind. Das bedeutet beispielsweise, dass zukünftig Menschen auch bei höherwertigen Aufgaben, wie Argumentation und Entscheidungsfindung, durch KI-Anwendungen unterstützt werden. Das könnte auch für die Arbeit von Wissenschaftlern/Wissenschaftlerinnen gelten, etwa im Bereich der Medizin bei der Diagnose und Therapie von Krankheiten. Hierbei ist anzumerken, dass in der Informatik zurzeit weltweit viele Arbeiten rund um das Thema „Explainable Artificial Intelligence“ durchgeführt werden (Adadi und Berrada 2018). Dies kann zukünftig helfen, Entscheidungen der KI besser zu erklären, was für die Akzeptanz der Lösung, beispielsweise für Ärzte, essenziell ist.

Vor diesem Hintergrund kommen viele Studien zu dem Ergebnis, dass KI zu einer Vernichtung von Arbeitsplätzen führen wird. So zeigt eine viel zitierte Studie der Oxford Universität, dass bis zu 47 Prozent aller US-amerikanischen Jobs stark von KI gefährdet seien (Frey und Osborne 2017). In diesem Zusammenhang warnt z. B. die OECD vor einem umfangreichen Stellenabbau. Rund 14 Prozent der Jobs in den Mitgliedsstaaten, darunter auch Deutschland, seien hochgradig automatisierbar, und 32 Prozent der Jobprofile könnten sich stark verändern (OECD 2018). Während diese Studien die Auswirkungen der KI auf den Arbeitsmarkt aufgabenbasiert erfassen, hat der Stanford-Forscher Michael Webb eine neue Methode entwickelt (Webb 2019): Er hat die Beschreibungen der aktuellen Tätigkeiten der Beschäftigten mit den Beschreibungen der Patentanmeldungen zur KI verglichen. Die Überlappungen zeigen das Substitutionspotenzial. Steht zum Beispiel in der Aufgabenbeschreibung eines Arztes „Diagnose des Patientenzustandes“, hat Webb die Patentanmeldungen mit dem Begriff „Zustandsdiagnose“ oder ähnlichen Wörtern gesucht. Die Häufigkeit der gefundenen Patente gibt Aufschluss über die Wahrscheinlichkeit einer KI-induzierten Automatisierung dieser Tätigkeit. Die Ergebnisse decken sich zu großen Teilen mit denen der aufgabenorientierten Studien, jedoch konnte er auch zeigen, dass von der KI innovative Tätigkeiten bzw. solche die Neuland betreten (zum Beispiel neue wissenschaftliche Projekte) weniger stark betroffen sind. (Webb 2019; Buxmann und Schmidt 2020).

Diese Angaben zu Ersetzungspotenzialen sind jedoch stets mit Vorsicht zu interpretieren, denn sie treffen notwendigerweise zahlreiche Annahmen, u. a. zu getätigten Investitionen in die Digitalisierung von Unternehmen, zur Beurteilung der Routinehaftigkeit von Berufen und zur Entwicklung der Fähigkeiten von Technologien. Zu beachten ist dabei ferner, dass das Substituierbarkeitspotenzial nicht zwangsläufig zu einer höheren Arbeitslosigkeit führt, sondern lediglich – zumindest auf Bundesebene – Veränderungen in der Struktur der Beschäftigung zu beobachten sein werden. So werden z. B. größere Zuwächse an Arbeitsplätzen in den Bereichen der geistes- und sozialwissenschaftlichen und künstlerischen Berufe und der Berufe in IT und Naturwissenschaften erwartet (vgl. Zika et al. 2018).

Speziell in Bezug auf die kostenabhängige Realisierung von Wissensarbeit merkte Hacker (2018) an, dass es auch zukünftig noch längere Zeit Nischen geben wird, in denen menschliche Wissensarbeit günstiger als digitale Lösungen ist. Jedoch gab er mit Blick auf die historische Entwicklung des globalen Arbeitsmarkts auch zu be-

denken, dass damit auch die verbleibenden Arbeitsplätze vermutlich in Länder abwandern würden, in denen diese Arbeit günstiger verrichtet werden könnte. Bereits heute befeuern neue, durch die Digitalisierung ermöglichte Formen der Arbeit, wie z. B. das Crowdsourcing, vor dem Hintergrund der Plattformökonomie den globalen Wettbewerb unter hochqualifizierten Arbeitskräften. Diese neuen Arbeitsformen gehen nicht selten auch mit prekären Beschäftigungsverhältnissen einher.

Doch es werden auch technologische Grenzen der Substituierbarkeit (zumindest in der näheren Zukunft) gesehen. In ihrer weltweit viel zitierten Studie zur Schätzung des Substituierbarkeitspotenzials in den USA gingen Frey und Osborne (2013) davon aus, dass drei Aspekte menschlicher Fähig- und Fertigkeiten in naher Zukunft nicht von Technologien ersetzt werden können, die sie unter den Begriffen Wahrnehmung und Feinmotorik, soziale Intelligenz und kreative Intelligenz zusammenfassten. Hacker (2018) sah ebenfalls eine Automatisierungsgrenze bei der kreativen Innovationsarbeit. Vor dem Hintergrund kognitionspsychologischer Überlegungen spezifizierte er, dass geistige Arbeit nicht automatisiert werden könnte, wenn diese die Suche nach Lösungen beinhaltet, bei denen weder Ziel noch der Weg zur Zielerreichung vorgegeben ist. Hingegen kann der Mensch dort ersetzt werden, „wo eindeutige Verarbeitungsregeln der Information (Algorithmen) vorliegen und prinzipiell auch im Übergangsfeld, wo definierbare Findehilfen (Heuristiken) angegeben werden können.“ (Hacker 2018, S. 9).

Selten ersetzen Technologien ganze Berufe, viel häufiger werden an Arbeitsplätzen nur einzelne Tätigkeitsbereiche automatisiert. Die Einführung automatisierter Systeme an Arbeitsplätzen bewirkt damit nicht nur eine Veränderung von Arbeitstätigkeiten der dort eingesetzten Menschen (vgl. Autor und Price 2013), sondern birgt auch Risiken, die es bei der Gestaltung dieser Arbeitsplätze und der Schulung von Personal zu berücksichtigen gilt. Diese Risiken liegen hauptsächlich in einer aus arbeitswissenschaftlichen Sicht unangemessenen Umsetzung des Automationsgrads im Falle einer Teilautomatisierung. Es ist davon auszugehen, dass diese auch teil-automatisierte wissenschaftliche Arbeit betreffen, insbesondere die Datenakquise und -verarbeitung.

Mit höheren Automationsgraden fällt es dem Menschen zunehmend schwerer, die agierenden Mechanismen und den Zustand des Systems nachzuvollziehen. Als Folge sinken die Prozesstransparenz und die Fähigkeit des Menschen, Fehler in den Handlungen oder den Handlungen zugrunde liegenden Annahmen zu detektieren oder gar zu korrigieren. Soll der Mensch bei teilautomatisierten Arbeitsaufgaben nicht nur die Entscheidungshoheit, sondern auch die Entscheidungskompetenz behalten, ist daher auf die Umsetzung einer menschenzentrierten Automation zu achten. Diese erfordert u. a. den Einsatz geeigneter Benutzungsschnittstellen und eine lernförderliche Arbeitsgestaltung.

Zusammenfassend lässt sich an dieser Stelle festhalten, dass auch wissenschaftliche Arbeit von der durch KI und anderen digitalen Technologien vorangetriebenen Automatisierung betroffen sein wird. Zum einen wird sie Arbeitstätigkeiten verändern und damit auch zweifellos eine Entlastung von ungeliebten und nicht wertschöpfenden Routinetätigkeiten (insbesondere im administrativen Bereich) bringen. Auf der anderen Seite wird sie jedoch auch den Arbeitsexport und globalen Wettbewerbsdruck in der Wissenschaft befördern. Bei der Einführung von neuen Auto-

mationstechnologien sollten Anwendende weiterhin für einen angemessenen Umgang mit diesen Technologien sensibilisiert und ein grundsätzliches Verständnis der Mechanismen und der Limitationen der Technologien geschaffen werden.

5.4 Empirische Befunde zu den Auswirkungen von KI und Digitalisierung auf die Arbeit und Arbeitsbedingungen von Forschenden

Eine im Jahr 2019 im Auftrag des IQIB-Instituts für qualifizierende Innovationsforschung und -beratung GmbH durchgeführte Umfrage lieferte erste Anhaltspunkte für die Bewertung möglicher Auswirkungen von Digitalisierung im Allgemeinen und KI-basierten Technologien im Speziellen auf die Arbeit und Arbeitsbedingungen von Forschenden (Rick et al. 2021).¹

Befragt wurden 166 Wissenschaftlerinnen und Wissenschaftler im Raum Nordrhein-Westfalen (s. Abb. 5.3 zur Beschreibung der Stichprobe), die zum Zeitpunkt der Befragung an Universitäten und Fachhochschulen, sowie außeruniversitären Forschungseinrichtungen (DLR, DFKI, verschiedene Fraunhofer Einrichtungen) beschäftigt waren. Die Teilnehmenden ordneten sich den Fachdisziplinen der Naturwissenschaften (15 %), der Ingenieurwissenschaft und Informatik (62 %) oder der Geistes- und Sozialwissenschaft (23 %) zu.

5.4.1 Auswirkungen der Digitalisierung auf den Forschungsalltag

Die Umfrage bestätigte, dass wesentliche Merkmale der Wissensarbeit auch auf die wissenschaftliche Arbeit zutrafen. So zeigte sich hier unter anderem, dass auch die wissenschaftliche Arbeit sehr kommunikationsintensiv ist und hierfür sehr häufig



Abb. 5.3 Beschreibung der Stichprobe (eigene Darstellung)

¹Die Begleitstudie wurde von Vera Rick (RWTH Aachen), Julia Distelrath (IQIB) und Verena Nitsch (RWTH Aachen) durchgeführt. Detailliertere Ausführungen zur Studie und den Ergebnissen wurden auf dem Publikationsserver der RWTH Aachen University veröffentlicht und sind frei zugänglich unter <https://publications.rwth-aachen.de/record/811477/files/811477.pdf>.

digitale Technologien genutzt wurden. Für das kollaborative wissenschaftliche Arbeiten wurden jedoch in erster Linie Technologien eingesetzt, mit denen man nur gelegentlich Kontakt zu anderen Personen hat, z. B. Videokonferenzdienste, oder mit denen Interaktionen eher indirekt vollzogen werden, z. B. über webbasierte Dienste für den Datenaustausch und die Wissensdokumentation. Weniger verbreitet hingegen waren Technologien zur direkten kollaborativen Arbeit, wie z. B. kollaborative Schreibprogramme und Projektmanagement-Software (s. Abb. 5.4).

Überraschenderweise scheint mobiles Arbeiten unter den befragten Wissenschaftlerinnen und Wissenschaftlern nicht so weit verbreitet zu sein, wie häufig angenommen. So gaben zwar rund zwei Drittel (68 %) der Befragten an, bei ihrer derzeitigen Tätigkeit Telearbeit, Homeoffice oder mobiles Arbeiten zu nutzen, von diesen Personen nahm jedoch nur ein Drittel (33 %) das Angebot ein- oder mehrmals pro Woche in Anspruch, während rund die Hälfte (55 %) dies nur einmal pro Monat oder seltener taten (s. Abb. 5.5). Es ist dabei jedoch anzumerken, dass die Umfrage vor der Covid-19 Pandemie stattfand. Aufgrund der pandemiebedingten Restriktionen haben sich in vielen deutschen Unternehmen technische, organisatorische und personelle Rahmenbedingungen zugunsten der Homeoffice-Arbeit entwickelt. Es lässt sich vermuten, dass diese Entwicklungen auch an vielen Forschungsstandorten stattfanden und damit das Homeoffice zukünftig auch mehr für wissenschaftliche Arbeit genutzt wird. Ob diese Umstellung nur von kurzer oder längerer Dauer ist, lässt sich zum Zeitpunkt dieser Veröffentlichung noch nicht abschätzen.

Obgleich zum Zeitpunkt der Befragung vergleichsweise wenig Personen häufig mobil arbeiteten, äußerten rund zwei Drittel der Befragten (67 %), dass der zunehmende Einsatz von digitalen Technologien die Grenzen zwischen Arbeits- und Privatleben auflöst. Wenige Personen verbanden damit jedoch ein erhöhtes Konfliktpotenzial oder Schwierigkeiten bei der Erledigung privater Aufgaben (s. Abb. 5.6).

Rund die Hälfte der befragten Wissenschaftlerinnen und Wissenschaftler schrieben dem Einsatz digitaler Technologien eine Vergrößerung ihres Handlungsspielraums zu. So gaben 48 % der Befragten an, dass digitale Technologien eine freiere

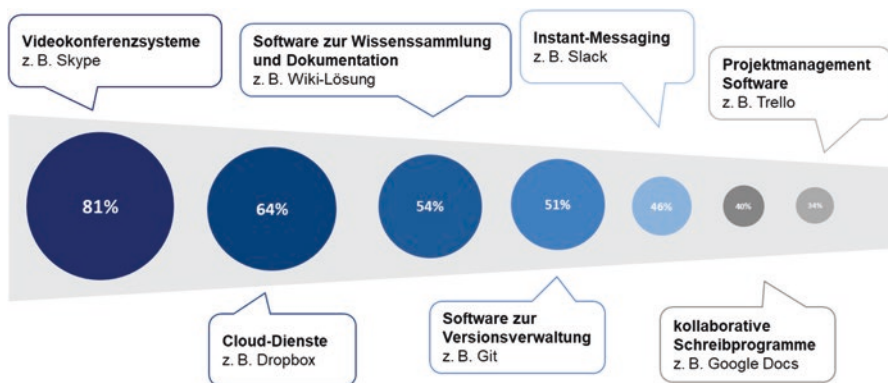


Abb. 5.4 Verwendete Technologien zur Unterstützung der kollaborativen wissenschaftlichen Arbeit (eigene Darstellung)

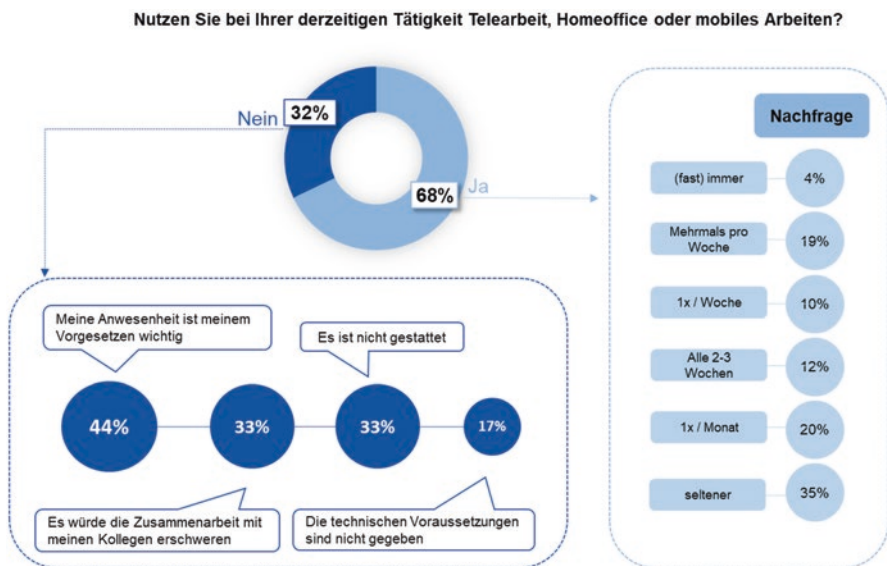
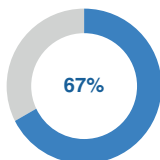


Abb. 5.5 Angaben zur Nutzung von Möglichkeiten des mobilen Arbeitens (eigene Darstellung)

Verwischung der Grenze zwischen Arbeits- und Privatleben durch digitale Technologien.



Digitale Technologien führen zu ...

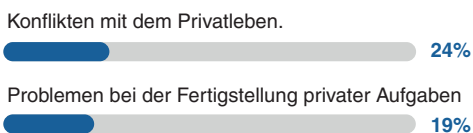


Abb. 5.6 Angaben zur technologiegetriebenen Entgrenzung von Arbeit (eigene Darstellung)

zeitliche Einteilung der Arbeit ermöglichen und 46 % stimmten der Aussage zu, dass digitale Technologien eine größere Entscheidungsautonomie geben.

Auch in Bezug auf Produktivität und Qualität der Arbeit wird der Einfluss von digitalen Technologien im Forschungsalltag positiv gesehen. So sahen z. B. 58 % der Befragten eine Verbesserung der Produktivität und ebenso viele eine qualitative Verbesserung der Arbeitsergebnisse durch den Einsatz von digitalen Technologien. Nur rund ein Viertel der befragten Personen waren der Ansicht, dass digitale Technologien den Zeitdruck bei der Arbeit erhöhten (Abb. 5.7). Fast alle (92 %) fühlten sich nach eigenen Angaben hinreichend kompetent, um die Potenziale digitaler Unterstützung zu nutzen. Weiterbildung im Umgang mit digitalen Technologien erfolgte bei den meisten Personen (91 %) in Eigenregie durch Ausprobieren; nur 15 % gaben an, dafür Schulungen und Weiterbildungen des Arbeitgebers in Anspruch zu nehmen.

Digitale Technologien führen in meinem Alltag zu...

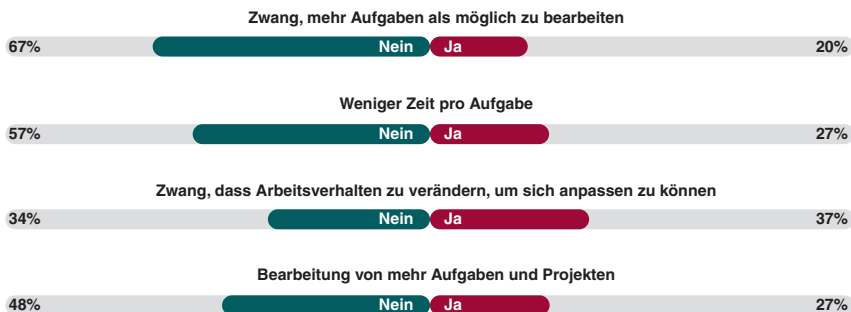


Abb. 5.7 Auswirkung digitaler Technologien auf den subjektiven Zeitdruck (eigene Darstellung)

5.4.2 Angaben zum Einsatz von KI-basierten Technologien

Zunächst wurden die Befragten gebeten, zu einer Vielzahl von Technologien die Einschätzung abzugeben, ob es sich ihrer Meinung nach dabei um eine KI-Technologie handelt. Das Ergebnis der Befragung ist in Abb. 5.8 dargestellt. Die Antworten deuten darauf hin, dass auch bei hochqualifizierten Personen mit selbst-eingeschätzter hoher Kompetenz im Umgang mit digitalen Technologien eine Unklarheit darüber zu herrschen scheint, was tatsächlich als „KI“ bezeichnet werden kann. So wurden z. B. Technologien wie Data Mining Tools und Suchmaschinen von der Mehrheit der Befragten nicht als KI-Technologie eingestuft. Die Begriffe Mustererkennung, maschinelles Lernen und Deep Learning hingegen wurden zwar von einer Mehrheit der Kategorie „KI“ zugeordnet, jedoch war auch hier jede 5. Person einer anderen Meinung (s. Abb. 5.8).

Um für die weitere Befragung eine gemeinsame Verständnisgrundlage zu schaffen, wurde anschließend KI folgendermaßen definiert:

„Künstliche Intelligenz (KI) beschreibt die Fähigkeit einer Computer-Anwendung, Eigenschaften menschlicher Intelligenz aufzuweisen, beispielsweise durch hören, sehen, lernen, sprechen, planen oder schlussfolgern. Demnach sind beispielsweise auf Programme zur Gesichtserkennung als KI-basierte Technologie einzustufen, da Eigenschaften wie sehen und lernen vorhanden sind.“

Dieser Definition folgend gaben nur rund ein Viertel der Befragten an, nicht mit KI Technologien zu arbeiten, allerdings wurden diese von den Befragten vergleichsweise wenig (<10 Std./Woche) eingesetzt, insbesondere von Geistes- und Sozialwissenschaftlerinnen und -wissenschaftlern.

Eine differenzierte Betrachtung der Einstellungen von Wissenschaftlern/Wissenschaftlerinnen gegenüber KI-Technologien in Abhängigkeit von der selbstbenannten Nutzung solcher Technologien zeigt einige Unterschiede auf. So stimmten Personen, die im Rahmen ihrer Arbeit nach eigenen Angaben viel Gebrauch von KI-Technologien machten (≥ 21 Std./Woche), der Aussage, dass KI die wissen-

Keine KI		KI	
PCs	100%	Maschinelles Lernen	82%
Videokonferenzdienste	100%	Deep Learning	81%
Textverarbeitungsprogramme	99%	Mustererkennung	80%
Webbasierte Kommunikation	99%	Problemlösung	76%
Fernzugang/VPN	99%	Übersetzungsprogramme	65%
Smartphones	97%	Digitale Assistenten	64%
Bildverarbeitung	95%	Bilderkennung	59%
Hochleistungsrechnensysteme	95%	Chatbots	50%
Literatursuchprogramme	90%	Big Data	33%
Suchmaschinen	65%		
Data Mining Tools	62%		

Abb. 5.8 KI-Klassifikation von verschiedenen Technologien nach Einschätzung der Befragten (eigene Darstellung)

schaftliche Selbstbestimmtheit oder die Standards guter wissenschaftlicher Praxis beeinträchtigt deutlich seltener zu als Personen, die weniger mit derartigen Technologien arbeiteten (s. Abb. 5.9). Hierbei ist jedoch auch zu beachten, dass sich die Größe der jeweiligen Stichproben auch stark unterscheidet, so dass ein Vergleich der relativen Häufigkeiten der Aussagen nur bedingt aussagekräftig ist.

Unabhängig von der KI-Nutzung zeigten die befragten Wissenschaftlerinnen und Wissenschaftler jedoch eine überwiegend positive Meinung von den Auswirkungen dieser Technologien auf den eigenen Forschungsalltag. Einzig in Bezug auf den durch die Technologien zu erwartenden Fortschritt zeigten sich hier größere Unterschiede zwischen den beiden Nutzergruppen (s. Abb. 5.10).

Doch den Befragten waren auch Limitationen der Technologien weitgehend bekannt. So war weniger als ein Viertel der Befragten (23 %) der Ansicht, dass KI-Technologien objektive Ergebnisse liefern. Immerhin 39 % der Befragten waren der Ansicht, dass damit Ergebnisse erzeugt werden, die von Ihnen nicht unmittelbar nachvollzogen werden können.

Mit offenen Fragen wurde nach absehbaren Folgen von KI-Technologien sowie nach Wünschen und Befürchtungen gefragt, die von etwas weniger als der Hälfte der Befragten beantwortet wurden. In Bezug auf absehbare Folgen sahen lediglich 5 % dieser Personen eine Zunahme der Bedeutung der Informatik in der Wissenschaft und 11 % eine Veränderung ihres Themenschwerpunkts. Die Mehrheit der

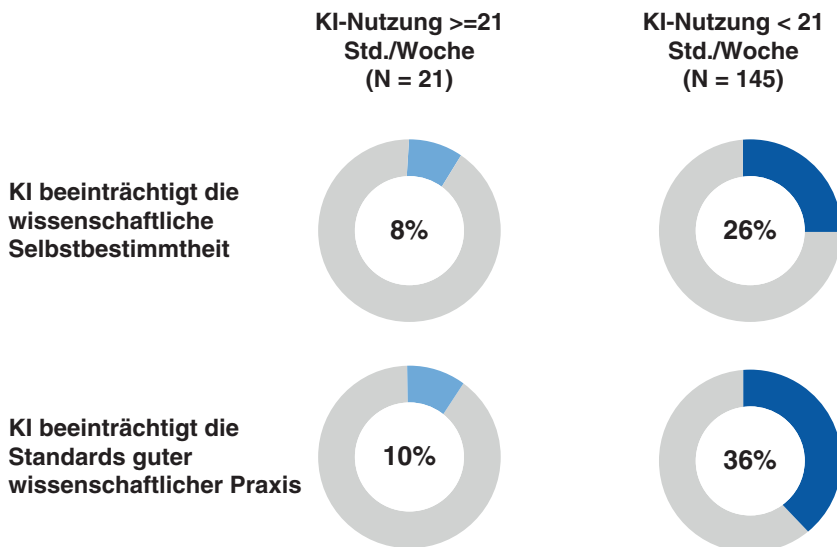


Abb. 5.9 Einschätzte Beeinträchtigungen des wissenschaftlichen Arbeitens in Abhängigkeit der Nutzung von KI-basierten Technologien (% der Teilnehmenden, die die jeweilige Aussage mit „trifft zu“ und „trifft voll und ganz zu“ bewerteten) (eigene Darstellung)

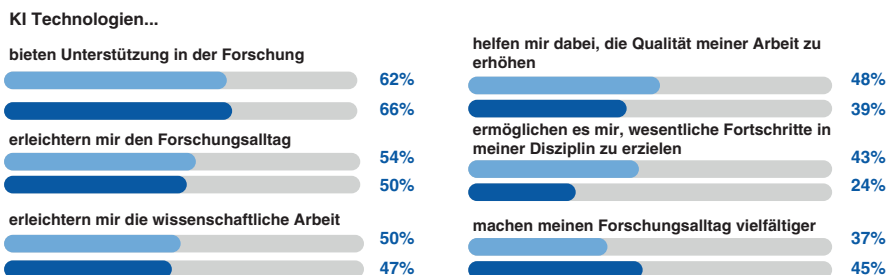


Abb. 5.10 Bewertung des Einflusses von KI Technologien auf den Forschungsalltag (hellblau: mindestens 21 Std./Woche KI-Nutzung; dunkelblau: höchstens 20 Std./Woche KI-Nutzung) (eigene Darstellung)

Befragten erhoffte sich jedoch Prozessoptimierungen durch KI-Technologien. Nur 3 % der Befragten waren der Ansicht, dass KI keine absehbaren Folgen haben würde. Lediglich 8 % der Befragten sahen einen Stellenabbau in der Wissenschaft voraus. Weitere Befürchtungen sind in Abb. 5.11 zusammengefasst dargestellt.

Wünsche an KI-Technologien wurden insbesondere in Bezug auf Nutzerfreundlichkeit (gemeinsame Standards, Transparenz, Zuverlässigkeit) geäußert (45 %), aber auch die Übernahme von Routinetätigkeiten hatten sich einige Personen (27 %) erhofft. Die Sicherstellung von Datensicherheit, Privacy und des ethisch vertretbaren Einsatzes wurde von 12 % der Befragten gefordert.

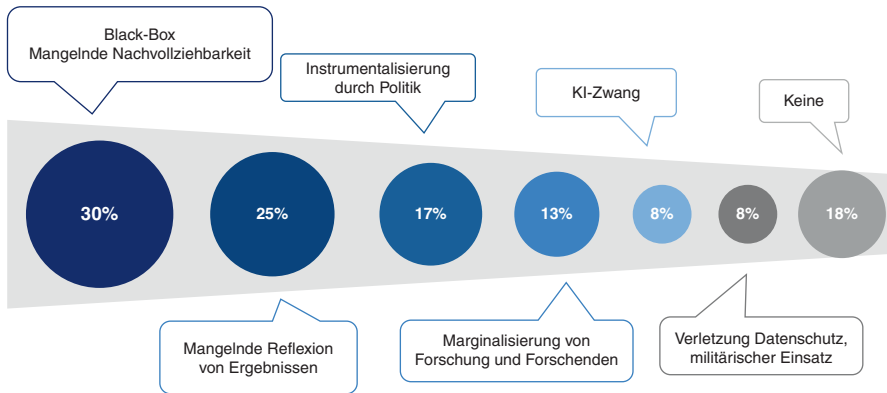


Abb. 5.11 Befürchtungen der befragten Wissenschaftlerinnen und Wissenschaftler gegenüber KI-Technologien (eigene Darstellung)

Die Ergebnisse der Umfrage sollten zweifellos mit Vorsicht interpretiert werden. Zum einen bilden sie nur eine Momentaufnahme ab, die aufgrund dynamischer technologischer, organisatorischer und kultureller Entwicklungen in Unternehmen eine kurze Halbwertszeit besitzt. Sowohl die Nutzung als auch die subjektive Bewertung von neuen Technologien werden stark von der „Expositionsdauer“ als auch einzelnen, bemerkenswerten Ereignissen beeinflusst (z. B. der Berichterstattung von verunfallten automatisierten Fahrzeugen oder populären Filmen). Zum anderen lassen sich aus der sehr selektiven Stichprobe von Befragten kaum Aussagen ableiten, die ohne weiteres auf eine ganze Berufsgruppe übertragbar wären. Ziel der Umfrage war es lediglich, erste Hinweise für eine Einschätzung möglicher positiver und negativer Auswirkungen von digitalen und KI-Technologien zu geben, die dann ggfs. in weiteren Untersuchungen fokussiert werden sollten.

Insgesamt deuten die Ergebnisse der Umfrage überwiegend positive Effekte von Digitalisierung und KI-Technologien auf wissenschaftliche Arbeitstätigkeiten an. Obgleich digitale Technologien – wie in vielen anderen Berufsfeldern auch – die Work-Life-Balance ins Schwanken bringen, scheint dies nur von wenigen Forschenden als problematisch bewertet zu werden. Digitale Technologien werden von Forschenden mehrheitlich als Hilfsmittel angesehen, die das Potenzial besitzen, sowohl die Produktivität als auch die Qualität der wissenschaftlichen Arbeit zu steigern.

Wie in vielen Forschungsarbeiten zu künstlicher Intelligenz zeigt auch diese Umfrage unter Forschenden, dass der Begriff „KI“ für verschiedene Personen unterschiedliche Bedeutungen hat. Technologien, die gemäß der in der Umfrage verwendeten Arbeitsdefinition als KI-Technologie bezeichnet werden könnten, kommen zwar bislang nur wenig (zumindest bewusst) zum Einsatz, jedoch werden auch diese eher selten negativ bewertet. Wie so häufig in Bezug auf Automationstechnologien beobachtet, wird auch beim Einsatz von KI-Technologien eine mangelnde

Transparenz und Nachvollziehbarkeit von Arbeitsprozessen und -ergebnissen befürchtet. Einschränkungen der wissenschaftlichen Selbstbestimmtheit oder des Forschungsalltags werden jedoch nur von wenigen befürchtet.

Es zeichnet sich ferner die Tendenz ab, dass Personen, die KI-Technologien häufiger einsetzen, den Technologien eher positive Auswirkungen zuschreiben. Dies könnte auf den Aspekt der freiwilligen Nutzung zurückzuführen sein: wer der Auffassung ist, dass KI-Technologien die Arbeit unterstützen, wird diese auch mit höherer Wahrscheinlichkeit bei sich einsetzen. Bei vielen digitalen Technologien ist diese Freiwilligkeit jedoch heutzutage nicht mehr gegeben: Forschende, die z. B. keine E-Mail nutzen oder nicht online publizieren, können ihre Aufgaben nicht im gleichen Maße erfüllen, wie Forschende, die diese Technologien nutzen. Sollte der Einsatz von KI-Technologien zukünftig für die wissenschaftliche Arbeit (und/oder Akzeptanz in der Community) zwingend notwendig werden, könnte dies auch eine negativere Bewertung solcher Technologien bewirken.

5.5 Fazit

Bislang existieren kaum arbeitswissenschaftliche Untersuchungen zu den möglichen Auswirkungen von Digitalisierungsmaßnahmen und dem Einsatz von KI-basierten Technologien auf die wissenschaftliche Arbeit und die Arbeitsbedingungen von Wissenschaftlerinnen und Wissenschaftlern; entsprechend unsicherheitsbehaftet sind deren Prognosen. Lassen sich jedoch arbeitswissenschaftliche Erkenntnisse zur Wissensarbeit übertragen, kann davon ausgegangen werden, dass wissenschaftlich Arbeitende von dem aktuellen Entwicklungstrend in vielerlei Hinsicht profitieren. Digitale Technologien und KI fördern sowohl die Arbeitsproduktivität als auch -qualität und schaffen mehr Freiräume für kreative Innovationsarbeit – insbesondere durch die Automatisierung von ungeliebten Routinetätigkeiten. Erfolgt diese Automatisierung jedoch ohne systematische Berücksichtigung menschlicher Fähigkeiten und Bedürfnisse, werden wissenschaftliche Erkenntnisse selbst für die Forschenden zunehmend schwieriger nachzuvollziehen. Mögliche Folgen: Ergebnisse können nicht mehr repliziert oder kritisch hinterfragt werden und Forschende verlieren nicht nur die Möglichkeit, an Forschungsprozessen zu lernen, sondern auch die Arbeitsmotivation aufgrund mangelnder Selbstwirksamkeit. Für den Erhalt der physischen und psychischen Gesundheit von Forschenden ist es ferner unabdingbar, auf menschenverträgliche Arbeitsbedingungen zu achten. Ergonomisch gestaltete Arbeitsplätze (auch im Homeoffice!) und regelmäßige Arbeitspausen von ausreichender Dauer sind eine wichtige Voraussetzung hierfür.

Aus gesamtwirtschaftlicher Perspektive kann festgehalten werden, dass die häufig in den Medien geäußerten Befürchtungen einer Massenarbeitslosigkeit durch

die Anwendung von KI-Systemen übertrieben erscheinen. Zwar werden KI-Anwendungen in Zukunft in immer mehr Berufen die Menschen unterstützen. Das gilt auch für den Beruf der Wissenschaftler/Wissenschaftlerinnen. Allerdings wird es nur in wenigen Berufsfeldern zu einer Substitution kommen. Zudem ist bei der Analyse der Nettoeffekte auf dem Arbeitsmarkt auch zu berücksichtigen, dass neue Jobs entstehen werden. Gerade im Wissenschaftsbereich brauchen sich Forscherinnen und Forscher also keine Sorgen zu machen, dass KI-Anwendungen zukünftig ihre Arbeiten übernehmen werden. Insbesondere kreatives Denken oder die Entwicklung von Innovationen werden auf absehbare Zeit weiterhin Domäne der Menschen bleiben. KI-Anwendungen können aber Wissenschaftlerinnen und Wissenschaftler bei ihrer Arbeit unterstützen. Damit ist die Hoffnung verbunden, dass dies zu einer Steigerung der Effizienz und Effektivität ihrer Forschungstätigkeiten führen kann. Jedoch sind auch politische Maßnahmen gefragt, die nicht nur prekären Beschäftigungsverhältnissen, sondern auch einer KI-begünstigten Entwertung wissenschaftlicher Arbeit und der damit verbundenen Verlagerung von Arbeitsplätzen in Niedriglohnländer entgegenwirken.

Die vorliegenden Ergebnisse zeigen weiterhin einen großen Forschungsbedarf für die Arbeitswissenschaft in den kommenden Jahren auf. So existieren bislang zum Beispiel kaum Kenntnisse zu den Effekten von KI auf innovative Gruppenarbeit, eine Arbeitsform, die insbesondere in einigen Wissenschaftsdisziplinen weit verbreitet ist. Unklar ist ebenso, welche Merkmale Mensch-Maschine Schnittstellen aufweisen sollten, um sowohl die Handlungs- als auch Entscheidungskompetenz von Beschäftigten im Umgang mit KI-Technologien zu fördern und zu präservieren. Weiterhin besteht noch Unklarheit über neue Arbeitsaufgaben und Arbeitsplätze, die durch KI-Technologien geschaffen werden. Dies hat u. a. zur Folge, dass weder Bildungseinrichtungen noch Unternehmen einschätzen können, welche Kompetenzen Beschäftigte in den kommenden Jahren benötigen, um auch den zukünftigen Arbeitsaufgaben in einer digitalisierten Arbeitswelt gerecht zu werden.

Literatur

- Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 6:52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Arlinghaus A (2017) Wissensarbeit: Aktuelle arbeitswissenschaftliche Erkenntnisse. Mitbestimmungsreport 35. Düsseldorf. <http://hdl.handle.net/10419/175264>. Zugegriffen am 13.10.2020
- Arntz M, Gregory T, Zierahn U (2019) Digitization and the future of work: macroeconomic consequences. In: Zimmermann KF (Hrsg) *Handbook of labor, human resources and population economics*. Springer International Publishing, Cham, S 1–29
- Autor DH, Price B (2013) The changing task composition of the US Labor Market: an update of Autor, Levy, and Murnane (2003). MIT, Arbeitspapier
- Buxmann P (2020) Künstliche Intelligenz – Wirtschaft und Innovation. Studie im Auftrag des Ministeriums für Digitales für das Land Hessen. Wiesbaden. (im Erscheinen)

- Buxmann P, Schmidt H (Hrsg) (2020) Künstliche Intelligenz. Mit Algorithmen zum wirtschaftlichen Erfolg, 2. Aufl. Springer Gabler, Berlin
- Drucker PF (1991) Management. An abridged and revised version of management: tasks, responsibilities, practices. Butterworth-Heinemann, Oxford
- Eichhorst W (2015) Der Wandel der Erwerbsformen in Deutschland. IZA Standpunkte 78. Forschungsinstitut zur Zukunft der Arbeit (IZA), Bonn
- Frey CB, Osborne MA (2013) The future of employment: how susceptible are jobs to computerisation? Oxford Martin 114. <https://doi.org/10.1016/j.techfore.2016.08.019>.
- Frey CB, Osborne MA (2017) The future of employment: how susceptible are jobs to computerisation? Technol Forecast Soc Chang 114:254–280
- Hacker W (2018) Menschengerechtes Arbeiten in der digitalisierten Welt. Eine Wissenschaftliche Handreichung, 1. Aufl. Mensch – Technik – Organisation, Bd 49. vdf Hochschulverlag, Zürich
- Hirsch-Kreinsen H, Karačić A (2019) Autonome Systeme und Arbeit: Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt. transcript, Bielefeld
- Hoppe A (2010) Komplexe Technik-Hilfe oder Risiko? Darstellung ausgewählter Ergebnisse einer Grundlagenuntersuchung zu Technikstress! In: Brandt C (Hrsg) Mobile Arbeit – Gute Arbeit? Arbeitsqualität und Gestaltungsansätze bei mobiler Arbeit. Gewerkschaft ver.di, Berlin, S 53–64
- Köchling A (1985) *Bildschirmarbeit. Gesundheitsregeln u. Gesundheitsschutz*. Wissenschaft im Arbeitnehmerinteresse, Bd 3. Bund, Köln
- May TY, Korczynski M, Frenkel SJ (2002) Organizational and occupational commitment: knowledge workers in large corporations. J Manag Stud 39(6):775–801
- Moser K, Preising K, Göritz AS, Paul K (2002) Steigende Informationsflut am Arbeitsplatz: belastungsgünstiger Umgang mit elektronischen Medien (E-Mail, Internet), 1. Aufl. Bremerhaven: Wirtschaftsverlag NW Verlag für neue Wissenschaft GmbH, S 228. (Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin: Forschungsbericht, Fb 967) ISBN: 3-89701-900-0, Projektnummer: F1814
- Müller JC (2015) Arbeit – Macht – Sinn: Zur Entgrenzung von Arbeit im Wissenschaftsbetrieb. In: Krentel F, Barthel K, Brand S, Friedrich A, Hoffmann AR, Meneghello L, Müller JC, Wilke C (Hrsg) Library Life: Werkstätten kulturwissenschaftlichen Forschens. meson press, Lüneburg, S 37–75
- Mütze-Niewöhner S, Nitsch V (2020) Arbeitswelt 4.0. In: Frenz W (Hrsg) Handbuch Industrie 4.0: Recht, Technik, Gesellschaft. Springer, Berlin/Heidelberg, S 1187–1217
- Nerdinger FW, Blickle G, Schaper N (2014) Arbeits- und Organisationspsychologie, 3. Aufl. Springer-Lehrbuch/Springer, Berlin/Heidelberg
- OECD (2018) Putting faces to the jobs at risk of automation. Policy brief on the future of work. OECD Publishing, Paris. <http://www.oecd.org/els/emp/future-of-work/Automation-policy-brief-2018.pdf>. Zugriffen am 13.10.2020
- Rick V, Distelrath J, Nitsch V (2021) Digitalisierung und KI in der Wissenschaft: Auswertung der Begleitstudie zum Projekt „Digitale Arbeitswelten in Forschung und Entwicklung“. DOI 10.18154/RWTH-2021-01212 . <https://publications.rwth-aachen.de/record/811477>. Zugriffen am 07.02.2021
- SAE International (2016) Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. https://doi.org/10.4271/J3016_201609
- Schmidt H (2018) Wie Maschinen die Arbeit übernehmen. Netzökonom. <https://www.netzoekonom.de/2018/10/13/wie-maschinen-die-arbeit-uebernehmen/>. Zugriffen am 13.10.2020
- Simpson CW, Prusak L (1995) Troubles with information overload – moving from quantity to quality in information provision. Int J Inf Manage 15(6):413–425
- Webb M (2019) The impact of artificial intelligence on the labor market. <https://doi.org/10.2139/ssrn.3482150>. Zugriffen am 13.10.2020
- Wilkesmann U (2005) Die Organisation von Wissensarbeit. Berl J Soziol 15(1):55–72
- Willke H (1998) Organisierte Wissensarbeit. Z Soziol 27(3):161–177

- World Economic Forum (2018) The future of jobs report 2018. Insight report. World Economic Forum, Cologne/Geneva
- Zika G, Helmrich R, Maier T, Weber E, Wolter MI (2018) Arbeitsmarkteffekte der Digitalisierung bis 2035: Regionale Branchenstruktur spielt eine wichtige Rolle. IAB-Kurzbericht 9/2018. Nürnberg. <http://hdl.handle.net/10419/185844>. Zugegriffen am 13.10.2020

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.





Indra Spiecker genannt Döhmann

6.1 Künstliche Intelligenz und BigData im Spannungsfeld von Wissenschaftsfreiheit und Forschungsevaluation

Forschung an und mit Daten und auf der Basis von Informationen zur Wissensgewinnung ist allgegenwärtig, nicht erst seit Künstliche Intelligenz und Big Data (siehe dazu Kap. 1 und 2 in diesem Band)¹ mediale Aufmerksamkeit und Forschungsetats treiben. Wird mit Daten geforscht, stehen rechtlich vor allem Belange der

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. In besonderer Weise hat der Text Schärfungen erfahren durch die Überlegungen von Carl Friedrich Gethmann und C. Jan Schmidt. Danke gebührt außerdem meinem Mitarbeiter Ass. iur. Jeremy Stevens für Input zur allgemeinen Regulierung von Künstlicher Intelligenz und stud. iur. Sophie Danner für Unterstützung bei Formalia und Recherche.

¹Die Diskussion um die Abgrenzung der Begrifflichkeiten von Daten, Information und Wissen ist ubiquitär und weiterhin nicht abgeschlossen. In diesem Text werden im Wesentlichen Daten und Information als gleich verstanden. Am Einführungssatz wird aber deutlich, dass grundsätzlich ein Unterschied besteht: Daten sind die Basis von Informationen, die den Kontext miteinbeziehen; Wissen ist die gezielte Verwendung von Informationen, z. B. durch Rekombination. Vgl. dazu Spiecker genannt Döhmann 2010, S. 250 ff.

Rechtlich wird allenfalls zwischen Daten und Information begrifflich unterschieden, wobei z. B. die Europäische Datenschutzgrundverordnung als wesentliches Datenschutzrecht oder die Informationsfreiheitsgesetze eine Vermengung der Begriffe von Datum und Information praktizieren und beides erfassen, vgl. Hamer 2020, S. 40; Karg 2019, Art. 4 Nr. 1 Rn. 25.

I. Spiecker genannt Döhmann (✉)
Goethe Universität Frankfurt a.M., Frankfurt, Deutschland
E-Mail: spiecker@jur.uni-frankfurt.de

Privatheit und des Datenschutzes im weiteren Sinne im Raum, rechtlich abgesichert etwa durch Datenschutzrecht, Urheberrecht, Betriebs- und Geschäftsgeheimnisrecht oder in jüngerer Zeit auch Wettbewerbsrecht. Zudem treten – rechtlich bisher kaum erfasst – weitere Problembereiche wie die Qualitätssicherung, Archivierung und Zugänglichkeit von Datenbeständen, die Zugehörigkeit von Daten zu Forscher/innen bzw. deren Forschungsinstitutionen oder die Teilbarkeit und Übertragbarkeit von Rohdaten in das Sichtfeld regulatorischer Überlegungen. Eingebettet ist die Suche nach Informationen zudem fast immer in die Problematik der Herstellung, Darstellung und Kontrolle von privaten und staatlichen Entscheidungen unter Unsicherheit (siehe zu letzterem etwa Spiecker genannt Döhmann 2021). Denn Informationsgewinnung, wie sie im Mittelpunkt des Einsatzes von Künstlicher Intelligenz in der Forschung steht, verweist auf den damit in Aussicht gestellten erwarteten Erkenntnisgewinn und die (vermeintlich) besser informierte Entscheidungslage. Das aber ruft Fragen auf, wie mit Informationen verschiedener Quellen und verschiedenartiger Bearbeitungsdichte und erst recht verschiedenartiger Gewissheit umzugehen ist, gerade dann, wenn staatliche Entscheidungen auf dieser Basis getroffen werden. Regulatorische Eingriffe in die Gewinnung, Nutzung und Verbreitung von Daten und des Einsatzes von Künstlicher Intelligenz können also aus vielerlei Perspektiven geboten sein.

Blickt man, wie es dieser Beitrag tut, in besonderem Maße auf die Wissenschaftler/innen und ihre Nutzung von Daten und dies im besonderen Anwendungsfeld der Forschung mit und an Künstlicher Intelligenz, ist zusätzlich die Forschungsfreiheit als ein zentraler rechtlicher Belang zu beachten. Dieser wird bisher nur wenig berücksichtigt in der Auseinandersetzung mit Künstlicher Intelligenz; ein Forschungsfreiheitsgesetz oder wenigstens allgemeine Richtlinien, die verfassungsrechtliche Vorgaben konkretisierten und Kollisionen mit anderen Rechtspositionen klärten, fehlen sowohl auf nationaler als auch auf internationaler/europäischer Ebene und selbst Leitlinien sind derzeit Mangelware. Auch der Entwurf einer KI-Verordnung, wie sie die EU jüngst vorgelegt hat (Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence – Artificial Intelligence Act), COM/2021/206 final), konzentriert sich auf die Anwendungsfelder und bezieht die eigentliche Forschungsperspektive kaum dezidiert ein.

Methoden des Machine Learnings und der Künstlichen Intelligenz werden in einer Vielzahl von Forschungsdisziplinen genutzt. Sie ermöglichen die Verarbeitung großer, bisher häufig als unbewältigten eingeschätzter Datenmengen, um hierdurch ggf. neue und unerwartete Muster und Zusammenhänge zu finden. Hierzu zählt beispielsweise die Suche nach sowie die Prognose der Eigenschaften neuer Materialien/Werkstoffe oder Arzneimittel, aber ebenso die Analyse von archäologischen Fundstücken bis hin zur Vorhersage menschlichen Verhaltens, z. B. bei Wahlen oder der Partner/innensuche. Die spezifische Art und Weise der Nutzung dieser Technologien unterscheidet sich je nach Anwendungsfall. Zum Teil erleichtert die Automation bestimmte Arbeitsschritte, vor allem die Untersuchung einer deutlich größeren Zahl an Datensätzen mit der Entwicklung erster Hypothesen, die dann von den Forschern ausgewertet, weiter bearbeitet und verfeinert werden können. Zum Teil basieren die Forschungsergebnisse jedoch derart auf den verwendeten Algorithmen, dass eine wesentliche Forschungsleistung vom algorithmischen System übernommen wird. Hier zeigt sich beim Einsatz von Künstlicher Intelligenz häufig das Problem, dass die Reproduzierbarkeit der Forschung erschwert wird, weil ein exakt gleich aufgebautes algorithmisches System mit ähnlichen Trainings- und Entwicklungsdaten nicht gegeben ist.

Die Verbreitung des Einsatzes von Methoden der Künstlichen Intelligenz und von Big Data in der Wissenschaft führt dazu, dass sich auch die Einschätzung von Wissenschaft verändert. Darauf basierende Methodik kommt zu anderen, bereits existenten und anerkannten Methoden der Beurteilung, Begutachtung und Bewertung hinzu oder ersetzt sogar bereits bestehende Vorgehensweisen. Zudem verändern sich die Möglichkeiten und das Spektrum der Beurteiler: Öffentlichkeit, Gesellschaft, Wirtschaft, Verlage, Publikationsorgane, Politik, die eigenen Community oder Forschungsinstitutionen benötigen möglicherweise andere Herangehensweisen, um die Qualität und Bedeutung von Wissenschaft unter Bedingungen der Künstlichen Intelligenz überhaupt einschätzen zu können, erhalten aber möglicherweise gleichzeitig auch andere Grundlagen und Ausgangspunkte zur Bewertung.

Dieser erste allgemeine Befund lässt erkennen, dass sich Wissenschaftler/innen nicht nur mit der Methodik von Künstlicher Intelligenz selbst befassen müssen, um sie überhaupt korrekt anwenden und für ihre Zwecke nutzen zu können, sondern dass sie ihrerseits auch das Verhältnis des Einsatzes von Künstlicher Intelligenz/Big Data im Verhältnis zu anderer Forschungsmethodik begründen und möglicherweise auch rechtfertigen müssen. Dies gilt ebenso auch für den Fall, dass sie dies *nicht* tun. Die Methodenfrage wird also erweitert.

Diese Methodenfrage stellt sich zum einen im Internum der Forschung, also bei der Ausgestaltung des eigentlichen Forschungsansatzes. Gewinnt die Verfolgung einer Forschungsfrage in besonderer Weise durch den Einsatz von Künstlicher Intelligenz? Oder sind die Kosten-Nutzen-Relationen zu ungünstig, die Ergebnisse dadurch besonderen Belastungen ausgesetzt, die zu erwartende Unsicherheit und die Schwierigkeiten der angemessenen Darstellung und Rezeption der Ergebnisse womöglich noch größer?

Diese Methodenfrage stellt sich aber auch im Externum der Forschung, also bei der Beurteilung von Forschungsleistungen und Forschungsaussagen. Diese erfolgt typischerweise wissenschaftsintern, also beispielsweise bei der Beurteilung von Publikationen durch Schriftleiter/Redakteure oder im begutachtenden Peer Review Verfahren. Zudem aber ist Forschung auch zunehmend ein Beurteilungsfaktor über den Forschenden selbst, also z. B. bei der Beantragung und Bewilligung von Drittmitteln, bei der Bewertung für wissenschaftsrelevante Tätigkeiten oder aber bei der Beurteilung seiner selbst im Rahmen der modernen Bezügeausgestaltung unter Leistungsanreizen der ökonomisierten Hochschule. Überall hier gilt es den Forschenden, seine Forschungsleistung und seinen Forschungsansatz – unter Einsatz von Künstlicher Intelligenz oder gerade nicht – adäquat zu beurteilen. Die Entscheidung über die Nutzung oder Nicht-Nutzung von Künstlicher Intelligenz und deren konkrete Ausgestaltung wird zu einem neuen Beurteilungsfaktor von Wissenschaft.

6.2 Regulierungsansätze von Künstlicher Intelligenz

So allgegenwärtig wie Künstliche Intelligenz derzeit in Wissenschaft, Wirtschaft und Öffentlichkeit zu sein scheint, sind konkrete Regulierungsansätze für den Bereich von Forschung und Wissenschaft dennoch noch wenige zu konstatieren. Dieser Befund erstreckt sich auch auf die Lehre und Bildung: Die Vermittlung von Kenntnissen zu deren

Einsatz verläuft weitgehend ohne Vorgaben und fehlt in den meisten Ausbildungsprofilen und Studiengängen. Dagegen gibt es bereits eine Reihe von Empfehlungen, Leitlinien, Vorschlägen und Einschätzungen verschiedenster Fachgesellschaften, Kommissionen, Verbänden und Gremien zu Einsatz und Nutzung von Künstlicher Intelligenz im Allgemeinen bis hin zum weltweit ersten Entwurf einer EU-Verordnung für Künstliche Intelligenz (Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence – Artificial Intelligence Act), COM/2021/206 final). Es ist, auch wenn das bisher noch kaum gesehen wird, zu erwarten, dass mit dem verbreiteten Einsatz von Künstlicher Intelligenz in vielen Lebensbereichen auch der Forschungsbereich zunehmend stärkerer Beobachtung und in der Folge auch stärkerer Setzung von Leitplanken ausgesetzt sein wird. Daher wird hier der Rechtsrahmen solcher potenziellen Regulierungsanstrengungen speziell für die Wissenschaft untersucht.

Der Regulierung, also der Beschränkung von Wissenschaft und Forschung, sind im nationalen Recht durch Art. 5 Abs. 3 S. 1 Grundgesetz (GG) bzw. auf europarechtlicher Ebene Art. 13 Grundrechte-Charta (GRCh) enge Grenzen gesetzt (siehe dazu Abschn. 6.4.2). Die Einschränkung der Forschungsfreiheit bedarf daher einer überzeugenden Begründung unter Einbeziehung entgegenstehender, überwiegender verfassungsrechtlich abgesicherter Rechtspositionen. Typische Fälle anerkannter forschungsrelevanter Regulierung sind die allgemeinen Anforderungen an eine ethische Forschung, Vorgaben hinsichtlich der Nutzung bestimmter als gefährlich eingestufte Materialien und Stoffe sowie die Regulierung aus ethischen Gesichtspunkten, insbesondere im Hinblick auf Forschung am Menschen und genetischen Materialien.

Forschungsvorhaben können in bestimmten Fällen unmittelbar Risiken für das Leben bzw. andere Rechtsgüter Einzelner oder größeren Gruppen von Personen bedeuten. So können bei Experimenten mit Gefahrstoffen, wie bestimmten Chemikalien oder radioaktiven Stoffen, negative Auswirkungen für die Beteiligten, aber auch für Dritte, entstehen. Verschiedene Vorschriften regeln daher den Umgang mit diesen Stoffen, die auch auf Forschungsvorhaben Anwendung finden (z. B. die deutsche Verordnung zum Schutz vor Gefahrstoffen (Gefahrstoffverordnung – GefStoffV), die europäische REACH-Verordnung für Chemikalien (Verordnung (EG) Nr. 1907/2006 zur Registrierung, Bewertung, Zulassung und Beschränkung chemischer Stoffe (REACH)) oder das deutsche Chemikaliengesetz (Gesetz zum Schutz vor gefährlichen Stoffen (Chemikaliengesetz – ChemG)).

Im Hinblick auf die Forschung an Menschen und Tieren spielen ethische Gesichtspunkte eine besondere Rolle. Diese Vorstellungen sind häufig auch rechtlich im Schutz der Würde des Menschen, Art. 1 Abs. 1 GG, seiner Persönlichkeitsrechte, Art. 2 Abs. 1 in Verbindung mit Art. 1 Abs. 1 GG, des Schutzes von Leib und Leben, Art. 2 Abs. 2 GG, oder dem Tierschutz, Art. 20a GG, verankert. Der Schutz der Grundrechte sowie der Tierschutz erfordern insofern verfahrensbezogene sowie inhaltliche, materiell-rechtlich wirkende Vorgaben zum Ausgleich der verschiedenen Rechtsgüter. Dies wird z. T. durch Einschränkung bestimmter Forschung, etwa von Tierversuchen nur zu bestimmten Zwecken und nach Einhaltung strikter Vorgaben, oder durch Vorgaben zur Begleitung und zum Monitoring dieser Forschung, etwa im Bereich der Arzneimittelzulassung für klinische Versuche, erreicht. Vielfach sind zudem weitere Sicherungen vorgesehen, z. B. Ethikkommissionen, die Forschungsvor-

haben proaktiv beurteilen und kontinuierlich begleiten sollen. Daneben erfolgt die Regulierung von Forschung teilweise auch im Hinblick auf die möglichen Folgen und Implikationen der Forschung. Dies betrifft beispielsweise die Forschung zur Schaffung neuer Waffen (z. B. Biowaffenkonvention (Gemeinsame Aktion 2006/184/GASP des Rates vom 27. Februar 2006 zur Unterstützung des Übereinkommens über das Verbot von biologischen Waffen und Toxinwaffen (BWÜ) im Rahmen der Strategie der EU gegen die Verbreitung von Massenvernichtungswaffen), die Forschung im Bereich der Präimplantationsdiagnostik (Gesetz zur Regelung der Präimplantationsdiagnostik (Präimplantationsdiagnostikgesetz – PräimpG)) oder die Forschung an bestimmten Forschungsgegenständen, etwa Stammzellen (Gesetz zur Sicherstellung des Embryonenschutzes im Zusammenhang mit Einfuhr und Verwendung menschlicher embryonaler Stammzellen (Stammzellgesetz – StZG)).

Die Notwendigkeit einer Regulierung von Forschung mit Einsatz von Künstlicher Intelligenz wird gegenwärtig überwiegend dahingehend diskutiert, dass der Einsatz solcher algorithmischer Systeme häufig zu Nachweis- und Reproduktionsproblemen führt: Nur ein exakt identischer Datensatz und damit die Kenntnis der verwendeten Daten ließe überhaupt eine genaue Reproduzierbarkeit und damit eine Kontrolle durch andere Forscher/innen erwarten; häufig sind die Ergebnisse aber aufgrund des Machine Learnings nicht nachvollziehbar, nicht transparent und nicht begründbar (Roßnagel 2019a, b, Art. 5 Rn. 148; Gausling 2019, S. 335 f.). Es besteht das Risiko, dass – insbesondere bei empirischer Forschung – die ursprüngliche Studie eventuell bloße Zufallsartefakte gemessen hat und die Ergebnisse daher keine Aussagekraft haben (siehe Huston 2018; Stewart 2018). Daher ist zu beobachten, dass – statt einer Restriktion solcher Forschung – zunächst eine Förderung von Forschung gefordert wird (und auch erfolgt) mit dem Ziel, erklärbare bzw. kontrollierbare Künstliche Intelligenz zu schaffen (High Level Expert Group 2019, S. 3). Dies soll jedenfalls zum Teil mit Mitteln der Künstlichen Intelligenz selbst erfolgen.

Ein weiterer Fokus der Debatte um den Einsatz Künstlicher Intelligenz liegt auf dem Einsatz in der Beurteilung von Menschen und den daraus resultierenden Problemen durch Manipulation, Dark Patterns, fehlende Begründbarkeit der Entscheidung und mögliche Diskriminierung (siehe z. B. Datenethikkommission 2019, S. 97 ff.; Spiecker genannt Döhmman 2020; Ruschemeier 2020). Weiterreichender sind Besorgnisse, dass die Entscheidungen auf der Basis von Künstlicher Intelligenz aufgrund unklarer Entscheidungsregeln und unklarer Dynamik normativ weitere nicht wünschenswerte Auswertungen und Ergebnisse produzieren, z. B. durch die Wiederholung und Verfestigung von Vorurteilen und Diskriminierungen (Sachverständigenkommission für den Dritten Gleichstellungsbericht der Bundesregierung 2021, S. 137; Fröhlich und Spiecker genannt Döhmman 2018), aber auch durch inhärente Vorgaben der Normativität (Sommerer 2020, S. 101 ff.). Gelegentlich rückt auch die Datenqualität und ihre -sicherung in den Mittelpunkt von Überlegungen zur Qualität Künstlicher Intelligenz (Stevens 2020, S. 74 ff.; Hoeren 2016, S. 9 ff.). Diese Aspekte betreffen allerdings zumeist weniger die Forschung mit und an Künstlicher Intelligenz als den Einsatz dieser Verfahren in Realbereichen.

Ein kohärentes Framework zum Umgang mit dem Einsatz von Künstlicher Intelligenz in der Forschung ist in der Konsequenz daher nicht bekannt. Nur hinsichtlich einiger Aspekte, wie eben der Reproduzierbarkeit oder Transparenz, beginnend mit

Diskriminierungen, existiert eine breitere und kenntnisreiche Diskussion, an die angeknüpft werden kann.² In weiten Bereichen adressiert diese allerdings weniger gezielt die Forschung, als dass sie generell Probleme im Umgang mit Künstlicher Intelligenz aufruft. Lösungsansätze sind überwiegend dadurch gekennzeichnet, dass durch Investitionen in Forschung und Entwicklung technische Lösungen für identifizierte spezielle Probleme des Einsatzes von Künstlicher Intelligenz gefunden werden sollen. Überlegungen zur Nutzung von Machine Learning zu Forschungszwecken und den Auswirkungen auf die Forschung selbst werden in diesem Szenario meistens nicht oder lediglich kursorisch thematisiert, selbst wenn Künstliche Intelligenz im Arbeitskontext analysiert wird.³

Insgesamt zeigen sich daher zwar erste Ansätze, warum der Einsatz von Künstlicher Intelligenz reguliert werden sollte und wie das geschehen kann, diese sind allerdings zumeist nicht forschungsspezifisch. In diesem Sinne bildet das Positionspapier des Wissenschaftsrats vom Herbst 2020 „Zum Wandel in den Wissenschaften durch datenintensive Forschung“ eine Ausnahme und einen ersten großen Schritt. Allerdings werden weniger Regulierungsansätze entwickelt, sondern zunächst einmal das Problemfeld sondiert und – häufig gerade auch zugunsten einer rechtssicheren Forschung mit Künstlicher Intelligenz – eher in bestimmten Bereichen das Fehlen von Regulierungsvorgaben und klaren rechtlichen Regeln moniert (WissRat 2020, S. 30 ff.).

6.3 Der Einsatz Künstlicher Intelligenz im allgemeinen Spannungsfeld von Wissenschaftsfreiheit und Datenschutz: Zwei Konstellationen

Künstliche Intelligenz – wie jedes algorithmische System – verarbeitet Daten zur Vorbereitung bzw. Durchführung von Entscheidungen. Damit können Einzelentscheidungen erfasst sein oder auch Teilschritte einer Gesamtentscheidung. Obliegt in letzterem Fall die finale Entscheidung bei einem Menschen oder einer menschlichen Institution, z. B. dem Staat oder einem Unternehmen, ist der Einsatz Künstlicher Intelligenz ein Mittel zur Assistenz menschlichen Entscheidens. In dieser stellen sich ihrerseits wiederum eigene Probleme nach dem Verhältnis von maschinellem und individuellem Entscheiden und der Verantwortung zwischen beiden

²Nicht näher eingegangen wird in diesem Kontext auf die Forschung an ethisch problematischen bzw. riskanten Anwendungsfällen, insbesondere autonome Waffensysteme (Lethal Autonomous Weapon Systems (LAWS)), siehe hierzu beispielsweise die 2012 gegründete Campaign to Stop Killer Robots sowie die Entschließung des Europäischen Parlaments 2018/2752(RSP); rechtswissenschaftliche Behandlung beispielsweise bei Dederer 2018, S. 380 ff.; Rauch 2020, S. 5 ff.

³So finden sich bei der *High Level Expert Group* nur Verweise auf die Relevanz von Forschung, aber keine konkreten Überlegungen zu Differenzierungen. Anders aber *Wissenschaftsrat*, Zum Wandel in den Wissenschaften durch datenintensive Forschung. Positionspapier, 2020, der eine spezifischere Perspektive einnimmt (WissRat 2020, S. 26 ff.). Ähnlich wenig aussagekräftig auch der Dritter Gleichstellungsbericht 2021, S. 99.

Systemen. Zudem wird hier virulent, dass maschinelles Entscheiden oftmals den Eindruck von (höherer) Kompetenz, Objektivität und Unvoreingenommenheit weckt und somit der menschliche Entscheider oftmals in Rechtfertigungslasten gerät, warum er von der – vorbereitenden – maschinellen Entscheidung abweicht. Diese Fragestellungen sind indes nicht spezifisch für den Einsatz von algorithmischen Systemen in der Forschung und für die Forschung.

Es sind zwei Konstellationen vorstellbar, in denen die Forschung an und mit Daten bei Künstlicher Intelligenz einer rechtlichen Regulierung unterworfen wird. Die erste und offenkundige Variante ist die Forschung an Daten durch Wissenschaftler. Der Zweck des Datenumgangs ist dann die Forschung selbst, die Daten sind Mittel zum forschersichen Zweck. Regulierung bemisst sich dann anhand der Forschung und ihrer Ziele; die Forschungsfreiheit widersetzt sich dem Bemühen um Einhegung und Eingrenzung. Der folgende Beitrag befasst sich mit diesem Problemfeld.

Der andere Fall entsteht dann, wenn Künstliche Intelligenz eingesetzt wird, um den Wissenschaftler zum Gegenstand von deren Anwendung zu machen, also z. B. zu Zwecken der Evaluation oder sonstiger Beurteilung und Steuerung von Forschungsinhalten, Publikationserfolgen oder zur Budgetierung. Der Zweck des Datenumgangs ist dann oftmals ein institutioneller Zweck – die Hochschule, welcher der Wissenschaftler/in angehört, möchte seine Tätigkeit vergleichbar machen, eine Governance effektiv gestalten, Schwerpunkte identifizieren und ausweisen, Forschungsleistungen und Rezeption beurteilen, die Nachfrage nach Forschungsinhalten und Forschenden näher bestimmen und vielleicht sogar die Arbeitszeit und -verteilung, etwa auf Lehre und Forschung, kontrollieren.

Auch hier bestimmt sich der anwendbare Rechtsrahmen vor allem aus diesen Zwecken der Forschungsinstitution. Der Rechtsrahmen entstammt daher vor allem dem Hochschul- und Forschungsrecht. Dieses als allgemeines Recht setzt (oft genug allerdings keine) Grenzen, weil es – vieles in die Regelungshoheit der Hochschulen selbst belassend – spezifische Regelungen für diesen Bereich zumeist nicht kennt. Daher muss zur Identifikation von Bindungen der Forschungsinstitutionen auf noch allgemeinere Regelungen zurückgegriffen werden. Diese bilden aber oftmals das hochschul- bzw. forschungsinstitutionell-spezifische Umfeld nicht ab. Da Künstliche Intelligenz auf Daten über Wissenschaftler/innen basiert und die daraus gewonnenen Erkenntnisse auf diese übertragen werden, ist in aller Regel das allgemeine Datenschutzrecht anwendbar.

Damit ergibt sich eine interessante multipolare Gemengelage, die der anderen Situation des Einsatzes von Künstlicher Intelligenz nicht entspricht. Nunmehr streitet die Forschungsfreiheit des Wissenschaftlers oftmals auf Seiten des Datenschutzrechts, während die gleichzeitig bestehende Forschungsfreiheit der Institution entgegengesetzte Interessen unterstützt. Für den Bereich der Publikation kommt dann noch möglicherweise die Pressefreiheit der Publikationsorgane als weiteres zu berücksichtigendes rechtliches Interesse hinzu.

Da in diesem Bereich Künstliche Intelligenz und Big Data bisher wenig eingesetzt werden, nicht zuletzt auch wegen der geringen Standardisierungsmöglichkeiten, soll diesem Themenfeld aber hier keine größere Bedeutung zugewiesen werden. Es

ist allerdings nicht ausgeschlossen, dass Anwendungsfelder für Problembereiche warten, sobald Künstliche Intelligenz auch hier Einzug hält: Man denke nur an die Beurteilung der Reichweite von Veröffentlichungen oder auch schon vorverlagert, die Abfassung und Kontrolle von Publikationen mit Hilfe von Künstlicher Intelligenz.

6.4 Der Einsatz Künstlicher Intelligenz als Mittel der Forschung

6.4.1 Der grundsätzliche Konflikt zwischen Datenschutzrecht und Forschungsfreiheit

Der anwendbare Rechtsrahmen für die Forschung mit Daten im Rahmen des Einsatzes Künstlicher Intelligenz bestimmt sich also zum einen aus der Verwendung und Zielrichtung dieser Daten heraus. Der Zweck bestimmt den Rechtsrahmen, etwa wenn die Daten einem Forschungszweck zugewendet werden sollen, der untersagt oder beschränkt ist. Zum anderen aber gibt es auch rechtliche Regelungen, die bereits an der Verfügbarkeit der Daten selbst ansetzen und daher nicht erst ihre Verwendung als regulatorischen Ansatzpunkt wählen. Sämtliche Rechtsregime, die Daten schützen, verfolgen diesen Ansatz, bei weitem nicht nur das Datenschutzrecht, das sich auf die personenbezogenen Daten bezieht. So werden möglicherweise auch Daten verwendet, die einem Betriebs- und Geschäftsgeheimnis unterliegen oder es werden Daten herangezogen, die aus anderen Gründen nur beschränkt verfügbar sein sollen.

Wissenschaft erfährt beim Einsatz von Künstlicher Intelligenz allerdings nicht nur Beschränkungen. Vielmehr stehen den Forschenden auch Rechte zu, allen voran die Wissenschaftsfreiheit. Darauf kann sich unter bestimmten Umständen auch die Institution, für die er oder sie forscht, berufen. Der Einsatz von Daten, wie er bei Künstlicher Intelligenz unentbehrlich ist, kann also gerade auch von Freiheitsrechten besonders geschützt sein und einen besonderen Freiraum auch rechtlich beanspruchen.

Im Folgenden wird der daraus entstehende Konflikt näher betrachtet, nämlich der – mögliche, nicht zwingende – Gegensatz zwischen Wissenschaftsfreiheit einerseits und Datenschutzrecht andererseits. Dabei werden die Konsequenzen für diejenigen betrachtet, deren Daten für wissenschaftliche Zwecke verwendet werden. Die spätere Übertragung der Erkenntnisse auf einzelne Personen, also die Anwendung der Forschungsergebnisse unter Einsatz von Künstlicher Intelligenz, soll nicht näher untersucht werden; auch dort stellen sich aber ähnliche Probleme und kann sich der Konflikt in ähnlicher Weise auftun.

Dieses Themenfeld, der Konflikt zwischen Datenschutzrecht und Wissenschaftsfreiheit, ist gegenwärtig eines der großen ungelösten Probleme beim Einsatz von Künstlicher Intelligenz und Big Data. Dies liegt nicht nur, aber ganz besonders

daran, dass die Zweckbestimmung wissenschaftlicher Arbeit mit Daten nicht immer klar zu bestimmen ist und dies teilweise auch explizit nicht gewünscht wird.

6.4.2 Wissenschaftsfreiheit

Die Wissenschaftsfreiheit ist in Europa umfänglich bekannt. Sie wird auf verfassungsrechtlicher Ebene geschützt. So formuliert etwa Art. 13 Abs. 1 EU-GR-Charta: „[...] und Forschung sind frei.“ Als Parallelnorm dazu sieht Art. 5 Abs. 3 GG für das nationale Recht vor: „[...] und Wissenschaft, Forschung und Lehre sind frei.“ Eine Besonderheit der Forschungsfreiheit liegt darin, dass sie ohne sog. Schranken gewährleistet ist und insoweit als sog. vorbehaltloses Grundrecht einen herausgehobenen Stellenwert einnimmt (Caspar 2019, Art. 98 Rn. 3).

Grundsätzlich verdrängt das EU-Recht nationales Recht, so dass vorrangig das Verständnis von Forschung und Wissenschaft des EU-Rechts auf die Beurteilung anzuwenden wäre. Allerdings sieht Art. 4 Abs. 3 AEUV Besonderheiten für den Gegenstand der „Forschung“ vor; hier existiert der Sonderfall, dass sich die Zuständigkeit der EU auf Unterstützungs-, Koordinierungs- und Ergänzungsmaßnahmen erstreckt (Calliess 2016, Art. 4 AEUV Rn. 21 unter Verweis auf Nettesheim 2004, S. 530). Daher ist das nationale Verständnis von Forschungsfreiheit in vielen Bereichen gleichwohl von Bedeutung.

Die Forschungsfreiheit erstreckt sich auf alle Elemente des forschersischen Tuns, also auch auf die angewandten Methoden. Dies schließt Künstliche Intelligenz und den Einsatz von Big Data grundsätzlich ein. Wie weit dieser Schutz aber konkret reicht, wird in der Abwägung mit dem Datenschutzrecht noch zu konkretisieren sein (siehe dazu Abschn. 6.4.4).

6.4.3 Datenschutzrecht

Hintergrund und Zielrichtung des Schutzes

Das Datenschutzrecht hält derzeit – neben dem Urheberrecht – wohl das umfassendste Informationsregulierungsrecht vor. Gleichzeitig ist es von der Schutzrichtung her wesentlich komplexer. Es schützt – aus nationalem Verständnis – die „informationelle Selbstbestimmung“ (Hornung und Spiecker genannt Döhmman 2019, Einleitung, Rn. 237 ff.), im internationalen Kontext undeutlicher „Privatheit“. Dahinter verbergen sich allerdings vielschichtige Konzepte, die den spezifischen Gefahren automatisierter Datenverarbeitung begegnen und somit die Verbindung von Daten und darauf basierenden Entscheidungen in den Blick nehmen. Daher beinhaltet Datenschutz auch Diskriminierungsschutz und Schutz vor Intransparenz (Hornung und Spiecker genannt Döhmman 2019, Einleitung, Rn. 10; Kühling und Buchner 2018, Art. 1 Rn. 14).

Während die Wissenschaftsfreiheit ein eher zurückhaltendes Dasein in der rechtlichen Beachtung fristet, hat das Datenschutzrecht in den letzten Jahren erheblich an Gewicht zugelegt. Das hängt zum einen damit zusammen, dass zunehmend erkannt worden ist, dass Digitalisierung durchaus Nebenwirkungen und unerwünschte Konsequenzen zeitigen kann. Zum anderen aber wirkt sich aus, dass gerade auf europäischer Ebene das Datenschutzrecht von einem „zahnlosen Tiger“ zu einem „scharfen Schwert“ gewandelt worden ist. Dies ist vor allem dem EuGH und der Datenschutzgrundverordnung (DSGVO) geschuldet. Beide beziehen sich dabei auf wesentliche Gewährleistungen im Europäischen Primärrecht, insbesondere Art. 7 EU-GR-Charta, wonach „Jede Person [...] ein Recht auf Achtung ihres Privat [...] lebens, ihrer Wohnung sowie ihrer Kommunikation“ hat, sowie Art. 8 Abs. 1 EU-GR-Charta, der „Jede Person hat ein Recht auf Schutz der sie betreffenden personenbezogenen Daten“ regelt.

Hinter dem Datenschutzrecht, das seit dem Aufkommen automatisierter Verarbeitungen in den 1960er-Jahren entwickelt wurde und u. a. in Hessen 1970 seine erste gesetzliche Fassung erhalten hat, stehen diverse Anliegen. Das Rechtsgebiet will verhindern, dass es aufgrund der Digitalisierungstechnik zu informationellen Machtasymmetrien kommt. Diese Gefahr wird u. a. dadurch gesehen, dass bei einer intensiven Datenverarbeitung Entscheidungen des Einzelnen dadurch vorherbestimmt werden können und somit der Einzelne Macht über die Umstände, Voraussetzungen und Folgen seiner persönlichen Entscheidungen verliert. In der Folge kommt es dann zu einem Freiheitsverlust und einer Selbstbeschränkung, weil der Einzelne von seinen Freiheiten keinen Gebrauch mehr macht, da er Sorge um die Konsequenzen aus der Beobachtung hat (Simitis et al. 2019, Einleitung, Rn. 12 ff.). Diese Grundzüge hat für das nationale Recht 1983 das Bundesverfassungsgericht im Volkszählungsurteil formuliert (BVerfGE 65, 1 (43)) und damit eine allgemeingültige Umschreibung gefunden. Auch der EuGH hat sich in verschiedenen Entscheidungen in ähnlicher Weise zu den Grundanliegen des Datenschutzes geäußert (EuGH, Urt. v. 21.12.2016 – C-203/15, C-698/1). Auf eine Kurzformel gebracht, bildet das Datenschutzrecht als ein Persönlichkeitsschutzrecht das Rückgrat von Demokratie, Individualität, Würde und Autonomie und damit den Eckpfeilern der freiheitlich-demokratischen Grundordnung (Spiecker genannt Döhmann 2018).

Die DSGVO hat in der Umsetzung dieser Anliegen eine Vielzahl an spezifischen Anforderungen für Datenverarbeiter aufgebracht; im Kern ist sie eine Fortführung des bisher geltenden Rechts und seines Regulierungsansatzes (Hornung und Spiecker genannt Döhmann 2019, Einleitung, Rn. 212; Paal und Pauly 2021, Einleitung, Rn. 7). Dieser sieht einige Kernanforderungen an Datenumgang vor, zu denen u. a. das Erfordernis einer Rechtsgrundlage für jegliche Datenverarbeitung (Gesetz oder Einwilligung) und ein insgesamt präventives Grundkonzept gehört.

Auf den Einsatz Künstlicher Intelligenz ist dies grundsätzlich unmittelbar anwendbar, weil die EU-Datenschutzgrundverordnung (Verordnung (EU) 2016/679 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung – (DSGVO)) jegliche Verarbeitung von personenbezogenen Daten nach Art. 4 Nr. 1 DSGVO ihrer Anwendung unterwirft. Wenn also personen-

bezogene Daten verarbeitet werden, sind erhebliche rechtliche Vorgaben zu beachten, wenn mit Künstlicher Intelligenz Daten verarbeitet werden. In der Regel knüpfen diese Vorgaben aber nicht an die Verwendung Künstlicher Intelligenz, sondern an die Verwendung von personenbezogenen Daten an (Karg 2019, Art. 4 Nr. 1 Rn. 3, 19; Ernst 2021, Art. 4 Rn. 20).

Personenbezogene Daten

Entsprechend umstritten ist der Begriff trotz der Legaldefinition des Art. 3 Nr. 1 DSGVO. Danach sind personenbezogene Daten „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann.“

Von wesentlicher Bedeutung ist also die Personenbeziehbarkeit, also ob die Verbindung zwischen einem scheinbar sachlichen Datum und einer Person hergestellt werden kann. Der EuGH hat den schon unter der Vorgängernorm des Art. 2 lit. a Datenschutz-Richtlinie (DSRL) (Richtlinie 95/46/EG zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr) schwelenden Streit darüber, auf wessen Zusatzwissen es für die Herstellung der Personenbeziehbarkeit bzw. Identifizierbarkeit ankommt (Karg 2019, Art. 4 Nr. 1 Rn. 6), inzwischen entschieden. Danach ist ein objektivierender Begriff zugrunde zu legen, d. h. grundsätzlich kommt es darauf an, dass überhaupt eine Verbindung hergestellt werden kann. Diese Grundüberlegung erfährt allerdings Einschränkungen. So ist u. a. einzubeziehen, wenn ein Dritter eine solche Verbindung herstellen kann (und nicht allein der Datenverarbeiter – dann handelt es sich um den subjektiven Begriff). Dies soll allerdings nur gelten, wenn diese Verbindung legal hergestellt werden kann (EuGH, Urt. v. 19.10.2016 – C-582/14).

Der Leitplankenansatz der DSGVO – insbesondere die Zweckbindung der Erhebung „sensibler“ Forschungsdaten

Ein zentrales Prinzip des Datenschutzrechts mit besonderer Bedeutung für den Einsatz künstlicher Intelligenz in der Forschung ist der Grundsatz der Zweckbindung, Art. 5 Abs. 1 lit. b) DSGVO. Dieser Grundsatz verlangt, dass jegliche Datenverarbeitung personenbezogener Daten für „festgelegte, eindeutige und legitime Zwecke“ erfolgen muss. Dieses Gebot erstreckt sich nicht nur auf die Datenerhebung selbst, sondern auch auf jegliche Weiterverarbeitung. Aufgegriffen wird damit eine Vorgabe aus dem primär-, d. h. verfassungsrechtlichen Gebot des Art. 8 Abs. 2 S. 1 EU-GR-Charta, wonach Daten nur „für festgelegte Zwecke“ verarbeitet werden dürfen (Roßnagel 2019a, b, Art. 5 Rn. 64).

Der Zweckbindungsgrundsatz ist also eine zentrale Einschränkung der Verarbeitung von Daten und beschränkt damit auch den Zugriff und die Weiterver-

arbeitung. Gleichzeitig wird damit auch die Verwendungsdauer reguliert, denn mit Zweckerreichung erlischt auch der Zweck und damit die Befugnis zur Verarbeitung (vgl. Roßnagel 2019a, b, Art. 5 Rn. 95). Das im nicht-europäischen Rechtskreis gelegentlich vertretene Prinzip, wonach man alle Daten nutzen kann, die man einmal auf rechtmäßigem Wege erhalten hat, (vgl. hierzu EuGH, Urt. v. 16.07.2020 – C-311/18) gilt also im europäischen Rechtskreis unter Geltung der DSGVO und der GR-Charta gerade nicht. Denn der Verarbeiter ist gefordert, für die konkrete Verarbeitung eine Zweckbestimmung zu formulieren. Da als Verarbeitung auch die Vorhaltung und Speicherung (Art. 4 Nr. 2 DSGVO) für noch nicht näher konkretisierte Zwecke gilt, ist eine Vorratshaltung von Daten grundsätzlich nicht legitim. Das hat auch der EuGH bereits mehrfach, u. a. für die sog. Vorratsdatenspeicherung, so entschieden (EuGH, Urt. v. 08.04.2014 – C-293/12, C-594/12).

Was allerdings als Beschränkung formuliert ist, ist gleichzeitig auch eine wesentliche Grundlage für die Sicherheit der Verarbeitung für den Datenverarbeiter. Denn der Verarbeiter gewinnt damit auch das Vertrauen desjenigen, dessen Daten verarbeitet werden. Der Zweckbindungsgrundsatz sorgt dafür, dass für den Betroffenen, das sog. Datensubjekt, vorhersehbar ist, was mit den Daten passiert; dies wiederum steigert die Bereitschaft, Daten preiszugeben und bestimmten Verarbeitungen zugänglich zu machen. Diese Funktionalität des Zweckbindungsgrundsatzes wird häufig verkannt.

Dreh- und Angelpunkt legitimer Datenverarbeitung ist damit der jeweilige und konkrete Zweck, der von der DSGVO und dem Datenschutzrecht allgemein als Beschränkung der Datenverarbeitung konstruiert ist. Ein Zweck, der so weit gefasst ist, dass er faktisch keine Begrenzungen vorsieht, kann also dem Zweckbindungsgrundsatz von vorneherein nicht genügen (vgl. Roßnagel 2019a, b, Art. 5 Rn. 70). Je nachdem, wie weit der Zweck gefasst ist, können mehr oder weniger Datenverarbeitungen vereinbar sein. Je präziser ein Zweck gefasst ist, umso mehr Datenverarbeitungen schließt er aus. Gleichzeitig steigt aber die Wahrscheinlichkeit, dass mit einer präzisen Zweckbestimmung die Zulässigkeit der Datenverarbeitung gegeben ist. Denn in diesem Fall ist der Maßstab der Beurteilung eher zu fassen, während Zweifel zu Lasten des Datenverarbeiters gehen. Grundsätzlich gilt, dass der Zweck umso konkreter gefasst sein muss, je schwerer der mit der Datenverarbeitung erfolgende Eingriff in die Rechte des Datensubjekts wiegt (Roßnagel 2019a, b, Art. 5 Rn. 71; Schantz 2020, Art. 5 Rn. 15). Dies entspricht dem Grundsatz der Verhältnismäßigkeit.

Daher ist als erste Schlussfolgerung für die Verwendung von Daten unter Einsatz künstlicher Intelligenz in der Forschung festzuhalten, dass die Verwendung von personenbezogenen Daten zu allgemeinen Zwecken der „Forschung“ den Anforderungen aus Art. 5 Abs. 1 lit. b) DSGVO nicht genügend kann. Dazu bedürfte es einer näheren Spezifikation, um welche Art von Forschung, durch wen, unter Beteiligung welcher weiteren Personen/Institutionen und welchen Weiterverwertungsstrategien es sich handelt. Umgekehrt lässt sich aber auch nicht abstrakt feststellen, wie hoch die jeweiligen Anforderungen an die Zweckbeschreibung sein müssen. Denn Forschungsanliegen unterscheiden sich: Eine Krebsforschung unter Einbeziehung genetischer Faktoren stellt einen schwerwiegenderen Eingriff in die

Rechte des Datensubjekts dar, als eine Forschung, die sich mit der Häufigkeit des Zahnarztbesuchs befasst und unterliegt entsprechend anderen Anforderungen.

Für die Verwendung von Daten im Umfeld von Künstlicher Intelligenz bedeutet der Zweckbindungsgrundsatz jedenfalls dann eine erhebliche Einschränkung für die Forschenden, wenn Daten zweckfrei verwendet und vorgehalten werden sollen. Eine Auswertung ohne Hypothesen oder vorherige Zielrichtung ist mit der DSGVO nicht vereinbar.

Typisch allerdings dürfte es im Bereich der Forschung sein, eine Vielzahl von Daten speichern zu wollen, ohne den späteren Verwendungszweck bereits zum Zeitpunkt und während des Verlaufs der Speicherung festlegen zu können. Dahinter steht der Gedanke, einmal verfügbare und einmal ermittelte Daten nicht wieder preisgeben bzw. löschen zu wollen, u. a. aus Effizienzerwägungen oder auch, weil diese Daten in der Zukunft nicht ohne weiteres replizierbar sein könnten. Ein solches Vorgehen ist mit der DSGVO nicht vereinbar, denn als bestimmter Zweck genügt es nicht, eine mögliche spätere Verwendung in einem noch unbekannten Kontext zu noch unbestimmten Zwecken anzugeben (Roßnagel 2019a, b, Art. 5 Rn. 72; Schantz 2020, Art. 5 Rn. 15 f.). Dies wäre eine zweckfreie Bevorratung von Daten, die den verfassungsrechtlichen Grundsätzen widerspricht.

Diese Einschränkung betrifft allerdings nur einen engen Teilbereich von Künstlicher Intelligenz, der zudem weniger den eigentlichen Vorgang der Auswertung erfasst als Vorbereitungshandlungen. Zudem wird davon eine wissenschaftlich umstrittene Vorgehensweise mittels Künstlicher Intelligenz, nämlich ohne Hypothese Zufallsergebnisse produzieren zu lassen, erfasst. Jedenfalls die letztere Einsatzmöglichkeit dieser speziellen Technologie kann im Hinblick auf ihre Schutzwürdigkeit durchaus kritisch betrachtet werden, weil von einem planmäßigen Vorgehen zu- meist wohl kaum die Rede sein kann.

Es ist aber zu betonen, dass die DSGVO einer Verwendung von Daten im Rahmen von selbstlernenden oder optimierenden Prozessen nicht prinzipiell entgegensteht – im Gegenteil. Das Datenschutzrecht ist – anders als gelegentlich behauptet wird – kein Verhinderungsrecht. Es ist vielmehr ein vielseitiges Regulierungsrecht, das, wie andere Regulierungsrechte auch, verschiedene Interessen einschließlich wirtschaftlicher Interessen in einen Ausgleich zu bringen versucht (Hornung und Spiecker genannt Döhmman 2019, Einleitung, Rn. 249).

Informationstechniken wie Künstliche Intelligenz müssen bei Verwendung personenbezogener Daten die vielfältigen Datenspuren der Nutzer/innen zusammenführen und Auswertungen von Verhalten erstellen können, um den Anforderungen an eine funktionsfähige (und möglicherweise auch kontrollierbare) Künstliche Intelligenz überhaupt genügen zu können.

Es gilt dann aber sehr genaue Strukturierungen und prozessuale Sicherungen in Ergänzung zu weiteren inhaltlichen Unterscheidungskriterien vorzusehen, um sicherzustellen, dass die Vorteile von neuen Technologien wie Künstlicher Intelligenz genutzt werden können, ohne aber die Befürchtungen des Datenschutzes vor Kontrollverlust und Aufgabe der Individualität wahr werden zu lassen (vgl. in diese Richtung auch Roßnagel 2019a, b, Art. 5 Rn. 112 f.). Dazu gehört gerade auch, Daten im Rahmen von Künstlicher Intelligenz nicht für jegliche Zwecke zur Ver-

wendung freizugeben. Vielmehr muss der Einsatz der Technologie als ein Mittel zur Erreichung eines Zwecks, für den weitere Datenerhebungen und -verwendungen erforderlich sind, verstanden und ausgestaltet werden. Dies kann dann dazu führen, dass zur Einhaltung des Zweckbindungsgrundsatzes die Zwecke der Datenverarbeitung als Sub-Zwecke einer funktionierenden Künstlichen Intelligenz in einem bestimmten Einsatzgebiet zu bestimmen sind.

Künstliche Intelligenz als Datenverarbeiter und Verantwortlicher i. S. v. Art. 26 DSGVO?

Die DSGVO ist, wie schon ihre Vorgängerin, die DSRL, davon geprägt, beteiligten Akteuren klare Verantwortlichkeiten zuzuschreiben. Dazu gehört auch, dass derjenige, der die Datenverarbeitung vornimmt, als Verantwortlicher einer Vielzahl von Verpflichtungen ausgesetzt ist. Darunter fallen sowohl primäre Pflichten, etwa die Einhaltung der Vorgaben des Art. 5 Abs. 1 DSGVO oder des Art. 6 Abs. 1 DSGVO zur Legitimität von Datenverarbeitungen. Zu beachten sind aber auch sekundäre Verpflichtungen, wie diejenige zu Schadensersatz nach Art. 82 DSGVO.

Angesichts der Möglichkeiten von Künstlicher Intelligenz lässt sich in Frage stellen, ob sich diese Zuschreibung klarer Verantwortlichkeiten aufrechterhalten lässt: Kann möglicherweise die Eigenständigkeit der Forschung, die mittels Künstlicher Intelligenz vorangetrieben wird, soweit reichen, dass sich die Verantwortlichkeit des Wissenschaftlers als Datenverarbeiter dahinter zurückzieht? Dies würde den Forschenden erheblich entlasten: Nicht er wäre verpflichtet, das Datenschutzrecht zu beachten, sondern die Künstliche Intelligenz, und bei Verstößen haftete nicht er, sondern die Künstliche Intelligenz.

Die Antwort auf diese Frage muss sich zunächst damit auseinandersetzen, ob man einem algorithmischen System eine Rechtspersönlichkeit zubilligen könnte. Denn nur in diesem Falle ließe sich eine Künstliche Intelligenz mit Rechten und – hier von besonderem Interesse – mit Pflichten denken. Obwohl diese Ansicht gelegentlich vertreten wird (z. B. Specht und Herold 2018, S. 43 f.; Schimer 2016, S. 664), ist nicht erkennbar, dass die Gründe, die gegen eine solche Annahme sprechen (siehe z. B. Datenethikkommission 2019, S. 219; Teubner 2018, S. 160 ff.), für den Einsatz Künstlicher Intelligenz in der Forschung anders zu bewerten wären. Zum anderen ist technisch jedenfalls gegenwärtig sehr fraglich, ob man einer Künstlichen Intelligenz so weitreichende Eigenständigkeit zubilligen kann, dass sie tatsächlich zum beherrschenden Akteur wird und ihre Funktion als ein Mittel vollständig aufgibt. Dies kann zumindest derzeit mit guten Gründen bezweifelt werden (siehe Kap. 2 in diesem Band).

Selbst wenn man davon ausgeht, dass algorithmischen Systemen eine Rechtspersönlichkeit zugeschrieben werden könnte, ginge damit noch keine vollständige Entlastung des Wissenschaftlers ein bzw. umgekehrt eine vollständige Verantwortlichkeit der Künstlichen Intelligenz. Denn die DSGVO kennt in Art. 26 DSGVO den sog. „gemeinsamen Verantwortlichen“ als Adressaten ihrer Vorgaben. Diese Vorschrift reagiert auf die zunehmend vernetzte Verarbeitung von Daten, die häufig arbeitsteilig erfolgt (Petri 2019, Art. 26 Rn. 1). Sie weist aber nicht etwa unter mehreren Datenverarbeitern einem die Alleinverantwortlichkeit zu und entlastet damit

alle anderen. Vielmehr obliegt es den Verantwortlichen, die konkrete Zuteilung zu regeln; nach außen bleiben alle Verantwortlicher (Petri 2019, Art. 26 Rn. 4). Damit wird vermieden, dass die arbeitsteiligen und vernetzten Strukturen zum Nachteil der Betroffenen wirken können, weil diese womöglich nicht sicher wissen, an wen sie ihre Ansprüche richten müssen (Petri 2019, Art. 26 Rn. 2). Aus der Rechtsprechung des EuGH (EuGH, Urt. v. 05.06.2018 – C-210/16) wird deutlich erkennbar, dass für die Annahme einer gemeinsamen Verantwortlichkeit keine erhebliche Einwirkungsmöglichkeit aller Beteiligten gegeben sein muss, sondern die Nutzung eines Angebots (etwa eines Plattformbetreibers) und damit die Nutzung einer Datenverarbeitung zu eigenen Zwecken ausreichen kann. Insofern ist schwer vorstellbar, dass die Wissenschaftler/innen, die eine Künstliche Intelligenz auf den Weg gebracht haben, nicht weiterhin wenigstens gemeinsam Verantwortlicher blieben. Zudem bliebe auch im Innenverhältnis fraglich, wer dann die Hauptlast trüge: Denn da der Forschende die Künstliche Intelligenz eingesetzt hat, bleibt eine erhebliche Einwirkungsmöglichkeit auf seiner Seite.

6.4.4 *Das Austarieren von Datenschutz und Wissenschaftsfreiheit nach dem Verständnis der DSGVO sowie des deutschen Rechts*

Forschungsfreiheit und Datenschutzrecht sind verfassungsrechtlich gewährleistete Rechte. Weder kann sich also das eine noch das andere vollständig durchsetzen; vielmehr müssen beide im Konfliktfall in einen Ausgleich gebracht werden. Dies ist zuvörderst Aufgabe des Gesetzgebers. In der DSGVO sind für diesen Fall Regelungen vorgesehen. Diese sind allerdings nicht ganz einfach zu verstehen und umzusetzen, wie im Folgenden gezeigt wird.

Einführung

Ist eine präzise und rechtlich tragfähige Bestimmung des Zwecks der Forschung mit Künstlicher Intelligenz schon nicht immer ganz einfach, erschwert die DSGVO mit einigen unklaren Formulierungen die Datenverarbeitung im Bereich der Künstlichen Intelligenz zusätzlich. Das liegt daran, dass die DSGVO selbst dem Anliegen der Wissenschaftsfreiheit aktiv Rechnung trägt, also bereits in ihren Vorschriften einen Ausgleich der grundsätzlich ähnlich wertigen Rechte des Datenschutzes und der Forschungsfreiheit herstellt. Dies geschieht allerdings unter Berücksichtigung der besonderen Kompetenzen der EU. Daher ist an dieser Stelle eine komplizierte rechtliche Gemengelage entstanden. In dieser greifen EU-Regelungen und nationale Regelungen ineinander, zum Teil können sogar entgegen der herkömmlichen Lehre die nationalen Regelungen den europäischen Vorgaben vorgehen.

Hintergrund dieses schwierigen rechtlichen Bestands sind mehrere Sonderregelungen innerhalb der DSGVO zum Thema „Forschung“ (vgl. zu einer Übersicht der verstreuten Vorschriften Caspar 2019, Art. 89 Rn. 31 ff.), allen voran

Art. 89 DSGVO und Art. 5 Abs. 1 lit. b) 2. HS DSGVO. Diese Vorschriften privilegieren auf unterschiedliche Weise die „Forschung“ gegenüber anderen Verarbeitungszwecken und Verarbeitungsschritten. U.a. kann nach Art. 5 Abs. 1 lit. b) 2. HS DSGVO eine Weiterverwendung von Daten, die zu einem anderen als einem Forschungszweck erhoben worden sind, mit derselben Rechtsgrundlage für Forschungszwecke erfolgen. Bei oberflächlicher Betrachtung könnte man damit die eben beschriebene Zweckbeschränkung jedenfalls für Forschungszwecke als aufgehoben betrachten. Und Art. 89 Abs. 2 DSGVO überlässt den Mitgliedstaaten in einer weiten Öffnungsklausel die Möglichkeit, für die Zwecke der Forschung Rechte der Betroffenen weitgehend zurückzuschneiden, solange Ausgleichsmaßnahmen anderer Art dafür vorgesehen sind.

Der Begriff der Forschung in der DSGVO

Angesichts solcher erheblichen Privilegierungen ist die Frage aufgerufen, was unter dem Begriff der „Forschung“ i. S. d. DSGVO überhaupt zu verstehen ist, wer und was also überhaupt von dieser erheblichen Privilegierung profitieren kann. Fest steht zunächst nur, dass die Berufung auf „Forschung“ nicht dazu führen kann, dass jegliche Bindung an das Datenschutzrecht aufgelöst wäre. Denn das führte dazu, dass die verfassungsrechtlich abgesicherte Forschungsfreiheit das ebenso verfassungsrechtlich abgesicherte Datenschutzrecht vollständig verdrängte und sich diesem gegenüber umfassend durchsetzte. Angesichts der weitreichenden Gleichrangigkeit beider Verfassungsrechte ist dies aber nicht zulässig.

Eine Legaldefinition des Forschungsbegriffs in der DSGVO fehlt, wie man ja auch von einem gesicherten Verständnis auf der verfassungs-/primärrechtlichen Ebene nicht ausgehen kann (Roßnagel 2019a, b, S. 158). Art. 13 EU-GR-Charta kennt den Forschungsbegriff und versteht ihn eng, nämlich unter Ausschluss der Lehre (Jarass 2021, Art. 13 Rn. 9, Bernsdorff 2019, Art. 13 Rn. 14). Die eigentliche Definition kann durchaus vergleichbar einer vom deutschen Bundesverfassungsgericht entwickelten Formel formuliert werden (Roßnagel 2019a, b, S. 158), zumal das deutsche Verfassungsrecht als Inspirationsquelle für die europäische Fassung diene (Ruffert 2016, Art. 13 GRCh Rn. 1). Danach ist Forschung alles, was „nach Inhalt und Form als ernsthafter planmäßiger Versuch zur Ermittlung von Wahrheit anzusehen ist“ (Zit. als: BVerfGE 47, 327 (367)). Geschützt ist also jeder als Person und alles als Gegenstand, was zum Zweck des Erkenntnisgewinns nach wissenschaftlichen Methoden unternommen wird (Ruffert 2016, Art. 13 GRCh Rn. 8). Die Anwendung ist davon nicht erfasst, d. h. Regulierung kann ohne Berücksichtigung der Forschungsfreiheit – dann aber möglicherweise unter Berücksichtigung anderer Grundrechte wie Berufs- und Eigentumsfreiheit – erwogen werden.

Dem Erwägungsgrund 159 der DSGVO lässt sich zudem für das Verständnis der DSGVO entnehmen, dass nicht nur die Grundlagen, sondern auch die anwendungsorientierte und auch die privatfinanzierte Forschung erfasst sein sollen. Ferner führt Erwägungsgrund 159 aus, dass der Begriff für die DSGVO „weit ausgelegt“ werden soll.

Diese Formulierungen der DSGVO verbergen, dass mit Art. 89 DSGVO und den weiteren Vorschriften zur Forschung ein verfassungsrechtlich verlangter Ausgleich

zweier Grundrechte vorgenommen wird. Denn auch wenn der Forschungsfreiheit des Art. 13 EU-GR-Charta und ebenso des Art. 5 Abs. 3 GG keine unmittelbare Beschränkung zu entnehmen ist, gilt trotzdem, dass im nationalen wie im europäischen Recht die Rechte anderer und die anerkannten Gemeinwohlbelange von Verfassungsrang bzw. aus Art. 52 Abs. 1 EU-GR-Charta entgegenstehen können (Ruffert 2016, Art. 13 GRCh Rn. 11). Die DSGVO mit ihren Privilegierungen für Forschungsanliegen stellt also einen Ausgleich zweier potenziell konfligierender Grundrechte dar und sorgt dafür, dass beide in einem angemessenen Verhältnis zueinander zur Geltung kommen können. Weder kann die Forschungsfreiheit noch das Datenschutzrecht das jeweils andere Grundrecht vollständig verdrängen.

Daher bedeutet die in Erwägungsgrund 159 DSGVO angesprochene „weite“ Lesart des Forschungsbegriffs keine vollständige Übernahme des verfassungsrechtlichen Begriffs und sie führt auch nicht dazu, dass angesichts des Vorliegens jeglicher Art von Forschungsanliegen der Datenschutz zurücktreten müsste. Eine weite und quasi unbeschränkte Lesart von Forschung im Rahmen der DSGVO würde dazu führen, dass nahezu alles, das irgendeinem Erkenntnisgewinn dient und egal von wem die Forschung und zu welchen Zwecken sie betrieben wird, in den Genuss der weitreichenden Privilegierungen der DSGVO käme. Damit würde die grundsätzliche Beschränkung, welche die DSGVO mit Prinzipien wie Zweckbindung und Datenminimierung in Art. 5 Abs. 1 DSGVO vorsieht, unterlaufen. Es träte ein Verdrängen der Anliegen der Datenschutzgrundrechte durch die Forschungsfreiheit ein. Sie entspräche auch nicht der Überschrift von Kap. 9 der DSGVO, wonach es um „besondere Verarbeitungssituationen“ gehen soll, also um Ausnahme- und nicht um Regeltatbestände (ebenso Caspar 2019, Art. 89 Rn. 10). Daher ist insgesamt aus dem Kontext der DSGVO, den allgemeinen Zielen, Zwecken und der Systematik eine restringierende Lesart zu entwickeln, welche dem Kernanliegen einen klugen Ausgleich zwischen Datenschutzrecht und Forschungsfreiheit herzustellen, gerecht wird, ohne aber damit gleichzeitig das Datenschutzrecht auszuhöhlen und leer laufen zu lassen. Das lässt sich auch aus Art. 89 Abs. 1 DSGVO auslesen, der nämlich bei Reduktion der Anforderungen der DSGVO aus Gründen der Forschungsfreiheit begleitende Garantien für die Rechte und Freiheiten fordert, also deutlich macht, dass ein Weniger an Datenschutz durch Forschung insgesamt nicht hingekommen werden soll.

Auch aus einer Metaperspektive auf die Bedeutung des Datenschutzrechts kann ein weiter Forschungsbegriff nicht zu einer vollständigen Verdrängung des Datenschutzrechts führen. Das Datenschutzrecht ist das ureigene, zentrale Recht der Informationsgesellschaft und der Regulierung der Informationstechnologie. Es befasst sich mit automatisierter Datenverarbeitung wie kein anderes Rechtsgebiet; es ist seit seinen Ursprüngen auf die Bewältigung der automatisierten Datenverarbeitung angelegt (Simitis et al. 2019, Einleitung, Rn. 1). Sein Regulierungsanliegen ist es also gerade, die technische Innovation und die damit verbundene massenhafte Auswertung von Daten zu beschränken, allerdings nicht grundsätzlich zu verbieten. Deshalb darf auch der Forschungsbegriff nicht zur Beliebigkeit verkümmern (Caspar 2019, Art. 89 Rn. 15).

Daher ist die Umschreibung der Privilegierung für Forschung, z. B. in Art. 5 Abs. 1 lit. b) DSGVO, durch den europäischen Gesetzgeber von „wissenschaftlicher Forschung“ (in Abgrenzung zu „historischer Forschung“) nur scheinbar tautologisch (Caspar 2019, Art. 89 Rn. 15). Daraus lässt sich nämlich folgern, dass gerade die Entwicklung eines Forschungsgegenstands und der hierzu verwendeten Methodik ihrerseits wiederum wissenschaftlichen Anforderungen und Standards genügen muss. Daher sind z. B. Korrelationen und Zufälligkeiten, die methodisch herangezogen werden, nicht ausreichend wissenschaftlich im Vergleich zu echter statistischer Auswertung und Belastbarkeit sowie Nachvollziehbarkeit und Kontrollierbarkeit der Verfahren und darauf basierenden Ergebnisse (Spiecker genannt Döhmman 2017, S. 60). Eine Künstliche Intelligenz, die sich diese Ansätze ohne wissenschaftliche Basierung und ohne wissenschaftliches Konzept zunutze machen wollte, hätte Schwierigkeiten vom Forschungsprivileg der DSGVO zu profitieren, das aus der Forschungsfreiheit erwächst.

Andererseits muss einer zu engen Lesart die Stirn geboten werden, die vorschnell unter Heranziehung des unklaren Wortlauts nur eine öffentlich finanzierte Forschung genügen lassen würde. Zum einen spricht Erwägungsgrund 159 ausdrücklich davon, dass auch privatfinanzierte Forschung privilegiert sein kann, wenngleich dies nicht als Zwangsläufigkeit („muss“) formuliert ist. Daher kann es sehr wohl Fälle geben, in denen eine privat finanzierte Forschung vom Forschungsprivileg ausgeschlossen wird – wie es eben auch Fälle geben kann, in denen eine öffentlich finanzierte Forschung ausgeschlossen wird, wie soeben erläutert. Vor allem aber entspricht die deutsche Übersetzung nicht dem englischen Text: Während die deutsche Fassung „im öffentlichen Interesse“ voranstellt, so dass man es auf alle drei genannten privilegierten Tatbestände (also Archivzwecke, wissenschaftliche und historische Forschungszwecke) beziehen kann, stellt die englische Fassung „in the public interest“ nach und zwar allein und ausschließlich den Archivzwecken.

Gleichwohl ist nicht jede privatfinanzierte Forschung eine privilegierte Forschung, wie sich aus dem Einschränkungsanliegen des 9. Kapitels der DSGVO folgern lässt. Forschungsfreiheit ist nicht um ihrer selbst geschützt – dafür hätte die Berufsfreiheit genügt. Sondern Forschungsfreiheit ist deshalb geschützt, weil die Wissensvermehrung grundsätzlich auch der Allgemeinheit zur Verfügung steht und ihr nützt. Es bedarf daher zur Anerkennung im Rahmen der DSGVO einer sozialen, allgemeinwohlorientierten Zwecksetzung, die nicht allein wirtschaftlichen Interessen dient. Rein kommerzielle Forschungen, die nicht auch der Allgemeinheit zu dienen bestimmt sind, sondern gerade darauf abzielen, durch Geheimhaltung und Verbergen der Erkenntnisse Vorsprung vor Konkurrent/innen zu gewinnen oder Defizite bei Kunden und Nutzer/innen auszunutzen, fallen nicht darunter (Caspar 2019, Art. 89 Rn. 16 ff.). Sie können die vielseitigen Allgemeinwohlbelange, die das Datenschutzrecht verfolgt, nicht aushebeln.

Das ergibt sich auch daraus, dass eine Abgrenzung der Berufsfreiheit und der Forschungsfreiheit zu treffen ist. Im Umkehrschluss kann man sagen, dass jedenfalls sog. drittmittelfinanzierte Forschung an öffentlichen Forschungseinrichtungen unter den privilegierenden Forschungsbegriff fallen dürfte (so wohl auch Caspar 2019, Art. 89 Rn. 17). So darf zwar das Thema durch Dritte vorgegeben werden,

nicht aber die Methodik. Und der Ablauf der Forschung darf nicht von Unternehmensentscheidungen oder den Entscheidungen Dritter abhängen, wie es aber für eine Auftragsforschung typisch ist. Die Grenzlinie ist also durchaus schwierig zu ziehen; sie ist graduell je nach Grad der Unabhängigkeit zu bestimmen.

Ferner ist zu belegen, dass die Forschung dem Allgemeinwohl zu dienen bestimmt ist, weil sie auf Veröffentlichung und damit auf Teilung der Ergebnisse und ihre Diskussion angelegt ist. Eine Forschung, die ihre Ergebnisse abgrenzen und allein für sich bzw. die Unternehmenszwecke nutzen will, genügt diesen Anforderungen nicht. Sie ist schon nicht wissenschaftlich, weil sie den Aspekt der Selbstkontrolle des Wissenschaftsapparats durch die fachliche Diskussion und Prüfung missachtet und damit einen wesentlichen Definitionsaspekt nicht entfaltet.

Zudem lässt sich auch eine anwendungsorientierte Forschung durchaus unter Art. 89 DSGVO und die weiteren Vorschriften der DSGVO zur Forschung subsumieren. Die Anwendung selbst ist allerdings nicht mehr erfasst, denn diese unterfällt nicht der Forschung selbst. Ansonsten unterliefe jede Verbesserung eines strukturell entwickelten algorithmenbasierten Informationssystems dem Forschungsprivileg der DSGVO (vgl. auch Caspar 2019, Art. 89 Rn. 17).

Um Missverständnisse vorzubeugen, sei darauf verwiesen, dass die Allgemeinwohlorientierung eine notwendige, aber nicht etwa hinreichende Bedingung ist: Denn der Forschungsbegriff umfasst weitere Kriterien, die zudem eingehalten sein müssen. So kann sich auch allgemeinwohlorientierte Forschung Methoden bedienen, die nicht mehr als wissenschaftlich einzustufen sind und damit aus dem Forschungsprivileg der DSGVO herausfallen.

Für eine Künstliche-Intelligenz-gestützte oder sogar zentral auf Künstliche Intelligenz ausgerichtete Forschung bedeutet dies Konsequenzen: Zunächst einmal ist zu identifizieren, wer die Forschung mittels Künstlicher Intelligenz betreibt und wie dies geschieht. Wenn es sich um eine Forschung handelt, die auf Öffentlichkeit ausgerichtet ist, also darauf, das Allgemeinwohl durch einen Erkenntnisgewinn zu steigern, der geteilt wird, greifen die Privilegierungen der DSGVO. Sie erreichen allerdings nur den Bereich, der tatsächlich der Entwicklung, also der eigentlichen Beforschung, dient, nicht aber der Anwendung und deren Verbesserung ohne fundamentale Neuentwicklung. Zudem ist darauf zu achten, dass der Einsatz Künstlicher Intelligenz nach allgemein anerkannten wissenschaftlichen Prinzipien erfolgt.

6.4.5 Die Lockerung der Zweckbindung in Art. 5 Abs. 1 lit. b 2. HS DSGVO

Unterfällt die Entwicklung und Verwendung einer Künstlichen Intelligenz dem Forschungsbegriff, ist eine entscheidende Privilegierung in Art. 5 Abs. 1 lit. b) 2. HS DSGVO enthalten: Danach können nämlich Daten, die der Entwickler der Künstlichen Intelligenz zur Verfügung hat, für die Zwecke dieser Forschung verwendet werden, auch wenn diese Daten ursprünglich zu einem ganz anderen Zweck erhoben und verarbeitet, z. B. gespeichert, worden sind. Das Grundproblem der

Zweckbindung ist damit für diesen besonderen Bereich weitgehend aufgehoben. Nur in Einzelfällen, bei besonders kritischen Daten oder besonders problematischen Anwendungen der Künstlichen Intelligenz, würde die Privilegierung des Art. 5 Abs. 1 lit. b) 2. HS DSGVO nicht greifen.

Einzige Einschränkung der erweiterten Nutzbarkeit von Daten über die Entbindung vom Zweckbindungsgrundsatz ist Art. 6 Abs. 4 DSGVO, wonach der Datenverarbeiter zu prüfen hat, ob der Zweck tatsächlich noch ein vereinbarter Zweck ist. Wegen der Sonderformel des Art. 5 Abs. 1 lit. b) 2. HS DSGVO ist diese Vorschrift allerdings in den meisten Fällen der Forschung gegeben. Art. 5 Abs. 1 lit. b) 2. HS DSGVO stellt dafür eine Vermutung auf, die allerdings widerlegt werden kann. Denn auch bei der Anwendung von Art. 6 Abs. 4 DSGVO im Zusammenspiel mit Art. 5 Abs. 1 lit. b) 2. HS DSGVO ist zu beachten, dass es sich um eine Regelung handelt, die einen Ausgleich zwischen Forschungsfreiheit und Datenschutzgrundrechten herstellt, also weder das eine noch das andere Recht unangemessen im Ergebnis privilegieren darf. Art. 7 Abs. 2 S. 2 EU-GR-Charta verlangt ausdrücklich die Einhaltung einer Zweckbindung und auch wenn diese zugunsten der Forschung zurückgenommen werden kann, darf sie doch nicht obsolet werden. Daher kann Art. 5 Abs. 1 lit. b) 2. HS DSGVO nicht dahingehend verstanden werden, dass jegliche Forschung jegliche Zweckbindung aufgeben kann, sondern dass die Beschränkungen des Art. 6 Abs. 4 DSGVO weiterhin gelten und jeweils zu prüfen sind (so auch Roßnagel 2019a, b, S. 162). Immer dann, wenn besonders eingriffsintensive Maßnahmen vorgenommen werden, darf der Verarbeiter nicht ohne weiteres von der Vereinbarkeit ausgehen, auch wenn der Wortlaut des Art. 5 Abs. 1 lit. b) 2. HS DSGVO dies vermuten lässt.

Für den Einsatz Künstlicher Intelligenz bedeutet dies auch in der Forschung eine hohe Aufmerksamkeit für die Einhaltung des Datenschutzrechts. Denn sie ist grundsätzlich mit ihrem großen und spezifischen Datenbedarf und den intensiven Verknüpfungsmöglichkeiten (auf diese Kriterien stellt auch Roßnagel 2019a, b, S. 162 *ab*) als besonders eingriffsintensiv einzuschätzen. Auch wenn sie als Forschung privilegiert ist, unterliegt sie dennoch besonderen Anforderungen. Das formuliert Art. 89 Abs. 1 DSGVO mit der Forderung nach zusätzlichen Garantien in besonderer Weise und dürfte angesichts der Schwierigkeiten an Nachvollziehbarkeit, Begründbarkeit und Wiederholbarkeit schwer umzusetzen sein. Art. 6 Abs. 4 DSGVO fordert hier mehr als nur eine standardisierte Betrachtung. Daher ist beispielsweise bzgl. der Herkunft der verwendeten Daten genau zu prüfen, ob es eine enge Verbindung mit der Forschung gibt. Zudem sind Beschränkungen der Weiterverwendung für andere Zwecke hilfreich, um eine Legalität der Verwendung Künstlicher Intelligenz zu sichern.

Dass im Übrigen die sonstigen Voraussetzungen der DSGVO eingehalten werden müssen, so z. B. eine Datenschutz-Folgeabschätzung nach Art. 35 DSGVO, soll hier nur als Hinweis aufgenommen werden. Dies bedeutet, dass der Einsatz von Künstlicher Intelligenz als Forschungsinstrument nicht leichtfertig eingesetzt werden sollte. Es erfordert einen aktiven planerischen Zugang auf die datenschutzrechtlichen Gewährleistungen einschließlich umfangreicher Sicherungsmaßnahmen, um Missbrauch und Manipulation, Diskriminierung und andere unerwünschte Effekte zu verhindern.

6.5 Datenschutzrecht und sonstige Anforderungen bei Kooperationen

Nicht unterschlagen werden soll eine weitere offene Flanke von und für Forschung mit Künstlicher Intelligenz, nämlich die unsichere Rechtslage für Kooperationen zwischen Forschungseinrichtungen.

Der erhebliche Ressourcenbedarf der Entwicklung und Begleitung Künstlicher Intelligenz setzt große Anreize, aber auch große Notwendigkeiten für eine Zusammenarbeit zwischen verschiedenen Wissenschaftlern und Forschungseinrichtungen. Dies ist bereits jetzt zu beobachten, wird sich aber noch verstärken, möglicherweise auch von Forschungsförderinstitutionen vorangetrieben (man denke nur an die DFG-Finanzierung des Höchstleistungsrechnens, dazu [WissRat 2012](#)). Ein wichtiger, auch wettbewerbsrelevanter Faktor ist dabei die Verfügbarkeit von Datensätzen: Ein Partner mit großen Sammlungen von Daten, auf die für den Einsatz der Künstlichen Intelligenz zurückgegriffen werden kann, kann für die Qualität der Ergebnisse von entscheidender Bedeutung sein.

Ein Kooperationsrecht für öffentliche Einrichtungen, gar für Hochschulen und Forschungseinrichtungen, gibt es allerdings nicht. Anders als im privatrechtlichen Bereich, wo sich seit Jahrhunderten mit dem Gesellschaftsrecht und dem Ehe recht ein umfassendes Kooperationsrecht herausgebildet hat, einschließlich eines teilweise bestehenden Formzwangs und der Möglichkeit der Gründung juristischer Personen für eigenständige Rechtspersönlichkeiten auf Dauer, kennt das öffentliche Recht ein solches Kooperationsrecht nur in Ansätzen. In der Folge gibt es kein Rechtsregime, auf das sich Kooperationspartner als Grundsatzregelung stützen können. Vielmehr bedarf es zu einer rechtssicheren Gestaltung umfassender, individuell auszuhandelnder Kooperationsverträge, die zum Teil (z. B. bei länderüberschreitender Kooperation und je nach Wesentlichkeit des Kooperationsgegenstands) sogar parlamentarisch beschlossen werden müssen. Das Spektrum der zu regelnden Bereiche ist groß. Es reicht von urheberrechtlichen Fragen bis hin zu Kontrollrechten und Budgethoheit, erfasst aber auch grundlegende Entscheidungen wie Abstimmungsrechte, Kündigungsregelungen und Weiterverwertung nach Beendigung oder Teilausstieg.⁴

Diese Probleme fehlender rechtlicher Regelungen steigern sich noch, wenn im Rahmen internationaler Forschung nicht nur europäische Partner zusammenkommen, sondern auch Partner anderer Staaten kooperieren, deren Rechtsvorstellungen erheblich divergieren können. Deutlich wird das gerade in den Informations- und Datenrechtsvorstellungen, ganz besonders im Urheber- und im Datenschutzrecht. Denn nur in wenigen Staaten weltweit kann von einem Datenschutzniveau ausgegangen werden, das demjenigen der DSGVO entspricht und das auch als solches anerkannt worden ist (zuletzt Japan Anfang 2019: Durchführungsbeschluss (EU) 2019/419 der Kommission vom 23. Januar 2019 über die Angemessenheit des Datenschutzniveaus in Japan; siehe die EuGH-Entscheidung

⁴Siehe zum Themenkomplex insgesamt Richter und Spiecker genannt Döhmman [2019](#), S. 181 ff.

Schrems II zum unzureichenden Datenschutz auf der Basis des Privacy Shield, EuGH, Urt. v. 16.07.2020 – C-311/18), so dass in diesen Kooperationen von einem rechtlich unproblematischen Transfer der Daten ausgegangen werden kann. Und auch die Zugriffsrechte auf urheberrechtlich geschützte Werke variieren erheblich, wie u. a. die Auseinandersetzung um Google Books – möglicherweise sogar auch für Forschungszwecke – zeigt (siehe dazu Gesmann-Nuissl und Wünsche 2012, S. 228 ff.).

Eine weitere ungelöste Schwierigkeit ergibt sich zudem durch eine jüngere Rechtsprechung des EuGH, in der dieser eine gemeinsame Verantwortlichkeit nach Art. 26 DSGVO (siehe dazu im Verhältnis zur Künstlichen Intelligenz unter Abschn. 6.4.3) und damit auch grundsätzlich eine gemeinsame Haftung für alle Verstöße gegen die DSGVO annahm, obwohl einer der Partner faktisch keinerlei Einflussmöglichkeit auf die – beanstandete – Datenverarbeitung des anderen Partners hatte. Der EuGH ließ es genügen, dass der geringere Partner von der Datenverarbeitung des bestimmenden Partners profitierte (EuGH, Urt. v. 05.06.2018 – C-210/16). Diese Situation dürfte sich auch bei Kooperationen regelmäßig stellen: Nicht alle Partner sind in gleicher Weise an allen Forschungsanliegen beteiligt, profitieren aber insgesamt – und sei es durch Förderzuwendungen oder Veröffentlichungen – von der Datenverarbeitung der anderen Partner. Verwendet also ein Beteiligter Daten unter Verstößen gegen die DSGVO, sind davon auch die sonstigen Datenverarbeitungen infiziert und jeder Partner muss fürchten, für die Verstöße anderer haftbar gemacht zu werden. Da Künstliche Intelligenz davon profitiert, dass Daten aus verschiedenen Quellen stammen, muss auf dieser Basis mangels rechtlicher Regelung wenigstens vertraglich sichergestellt sein, dass strikte Datenschutz- und IT-Sicherheitsvorkehrungen getroffen werden. Sind in anderen Staaten allerdings weitreichende Zugriffsrechte des Staates ohne besondere Zweckbindungen zulässig, ist dies schwierig (vgl. EuGH, Urt. v. 16.07.2020 – C-311/18).

Schließlich ist damit umzugehen, in welcher Weise sich Rechte und Pflichten nicht nur, aber insbesondere aus der DSGVO verändern, wenn ein Wissenschaftler seine institutionelle Anbindung wechselt, erst recht, wenn dies in einer Kooperation der Fall ist. Unklar ist, ob und wie der Datenzugriff dann weiterhin möglich ist, wer also Zugriffsrechte behält (siehe auch WissRat 2020, S. 33) und wie eine Kontrolle über die Daten gewährleistet werden kann, durch wen also beispielsweise Pflichten auf Löschung oder Berichtigung nachzukommen ist. Es entstehen darüber möglicherweise Datenpools bei Institutionen, die losgelöst von den individuellen Forschern existieren und damit – entgegen des Zweckbindungsgrundsatzes – zweckfrei für allgemeine Forschung vorgehalten werden.

6.6 Fazit und Ausblick

Wie bereits kenntlich gemacht, sind die Regulierungsanstrengungen im Bereich des Einsatzes Künstlicher Intelligenz bisher begrenzt; im Bereich der Forschung mit und an Künstlicher Intelligenz gilt dies erst recht. Am ehesten lässt sich noch der

Konflikt zwischen Wissenschaftsfreiheit und Datenschutzrecht als vom geltenden Recht beeinflusst herausarbeiten. Er prägt die Nutzung künstlicher Intelligenz im Bereich der Forschung erheblich mehr als vielen Anwendern dieser Technologie bisher bewusst ist, denn Künstliche Intelligenz basiert auf der weitreichenden und oftmals zweckentfremdenden Nutzung von Datenmengen und dazu gehören gerade auch personenbezogene Daten. Angesichts fast gänzlich fehlender Regelungen wäre es wünschenswert, wenn es Klarstellungen für Forschende gäbe. Dies kann in den hochschulrechtlichen Bestimmungen geschehen, dies kann und muss auch im Beamten- und Arbeitsrecht festgelegt werden. Denn Forschung findet eben nicht allein im Schutzzumfeld öffentlich geförderter Institutionen statt, sondern ganz wesentlich auch im privaten Bereich und unter Kooperationen beider.

Das bedeutet nicht, dass nicht eine Reihe von weiteren Rechtsfragen aufgeworfen werden können und aufgeworfen werden. Dies betrifft beispielsweise urheberrechtliche Fragestellungen: Wer ist Urheber eines Werks, das mit Hilfe von Künstlicher Intelligenz gefertigt worden ist – greift der Werkbegriff hier überhaupt? Am eindringlichsten mag dieses Problem sich illustrieren lassen, wenn Künstliche Intelligenz eigene Texte verfasst.

Eng damit verknüpft sind Fragen der Ergebniskontrolle und der Sicherung der Qualität von Forschung; so sind z. B. Publikationsorgane dazu übergegangen, von den Autoren/innen zu verlangen, ihre Datensätze zur Verfügung zu stellen, um eine Nachkontrolle – auch z. B. im Peer Review Verfahren durch unbekannte Dritte – zu ermöglichen. Dies schafft allerdings auch Möglichkeiten der Weiterverwertung dieser Daten für eine Vielzahl weiterer Personen und Institutionen über den unmittelbaren Forschungsbereich hinaus.

Will man solche weitreichenden Einflussnahmen auf Forschung durch Künstliche Intelligenz wegen der mangelnden Determination des Ergebnisses nicht vollständig ausschließen, kommt auch die Frage nach der Rechtsstellung schnell auf. Wird in absehbarer Zeit der/die Wissenschaftler/in womöglich der/diejenige sein, der/die Forschungsergebnisse der Künstlichen Intelligenz nur noch aufbereitet und präsentiert und die eigenständige Forschungsleistung maschinell erbracht?

Das ruft Fragen auf, die im Kontext Künstlicher Intelligenz ohnehin diskutiert werden, nämlich nach Nachvollziehbarkeit, Transparenz und Begründung/Rationalisierung des Einsatzes dieser Technologie. Sie sind im Forschungsbereich etwas anders gestellt und sie werfen zum Teil auch eigenständige Folgefragen auf – etwa, wenn zur Qualitätssicherung von Forschungsergebnissen und zur unabhängigen Kontrolle die vollständigen Datensätze und ein Duplikat des algorithmischen Systems an ein Publikationsorgan geliefert werden sollen und darüber – z. B. in Peer Review Verfahren – unmittelbare Konkurrenten Zugriff auf diese Daten und Forschungsansätze erhalten.

Auch der Transfer von Forschung mit und an Künstlicher Intelligenz in der Anwendung bleibt problematisch: Schon jetzt wird diskutiert, wie sichergestellt werden kann, dass und wie eine algorithmisch unterstützte Entscheidungsfindung weiterhin verantwortbar ausgestaltet werden kann. Dies trifft auch die Forschung selbst und es trifft sie zusätzlich, wenn sie den Transfer von der Wissenschaft in die Innovation und Marktfähigkeit bewältigen muss.

Letztlich treten damit die großen Fragen des Einsatzes von Künstlicher Intelligenz bereits in der Entwicklung im Umfeld der Forschenden auf: Wie kann die Verantwortung für die Durchführung und die Ergebnisse einer Forschung mit und von Künstlicher Intelligenz sichergestellt werden, wie kann eine Qualitätskontrolle gesichert werden und wie können Fehler und Ungenauigkeiten nachgewiesen und in eine Haftung für die Auswirkungen und den Einsatz überführt werden?

Die Forschungsfreiheit verlangt für den der Anwendung und Innovation vorgelagerten Bereich noch einmal besondere Freiheiten, gleichzeitig aber auch besondere Verantwortlichkeiten. Insoweit stellen sich für den Einsatz Künstlicher Intelligenz die grundlegenden Fragen der Regulierung von Forschung einmal mehr und einmal anders, aber nicht gänzlich neu. Es bleibt dabei, dass die Auswertung von Daten, gerade in großem Umfang, vielfältige Ergebnisse generieren kann. Wofür diese eingesetzt, weiterverfolgt, weiterentwickelt und letztlich verwendet werden, kann durch die Forschung oftmals selbst gar nicht bestimmt werden. Diese Weiterverwendungs- und Zugangsmöglichkeiten sind aber mitzudenken und sie verlangen eine nicht nur beiläufige beständige datenschutzrechtliche Begleitung bis hin zur Weiterentwicklung auf technischer Ebene. Künstliche Intelligenz verändert also nicht das Verständnis von Forschung, aber es führt dazu, dass – jedenfalls aus datenschutzrechtlicher Sicht – einige Fragen mit neuer Stoßrichtung und anderer Präzision gestellt werden und in der Konsequenz Künstliche Intelligenz auch anders ausgerichtet werden muss

Literatur

- Bernsdorff N (2019) Art. 13. In: Meyer J, Hölscheidt S (Hrsg) Charta der Grundrechte der Europäischen Union, 5. Aufl. Nomos, Baden-Baden
- Calliess C (2016) Art. 4 AEUV. In: Ruffert M (Hrsg) EUV, AEUV: das Verfassungsrecht der Europäischen Union mit Europäischer Grundrechtecharta. Kommentar, 5. Aufl. C.H. Beck, München
- Caspar J (2019) Art. 98. In: Simitis S, Hornung G, Indra Spiecker genannt Döhmann (Hrsg) Kommentar Datenschutzrecht. DSGVO mit BDSG, 1. Aufl. Nomos, Baden-Baden
- Datenethikkommission (2019) Gutachten der Datenethikkommission. https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf;jsessionid=F8BF886A15BAEEF0C4117E68CCEACABD.2_cid373?__blob=publicationFile&v=6. Zugriffen am 18.02.2021
- Dederer H-G (2018) Letale Autonome Waffensysteme (LAWS). Z Rechtswiss Forsch 9(4):380–413
- Ernst S (2021) Art. 4. In: Paal BP, Pauly DA (Hrsg) Kommentar Datenschutz-Grundverordnung. Bundesdatenschutzgesetz, 3. Aufl. C.H. Beck, München
- Fröhlich W, Spiecker genannt Döhmann I (2018) Können Algorithmen diskriminieren? <https://verfassungsblog.de/koennen-algorithmen-diskriminieren/>. Zugriffen am 18.02.2021
- Gausling T (2019) Künstliche Intelligenz im digitalen Marketing. Datenschutzrechtliche Bewertung KI-gestützter Kommunikations-Tools und Profiling-Maßnahmen. Z Datenschutz 9(8):335–341
- Gesmann-Nuissl D, Wünsche K (2012) Neue Ansätze zur Bekämpfung der Internetpiraterie – ein Blick über die Grenzen. Gewerbl Rechtsschutz Urheberrecht Int 61(3):225–234
- Hamer M (2020) Informatisierung des Verwaltungsprozesses: Berücksichtigung der Möglichkeiten moderner Informations- und Kommunikationstechniken bei der Gestaltung und Aus-

- legung der Normen des Verwaltungsprozeßrechts. In: Schriften zum Prozessrecht 167. Duncker und Humblot, Berlin
- High Level Expert Group (2019) Ethik-Leitlinien für eine vertrauenswürdige KI. <https://www.demographie-netzwerk.de/site/assets/files/5064/ethicsguidelinesfortrustworthyai-depdf.pdf>. Zugriffen am 17.02.2021
- Hoeren T (2016) Thesen zum Verhältnis von Big Data und Datenqualität. Erstes Raster zum Erstellen juristischer Standards. Z IT-Recht Recht Digital 19(1):8–11
- Hornung G, Spiecker genannt Döhmman I (2019) Einleitung, Rn. 155–273. In Kommentar Datenschutzrecht. DSGVO mit BDSG, hrsg. Spiros Simitis und dies, 1. Aufl. Nomos, Baden-Baden
- Huston M (2018) AI researchers allege that machine learning is alchemy. <https://www.sciencemag.org/news/2018/05/ai-researchers-allege-machine-learning-alchemy>. Zugriffen am 16.02.2021
- Jarass H (2021) Art. 13. In Charta der Europäischen Union unter Einbeziehung der sonstigen Grundrechtsregelungen des Primärrechts und der EMRK, hrsg. ders, 4. Aufl. C.H. Beck, München
- Karg M (2019) Art. 4. In: Simitis S, Hornung G, Spiecker Genannt Döhmman I (Hrsg) Kommentar Datenschutzrecht. DSGVO mit BDSG, 1. Aufl. Nomos, Baden-Baden
- Kühling J, Buchner B (2018) Art. 1. In Datenschutz-Grundverordnung/BDSG: Kommentar, hrsg. dies, 2. Aufl. C.H. Beck, München
- Nettesheim M (2004) Die Kompetenzordnung im Vertrag über eine Verfassung für Europa. Europarecht 39(4):511–546
- Paal BP, Pauly DA (2021) Einleitung. In Kommentar Datenschutz-Grundverordnung. Bundesdatenschutzgesetz, hrsg. dies, 3. Aufl. C.H. Beck, München
- Petri T (2019) Art. 26. In: Simitis S, Hornung G, Spiecker genannt Döhmman I (Hrsg) Kommentar Datenschutzrecht. DSGVO mit BDSG, 1. Aufl. Nomos, Baden-Baden
- Rauch A (2020) Autonome Waffensysteme und Völkerrecht: Stand der wissenschaftlichen Diskussion, offene Forschungsfragen und aktuelle Regelungsanstrengungen im internationalen System. In C. Tietje (Hrsg) Beiträge zum Europa- und Völkerrecht 20. https://tietje.jura.uni-halle.de/sites/default/files/BeitraegeEVR/Heft%2020_Rauch.pdf. Zugriffen am 17.02.2021
- Richter E, Spiecker genannt Döhmman I (2019) Rechtliche Gestaltung von Verwaltungskooperationen. In: Durner W, Reimer F, Spiecker genannt Döhmman I, Wallrabenstein A (Hrsg) *Das sinnvoll Denkbare denken, das davon Machbare machen, Gedächtnisschrift für Arndt Schmehl*. Duncker und Humblot, Berlin, S 179–198
- Roßnagel A (2019a) Datenschutz in der Forschung. Die neuen Datenschutzregelungen in der Forschungspraxis von Hochschulen. Z Datenschutz 9(4):157–164
- Roßnagel A (2019b) Art. 5. In: Simitis S, Hornung G, Spiecker genannt Döhmman I (Hrsg) Kommentar Datenschutzrecht. DSGVO mit BDSG, 1. Aufl. Nomos, Baden-Baden
- Ruffert M (2016) Art. 13 GRCh. In EUV, AEUV: das Verfassungsrecht der Europäischen Union mit Europäischer Grundrechtecharta. Kommentar, hrsg. Christian Calliess und ders, 5. Aufl. C.H. Beck, München
- Ruscheimer H (2020) 9. Speyerer Forum zur digitalen Lebenswelt: Regulierung Künstlicher Intelligenz in der Europäischen Union zwischen Recht und Ethik. NVwZ 39(7):446–448
- Sachverständigenkommission für den Dritten Gleichstellungsbericht der Bundesregierung (2021) Digitalisierung geschlechtergerecht gestalten. Gutachten für den Dritten Gleichstellungsbericht der Bundesregierung. Berlin: Geschäftsstelle Dritter Gleichstellungsbericht. www.dritter-gleichstellungsbericht.de/gutachten3gleichstellungsbericht.pdf. Zugriffen am 18.02.2021
- Schantz P (2020) Art. 5. In: Brink S, Wolff H A (Hrsg) BeckOK Datenschutzrecht, 34. Aufl. C.H. Beck, München
- Schirmer J-E (2016) Rechtsfähige Roboter? JuristenZeitung 13:660–666. <https://doi.org/10.1628/002268816X14615987983565>
- Simitis S, Hornung G, Spiecker genannt Döhmman I (2019) Einleitung, Rn. 1–154. In Kommentar Datenschutzrecht. DSGVO mit BDSG, hrsg. dies, 1. Aufl. Nomos, Baden-Baden
- Sommerer LM (2020) Personenbezogenes Predictive Policing Kriminalwissenschaftliche Untersuchung über die Automatisierung der Kriminalprognose. In: Schriften zur Kriminologie 19, 1. Aufl. Nomos, Baden-Baden

- Specht L, Herold S (2018) Roboter als Vertragspartner? Gedanken zu Vertragsabschlüssen unter Einbeziehung automatisiert und autonom agierender Systeme. Z IT-Recht Recht Digital 21(1):40–44
- Spiecker genannt Döhmman I (2010) Wissensverarbeitung im Öffentlichen Recht. Z Rechtswiss Forsch 3:247–282
- Spiecker genannt Döhmman I (2017) Big und Smart Data: Zweckbindung zwecklos? Spektrum Wiss 1:56–62
- Spiecker genannt Döhmman I (2018) Fragmentierungen. Kontexte der Demokratie – Parteien, Medien, Sozialstrukturen. VVDStRL 77:9–65
- Spiecker genannt Döhmman I (2020) „Ich habe nichts zu verbergen“ – wirklich nicht?: Unterscheidung versus Diskriminierung durch KI, Algorithmen und die digitalen Dienste. Forsch Frankf 37(1):38–41
- Spiecker genannt Döhmman I (2021) Staatliche Entscheidungen unter Unsicherheit. Mohr Siebeck (i.E.), Tübingen
- Stevens J (2020) Datenqualität bei algorithmischen Entscheidungen. Überlegungen aus Anlass des Gutachtens der Datenethikkommission. Comput Recht 36:73–79
- Stewart M (2018) The machine learning crisis in scientific research. <https://towardsdatascience.com/the-machine-learning-crisis-in-scientific-research-91e61691ae76>. Zugegriffen am 16.02.2021
- Teubner G (2018) Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten. Archiv civilis Prax 218:155–205
- Wissenschaftsrat (2012) Strategische Weiterentwicklung des Hoch- und Höchstleistungsrechnens in Deutschland. Positionspapier. https://www.wissenschaftsrat.de/download/archiv/1838-12.pdf?__blob=publicationFile&v=1. Zugegriffen am 17.02.2021
- Wissenschaftsrat (2020) Zum Wandel in den Wissenschaften durch datenintensive Forschung. Positionspapier. https://www.wissenschaftsrat.de/download/2020/8667-20.pdf?__blob=publicationFile&v=5. Zugegriffen am 18.02.2021

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.





**Carl Friedrich Gethmann, Peter Buxmann, Julia Distelrath,
Bernhard G. Humm, Stephan Lingner, Verena Nitsch, Jan C. Schmidt
und Indra Spiecker genannt Döhmann**

7.1 Zentrale Ergebnisse

Die Digitalisierung von Arbeitswelten erreicht alle Bereiche öffentlich oder privat finanzierter Forschung und Entwicklung. Insbesondere lassen neuartige Ent-

C. F. Gethmann

Universität Siegen, Siegen, Deutschland

E-Mail: carl.gethmann@uni-siegen.de

P. Buxmann

Fachgebiet Wirtschaftsinformatik, Technische Universität Darmstadt, Darmstadt, Deutschland

E-Mail: buxmann@is.tu-darmstadt.de

J. Distelrath

Schalkenbach, Deutschland

E-Mail: julia.distelrath@gmx.de

B. G. Humm

Hochschule Darmstadt, Darmstadt, Deutschland

E-Mail: bernhard.humm@h-da.de

S. Lingner (✉)

Institut für qualifizierende Innovationsforschung und -beratung GmbH (IQIB), Bad Neuenahr-Ahrweiler, Deutschland

E-Mail: Stephan.Lingner@iqib.de

V. Nitsch

Institut für Arbeitswissenschaft, RWTH Aachen, Aachen, Deutschland

E-Mail: v.nitsch@iaw.rwth-aachen.de

J. C. Schmidt

Department of Social Sciences Hochschule Darmstadt, Darmstadt, Deutschland

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

I. Spiecker genannt Döhmann

Goethe Universität Frankfurt a.M., Frankfurt am Main, Deutschland

E-Mail: spiecker@jur.uni-frankfurt.de

wicklungen der künstlichen Intelligenz große Veränderungen im Alltag grundlagen- und anwendungsorientierter Forschung erwarten. Dennoch ist wissenschaftliche Forschung auch im Zeitalter von KI, Machine Learning und Big Data weiterhin auf den Menschen als Wissenschaftler/Wissenschaftlerin und Forscher/Forscherin angewiesen. *Der Mensch bleibt daher als handelnder Akteur im Zentrum von Wissenschaft und Forschung.* Ein etwaiger Ersatz des Menschen durch „starke KI“ ist dagegen nicht zu erwarten, da diese zumindest auf absehbare Zeit nicht erreichbar erscheint. KI-Systeme können daher zunächst eher als hilfreiche Werkzeuge in der Forschungspraxis verstanden werden: Auf den ersten Blick scheinen sie sich an technische Inventionen der Vergangenheit anzuschließen, wie die des Mikroskops oder des Fernrohrs, welche die wissenschaftliche Forschung und Erkenntnis bereits fundamental verändert haben. Ähnliches könnte auch für KI-getriebene Forschung erwartet werden. *Allerdings führt die Selbstorganisationsfähigkeit von KI-Systemen im Modus des maschinellen Lernens (ML) zu einer Intransparenz von Forschungsprozessen,* die eine Kontrolle, Validierung und Anerkennung abgeleiteter Aussagen, Modelle und Theorien erschweren oder gar unmöglich machen. Insofern kann der Wandel in der KI-getriebenen Forschung in diesem Modus sogar als disruptiv betrachtet werden. Ein allzu einfaches Werkzeug- und Instrumentenverständnis von KI führt hier also in die Irre; vielmehr sind ML-basierte KI-Systeme aufgrund ihrer verborgenen Dynamik und ihrer geringen Kontrollierbarkeit gewissermaßen als Instrumente zweiter Ordnung zu werten.

Für einen Wandel durch den Einsatz von KI-Systemen in Wissenschaft und Gesellschaft sprechen nicht nur Effizienzgewinne (im Sinne von mehr, schneller, billiger); sie stellen auch Ermöglichungs- und Infrastrukturtechniken dar, die qualitativ neue Zugänge zu Forschungsobjekten und -leistungen z. B. durch hohe Datenverfügbarkeit, Rechen- und sensorische Kapazitäten sowie durch schnelle Tools (semantic mining etc. ...) erlauben. Dadurch, dass sie sich oft im Nachhinein als „multi-purpose tools“ erweisen, sind produktive, aber oft auch zweischneidige und nicht-vorhersehbare Zwecksetzungen und Nutzungskontexte mit großer Tragweite für die Anwenderseite möglich. *KI führt so in neue, auch disruptive Nutzungskontexte.* Insofern dringt KI sogar in die Zweckdimension forschenden Handelns ein, in dem sie neue Zwecksetzungen „absichtslos“ induziert.

Was sich nun auf Seiten der wissenschaftlichen Erkenntnisbildung (disruptiv) ändert bzw. abzeichnet, ist ein *Rückgang der auf Kausalität beruhenden Erklärungsleistung KI-vermittelter Forschung* hin zu einem Forschungsmodus, der auf die Erkennung von Zusammenhängen gerichtet ist, sowie auf daraus abgeleitete Vorausagen. Dieser Modus kann zwar die Generierung neuer wissenschaftlicher Hypothesen unterstützen, ist aber selbst weniger theorieabhängig. Damit nimmt die Relevanz von Experimenten für die Widerlegung (Falsifikation) oder Bestätigung wissenschaftlicher Hypothesen ab; indirekte Formen der Falsifikation und Fragen der Kohärenz von Erklärungen spielen dann eine stärkere Rolle. Die o. g. Intransparenz von Forschung im Modus des maschinellen Lernens erschwert zusätzlich die Nachvollziehbarkeit von Forschungsprozessen und -ergebnissen. Im Detail sind hier auch etwaige Verzerrungen (Biases) in den Resultaten schwerer nachzuweisen und einzuordnen als mit herkömmlichen Forschungsmethoden. Es ist daher damit

zu rechnen, dass mangels Überprüfbarkeit *Vertrauensdefizite von entsprechenden KI-basierten Forschungsleistungen* auftreten können, soweit ihnen nicht durch geeignete neuartige Validierungsverfahren entgegengewirkt werden kann.

Auf der anderen Seite wird die o. g. besondere Leistungsfähigkeit von KI-Systemen neue, bislang kaum zugängliche Forschungsfelder erschließbar machen, die insbesondere durch systemische Komplexität, Nichtlinearität und Instabilität ihrer Komponenten und Prozesse gekennzeichnet sind – und somit als schwach-kausal bezeichnet werden können. Weitergehend betrifft dies auch die Erweiterung bestehender Forschungsfelder und -gegenstände. Es handelt sich hierbei um Objekte und Systeme, die bislang als zu komplex galten, die aufgrund von hohem Energie-, Stoff- und Informationsaustausch mit ihrer Umwelt schwer zu isolieren waren und deren Dynamik insgesamt groß ist. Dies ist typisch für Biosysteme und Objekte der Lebenswissenschaften, aber auch für physikalische Systeme der Hochenergiephysik, der Kosmologie und Astronomie sowie der Festkörperphysik, ferner für technische bzw. soziotechnische Systeme.

Insgesamt stellen KI, Machine Learning und Big Data Katalysatoren der Transformation des forschenden Wissenschaftssystems in Richtung Anwendungsorientierung dar. Dabei zeigt sich ein Übergang von einem erkenntnis-, theorie- und erklärungsorientierten Regime der Forschung zu einem, das nutzenorientierter ist. Diese Transformation spiegelt sich auch in der öffentlichen und innerwissenschaftlichen Legitimationsrhetorik wissenschaftlichen Erkenntnishandelns wider. Allgemein ist diese Richtung von Forschung enger mit einer technik- und ingenieurwissenschaftlichen Wissenschaftshaltung verbunden als mit einer grundlagenorientierten. Sie ist stärker phänomenologisch und weniger fundamental ausgerichtet. Es geht ihr mehr um Technik als um Theorie, mehr um Werke als um Wahrheit, mehr um Gestaltung als um Erkennen. Dabei bleibt das erkenntnis-, theorie- und erklärungsorientierte Regime der Forschung weiterhin existent, es tritt nur gegenüber dem anwendungs- und nutzenorientierten Regime in den Hintergrund. KI, Machine Learning und Big Data fördern somit den Übergang von „Science“ zu „Techno-Science“. In gewisser Hinsicht rückt dann Wissenschaft – in aller Ambivalenz – näher an die (nun verwissenschaftlichte) Gesellschaft heran.

In der Arbeitswelt der Forschenden führt der Trend der Digitalisierung und KI-gestützten Automatisierung zu moderaten Anpassungsanforderungen an die Forschenden. Erforderliche Kompetenzen und Fragen der Work-life Balance werden von den Betroffenen zumeist als wenig kritisch wahrgenommen. KI wird vielmehr als willkommene Unterstützung und Befreiung von administrativen Routineaufgaben im Arbeitsalltag gesehen. Als kritische Punkte für verwendete KI-Systeme werden dagegen von den Betroffenen genannt: unzureichende Standards und mangelnde Transparenz (Black-Box-Probleme, Biases, fehlende Reflexion), Missbrauch und Datenschutzprobleme sowie vereinzelt auch ethische Fragen (befürchtete Autonomieverluste und Technikabhängigkeit). Insgesamt steigt für die Forschenden die psychische und physische Belastung am Computerarbeitsplatz, sofern von Arbeitgeberseite nicht stärker auf eine menschengerechte Arbeits(platz)gestaltung geachtet wird. Für den deutschen Arbeitsmarkt könnte die KI-getriebene Automatisierung und die damit einhergehende Zunahme der Globalisierung von

Wissensarbeit in der Forschung eine Abwanderung von entsprechenden Arbeitsplätzen in Niedriglohnländer nach sich ziehen.

Aus rechtswissenschaftlicher Sicht sieht sich KI-basierte Forschung insbesondere mit den Herausforderungen der Forschungsevaluation und des Datenschutzes konfrontiert. Der erste Punkt könnte in einen Konflikt zwischen KI und Big Data einerseits und fairer Forschungsevaluation andererseits münden. So würde KI-basierte Forschung durch ihre „Datenfixiertheit“ voraussichtlich neue und ggf. einseitige Benchmarks für die Evaluation von Forschungsleistungen nach sich ziehen, die Fachkulturen nicht-empirisch arbeitender Disziplinen benachteiligen könnten. Hier stellt sich die Frage, wie künftig für einen fairen Wettbewerb zwischen den verschiedenen wissenschaftlichen Disziplinen gesorgt werden kann und konkret, in welcher Weise entsprechende Evaluationsmethoden ggf. angepasst werden sollten, um den Besonderheiten der Fachkulturen weiterhin Rechnung zu tragen. Diese Position wird auch vom Wissenschaftsrat unterstützt (Wissenschaftsrat 2020).

Ein anderer Punkt betrifft den grundsätzlich bestehenden *Konflikt zwischen Wissenschaftsfreiheit und Datenschutz*: Dieser Konflikt wird sich angesichts des hohen Datenbedarfs KI-basierter Forschung einerseits und dem Gebot der Datensparsamkeit der Europäischen Datenschutz-Grundverordnung DSGVO andererseits voraussichtlich verschärfen – zumindest auf „sensiblen“ Forschungsfeldern. Zwar bietet die Zweckbindung aus der Verordnung eine mögliche Orientierung für die rechtmäßige Datenerhebung und -nutzung in der klassischen Forschung; sie wird aber z. B. dem explorativ-suchenden Modus von Big-Data-basierter KI für die Mustererkennung in unstrukturierten Datensätzen nicht gerecht. Für einen angemessenen Datenzugang in der KI-unterstützten Forschung bedarf es daher neuer Vorgaben zur besseren Ausgestaltung des Umgangs mit Forschungsdaten, die über das derzeit geltende Datenschutzrecht hinausgehen (vergl. auch Deutscher Ethikrat 2017). Hierzu sind geeignete Kriterien und Standards zu entwickeln und zu rechtfertigen.

Das Minimierungsgebot und die Zweckbindungserfordernis der DSGVO sind gleichwohl nicht grundsätzlich forschungsfeindlich, da die Verordnung Öffnungsklauseln vorsieht und in der Rechtsprechung Raum für wohlwollende Auslegungen zulässt. Dennoch wären klarere, *forschungsfreundliche Regelungen wünschbar*, die den Forschenden im Zeitalter von KI und Big Data Ungewissheiten und Auslegungsfragen rechtmäßiger Datenverwendung ersparen würden.

7.2 Wünschbare Forschungsbedingungen im Zeitalter von KI

7.2.1 Strategievorschläge an das Wissenschaftssystem

Die weitere Gestaltung des forschenden Wissenschaftssystems sollte sich an bewährten Qualitätsmaßen moderner Forschung orientieren. Dabei soll KI-basierte Forschung unter Bedingungen erfolgen, unter denen sie ihre wissenschaftlichen

Geltungsansprüche hinreichend begründet, ihre *Prämissen offenlegt* und ihre Resultate im wissenschaftlichen und ggf. auch im weiteren gesellschaftlichen Diskurs validiert. Nur so kann ein intersubjektiver Status und folglich auch eine *wissenschaftliche Robustheit* von Ergebnissen KI-basierter Forschung erreicht und sichergestellt werden. Damit wissenschaftliche Ergebnisse nachvollziehbar sind, sollen die zugrunde liegenden Daten zugänglich sein. Des Weiteren sollen die eingesetzten Berechnungsverfahren benannt werden.

Hierzu sind zunächst vertrauen-schaffende Mechanismen zu implementieren, die die technisch-induzierte Intransparenz von KI-Systemen und ihrer Beiträge zu entsprechenden Forschungsleistungen weitestgehend reduzieren. Weiterhin sollten die den in der Forschung eingesetzten KI-Systemen zugrunde liegenden *wissenschaftlich-technischen Normen* offengelegt werden, um z. B. mögliche technisch-bedingte Verzerrungen (Biases) von Forschungsergebnissen erkennbar zu machen. Schließlich sind auch flankierende Maßnahmen zur Offenlegung (und ggf. Legitimation) des jeweiligen *Erkenntnisinteresses an KI-basierter Forschung* zu ergreifen, um die Resultate in Wissenschaft und Gesellschaft besser einbetten zu können.

Entsprechende Maßnahmen sind möglichst frühzeitig zu implementieren, um die Forscher/Forscherinnen für epistemische Risiken, die mit der partiellen Intransparenz, Eigendynamik und Unkontrollierbarkeit von KI, insbesondere im Modus des Machine Learning verbunden sind, rechtzeitig zu sensibilisieren und in ihre wissenschaftlichen Interpretationen der Gesamtergebnisse einfließen zu lassen. So wird vermieden, dass sich im Umgang mit KI in der Forschung eine mögliche „Sorglosigkeit“ herausbildet und verfestigt. Den Forschenden muss bewusst werden, dass KI im Modus des maschinellen Lernens teilweise undurchschaubar ist und daher als Forschungsinstrument nur bedingt vertraut werden kann. Vor diesem Hintergrund sollte das Forschungsfeld der so genannten „erklärbaren KI“ (engl. Explainable AI (XAI)) verstärkt gefördert werden, das darauf abzielt, die Intransparenz von KI-Systemen zumindest zu reduzieren und Nachvollziehbarkeit und Interpretierbarkeit zu ermöglichen. Die Forderung nach mehr Transparenz von KI-Prozessen wird auch in außerwissenschaftlichen Kontexten erhoben (Heesen et al. 2020).

Forschungsprojekte, in denen KI-Systeme eingesetzt werden, sollten daher auch stets personelle Informatik-Kompetenz vorhalten, um den wissenschaftlichen Risiken naiver KI-Verwendung zu begegnen. Darüber hinaus sind in den jeweiligen Forschungskontexten „Gütemaße“ ggf. einzusetzender KI-Systeme festzulegen, die den Umfang ihres jeweiligen Einsatzes für projektspezifische Aufgaben sinnvollerweise bestimmen bzw. begrenzen. Auch hierfür wird informationstechnische Beratungskompetenz benötigt.

Forscher/Forscherinnen sollten sich auch im Zeitalter von KI weiterhin als Urheber des Forschungshandelns begreifen. Sie bleiben trotz „autonomer“ informationstechnischer Teilprozesse treibende Kraft in der Forschung – nicht zuletzt wegen der besonderen Reflexionserfordernisse zuweilen intransparenter KI-basierter Verfahren und im Interesse zustimmungsfähiger Wissenschaft.

7.2.2 Empfehlungen für den Ausbildungs- und Weiterbildungsbereich

Heutigen und zukünftigen Forschern/Forscherinnen sollten verpflichtende Aus- und Weiterbildungsangebote gemacht werden, die ihnen die notwendigen theoretischen und praktischen Grundkompetenzen KI-basierter Forschung inklusive entsprechender Erfahrungen und Fertigkeiten vermitteln. Hierfür sind frühzeitig entsprechende Curricula zu entwickeln oder zu ergänzen und ggf. um nicht mehr im Forschungsalltag benötigte Kompetenzen zu verschlanken.

Diese Bildungsmaßnahmen sollen nicht nur das technische und das praktische Verständnis über KI-Systeme in der Forschung stärken, sondern auch dazu beitragen, den „KI-Hype“ auf die Sachebene hin zu entmystifizieren. Sie sollen dabei die Reflexivität und Urteilsfähigkeit von Forschenden in neuartigen „kollaborativen“ KI-Umgebungen ermöglichen bzw. erhalten. *Dabei sind neue wissenschaftliche Risiken aber auch neue erkenntnisbildende Chancen durch KI-basierte Forschung genauso zu vermitteln, wie auch herkömmliche Methoden, die im Zeitalter von KI mehr an Bedeutung gewinnen können* (Statistik, Wahrscheinlichkeitstheorie etc.). Ferner sind die Forscher/Forscherinnen in diesem Zusammenhang weiterhin auf die bewährten wissenschaftskonstituierenden Normen, wie die Sicherung der Falsifizierbarkeit und Reproduzierbarkeit ihrer Forschungsergebnisse hin zu sensibilisieren und zu verpflichten.

7.2.3 Wünsche an die Regulierung von KI-basierter Forschung

Die Verfahren der Forschungsevaluation sollten auch im Zeitalter von KI und Big Data einen fairen Wettbewerb zwischen den unterschiedlichen Fachkulturen erlauben und sollten daher insbesondere wenig datengenerierende Disziplinen z. B. im Wettbewerb um öffentliche Fördermittel nicht benachteiligen. Hierzu sind wissenschaftliche Begutachtungsverfahren ggf. anzupassen oder zu ergänzen.

Umgekehrt sind für öffentlich finanzierte, datenintensive Forschungen *klarere Regelungen der Europäischen Datenschutz-Grundverordnung wünschbar, die das Minimierungsgebot und die Zweckbindung der Datenverwendung eindeutiger und forschungsfreundlicher gestalten*. Nur so lässt sich dem explorativen Forschungsmodus von KI in Verbindung mit Big Data gerecht werden.

Literatur

Deutscher Ethikrat (Hrsg) (2017) Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Stellungnahme. Deutscher Ethikrat, Berlin. <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf>. Zugriffen am 20.11.2020

- Heesen J, Grunwald A, Matzner T, Roßnagel A (2020) Ethik-Briefing. Leitfaden für eine verantwortungsvolle Entwicklung und Anwendung von KI-Systemen. Whitepaper. München: Plattform Lernende Systeme. https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG3_Whitepaper_EB_200831.pdf. Zugegriffen am 17.12.2020
- Wissenschaftsrat (2020) Zum Wandel in den Wissenschaften durch datenintensive Forschung. Drs. 8667-20. Wissenschaftsrat, Köln. https://www.wissenschaftsrat.de/download/2020/8667-20.pdf?__blob=publicationFile&v=5. Zugegriffen am 20.11.2020

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.

