Künstliche Intelligenz in der Gesellschaft



OECD

Künstliche Intelligenz in der Gesellschaft



Das vorliegende Dokument wird unter der Verantwortung des Generalsekretärs der OECD veröffentlicht. Die darin zum Ausdruck gebrachten Meinungen und Argumente spiegeln nicht zwangsläufig die offizielle Einstellung der Organisation oder der Regierungen ihrer Mitgliedstaaten wider.

Dieses Dokument sowie die darin enthaltenen Daten und Karten berühren weder den völkerrechtlichen Status von Territorien noch die Souveränität über Territorien, den Verlauf internationaler Grenzen und Grenzlinien sowie den Namen von Territorien, Städten oder Gebieten.

Die statistischen Daten für Israel wurden von den zuständigen israelischen Stellen bereitgestellt, die für sie verantwortlich zeichnen. Die Verwendung dieser Daten durch die OECD erfolgt unbeschadet des völkerrechtlichen Status der Golanhöhen, von Ost-Jerusalem und der israelischen Siedlungen im Westjordanland.

Bitte zitieren Sie diese Publikation wie folgt:
OECD (2020), *Künstliche Intelligenz in der Gesellschaft*, OECD Publishing, Paris, *https://doi.org/10.1787/6b89dea3-de*.

ISBN 978-92-64-18146-5 (Print)
ISBN 978-92-64-43961-0 (PDF)

Originaltitel: *Artificial Intelligence in Society*
Übersetzung durch den Deutschen Übersetzungsdienst der OECD.

Foto(s): Cover © Adobe Stock.

Korrigenda zu Veröffentlichungen sind verfügbar unter: *www.oecd.org/about/publishing/corrigenda.htm*.

© OECD 2020

Die Verwendung dieser Arbeiten, sei es in digitaler oder gedruckter Form, unterliegt den Nutzungsbedingungen unter: *http://www.oecd.org/termsandconditions*.

Geleitwort

Künstliche Intelligenz (KI) verändert das Gesicht unserer Volkswirtschaften. Sie ermöglicht Produktivitäts- und Effizienzsteigerungen sowie Kostensenkungen. Sie kann die Lebensqualität steigern, die Genauigkeit von Vorhersagen erhöhen und die Entscheidungsfindung verbessern. Die KI-Technologien stecken jedoch noch in den Kinderschuhen. Sie bergen noch großes Potenzial zur Bewältigung globaler Herausforderungen sowie zur Steigerung von Innovationstätigkeit und Wachstum. KI hat Auswirkungen auf alle Bereiche unserer Gesellschaft. Ihre transformative Kraft muss daher in den Dienst des Menschen und des Planeten gestellt werden.

KI weckt aber zugleich Ängste und ethische Bedenken. So stellt sich beispielsweise die Frage der Vertrauenswürdigkeit von KI-Systemen. Dabei geht es u. a. um die Gefahr einer Kodifizierung und möglichen Verstärkung bestehender unbewusster Vorurteile – etwa in Bezug auf Geschlecht oder ethnische Herkunft – oder der Verletzung von Menschenrechten oder Werten wie dem Schutz der Privatsphäre. Auch wächst die Besorgnis, dass KI-Systeme Ungleichheiten, Klimawandel, Marktkonzentrationen und digitale Gräben verstärken könnten. Kein Land und kein Akteur hat alle Antworten auf diese Fragen. Nötig sind daher internationale Zusammenarbeit und gemeinsame Lösungen verschiedener Akteure, um die Entwicklung und Nutzung von KI so zu steuern, dass sie allen zugutekommt.

Unter dem Titel *Künstliche Intelligenz in der Gesellschaft* untersucht diese Publikation die KI-Landschaft und befasst sich mit einigen zentralen Fragen für die Politik. Ziel ist es, ein gemeinsames Verständnis von künstlicher Intelligenz – aktuell und auf kurze Sicht – zu entwickeln und einen breiten Dialog über wichtige Politikfragen anzustoßen, z. B. zu Themen wie Arbeitsmarktentwicklung und Weiterbildung für das digitale Zeitalter, Datenschutz, Rechenschaftspflicht für KI-gestützte Entscheidungen, Haftung sowie Sicherheit und Gefahrenabwehr.

Diese Publikation beruht auf den Arbeiten der Sachverständigengruppe der OECD für künstliche Intelligenz (AIGO), die 2018 eingerichtet wurde, um Grundsätze zur Förderung von Innovation, Technologieeinführung und Vertrauen im KI-Bereich zu erörtern. Die Ergebnisse dieser Diskussionen flossen in die *Empfehlung des Rats zu künstlicher Intelligenz* ein, die die ersten zwischenstaatlich vereinbarten Standards zu KI enthält. Sie wurde am 22. Mai 2019 von sämtlichen OECD-Mitgliedstaaten sowie mehreren Partnerländern verabschiedet. Besonders betont wurde darin die Notwendigkeit internationaler Zusammenarbeit, um politische Rahmenbedingungen zu schaffen, die das Vertrauen in KI-Technologien festigen und ihre Einführung fördern.

Wir müssen KI-bezogene technische, ethische und rechtliche Fragen mit Blick auf die Zukunft gemeinsam lösen, um die Abstimmung von Standards und Verhaltenskodizes sowie die Vereinbarkeit der geltenden Gesetze und sonstigen Rechtsvorschriften zu gewährleisten. Angesichts des Tempos der technologischen Entwicklungen und des breiten Spektrums möglicher Anwendungen darf dabei keine Zeit verloren werden. Daher steht KI inzwischen weit oben auf der Tagesordnung nationaler und internationaler Gremien wie der G7 und der G20.

Die Verabschiedung der Empfehlung und die Einleitung eines globalen Dialogs sind wesentliche erste Schritte. Es muss jedoch noch viel mehr getan werden. Mit der Einrichtung des KI-Observatoriums der OECD wollen wir unser Analyse-, Statistik- und Politikwissen einbringen, um hier Neuland zu erschließen. Das Observatorium soll den Staaten als inklusiver Wissenshub zur KI-Politik helfen, eine verantwortungsvolle Entwicklung vertrauenswürdiger KI-Systeme im Dienst der Gesellschaft zu fördern, zu unterstützen und zu begleiten.

Die OECD wird in der kommenden Zeit ihre Anstrengungen verstärken, damit allgemeine Grundsätze in konkretes Handeln umgesetzt werden können. Wir sind entschlossen, den Ländern bei der Umsetzung der Empfehlung zu helfen, um sicherzustellen, dass die Vorteile von KI in Gesellschaft und Wirtschaft voll ausgeschöpft werden, dass alle an ihnen teilhaben und dass niemand ins Abseits gerät – heute und auch in Zukunft.



Angel Gurría

Generalsekretär der OECD

Vorwort

Diese Publikation soll helfen, ein gemeinsames Verständnis von künstlicher Intelligenz (KI) aktuell und auf kurze Sicht zu entwickeln. Sie befasst sich mit den wirtschaftlichen und gesellschaftlichen Auswirkungen von KI-Technologien und -Anwendungen sowie den sich daraus ergebenden Konsequenzen für die Politik. Dabei werden auch konkrete Anwendungsfälle beschrieben und Politikoptionen aufgezeigt. Zudem soll sie die Koordination und Kohärenz der Diskussionen fördern, die in verschiedenen internationalen Foren geführt werden, insbesondere der G7, der G20, der Europäischen Union und den Vereinten Nationen.

Die Publikation stützt sich auf die Ergebnisse der OECD-Konferenz „AI: Intelligent Machines, Smart Policies" von Oktober 2017 (http://oe.cd/ai2017), die Aktivitäten und Diskussionen, die zwischen September 2018 und Februar 2019 in der OECD-Sachverständigengruppe für künstliche Intelligenz (AIGO) stattfanden, sowie die Empfehlung des Rats der OECD zu künstlicher Intelligenz. Zugleich trägt sie zum OECD-Projekt „Going Digital" bei, zu dem u. a. die Publikation *Going Digital: Den digitalen Wandel gestalten, das Leben verbessern* erschienen ist.

Kapitel 1, „Technische Grundlagen", bietet einen historischen Überblick über die Entwicklung der künstlichen Intelligenz von der symbolischen KI der 1950er Jahre bis zu den jüngsten Errungenschaften auf dem Gebiet des maschinellen Lernens. Dabei werden auch die Arbeiten der AIGO zur Definition von KI-Systemen – Systeme, die Ergebnisse zur Beeinflussung der Umgebung vorhersagen, empfehlen oder bestimmen – und zur Beschreibung ihres Lebenszyklus vorgestellt. Kapitel 1 enthält zudem eine Klassifizierung verschiedener Forschungszweige, die Politikverantwortlichen helfen soll, KI-Trends zu verstehen und wichtige Fragen zu erkennen, denen sie sich stellen müssen.

Kapitel 2, „Das wirtschaftliche Umfeld", erörtert die Rolle von KI als neue Universaltechnologie, die die Vorhersagekosten senken und die Entscheidungsfindung verbessern kann. Um das zu ermöglichen, bedarf es ergänzender Investitionen in Daten, Kompetenzen und digitale Arbeitsabläufe. Wichtig ist zudem die Fähigkeit zur Anpassung von Organisationsabläufen. Außerdem werden Trends in Bezug auf Private-Equity-Investitionen in KI-Start-ups beleuchtet.

In Kapitel 3, „KI-Anwendungen", werden zehn Bereiche untersucht, in denen die Einführung von KI-Technologien rasch voranschreitet: Verkehr, Landwirtschaft, Finanzbranche, Marketing und Werbung, Wissenschaft, Gesundheitswesen, Strafverfolgung, Sicherheitswesen, öffentlicher Sektor sowie Augmented Reality und Virtuelle Realität. In diesen Bereichen ermöglicht KI u. a. eine effizientere Entscheidungsfindung, Kosteneinsparungen und eine bessere Ressourcenallokation.

Kapitel 4, „Überlegungen zur Politikgestaltung", befasst sich mit wichtigen Fragen, denen sich die Politik im Hinblick auf die Verbreitung von KI-Technologien stellen muss. Das Kapitel baut auf den KI-Grundsätzen der OECD auf, die im Mai 2019 verabschiedet wurden. Dabei geht es an erster Stelle um Werte: inklusives Wachstum, nachhaltige Entwicklung und Lebensqualität, menschenzentrierte Werte und Fairness, Transparenz und

Nachvollziehbarkeit, Robustheit und Sicherheit sowie Rechenschaftspflicht. Zweitens werden Elemente erörtert, durch die vertrauenswürdige KI-Systeme auf nationaler Ebene gefördert werden können: Investitionen in eine verantwortungsvolle KI-Forschung und -Entwicklung, Maßnahmen zugunsten eines digitalen Ökosystems für KI, günstige politische Rahmenbedingungen für die KI-Entwicklung, Vorbereitung der Erwerbsbevölkerung auf sich verändernde Arbeitsplatzanforderungen, insbesondere durch Kompetenzentwicklung, sowie Fortschrittsmessung.

Kapitel 5, „KI-Maßnahmen und -Initiativen", zeigt anhand konkreter Beispiele, welche wachsende Bedeutung KI in der Agenda von Akteuren auf nationaler und internationaler Ebene einnimmt. Alle diese Akteure – Staaten und zwischenstaatliche Organisationen, Unternehmen, Fachverbände, Wissenschaft, Zivilgesellschaft und Gewerkschaften – engagieren sich aktiv in Diskussionen über die Frage, wie die Entwicklung und Einführung von KI-Technologien in die richtigen Bahnen gelenkt werden kann, um der Gesellschaft als Ganzes zugutezukommen.

Dieser Bericht wurde vom OECD-Ausschuss für digitale Wirtschaft (Committee on Digital Economy Policy – CDEP) am 10. April 2019 schriftlich freigegeben und anschließend vom OECD-Sekretariat für die Veröffentlichung vorbereitet.

Dank

Die unter dem Titel *Artificial Intelligence in Society* veröffentlichte englische Originalfassung dieser Publikation wurde unter der Federführung des OECD-Ausschusses für digitale Wirtschaft (CDEP) und unter Mitwirkung seiner Arbeitsgruppen erstellt. Die im CDEP vertretenen Delegierten leisteten dabei einen wesentlichen Beitrag, indem sie Kommentare und Änderungsvorschläge beisteuerten und sich miteinander über die KI-Strategien ihrer Länder austauschten.

Die Hauptautoren waren Karine Perset, Nobuhisa Nishigata und Luis Aranda von der OECD-Abteilung Digitale Wirtschaftspolitik. Karine Perset war für die Gesamtredaktion und Koordination zuständig. Anne Carblanc, Leiterin der Abteilung digitale Wirtschaftspolitik, Andrew Wyckoff und Dirk Pilat, Leiter bzw. stellvertretender Leiter der Direktion Wissenschaft, Technologie und Innovation, gaben die Gesamtausrichtung vor und führten die Aufsicht über das Projekt. Die Recherche für Teile dieser Publikation sowie deren Redaktion übernahmen Doaa Abu Elyounes, Gallia Daor, Lawrence Pacewicz, Alistair Nolan, Elettra Ronchi, Carlo Menon und Christian Reimsbach-Kounatze. Fachleute aus verschiedenen Bereichen der OECD steuerten Anregungen und Kommentare bei, darunter Laurent Bernat, Dries Cuijpers, Marie-Agnes Jouanjean, Luke Slawomirski, Mariagrazia Squicciarini, Barbara Ubaldi und Joao Vasconcelos.

Wesentliche Beiträge leisteten ferner Taylor Reynolds und Jonathan Frankle von der MIT Internet Policy Research Initiative, Douglas Frantz, freier Mitarbeiter, Avi Goldfarb von der University of Toronto, Karen Scott von der Princeton University, der Gewerkschaftliche Beratungsausschuss bei der OECD (TUAC), Amar Ashar, Ryan Budish, Sandra Cortesi, Finale Doshi-Velez, Mason Kortz und Jessi Whitby vom Berkman Klein Center for Internet and Society an der Harvard University sowie die Mitglieder der OECD-Sachverständigengruppe für künstliche Intelligenz (AIGO). Besonders danken möchte das Reaktionsteam auch Nozha Boujemaa, Marko Grobelnik, James Kurose, Michel Morvan, Carolyn Nguyen, Javier Juárez Mojica und Matt Chensen, die seine Arbeit durch ihre wertvollen Anregungen und ihr Feedback unterstützt haben.

Darüber hinaus profitierte die Publikation von anderen Arbeiten, die derzeit in verschiedenen Bereichen der OECD laufen, so u. a. im Ausschuss für Wissenschafts- und Technologiepolitik und dessen Arbeitsgruppe zu Innovations- und Technologiepolitik, im Ausschuss für Verbraucherpolitik und seiner Arbeitsgruppe zur Sicherheit von Verbraucherprodukten, im Ausschuss für Industrie, Innovation und Unternehmertum und seiner Arbeitsgruppe Industrieanalyse, im Ausschuss für Beschäftigung, Arbeit und Sozialfragen, im Ausschuss für Bildungspolitik und in der E-Leader-Initiative des Ausschusses für öffentliche Governance sowie natürlich im Ausschuss für digitale Wirtschaft und dessen Arbeitsgruppen, insbesondere der Arbeitsgruppe Sicherheit und Datenschutz in der digitalen Wirtschaft.

Großer Dank gebührt zudem dem japanischen Ministerium für Information und Kommunikation, das dieses Projekt unterstützt hat.

Inhaltsverzeichnis

Geleitwort ... **3**

Vorwort ... **5**

Dank ... **7**

Abkürzungsverzeichnis ... **13**

Zusammenfassung ... **15**

Die jüngsten Fortschritte in der künstlichen Intelligenz wurden durch maschinelles Lernen, Big Data und wachsende Rechenkapazitäten ermöglicht ... 15
KI-Systeme erstellen Prognosen, geben Empfehlungen ab oder treffen Entscheidungen zur Umgebungsbeeinflussung ... 15
KI kann die Produktivität erhöhen und zur Lösung komplexer Probleme beitragen ... 15
KI-Investitionen und -Unternehmen verzeichnen ein rasches Wachstum ... 16
Es gibt eine Vielzahl von KI-Anwendungen – von Verkehrslösungen bis hin zu Gesundheitsanwendungen ... 16
Eine vertrauenswürdige KI ist Voraussetzung, um das Potenzial dieser Technologien voll auszuschöpfen ... 17
KI wird für alle Akteure zu einem immer wichtigeren politischen Handlungsfeld ... 17

1. Technische Grundlagen ... **19**

Eine kurze Geschichte der künstlichen Intelligenz ... 20
Was ist KI? ... 23
Lebenszyklus eines KI-Systems ... 26
KI-Forschung ... 27
Literaturverzeichnis ... 35
Anmerkungen ... 37

2. Das wirtschaftliche Umfeld ... **39**

Wirtschaftliche Aspekte der künstlichen Intelligenz ... 40
Private-Equity-Investitionen in KI-Start-ups ... 42
Allgemeinere Trends in der Entwicklung und Verbreitung von KI ... 49
Literaturverzeichnis ... 51
Anmerkungen ... 52

3. KI-Anwendungen ... **53**

KI im Verkehr: autonome Fahrzeuge ... 54
KI in der Landwirtschaft ... 59
KI im Finanzsektor ... 62
KI in Marketing und Werbung ... 66
KI in der Wissenschaft ... 68
KI im Gesundheitswesen ... 70
KI in der Strafverfolgung ... 74

KI im Sicherheitsbereich ... 77
KI im öffentlichen Sektor ... 81
KI-Anwendungen in Verbindung mit erweiterter und virtueller Realität ... 81
Literaturverzeichnis ... 83
Anmerkungen ... 92

4. Überlegungen zur Politikgestaltung ... 93

Menschenzentrierte KI ... 94
Inklusives und nachhaltiges Wachstum und Lebensqualität ... 95
Menschenzentrierte Werte und Fairness ... 96
Transparenz und Nachvollziehbarkeit ... 106
Robustheit und Sicherheit ... 111
Rechenschaftspflicht ... 116
Politikumfeld für künstliche Intelligenz ... 117
In KI-Forschung und -Entwicklung investieren ... 117
Ein digitales Ökosystem für KI fördern ... 118
Günstige Rahmenbedingungen für KI-Innovationen schaffen ... 124
Wandel der Arbeitswelt und Kompetenzentwicklung ... 125
Messung ... 133
Literaturverzeichnis ... 134
Anmerkungen ... 143

5. KI-Maßnahmen und -Initiativen ... 145

Künstliche Intelligenz für wirtschaftliche Wettbewerbsfähigkeit: Strategien und Aktionspläne ... 146
Grundsätze für KI in der Gesellschaft ... 147
Nationale Initiativen ... 149
Zwischenstaatliche Initiativen ... 164
Initiativen privater Akteure ... 169
Literaturverzeichnis ... 173
Anmerkungen ... 178

Tabellen

Tabelle 1.1. Cluster 1: ML-Anwendungen ... 30
Tabelle 1.2. Cluster 2: ML-Techniken ... 31
Tabelle 1.3. Cluster 3: ML-Verbesserung/-Optimierung ... 32
Tabelle 1.4. Cluster 4: ML-Kontext ... 34
Tabelle 2.1. Durchschnittlicher Betrag pro Transaktion, für Investitionen bis 100 Mio. USD ... 48
Tabelle 2.2. Durchschnittlicher Betrag pro Transaktion, für alle KI-Investitionen ... 48
Tabelle 3.1. KI-Start-ups in der Landwirtschaft (Auswahl) ... 61
Tabelle 4.1. Ansätze zur Erhöhung der Transparenz und Rechenschaftspflicht von KI-Systemen ... 107
Tabelle 5.1. Von verschiedenen Gremien und Akteuren aufgestellte KI-Leitlinien (nicht erschöpfende Auswahl) ... 148
Tabelle 5.2. Grundsätze für FuE im KI-Bereich ... 156
Tabelle 5.3. Grundprinzipien der IEEE-Leitlinien *Ethically Aligned Design (Version 2)* ... 169
Tabelle 5.4. Die KI-Leitsätze von Asilomar (Auszug) ... 170
Tabelle 5.5. Die KI-Politikgrundsätze des ITI ... 171
Tabelle 5.6. Die zehn wichtigsten Grundsätze für ethische künstliche Intelligenz (UNI Global Union) ... 172

Abbildungen

Abbildung 1.1. Die Anfänge der KI (1950er Jahre bis 2000) 20
Abbildung 1.2. AlphaGo – durch Selbstlernen in 40 Tagen zum weltbesten Go-Spieler 21
Abbildung 1.3. Grundstruktur eines KI-Systems 23
Abbildung 1.4. Detaillierte Darstellung eines KI-Systems 24
Abbildung 1.5. Lebenszyklus eines KI-Systems 27
Abbildung 1.6. KI und ML 28
Abbildung 1.7. Training einer Maschine mithilfe einer Computerkamera 32
Abbildung 2.1. Gesamtinvestitionen in KI-Start-ups, 2011-2017 und erstes Halbjahr 2018, geschätzte Zahlen 43
Abbildung 2.2. KI-Anteil an den Private-Equity-Investitionen in Start-ups, 2011 bis 2017 und erstes Halbjahr 2018 45
Abbildung 2.3. Private-Equity-Investitionen in KI-Start-ups mit Sitz in der Europäischen Union, 2011 bis Mitte 2018 46
Abbildung 2.4. Zahl der Private-Equity-Investitionen in KI-Start-ups, nach Standort 47
Abbildung 2.5. Umfang der Investitionstransaktionen, 2012-2017 und erstes Halbjahr 2018 48
Abbildung 3.1. Vergleich der Kosten verschiedener Verkehrsmittel mit und ohne AF-Technologie .. 54
Abbildung 3.2. Patentanmeldungen im Bereich autonomes Fahren nach Unternehmen, 2011-2016 ... 56
Abbildung 3.3. Beispiele für die Nutzung von Satellitendaten für ein besseres Monitoring 60
Abbildung 3.4. Veranschaulichung einer Gesichtserkennungssoftware 80
Abbildung 4.1. Datenvisualisierungsinstrumente zur Verbesserung der Erklärbarkeit..... 110
Abbildung 4.2. Durch eine kleine Veränderung wird ein Algorithmus so getäuscht, dass er einen Panda als Gibbon klassifiziert 114

Kästen

Kasten 1.1. Angewandte vs. allgemeine künstliche Intelligenz 22
Kasten 1.2. Teachable Machine..... 32
Kasten 2.1. Anmerkung zur Methodik 44
Kasten 3.1. KI zur Steuerung digitaler Sicherheitsrisiken in Geschäftsumgebungen 78
Kasten 3.2. Überwachung mit „intelligenten“ Kameras..... 79
Kasten 3.3. Gesichtserkennung als Instrument der Überwachung 80
Kasten 4.1. „Black-Box“-KI-Systeme lassen neue Herausforderungen entstehen..... 94
Kasten 4.2. Menschenrechte und KI..... 96
Kasten 4.3. Menschenrechtsverträglichkeitsprüfungen..... 99
Kasten 4.4. Die OECD-Leitlinien zum Datenschutz..... 102
Kasten 4.5. Nachvollziehbarkeitsprobleme durch besser gestaltete Benutzeroberflächen angehen ... 110
Kasten 4.6. Die Gefahr von Adversarial Examples für maschinelles Lernen 113
Kasten 4.7. Synthetische Daten für eine sicherere und genauere KI: autonome Fahrzeuge 115
Kasten 4.8. Neue kryptografische Werkzeuge ermöglichen datenschutzgerechte Berechnungen 121
Kasten 4.9. KI-basierte datenschutzgerechte Identitätsprüfung dank Blockchain-Technologie 122
Kasten 5.1. Wie versuchen die Länder, einen Wettbewerbsvorteil im KI-Bereich zu erreichen? 149

Abkürzungsverzeichnis

AF	Autonomes Fahren
AKI	Allgemeine künstliche Intelligenz
AIGO	Sachverständigengruppe für künstliche Intelligenz (OECD)
AIS	Autonomes Intelligentes System
AR	Augmented Reality
DSGVO	Datenschutz-Grundverordnung (Europäische Union)
ePA	Elektronische Patientenakte
EWSA	Europäischer Wirtschafts- und Sozialausschuss
FuE	Forschung und Entwicklung
GM	General Motors
HEG-KI	Hochrangige Expertengruppe für künstliche Intelligenz (Europäische Kommission)
HoME	Household Multimodal Environment
IEC	International Electrotechnical Commission
IEEE	Institute for Electrical and Electronics Engineers
IKT	Informations- und Kommunikationstechnologien
IoT	Internet der Dinge (*Internet of Things*)
IP	Geistiges Eigentum (*intellectual property*)
IPRI	Internet Policy Research Initiative (Massachusetts Institute of Technology)
ISO	Internationale Organisation für Normung
ITI	Information Technology Industry Council
KI	Künstliche Intelligenz
KMU	Kleine und mittlere Unternehmen
MINT	Mathematik, Informatik, Naturwissenschaften und Technik
MIT	Massachusetts Institute of Technology
ML	Maschinelles Lernen
MPC	Multi-Party Computation (Mehrparteienberechnung)
MPI	Max Planck Institut
NLP	Natural Language Processing (Verarbeitung natürlicher Sprache)
OAI	Office for Artificial Intelligence (Vereinigtes Königreich)

PAI	Partnership on Artificial Intelligence to Benefit People and Society
PIAAC	Internationale Vergleichsstudie der Kompetenzen Erwachsener – OECD (Programme for the International Assessment of Adult Competencies)
SAE	Society of Automotive Engineers
SDG	Ziele für Nachhaltige Entwicklung (Vereinte Nationen) (Sustainable Development Goals)
UGAI	Universal Guidelines on Artificial Intelligence
VR	Virtuelle Realität

Zusammenfassung

Die jüngsten Fortschritte in der künstlichen Intelligenz wurden durch maschinelles Lernen, Big Data und wachsende Rechenkapazitäten ermöglicht

1950 stellte Alan Turing erstmals die Frage, ob Maschinen denken können. Seitdem haben sich die technischen Grundlagen der künstlichen Intelligenz (KI) – ein Begriff, der 1956 geprägt wurde – grundlegend verändert: Auf die sogenannte symbolische KI und ihre dem menschlichen Denken nachempfundenen, logikbasierten Systeme folgte eine Phase der Ernüchterung, der „KI-Winter“ der 1970er Jahre. In den 1990er Jahren wurde dann der Schachcomputer Deep Blue entwickelt. Ab 2011 wurden bahnbrechende Fortschritte beim sogenannten maschinellen Lernen (ML) erzielt, einem auf einem statistischen Ansatz beruhenden Teilbereich der KI. Dadurch verbesserte sich die Fähigkeit von Maschinen, aus historischen Daten Prognosen abzuleiten. Zu verdanken war dies der zunehmenden Ausgereiftheit einer Modellierungstechnik des maschinellen Lernens, die unter dem Begriff „neuronale Netze“ bekannt wurde. Eine wichtige Rolle spielten zudem die größeren Datensätze und Rechenkapazitäten.

KI-Systeme erstellen Prognosen, geben Empfehlungen ab oder treffen Entscheidungen zur Umgebungsbeeinflussung

Ein KI-System, so die Definition der OECD-Sachverständigengruppe für KI (AIGO), ist

> *ein maschinenbasiertes System, das für bestimmte von Menschen definierte Ziele Vorhersagen anstellen, Empfehlungen abgeben oder Entscheidungen treffen kann. Es nutzt maschinelle und/oder von Menschen generierte Inputs, um ein reales und/oder virtuelles Umfeld zu erfassen, davon ausgehend (automatisch, z. B. mithilfe von ML, oder manuell) Modelle zu erstellen und mittels Modellinferenz Informations- oder Handlungsoptionen zu ermitteln. KI-Systeme können mit einem unterschiedlichen Grad an Autonomie ausgestattet sein.*

Der Lebenszyklus eines KI-Systems besteht aus: 1. Planung und Design, Datensammlung und -verarbeitung sowie Modellierung und Auswertung, 2. Verifizierung und Validierung, 3. Einführung sowie 4. Betrieb und Monitoring. Die KI-Forschung unterteilt sich einer gängigen Klassifizierung zufolge in die Entwicklung von KI-Anwendungen (z. B. Natural Language Processing – NPL), Techniken zum Trainieren von KI-Systemen (z. B. neuronalen Netzen), Optimierungsarbeiten (z. B. One-Shot-Learning) und Untersuchungen zu gesellschaftlichen Fragen (z. B. Transparenz).

KI kann die Produktivität erhöhen und zur Lösung komplexer Probleme beitragen

KI entwickelt sich zu einer Universaltechnologie und verändert damit auch den wirtschaftlichen Kontext. Durch kostengünstigere und präzisere Prognosen, Empfehlungen und Entscheidungen verspricht KI die Produktivität zu steigern, die Lebensqualität zu erhöhen und zur Bewältigung komplexer Herausforderungen beizutragen. Um das Potenzial von KI

zu erschließen, bedarf es ergänzender Investitionen in Daten, Kompetenzen und digitalisierte Arbeitsabläufe sowie Änderungen der organisatorischen Abläufe. Daher wird in den einzelnen Unternehmen und Branchen in unterschiedlichem Maße auf KI zurückgegriffen.

KI-Investitionen und -Unternehmen verzeichnen ein rasches Wachstum

Die Private-Equity-Investitionen in KI-Start-ups haben seit 2016 stark angezogen, nachdem sie zuvor bereits fünf Jahre kontinuierlich gestiegen waren. Im Zeitraum 2016-2017 verdoppelten sie sich auf 16 Mrd. USD. Im ersten Halbjahr 2018 mobilisierten im KI-Bereich tätige Start-ups 12 % der weltweiten Private-Equity-Investitionen. Dies ist ein beträchtlicher Anstieg gegenüber 2011, als dieser Anteil lediglich 3 % betrug. Zudem war dieser Trend in allen großen Volkswirtschaften zu beobachten. Bei diesen Investitionen handelt es sich in der Regel um große Summen von mehreren Millionen US-Dollar. Da die Technologien und Geschäftsmodelle zunehmend ausgereift sind, steuert KI auf eine breite Einführung zu.

Es gibt eine Vielzahl von KI-Anwendungen – von Verkehrslösungen bis hin zu Gesundheitsanwendungen

Besonders rasch werden KI-Anwendungen in Branchen eingeführt, in denen sie zur Mustererkennung in großen Datenmengen und zur Modellierung komplexer, interdependenter Systeme genutzt werden können, um Entscheidungsprozesse zu verbessern und Kosten einzusparen.

- Im Verkehrssektor versprechen autonome Fahrzeuge mit virtuellen Fahrersystemen, HD-Karten und optimierten Verkehrsrouten Vorteile im Hinblick auf Kosten, Sicherheit, Lebensqualität und Umwelt.
- In der wissenschaftlichen Forschung wird KI genutzt, um große Datenmengen zu sammeln und zu verarbeiten, um Experimente zu reproduzieren und ihre Kosten zu senken und um den Prozess der wissenschaftlichen Entdeckung zu beschleunigen.
- Im Gesundheitsbereich erleichtern KI-Systeme die Diagnose und Prävention von Krankheiten, die frühzeitige Bewältigung von Krankheitsausbrüchen sowie die Entwicklung von Therapien und Arzneimitteln; sie machen zielgenaue Interventionen möglich und gestatten die Beobachtung der eigenen Körperdaten mithilfe von Self-Trackern.
- In der Strafverfolgung kommt KI z. B. im Predictive Policing („vorausschauende Polizeiarbeit“) und bei der Bewertung des Rückfallrisikos von Straftätern zum Einsatz.
- In digitalen Sicherheitsanwendungen werden KI-Systeme zur automatisierten Erkennung und Abwehr von Bedrohungen genutzt, was zunehmend in Echtzeit geschieht.
- In der Landwirtschaft gibt es u. a. KI-Anwendungen zur Überwachung der Pflanzen- und Bodengesundheit sowie für Prognosen über die Auswirkungen von Umweltfaktoren auf den Ernteertrag.
- Im Bereich der Finanzdienstleistungen wird KI zur Betrugsaufdeckung, Kreditwürdigkeitsbewertung, Senkung der Kosten der Kundenbetreuung, Automatisierung des Handels und Förderung der Rechtskonformität eingesetzt.

- In Marketing und Werbung dient KI zur Auswertung von Daten zum Kundenverhalten, um Inhalte, Werbung, Produkte, Empfehlungen und Preise zielgenau abzustimmen und zu personalisieren.

Eine vertrauenswürdige KI ist Voraussetzung, um das Potenzial dieser Technologien voll auszuschöpfen

KI bringt Vorteile, wirft jedoch auch Fragen für die Politik auf. Es muss sichergestellt werden, dass KI-Systeme vertrauenswürdig und menschenzentriert sind. KI und insbesondere manche Formen maschinellen Lernens werfen neue Ethik- und Gerechtigkeitsfragen auf. Im Mittelpunkt steht dabei die Achtung der Menschenrechte und der demokratischen Werte. Zudem muss dem Risiko begegnet werden, dass Voreingenommenheiten bzw. verzerrte Wahrnehmungen, sogenannte Biases, aus der analogen in die digitale Welt übertragen werden. Manche KI-Systeme sind so komplex, dass ihre Entscheidungen u. U. nicht erklärt werden können. Daher ist es von zentraler Bedeutung, Systeme zu entwickeln, die KI transparent nutzen, und über deren Ergebnisse Rechenschaft abgelegt werden kann. KI-Systeme müssen ordnungsgemäß und sicher funktionieren.

Es bedarf nationaler Maßnahmen zur Förderung vertrauenswürdiger KI-Systeme, einschließlich Anreizen für Investitionen in verantwortungsvolle Forschung und Entwicklung im KI-Bereich. Zusätzlich zu Technologielösungen und großen Rechenkapazitäten benötigt KI auch große Datenmengen. Dringend notwendig ist daher ein digitales Umfeld, das den Zugriff auf Daten ermöglicht, zugleich aber ein hohes Maß an Datenschutz und Schutz der Privatsphäre gewährleistet. KI-freundliche Rahmenbedingungen können außerdem kleine und mittlere Unternehmen bei der Einführung von KI unterstützen und für ein wettbewerbsorientiertes Umfeld sorgen.

KI-Technologien lösen einen Wandel in der Arbeitswelt aus, da sie menschliche Arbeitsleistungen ersetzen und verändern. Daher müssen Maßnahmen ergriffen werden, die gegebenenfalls den Arbeitsplatzwechsel erleichtern und Fort- und Weiterbildung sowie Kompetenzentwicklung sicherstellen.

KI wird für alle Akteure zu einem immer wichtigeren politischen Handlungsfeld

In Anbetracht der Vorteile und Risiken, die mit dem tiefgreifenden, durch KI angestoßenen Wandel einhergehen, wird KI für alle Akteure zu einem immer wichtigeren politischen Handlungsfeld. Viele Länder verfügen über spezielle KI-Strategien: Ziel ist es, KI als Motor für Wachstum und Lebensqualität zu nutzen, die Ausbildung und Rekrutierung wissenschaftlicher Nachwuchskräfte zu verbessern und optimale Lösungen für mit KI verbundene Herausforderungen zu entwickeln. Auch nichtstaatliche Akteure – Unternehmen, Fachorganisationen, Universitäten, Zivilgesellschaft und Gewerkschaften – sowie internationale Foren wie die G7, die G20, die OECD, die Europäische Kommission und die Vereinten Nationen werden in diesem Bereich aktiv.

Im Mai 2019 wurden unter Leitung einer Multi-Stakeholder-Sachverständigengruppe die OECD-Grundsätze für künstliche Intelligenz verabschiedet. Dabei handelt es sich um die ersten auf Regierungsebene vereinbarten, internationalen Standards für eine verantwortungsvolle Steuerung vertrauenswürdiger KI.

1. Technische Grundlagen

Dieses Kapitel beschäftigt sich mit den technischen Grundlagen künstlicher Intelligenz (KI). Diese haben sich tiefgreifend verändert, seit Alan Turing 1950 erstmals die Frage stellte, ob Maschinen denken können. Seit 2011 wurden insbesondere beim sogenannten maschinellen Lernen (ML) bahnbrechende Fortschritte erzielt. ML ist ein auf einem statistischen Ansatz beruhender Teilbereich der KI, bei dem Maschinen aus historischen Daten lernen, um in neuen Situationen Vorhersagen zu machen. Die aktuelle KI-Welle verdankt sich ausgereiften Techniken des maschinellen Lernens, großen Datensätzen und immer größeren Rechenkapazitäten. Das Kapitel enthält eine detaillierte Beschreibung von KI-Systemen, d. h. Systemen, die zur Umgebungsbeeinflussung Vorhersagen machen, Empfehlungen abgeben oder Entscheidungen treffen. Dabei wird auch der typische Lebenszyklus von KI-Systemen skizziert, der sich in verschiedene Phasen gliedert: 1. Design, Daten und Modelle (was Planung und Design, Datensammlung und -verarbeitung sowie Modellierung und Interpretation umfasst), 2. Verifizierung und Validierung, 3. Einführung sowie 4. Betrieb und Monitoring. Abschließend wird eine Klassifizierung der KI-Forschungsbereiche vorgeschlagen, die politischen Entscheidungsträgern Orientierungshilfen bieten soll.

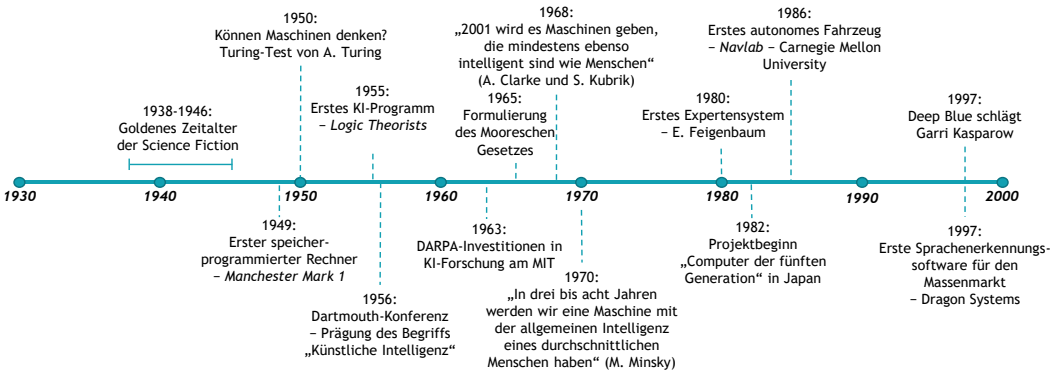
Eine kurze Geschichte der künstlichen Intelligenz

Im Jahr 1950 veröffentlichte der britische Mathematiker Alan Turing einen Artikel zum Thema Rechenmaschinen und Intelligenz (Turing, 1950[1]), in dem er der Frage nachging, ob Maschinen denken können. Zur Beantwortung dieser Frage entwickelte er ein einfaches Testverfahren: Kann ein Computer ein Gespräch führen und Fragen so beantworten, dass sein menschliches Gegenüber den Eindruck gewinnt, er hätte es mit einem anderen Menschen zu tun?[1] Dieser Turing-Test wird auch heute noch verwendet. Ebenfalls 1950 schlug Claude Shannon vor, eine Maschine zu entwickeln, der man das Schachspielen beibringen könne (Shannon, 1950[2]). Trainiert werden könne eine solche Maschine mit der sogenannten Brute-Force-Methode, die auf der Berechnung aller möglichen Züge beruht, oder durch eine Bewertung einiger weniger strategischer Schachzüge des Gegners (UW, 2006[3]).

Als eigentliche Geburtsstunde der künstlichen Intelligenz (KI) erachten viele die Konferenz „Dartmouth Summer Research Project", die im Sommer 1956 stattfand. John McCarthy, Alan Newell, Arthur Samuel, Herbert Simon und Marvin Minsky umrissen dort das Konzept der künstlichen Intelligenz. Seither wurden in der KI-Forschung kontinuierlich Fortschritte erzielt. Die Versprechen der KI-Pioniere erwiesen sich jedoch als zu optimistisch. Dies führte in den 1970er Jahren zum sogenannten KI-Winter, in dem die KI-Forschung bedeutende Mittel- und Attraktivitätseinbußen hinnehmen musste.

In den 1990er Jahren wurden dann allerdings große Fortschritte bei der Rechenleistung erzielt. Dies brachte neue Fördermittel und ließ das Interesse an KI wieder aufflammen (UW, 2006[3]). Abbildung 1.1 bietet einen Überblick über die Anfänge der KI.

Abbildung 1.1. Die Anfänge der KI (1950er Jahre bis 2000)

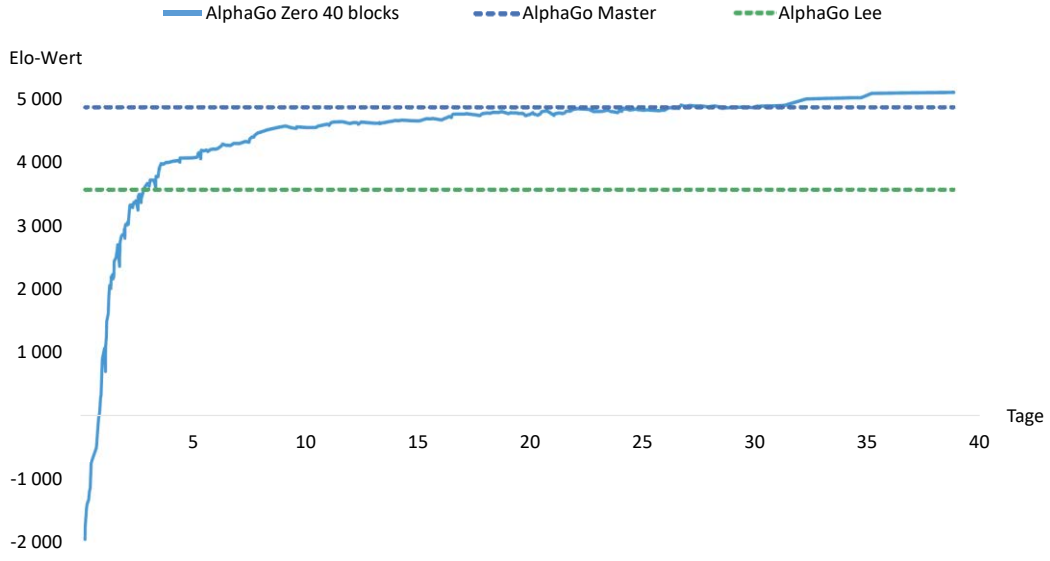


Quelle: Nach Anyoha (2017[4]), "The history of artificial intelligence", http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/.

Dank den in den 1990er Jahren bei Rechenleistung und Datenspeicherung erzielten Fortschritten konnten nun komplexe Aufgaben bewältigt werden. Damit endete der KI-Winter. Ein bedeutender Durchbruch gelang Richard Wallace 1995 mit der Entwicklung des Chatbots A.L.I.C.E (Artificial Linguistic Internet Computer Entity), der in der Lage war, einfache Gespräche zu führen. Ebenfalls in die 1990er Jahre fällt die Entwicklung des IBM-Computers Deep Blue, der gegen den Schachweltmeister Garri Kasparow antreten sollte. Dieser auf einem Brute-Force-Ansatz basierende Schachcomputer antizipierte jeweils mindestens 6 Züge und konnte 330 Millionen Stellungen pro Sekunde berechnen (Somers, 2013[5]). 1996 verlor Deep Blue noch gegen Kasparow, die Revanche im folgenden Jahr konnte er jedoch schon für sich entscheiden.

2015 entwickelte die Alphabet-Tochtergesellschaft DeepMind eine Software, die bei dem traditionsreichen Brettspiel Go gegen die besten Spieler der Welt antreten sollte. AlphaGo beruhte auf einem künstlichen neuronalen Netz, das mit tausenden Partien menschlicher Amateur- und Profispieler trainiert worden war. Im Jahr 2016 besiegte AlphaGo den damals weltbesten Go-Spieler Lee Sedol mit 4:1. Anschließend ließen die Entwickler das Programm ausgehend von einem zufälligen Spiel und einigen simplen Regeln nach dem Versuch-Irrtum-Prinzip gegen sich selbst spielen. So entstand ein neues Programm – AlphaGo Zero –, das sich selbst schneller trainierte und das Vorgängerprogramm Alpha Go mit 100:0 besiegte. Allein durch das Spielen gegen sich selbst gelang es AlphaGo Zero binnen vierzig Tagen – ohne menschliche Eingriffe oder historische Daten – sämtliche AlphaGo-Versionen zu schlagen (Silver et al., 2017[6]) (Abbildung 1.2).

Abbildung 1.2. AlphaGo – durch Selbstlernen in 40 Tagen zum weltbesten Go-Spieler



Quelle: Nach Silver et al. (2017[6]) „Mastering the game of Go without human knowledge", http://dx.doi.org/10.1038/nature24270.

Aktueller Stand der Technik

Durch Big Data, Cloud Computing, die damit verbundene Rechen- sowie Speicherkapazität und die bahnbrechenden Fortschritte im Bereich des maschinellen Lernens haben sich die Leistung, Verfügbarkeit, Expansion und Wirkung von KI drastisch erhöht.

Dank des stetigen technischen Fortschritts können zudem bessere und günstigere Sensoren entwickelt werden, die zuverlässigere Daten für KI-Systeme liefern. Außerdem stehen immer mehr Daten für KI-Systeme zur Verfügung, da diese Sensoren kleiner und kostengünstiger werden. Dies ermöglicht beträchtliche Fortschritte in vielen zentralen KI-Forschungsbereichen wie:

- maschinelle Sprachverarbeitung
- autonome Fahrzeuge und Robotik
- maschinelles Sehen
- Sprachenlernen

Einige der interessantesten KI-Entwicklungen sind nicht in der Informatik, sondern in Bereichen wie Gesundheit, Medizin, Biologie und Finanzwirtschaft zu beobachten. Der Übergang zu KI erinnert in mehrfacher Hinsicht an die Einführung der Computer, die anfangs nur in einigen wenigen spezialisierten Unternehmen eingesetzt wurden, in den 1990er Jahren dann aber überall in Wirtschaft und Gesellschaft Einzug hielten. Ähnliches war auch beim Internetzugang zu beobachten: Nutzten zunächst nur einige multinationale Unternehmen das Internet, war dies in den 2000er Jahren in vielen Ländern bald schon für den Großteil der Bevölkerung der Fall. Mit der zunehmenden Verbreitung von KI wird auch der volkswirtschaftliche Bedarf an branchenspezifischer „Zweisprachigkeit" steigen, d. h. der Bedarf an Personen, die auf einen Fachbereich (z. B. Wirtschaft, Biologie oder Recht) spezialisiert sind, aber auch KI-Techniken wie ML beherrschen. Im Fokus dieses Kapitels stehen nicht etwa mögliche längerfristige Entwicklungen wie eine allgemeine künstliche Intelligenz (AKI), sondern Anwendungen, die bereits genutzt werden oder kurz- bzw. mittelfristig zu erwarten sind (Kasten 1.1).

Kasten 1.1. Angewandte vs. allgemeine künstliche Intelligenz

Die angewandte bzw. schwache künstliche Intelligenz dient der Bewältigung konkreter Problemlöse- und Analyseaufgaben. Sie ist Stand der Technik. Gegenwärtig sind selbst die fortschrittlichsten KI-Systeme wie Googles AlphaGo dieser Kategorie zuzuordnen. Solche Systeme können Muster erkennen und das dabei gewonnene Wissen bis zu einem gewissen Grad verallgemeinern, indem sie z. B. Ergebnisse der Bilderkennung auf die Spracherkennung übertragen. Der menschliche Verstand ist jedoch wesentlich vielseitiger.

Die angewandte KI wird häufig von der (hypothetischen) allgemeinen bzw. starken künstlichen Intelligenz (AKI) abgegrenzt. Starke künstliche Intelligenz würde Maschinen zu Autonomie und allgemeinem intelligentem Handeln befähigen. Maschinen könnten dann so wie Menschen durch diverse kognitive Funktionen gewonnene Erkenntnisse verallgemeinern bzw. abstrahieren. Dies setzt ein starkes assoziatives Gedächtnis und Urteils- sowie Entscheidungsfähigkeit voraus. Allgemeine künstliche Intelligenz könnte vielschichtige Probleme lösen, wahrnehmungs- und erfahrungsbasiert lernen, Konzepte entwickeln, sich selbst und die Welt wahrnehmen, erfinden, kreieren, in komplexen Umgebungen auf Unerwartetes reagieren und antizipieren. Bei der Frage, ob eine allgemeine künstliche Intelligenz möglich ist, gehen die Meinungen indessen stark auseinander. Experten unterstreichen, dass in diesbezüglichen Diskussionen von einem realistischen Zeithorizont ausgegangen werden sollte. Weitgehende Einigkeit besteht darüber, dass die angewandte künstliche Intelligenz große Chancen, Risiken und Herausforderungen mit sich bringen wird. Dass diese Herausforderungen durch die Entwicklung einer allgemeinen künstlichen Intelligenz im Lauf des 21. Jahrhunderts maßgeblich verstärkt würden, ist ebenfalls Konsens.

Quelle: OECD (2017[7]), *OECD Digital Economy Outlook 2017*, http://dx.doi.org/10.1787/9789264276284-en.

Was ist KI?

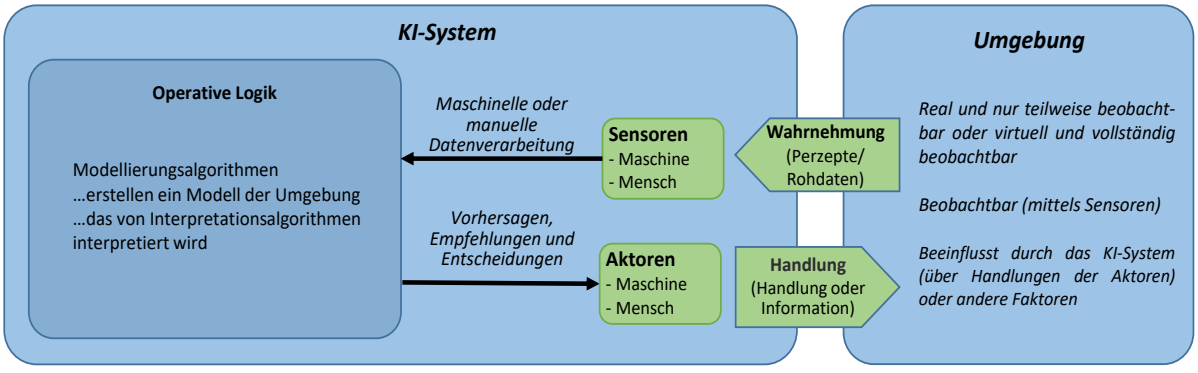
Eine allgemein anerkannte Definition von KI gibt es nicht. Die OECD-Sachverständigengruppe für KI (AI Group of Experts at the OECD – AIGO) beauftragte im November 2018 eine Arbeitsgruppe damit zu beschreiben, was ein KI-System ausmacht. Diese Beschreibung sollte verständlich, technisch korrekt und technologieneutral sein und auf lange Sicht Gültigkeit beanspruchen können. Das Resultat dieser Arbeiten war eine Beschreibung, die hinreichend allgemein ist, um einer Vielzahl der in Wissenschaft, Wirtschaft und Politik gebräuchlichen KI-Definitionen gerecht zu werden. Auf dieser Beschreibung basiert auch die *OECD-Empfehlung zu künstlicher Intelligenz* (OECD, 2019[8]).

KI-Systeme – Begriffsbestimmung

Der nachstehenden Beschreibung eines KI-Systems liegt das in der Studie *Künstliche Intelligenz: Ein moderner Ansatz* (Russel, S. und P. Norvig, 2009[9]) erläuterte Begriffsverständnis von KI zugrunde. Es steht im Einklang mit der gängigen Definition von KI als „Studium der Berechnungen, die Wahrnehmung, Schlussfolgerung und Handeln ermöglichen“ (Winston, 1992[10]) bzw. mit ähnlich allgemeinen Definitionen (Gringsjord, S. und N. Govindarajulu, 2018[11]).

In einem ersten Schritt soll hier der KI-Begriff anhand der Grundstruktur eines klassischen KI-Systems (bzw. „intelligenten Agenten“) abgesteckt werden (Abbildung 1.3). Ein KI-System besteht aus drei Hauptelementen, aus Sensoren, einer operativen Logik und Aktoren. Die Sensoren sammeln Rohdaten über die Umgebung und die Aktoren beeinflussen den Zustand der Umgebung. Die operative Logik ist das wichtigste Element eines KI-Systems. Sie liefert für bestimmte Ziele ausgehend vom Daten-Input der Sensoren Output für die Aktoren. Dies können Empfehlungen, Vorhersagen oder Entscheidungen zur Beeinflussung der Umgebung sein.

Abbildung 1.3. Grundstruktur eines KI-Systems

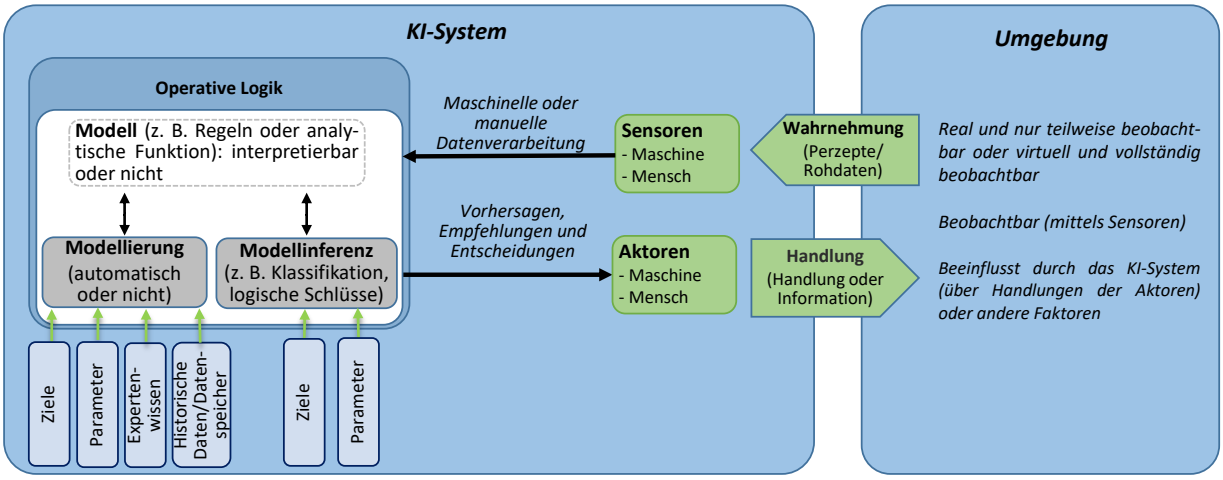


Quelle: Wie von der Sachverständigengruppe AIGO im Februar 2019 festgelegt und beschlossen.

In der nachstehenden detaillierteren Darstellung sind die wichtigsten politikrelevanten Elemente eines KI-Systems erfasst (Abbildung 1.4). Um verschiedenen Arten von KI-Systemen und Szenarien Rechnung zu tragen, wird in der Abbildung zwischen dem Prozess der Modellierung (z. B. durch ML) und dem Modell an sich differenziert. Der Modellierungsprozess ist zudem vom Prozess der Modellinterpretation zu unterscheiden, bei dem

mithilfe des jeweiligen Modells Vorhersagen gemacht, Empfehlungen abgegeben oder Entscheidungen getroffen werden. Auf diese Outputs stützen sich die Aktoren, um die Umgebung zu beeinflussen.

Abbildung 1.4. Detaillierte Darstellung eines KI-Systems



Quelle: Wie von der Sachverständigengruppe AIGO im Februar 2019 festgelegt und beschlossen.

Umgebung

Die Umgebung eines KI-Systems ist ein Raum, der (mittels Sensoren) beobachtet und durch Handlungen (von Aktoren) beeinflusst werden kann. Bei den Sensoren und Aktoren kann es sich um Maschinen oder um Menschen handeln. Eine Umgebung ist entweder real (z. B. physisch, sozial, mental) und dann in der Regel nur teilweise beobachtbar oder virtuell (z. B. Brettspiele) und dann für gewöhnlich vollständig beobachtbar.

KI-System

Ein KI-System ist ein maschinenbasiertes System, das für bestimmte von Menschen definierte Ziele Vorhersagen machen, Empfehlungen abgeben oder Entscheidungen treffen kann, um eine reale oder virtuelle Umgebung zu beeinflussen. Dabei bilden maschinelle und/oder von Menschen vorgegebene Inputs die Grundlage für 1. die Erfassung realer und/oder virtueller Umgebungen, 2. die automatische (z. B. durch ML) oder manuelle Erstellung von Modellen auf Basis dieser Beobachtungen mithilfe von Analysen und 3. die Ermittlung von Informations- oder Handlungsoptionen mittels Modellinferenz. KI-Systeme können mit einem unterschiedlichen Grad an Autonomie ausgestattet sein.

Modell, Modellierung und Interpretation

Kernstück eines KI-Systems ist das KI-Modell, das die Struktur und/oder Dynamik der gesamten oder eines Teils der Umgebung des Systems abbildet. Es basiert auf Expertenwissen und/oder Daten, die von Menschen und/oder automatisierten Instrumenten (z. B. ML-Algorithmen) bereitgestellt werden. Bei der Modellierung werden außerdem bestimmte Ziele (z. B. Outputvariablen) und Parameter (z. B. Genauigkeit, Trainingsressourcen, Repräsentativität des Datensatzes) berücksichtigt. Als Modellinferenz bezeichnet man das Verfahren, mit dem Menschen und/oder automatisierte Instrumente aus einem Modell

Ergebnisse ableiten. Dabei kann es sich um Empfehlungen, Vorhersagen oder Entscheidungen handeln. Die Ausführung orientiert sich ebenfalls an Zielen und Parametern. Manche Modelle (z. B. deterministische) liefern eine einzige Empfehlung, andere (z. B. probabilistische) bieten mehrere Empfehlungen an. Diese entsprechen dann unterschiedlichen Niveaus bestimmter Parameter wie Zuverlässigkeit, Robustheit oder Risiko. In einigen Fällen lässt sich beim Interpretationsprozess klären, warum bestimmte Empfehlungen gemacht wurden. In anderen Fällen ist eine Erklärung so gut wie unmöglich.

Beispiele für KI-Systeme

Kredit-Scoring-System

Ein Kredit-Scoring-System ist ein maschinenbasiertes System, das die Umgebung beeinflusst (dahingehend, ob jemandem ein Kredit gewährt wird). Es gibt Empfehlungen (Kredit-Scores) im Hinblick auf bestimmte Ziele (Kreditwürdigkeit) ab. Dazu stützt sich das System sowohl auf maschinelle Inputs (historische Daten zu Personen bzw. deren Kreditrückzahlungen) als auch auf von Menschen vorgegebene (Regeln). Ausgehend von diesen Inputs erfasst das System reale Umgebungen (ob Personen gegenwärtig ihre Kredite zurückzahlen). Auf Basis dieser Beobachtungen werden dann automatisch Modelle erstellt. Ein Kredit-Scoring-Algorithmus könnte z. B. auf einem statistischen Modell beruhen. Mittels Modellinferenz (Kredit-Scoring-Algorithmus) wird schließlich eine Empfehlung (Kredit-Score) im Hinblick auf die Handlungsoptionen (Kreditgewährung oder -ablehnung) abgegeben.

Assistent für Menschen mit Sehbehinderung

Anhand eines Assistenten für Menschen mit Sehbehinderung lässt sich ebenfalls veranschaulichen, wie maschinenbasierte Systeme die Umgebung beeinflussen. Ein solcher Assistent liefert Empfehlungen (z. B. wie einem Hindernis ausgewichen oder die Straße überquert werden kann) im Hinblick auf bestimmte Ziele (Bewegung von einem Ort zu einem anderen). Dies geschieht in drei Schritten, bei denen maschinelle und/oder von Menschen generierte Inputs (große Bilddatenbanken mit getaggten Objekten, Wörtern und sogar Gesichtern) genutzt werden: Zunächst werden Bilder der Umgebung erfasst (eine Kamera fotografiert, was sich vor der betreffenden Person befindet, und schickt das Bild an eine Anwendung). Auf Basis dieser Bilder werden automatisch Modelle erstellt (Objekterkennungsalgorithmen, die Verkehrsampeln, Fahrzeuge oder Hindernisse auf Gehwegen erkennen können). Mittels Modellinferenz werden dann Empfehlungen zu Handlungsoptionen formuliert (Audiobeschreibung der in der Umgebung registrierten Objekte). Der Assistent beeinflusst die Umgebung also, indem er seinem Nutzer Informationsgrundlagen für Entscheidungen über sein Verhalten liefert.

AlphaGo Zero

AlphaGo Zero ist ein KI-System, das das Brettspiel Go besser beherrscht als professionelle Go-Spieler. Ein Brettspiel ist eine virtuelle und vollständig beobachtbare Umgebung. Die Zahl möglicher Spielpositionen ist durch die Ziele und Regeln des Spiels begrenzt. Das System basiert sowohl auf von Menschen vorgegebenen Inputs (Go-Regeln) als auch auf maschinellen (Selbsttraining durch wiederholtes Spielen gegen sich selbst, ausgehend von einem zufälligen Spiel). Auf Basis der beim Spielen gewonnenen Daten wird ein (stochastisches) Modell von Handlungen (Spielzügen) erstellt und mittels Reinforcement-Learning bzw. bestärkendem Lernen trainiert. Auf dieses Modell stützt sich das System dann, um je nach Spielstand die passenden Züge vorzuschlagen.

Autonome Fahrsysteme

Auch ein autonomes Fahrsystem ist ein maschinenbasiertes System, das die Umgebung beeinflussen kann (dahingehend, ob ein Fahrzeug beschleunigt, bremst oder abbiegt). Es macht Vorhersagen (darüber, ob ein Objekt oder Schild ein Hindernis darstellt oder Informationen liefert) und/oder trifft Entscheidungen (beschleunigen, bremsen usw.) im Hinblick auf bestimmte Ziele (in möglichst kurzer Zeit sicher von A nach B gelangen). Dabei stützt sich das System sowohl auf maschinelle Inputs (historische Fahrdaten) als auch auf menschliche (Verkehrsregeln). Auf Basis dieser Inputs wird ein Modell des Fahrzeugs und seiner Umgebung erstellt. Dies ermöglicht dann 1. die Erfassung realer Umgebungen (mithilfe von Sensoren wie Kameras oder Sonargeräten), 2. die automatische Erstellung von Modellen (einschließlich Objekterkennung, Ermittlung von Geschwindigkeit und Bewegungsbahn sowie standortbezogener Daten) auf Basis dieser Beobachtungen sowie 3. Modellinferenz. Der Selbstfahr-Algorithmus kann z. B. aus zahlreichen Simulationen kurzfristiger Zukunftsszenarien für Fahrzeug und Umgebung bestehen. Dadurch kann das System Handlungsoptionen empfehlen (halten oder weiterfahren).

Lebenszyklus eines KI-Systems

Im November 2018 richtete die AIGO eine Arbeitsgruppe ein, die den Lebenszyklus eines KI-Systems für die OECD-Ratsempfehlung zu künstlicher Intelligenz (OECD, 2019[8]) im Detail beschreiben sollte. Dieses Lebenszyklusmodell stellt keinen neuen Standard für den Lebenszyklus von KI-Systemen[2] dar und ist nicht präskriptiv. Es kann jedoch zur Kontextualisierung anderer internationaler Initiativen zur Erarbeitung von KI-Grundsätzen beitragen.[3]

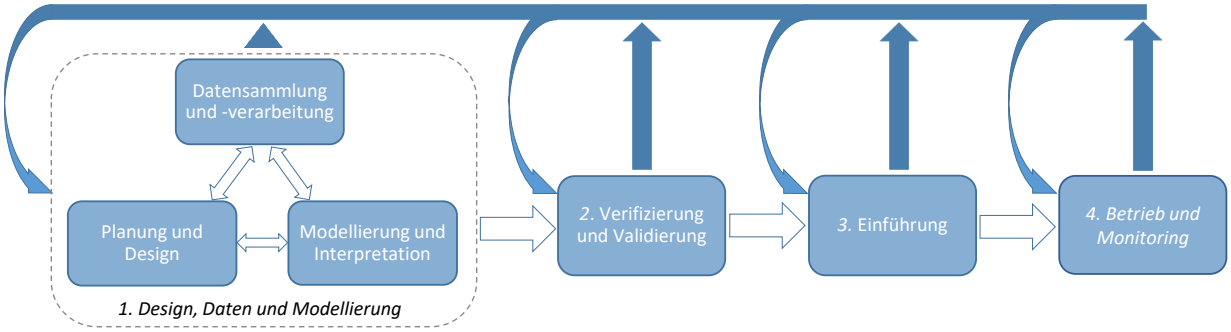
Der Lebenszyklus eines KI-Systems umfasst mehrere Phasen, die auch für den Lebenszyklus traditioneller Software und Systeme kennzeichnend sind. Beim Lebenszyklus eines KI-Systems werden in der Regel vier Phasen unterschieden: Die Phase „Design, Daten und Modelle“ ist eine kontextabhängige Abfolge der Schritte „Planung und Design“, „Datensammlung und -verarbeitung“ sowie „Modellierung und Interpretation“. Daran schließen sich die Phasen „Verifizierung und Validierung“, „Einführung“ sowie „Betrieb und Monitoring“ an (Abbildung 1.5. Lebenszyklus eines KI-Systems). Diese Phasen sind häufig iterativ und nicht zwangsläufig sequenziell. Die Entscheidung, ein KI-System außer Betrieb zu nehmen, kann zu jedem Zeitpunkt der Betriebs- oder Monitoringphase erfolgen.

Die Lebenszyklusphasen eines KI-Systems lassen sich wie folgt beschreiben:

1. **Design, Daten** und **Modellierung** besteht aus mehreren Schritten, deren Abfolge variieren kann:
    - **Planung und Design des KI-Systems:** Beschreibung von Konzept, Zielen, Prämissen, Kontext und Anforderungen sowie eventuell Entwicklung eines Prototyps.
    - **Datensammlung und -verarbeitung**: Sammlung und Bereinigung der Daten, Vollständigkeits- und Qualitätsprüfungen sowie Dokumentation von Metadaten bzw. Merkmalen des Datensatzes. Hierzu zählen Angaben zur Erstellung, Zusammensetzung, beabsichtigten Verwendung und Pflege des Datensatzes.
    - **Modellierung und Interpretation**: Erstellung oder Auswahl von Modellen bzw. Algorithmen, deren Kalibrierung und/oder Training sowie Interpretation.

2. **Verifizierung und Validierung** umfasst die Ausführung und Optimierung der Modelle. Dabei werden Tests durchgeführt, um die Leistungsfähigkeit in verschiedenen Bereichen und unter mehreren Gesichtspunkten zu prüfen.
3. **Einführung** im Produktivbetrieb setzt Pilotprojekte, eine Überprüfung der Kompatibilität mit bestehenden Systemen, die Gewährleistung der Einhaltung der geltenden Vorschriften, organisatorische Veränderungen sowie eine Evaluierung des Nutzererlebnisses voraus.
4. **Betrieb und Monitoring** eines KI-Systems umfasst den Betrieb des KI-Systems sowie die laufende Evaluierung seiner Empfehlungen und (beabsichtigten und unbeabsichtigten) Effekte unter Berücksichtigung der Ziele sowie ethischer Gesichtspunkte. In dieser Phase sollen Probleme erkannt werden, zu deren Behebung dann Anpassungen in früheren Phasen vorgenommen werden können; unter Umständen kann es nötig sein, das System aus dem Produktivbetrieb zu nehmen.

Abbildung 1.5. Lebenszyklus eines KI-Systems



Quelle: Wie von der Sachverständigengruppe AIGO im Februar 2019 festgelegt und beschlossen.

Die zentrale Bedeutung von Daten bzw. von Modellen, für deren Training und Evaluierung Daten erforderlich sind, unterscheidet den Lebenszyklus vieler KI-Systeme von dem traditioneller Systeme. Einige ML-basierte KI-Systeme können zudem Iterationen durchlaufen und sich im Lauf der Zeit weiterentwickeln.

KI-Forschung

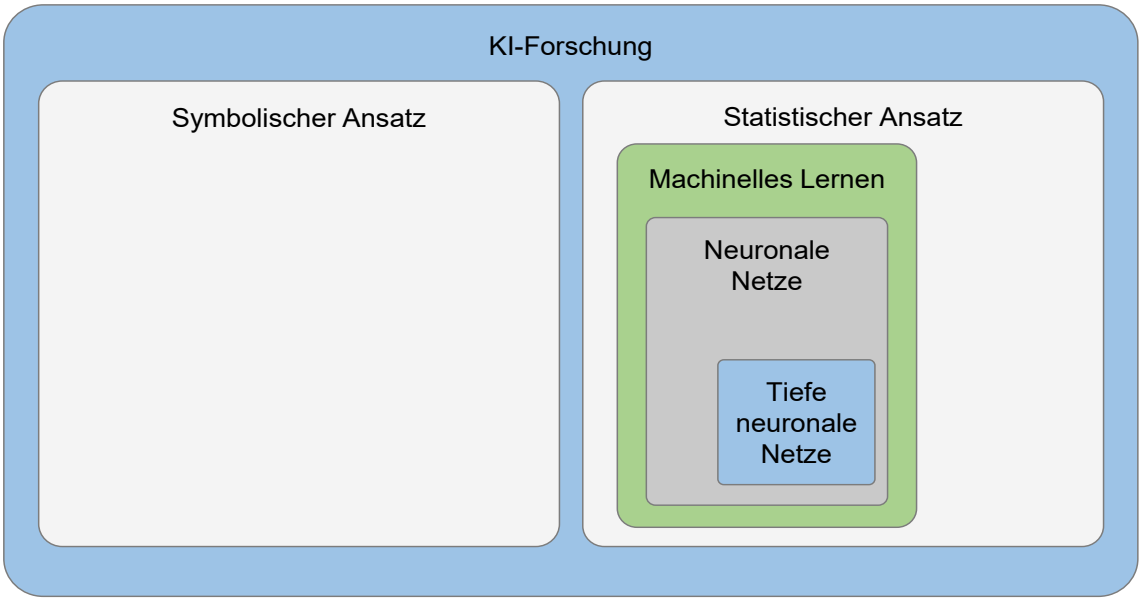
Dieser Abschnitt beschäftigt sich mit Entwicklungen in der KI-Forschung an Hochschulen und im privaten Sektor, die die Einführung von KI ermöglichen. KI ist ein aktives Forschungsgebiet der Informatik, insbesondere der Teilbereich des maschinellen Lernens. KI-Techniken werden in zahlreichen wissenschaftlichen Disziplinen für eine Vielzahl von Anwendungen genutzt.

Ein allgemein anerkanntes Klassifikationssystem für KI-Forschungsbereiche, das z. B. mit jenem der 20 großen wirtschaftswissenschaftlichen Forschungsbereiche des *Journal of Economic Literature* vergleichbar wäre, gibt es bislang nicht. In diesem Abschnitt wird eine solche Klassifizierung vorgeschlagen. Sie soll Politikverantwortlichen helfen, einige der jüngsten KI-Trends zu verstehen, und Fragen aufzeigen, mit denen sich die Politik auseinandersetzen muss.

In der Forschung wurde von Anfang an zwischen symbolischer und statistischer KI unterschieden. Die symbolische KI basiert auf logischen Repräsentationen, aus denen anhand eines vorgegebenen Regelsatzes Schlussfolgerungen abgeleitet werden. Dazu müssen detaillierte und nachvollziehbare Entscheidungsstrukturen nachgebildet werden, um der Komplexität der Realität gerecht zu werden und Maschinen zu befähigen, Entscheidungen zu treffen, die denen von Menschen ähneln. Symbolische KI wird nach wie vor häufig eingesetzt, z. B. für Optimierungs- und Planungsinstrumente. In jüngerer Zeit findet die statistische KI, bei der Maschinen Muster erkennen und daraus ein Modell erstellen, immer stärkere Verbreitung. Einige Anwendungen beruhen auf Kombinationen symbolischer und statistischer Ansätze. Bei Algorithmen für die maschinelle Sprachverarbeitung (NLP-Algorithmen) z. B. werden statistische Ansätze (die sich auf große Datenmengen stützen) häufig mit symbolischen (die beispielsweise Grammatikregeln berücksichtigen) kombiniert. Kombinierte Modelle, die sowohl auf Daten als auch auf menschlicher Expertise beruhen, gelten als vielversprechende Entwicklung, mit der die Unzulänglichkeiten beider Ansätze überwunden werden können.

KI-Systeme basieren zunehmend auf maschinellem Lernen (ML). Dieser Begriff steht für eine Reihe von Techniken, bei denen Maschinen ohne explizite Programmierung anhand von Mustern und Inferenzen selbstständig lernen. Viele ML-Ansätze beruhen darauf, dass Maschinen anhand einer Vielzahl von Beispielen darauf trainiert werden, ein bestimmtes Ergebnis zu erzielen. Bei anderen Ansätzen werden Regeln vorgegeben und die Maschine lernt durch Versuch und Irrtum. ML wird im Allgemeinen bei der Modellierung oder Modellanpassung eingesetzt, kann aber auch bei der Interpretation der Ergebnisse eines Modells zur Anwendung kommen (Abbildung 1.6). Viele Techniken des maschinellen Lernens werden bereits seit Langem von Ökonomen, Forschern und Technologen genutzt. Dies gilt etwa für die lineare und logistische Regression, Entscheidungsbäume, die Hauptkomponentenanalyse oder tiefe neuronale Netze.

Abbildung 1.6. KI und ML



Quelle: Zur Verfügung gestellt von der Internet Policy Research Initiative (IPRI) des Massachusetts Institute of Technology (MIT).

Wenn in den Wirtschaftswissenschaften mit Regressionsmodellen auf Basis von Inputdaten Prognosen erstellt werden, können Forscher die Koeffizienten (Gewichte) der Inputvariablen interpretieren. Dies geschieht z. B. häufig zur Untersuchung von Politikfragen. Beim ML sind die Modelle dagegen oft nicht nachvollziehbar. Hinzu kommt, dass bei ML-Problemstellungen häufig mit wesentlich mehr Variablen gearbeitet wird, als dies in den Wirtschaftswissenschaften gemeinhin der Fall ist. In der Regel sind es tausende Variablen (sogenannte „Merkmale“) oder mehr. Größere Datensätze können zwischen Zehntausenden und Hunderten Millionen Beobachtungen umfassen. Bei dieser Größenordnung greifen Forscher zur Erstellung von Prognosen auf komplexere, in geringerem Maße nachvollziehbare Techniken wie neuronale Netze zurück. Ein zentraler Forschungsbereich des ML befasst sich daher auch mit der Frage, wie selbst bei derart umfassenden Modellen das in den Wirtschaftswissenschaften übliche Maß an Erklärbarkeit erreicht werden kann (vgl. Cluster 4 unten).

Grundlage für die Vielzahl neuer ML-Anwendungen ist die Technologie der neuronalen Netze – ein komplexes statistisches Modellierungsverfahren. Die Voraussetzungen dafür wurden durch die zunehmende Rechenleistung und die Verfügbarkeit riesiger Datensätze (sogenannter Big Data) geschaffen. In neuronalen Netzen werden Tausende bzw. Millionen einfacher Transformationen immer wieder verknüpft, so dass ein umfangreicheres statistisches Modell entsteht, das komplexe Zusammenhänge zwischen In- und Outputdaten erfassen kann. Anders ausgedrückt: Neuronale Netze modifizieren ihren eigenen Code, um Beziehungen zwischen Input- und Outputdaten zu ermitteln und die Ergebnisse zu optimieren. Der Begriff Deep Learning bzw. tiefes Lernen bezieht sich auf besonders große neuronale Netze. Ein Schwellenwert, ab dem ein neuronales Netz als „tief“ gilt, wurde allerdings nicht festgelegt.

Diese Entwicklungstendenzen in der KI-Forschung gehen mit stetigen Fortschritten bei der Rechenkapazität, der Datenverfügbarkeit und dem Design neuronaler Netze einher. Dies lässt darauf schließen, dass dem statistischen Ansatz in der KI-Forschung auf kurze Sicht weiterhin eine zentrale Bedeutung zukommen wird. Die Politikverantwortlichen sollten sich auf die KI-Entwicklungen konzentrieren, die in den kommenden Jahren voraussichtlich die größten Auswirkungen haben und die größten Politikherausforderungen mit sich bringen werden. Zu diesen Herausforderungen zählt u. a. die Verbesserung der Erklärbarkeit und Nachvollziehbarkeit der Entscheidungen von KI-Systemen sowie der Transparenz der Entscheidungsprozesse. Zu bedenken gilt auch, dass die meisten dynamischen Ansätze im Bereich KI – statistische KI und insbesondere neuronale Netze – nicht für alle Arten von Problemstellungen geeignet sind. Andere KI-Ansätze sowie Kombinationen symbolischer und statistischer Methoden werden also weiterhin eine wichtige Rolle spielen.

Eine allgemein anerkannte Klassifizierung der KI-Forschungsbereiche oder des Unterbereichs ML gibt es bislang nicht. Die im Folgenden vorgeschlagene Klassifizierung umfasst 25 KI-Forschungsbereiche. Diese Bereiche wurden in 4 Haupt- und 9 Unterkategorien gegliedert, deren Fokus in erster Linie auf ML liegt. In den Wirtschaftswissenschaften spezialisieren sich Wissenschaftler meist auf einen Forschungsbereich. In der KI-Forschung dagegen arbeitet man zur Klärung von Forschungsfragen in der Regel in mehreren Clustern gleichzeitig.

Cluster 1: ML-Anwendungen

Bei der ersten Hauptkategorie geht es um die Anwendung von ML-Methoden zur Bewältigung verschiedener konkreter Herausforderungen in Wirtschaft und Gesellschaft.

Anwendungen maschinellen Lernens gewinnen zunehmend an Bedeutung, ähnlich wie einst das Internet, das zunächst einige Branchen und schließlich die gesamte Wirtschaft grundlegend veränderte. In Kapitel 3 werden Beispiele solcher KI-Anwendungen im OECD-Raum beschrieben. Die bedeutendsten Forschungsbereiche, die sich mit der Entwicklung praktischer Anwendungen befassen, sind in Tabelle 1.1 aufgelistet.

Tabelle 1.1. Cluster 1: ML-Anwendungen

ML-Anwendungen	Nutzung von ML	Maschinelle Sprachverarbeitung
		Maschinelles Sehen
		Roboternavigation
		Sprachenlernen
	Kontextualisierung von ML	Algorithmische Spieltheorie und Computational Social Choice
		Kollaborative Systeme

Quelle: Bereitgestellt von der Internet Policy Research Initiative (IPRI) des MIT.

Zu den wichtigsten Forschungsbereichen, in denen ML angewandt wird, zählen die maschinelle Sprachverarbeitung, das maschinelle Sehen und die Roboternavigation – drei vielfältige und wachsende Forschungsfelder. Die zu bewältigenden Forschungsherausforderungen können einen oder mehrere Bereiche betreffen. Bei Anwendungen für das Brustkrebs-Screening etwa stützen sich Forscher in den Vereinigten Staaten sowohl auf eine maschinelle Sprachverarbeitung der Freitext-Kommentare und pathologischen Befunde als auch auf maschinelles Sehen (Computervision) (Yala et al., 2017[12]).

In zwei Forschungsbereichen dieses Clusters liegt der Fokus auf der Kontextualisierung von ML: Die algorithmische Spieltheorie ist im Schnittbereich von Wirtschaftswissenschaft, Spieltheorie und Informatik angesiedelt. Sie befasst sich mit der Analyse und Optimierung dynamischer Spiele mithilfe von Algorithmen. Der Bereich Kollaborative Systeme zielt auf die Bewältigung umfassender Herausforderungen ab. Dabei werden mehrere ML-Systeme kombiniert, um Lösungen für die verschiedenen Aspekte eines komplexen Problems zu liefern.

Cluster 1: Politikrelevanz

KI-Anwendungen werfen mehrere politische Grundsatzfragen auf, etwa im Hinblick auf die Zukunft der Arbeit, mögliche gesellschaftliche Auswirkungen von KI sowie Kompetenzangebot und -entwicklung. Zudem muss geklärt werden, in welchen Situationen KI-Anwendungen in sensiblen Bereichen angemessen sind und in welchen nicht. Weitere wichtige Politikfragen ergeben sich im Zusammenhang mit den Auswirkungen von KI auf Wirtschaftsakteure und Wirtschaftsdynamik, staatlichen Open-Data-Strategien, Vorschriften zur Roboternavigation sowie Datenschutzbestimmungen.

Cluster 2: ML-Techniken

Im Fokus der zweiten Hauptkategorie stehen die Techniken und Paradigmen des ML. Hier geht es, ähnlich wie im Forschungszweig „Quantitative Methoden“ der Sozialwissenschaften, um die Entwicklung der technischen Instrumente und der Ansätze, auf denen die ML-Anwendungen beruhen (Tabelle 1.2).

In dieser Kategorie steht vor allem die Technik der neuronalen Netze (mit dem Teilbereich „Deep Learning“) im Vordergrund. Auf ihr basiert heute ein Großteil des ML. Außer um die Techniken geht es in diesem Cluster auch um die Paradigmen, die dem maschinellen

Lernen zugrunde liegen. Hierzu zählt z. B. das Reinforcement-Learning bzw. bestärkende Lernen, bei dem sich das Training des Systems am menschlichen Lernen durch Versuch und Irrtum orientiert. Die Algorithmen erhalten dabei keine expliziten Aufgabenstellungen, sondern lernen, indem sie in rascher Abfolge verschiedene Optionen erproben. Dabei passen sie ihr Verhalten an, je nachdem, ob die vorherigen Schritte zu „Belohnung" oder „Bestrafung" führten. Daher wurde diese Lerntechnik auch als „unablässiges Experimentieren" bezeichnet (Knight, 2017[13]).

Tabelle 1.2. Cluster 2: ML-Techniken

ML-Techniken	Techniken	Deep Learning
		Simulationsbasiertes Lernen
		Crowdsourcing und Human Computation
		Evolutionäre Algorithmen
		Techniken ohne neuronale Netze
	Paradigmen	Überwachtes Lernen
		Bestärkendes Lernen
		Generative Modelle/GANs

Quelle: Bereitgestellt von der Internet Policy Research Initiative (IPRI) des MIT.

Mit generativen Modellen und insbesondere generativen gegnerischen Netzen (*generative adversarial networks* – GAN) wird ein System darauf trainiert, neue Daten zu generieren, die einem bestehenden Datensatz entsprechen. Dies ist ein besonders spannender Forschungsbereich der KI. Der Ansatz besteht darin, zwei oder mehr unüberwachte neuronale Netze in einem Nullsummenspiel gegeneinander antreten zu lassen. Unter spieltheoretischen Gesichtspunkten betrachtet, beruhen diese Systeme bzw. ihr Lernverhalten auf einer Reihe rasch wiederholter Spiele. Indem die neuronalen Netze mit hoher Rechenleistung gegeneinander arbeiten, kann das System erfolgreiche Strategien lernen. Dies gilt insbesondere für strukturierte Umgebungen mit klaren Regeln, wie z. B. das Brettspiel Go und AlphaGo Zero.

Cluster 2: Politikrelevanz

Bei der Entwicklung und Einführung von ML-Techniken stellen sich mehrere Fragen, die für die Politikgestaltung relevant sind. Dabei geht es u. a. um die Förderung besserer Trainingsdatensätze, die Finanzierung von Hochschul- und Grundlagenforschung, die Förderung der „Zweisprachigkeit" von Arbeitskräften, damit diese neben Fach- auch über KI-Kompetenzen verfügen, sowie die Informatikausbildung. In Kanada z. B. machte die staatliche Forschungsförderung bahnbrechende Fortschritte möglich, durch die die Voraussetzungen für den außerordentlichen Erfolg moderner neuronaler Netze geschaffen wurden (Allen, 2015[14]).

Cluster 3: ML-Verbesserung/-Optimierung

Der Fokus der dritten Hauptkategorie liegt auf Strategien zur Verbesserung und Optimierung von ML-Instrumenten. Hier lassen sich je nach Zeithorizont drei Bereiche unterscheiden (kurz-, mittel- und langfristige Forschung) (Tabelle 1.3). Das Hauptaugenmerk der kurzfristigen Forschung gilt der Beschleunigung von Deep Learning. Erreicht werden soll dies durch eine bessere Datensammlung oder durch die Nutzung verteilter Rechnersysteme zum Training der Algorithmen.

Tabelle 1.3. Cluster 3: ML-Verbesserung/-Optimierung

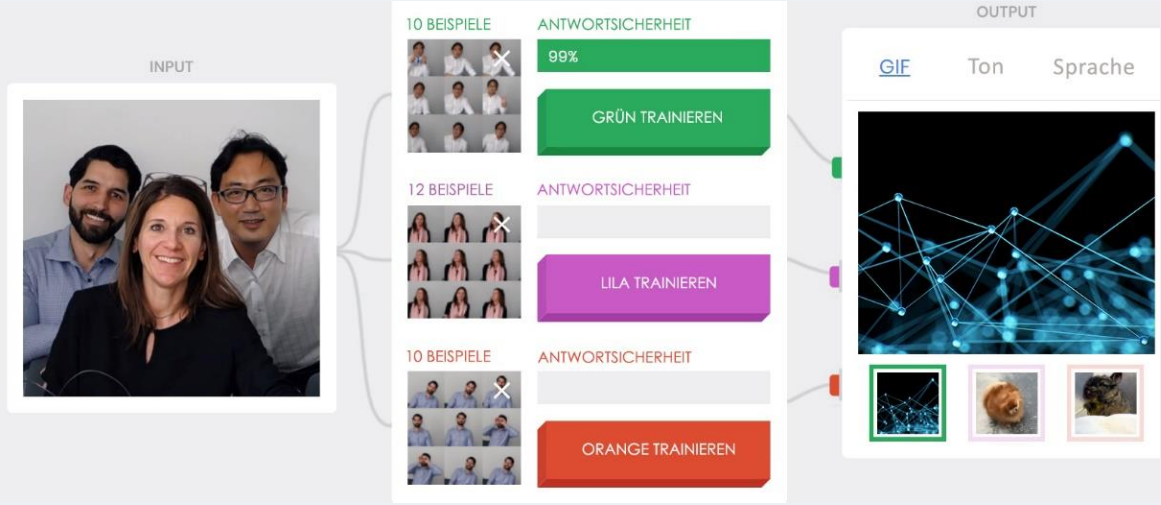
ML-Verbesserung	Kurzfristige Forschung	Schnelleres Deep Learning
		Bessere Datensammlung
		Verteilte Trainingsalgorithmen
	Mittelfristige Forschung	ML auf Geräten mit geringer Rechenkapazität
		Lernen lernen/Metalernen
		KI-Entwicklertools
	Langfristige Forschung	Verständnis neuronaler Netze
		One-Shot-Lernen

Quelle: Bereitgestellt von der Internet Policy Research Initiative (IPRI) des MIT.

Kasten 1.2. Teachable Machine

Das Google-Experiment Teachable Machine ermöglicht es Nutzern, eine Maschine mit der Kamera ihres Smartphones oder Computers darauf zu trainieren, bestimmte Szenarien zu erkennen. Um die lernfähige Maschine zu trainieren, macht der Nutzer mehrere Aufnahmen von drei unterschiedlichen Szenarien, z. B. von unterschiedlichen Gesichtsausdrücken. Die Maschine analysiert die Fotos dieses Trainingsdatensatzes und kann sie somit nutzen, um die verschiedenen Szenarien zu erkennen. Sie produziert dann z. B. ein bestimmtes Geräusch, wenn die Person in die Kamera lächelt. Dieses ML-Tool ist insofern einzigartig, als das neuronale Netz ausschließlich über den Browser des Nutzers läuft und keine externen Rechenkapazitäten oder Datenspeicher erforderlich sind (Abbildung 1.7).

Abbildung 1.7. Training einer Maschine mithilfe einer Computerkamera



Quelle: https://experiments.withgoogle.com/ai/teachable-machine.

Einen weiteren Forschungsschwerpunkt stellt die Nutzung von ML auf Geräten mit geringer Rechenleistung wie Mobiltelefonen und anderen vernetzten Geräten dar. In diesem Bereich wurden bereits bedeutende Fortschritte erzielt. Durch Projekte wie Googles Teachable Machine stehen inzwischen Open-Source-Tools für ML zur Verfügung, die auch in einem Browser verwendet werden können (Kasten 1.2). Teachable Machine ist lediglich

ein Beispiel für die neuen KI-Entwicklertools, die eine breitere Nutzung und höhere Effizienz von ML gewährleisten sollen. Bei der Entwicklung von KI-Chips für mobile Geräte wurden ebenfalls erhebliche Fortschritte erzielt.

Die längerfristige ML-Forschung beschäftigt sich u. a. mit den Mechanismen, die ausschlaggebend dafür sind, dass neuronale Netze effizient lernen. Neuronale Netze haben sich zwar als leistungsstarke ML-Technik erwiesen, ihre Funktionsweise wird jedoch nach wie vor nicht hinreichend verstanden. Ein besseres Verständnis dieser Prozesse würde die Entwicklung tieferer neuronaler Netze ermöglichen. Die längerfristige Forschung untersucht auch, wie neuronale Netze mit wesentlich kleineren Datensätzen trainiert werden können, was als „One-Shot-Lernen" bezeichnet wird. Außerdem gibt es Bemühungen, die Effizienz des Trainingsprozesses generell zu erhöhen. Größere Modelle erfordern nämlich u. U. ein mehrwöchiges oder sogar mehrmonatiges Training mit Hunderten Millionen Beispielen.

Cluster 3: Politikrelevanz

Politikrelevant sind hierbei u. a. die Konsequenzen der Nutzung von ML auf Einzelgeräten, also ohne dass Daten in einer Cloud gespeichert werden müssen. Wichtig ist auch die Frage, welche Möglichkeiten sich zur Verringerung des Energieverbrauchs bieten und was zur Entwicklung besserer KI-Tools getan werden muss, von denen ein breiterer Kreis von Nutzern profitieren kann.

Cluster 4: ML-Kontext

Die vierte Hauptkategorie beleuchtet den technischen, rechtlichen und gesellschaftlichen Kontext von ML. ML-Systeme treffen wesentliche Entscheidungen, die zunehmend auf Algorithmen basieren. Daher ist es wichtig zu verstehen, wie algorithmische Voreingenommenheit entstehen kann und verstärkt wird und wie sich die aus ihr resultierenden Verzerrungen in den Ergebnissen beheben lassen. Einer der aktivsten ML-Forschungsbereiche befasst sich mit der Transparenz und der Verantwortlichkeit von KI-Systemen (Tabelle 1.4). Bei statistischen KI-Ansätzen sind die Berechnungen, die den algorithmischen Entscheidungen zugrunde liegen, kaum nachvollziehbar. Diese Entscheidungen können jedoch schwerwiegende Auswirkungen auf das Leben von Menschen haben. Sie können z. B. zur Folge haben, dass ein Kredit verweigert wird oder dass ein Straftäter vorzeitig aus der Haft entlassen wird (Angwin et al., 2016[15]). Ein weiterer Schwerpunkt der kontextbezogenen ML-Forschung ist die Gewährleistung der Sicherheit und Integrität der Systeme. Man weiß noch immer nur wenig darüber, wie neuronale Netze Entscheidungen treffen. Oft können solche Netze mit einfachen Methoden getäuscht werden, z. B. indem ein paar Pixel in einem Bild verändert werden (Ilyas et al., 2018[16]). Ziel der Forschung ist es, Systeme vor dem Eindringen unerwünschter Daten sowie feindlichen Angriffen zu schützen. Ein weiterer Forschungsbereich beschäftigt sich mit der Überprüfung der Integrität von ML-Systemen.

Cluster 4: Politikrelevanz

Die kontextbezogenen Aspekte des ML werfen einige wichtige Fragen für die Politik auf. Diese betreffen u. a. die Voraussetzungen algorithmischer Verantwortlichkeit, die Bekämpfung algorithmischer Voreingenommenheit, die gesellschaftlichen Auswirkungen von ML-Systemen, Produktsicherheit und Haftung sowie Sicherheit insgesamt (OECD, 2019[8]).

Tabelle 1.4. Cluster 4: ML-Kontext

ML-Kontext	Erklärbarkeit	Transparenz und Verantwortlichkeit
		Erklärung von Einzelentscheidungen
		Vereinfachung in nachvollziehbare Algorithmen
		Fairness/Voreingenommenheit
		Debugging-Fähigkeit
	Sicherheit und Zuverlässigkeit	Feindliche Angriffe
		Überprüfung
		Angriffe anderer Art

Quelle: Bereitgestellt von der Internet Policy Research Initiative (IPRI) des MIT.

Literaturverzeichnis

Allen, K. (2015), "How a Toronto professor's research revolutionized artificial intelligence,", *The Star,* 17 April, https://www.thestar.com/news/world/2015/04/17/how-a-toronto-professors-research-revolutionized-artificial-intelligence.html. [14]

Angwin, J. et al. (2016), "Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks", *ProPublica*, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. [15]

Anyoha, R. (2017), "The history of artificial intelligence", Harvard University Graduate School of Arts and Sciences blog, 28. August, http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/. [4]

Gringsjord, S. und N. Govindarajulu (2018), *Artificial Intelligence*, The Stanford Encyclopedia of Philosophy Archive, https://plato.stanford.edu/archives/fall2018/entries/artificial-intelligence/. [11]

Ilyas, A. et al. (2018), "Blackbox Adversarial Attacks with Limited Queries and Information" Paper für die 35. International Conference on Machine Learning, Stockholmsmässan, Stockholm, 10.-15. Juli, Vol. 80, S. 2137-2146, http://proceedings.mlr.press/v80/ilyas18a/ilyas18a.pdf. [16]

Knight, W. (2017), "5 big predictions for artificial intelligence in 2017", *MIT Technology Review,* 4. Januar, https://www.technologyreview.com/s/603216/5-big-predictions-for-artificial-intelligence-in-2017/. [13]

OECD (2019), *Empfehlung des Rats zu künstlicher Intelligenz*, OECD, Paris, http://www.oecd.org/berlin/presse/Empfehlung-des-Rats-zu-kuenstlicher-Intelligenz.pdf. [8]

OECD (2017), *OECD Digital Economy Outlook 2017*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264276284-en. [7]

Russel, S. und P. Norvig (2009), *Künstliche Intelligenz: Ein moderner Ansatz*, 3. Auflage, Pearson, München. [9]

Shannon, C. (1950), "XXII. Programming a computer for playing chess", *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, Vol. 41/314, S. 256-275, https://doi.org/10.1080/14786445008521796. [2]

Silver, D. et al. (2017), "Mastering the game of Go without human knowledge", *Nature*, Vol. 550/7676, S. 354-359, http://dx.doi.org/10.1038/nature24270. [6]

Somers, J. (2013), "The man who would teach machines to think", *The Atlantic,* November, https://www.theatlantic.com/magazine/archive/2013/11/the-man-who-would-teach-machines-to-think/309529/. [5]

Turing, A. (1950), "Computing machinery and intelligence", in *Parsing the Turing Test*, Springer, Dordrecht, S. 23-65, http://http:/doi.org/10.1007/978-1-4020-6710-5. [1]

UW (2006), *The History of Artificial Intelligence*, University of Washington, History of Computing Course (CSEP 590A), https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf. [3]

Winston, P. (1992), *Artificial Intelligence*, Addison-Wesley, Reading, MA, https://courses.csail.mit.edu/6.034f/ai3/rest.pdf. [10]

Yala, A. et al. (2017), "Using machine learning to parse breast pathology reports", *Breast Cancer Research and Treatment*, Vol. 161/2, S. 201-211, http://doi.org/10.1007/s10549-016-4035-1. [12]

Anmerkungen

[1] Dieser Test wurde nicht mündlich, sondern schriftlich durchgeführt.

[2] Arbeiten zum Systemlebenszyklus wurden u. a. vom National Institute of Standards vorgelegt. In jüngster Zeit haben Normungsorganisationen wie die Internationale Organisation für Normung (ISO) SC 42 begonnen, sich mit dem KI-Lebenszyklus zu befassen.

[3] Dies gilt z. B. für die globale Initiative zur Ethik autonomer und intelligenter Systeme des Institute for Electrical and Electronics Engineers.

2. Das wirtschaftliche Umfeld

Dieses Kapitel beschreibt die wirtschaftlichen Merkmale der künstlichen Intelligenz (KI) und definiert KI als eine neue Universaltechnologie, die die Vorhersagekosten senken und die Entscheidungsfindung verbessern könnte. Durch kostengünstigere und präzisere Vorhersagen, Empfehlungen und Entscheidungen verspricht KI die Produktivität zu steigern, die Lebensqualität zu erhöhen und zur Bewältigung komplexer Herausforderungen beizutragen. Die Einführung von KI erfolgt in den einzelnen Unternehmen und Branchen in unterschiedlichem Tempo, weil die Nutzung von KI zusätzliche Investitionen in Daten, Kompetenzen, die Digitalisierung von Arbeitsabläufen und die Anpassung organisatorischer Prozesse erfordert. KI ist außerdem verstärkt das Ziel von Investitionen und eröffnet neue Geschäftsmöglichkeiten. Die Private-Equity-Investitionen in KI-Start-ups steigen seit 2016 deutlich. Zwischen 2016 und 2017 haben sie sich auf 16 Mrd. USD verdoppelt. Im ersten Halbjahr 2018 mobilisierten im KI-Bereich tätige Start-ups 12% der weltweiten Private-Equity-Investitionen – ein beträchtlicher Anstieg gegenüber 2011, als ihr Anteil lediglich 3% betrug. Die Investitionen in KI-Technologien werden angesichts des zunehmenden Reifegrads dieser Technologien voraussichtlich ihren Aufwärtstrend fortsetzen.

Die statistischen Daten für Israel wurden von den zuständigen israelischen Stellen bereitgestellt, die für sie verantwortlich zeichnen. Die Verwendung dieser Daten durch die OECD erfolgt unbeschadet des völkerrechtlichen Status der Golanhöhen, von Ost-Jerusalem und der israelischen Siedlungen im Westjordanland.

Wirtschaftliche Aspekte der künstlichen Intelligenz

Künstliche Intelligenz erleichtert Vorhersagen

Aus wirtschaftlicher Sicht sind die jüngsten Fortschritte im Bereich der künstlichen Intelligenz (KI) von Vorteil, weil sie einen Rückgang der Vorhersagekosten bzw. Anstieg der Qualität der verfügbaren Vorhersagen bei gleichen Kosten ermöglichen. Viele Aspekte der Entscheidungsfindung sind nicht mit Vorhersagen verbunden. Bessere, billigere und allgemein zugängliche KI-gestützte Vorhersagen könnten jedoch transformativ wirken, weil Vorhersagen in viele menschliche Aktivitäten einfließen.

Die sinkenden Kosten von KI-gestützten Vorhersagen haben die Möglichkeiten ihrer Nutzung erweitert, was ähnlich tiefgreifende Effekte haben dürfte wie seinerzeit die zunehmende Verbreitung der Computer. Bereits die ersten KI-Anwendungen waren auf Vorhersageprobleme ausgerichtet. Maschinelles Lernen (ML) wird beispielsweise genutzt, um Kreditausfälle und Versicherungsrisiken vorherzusagen. Da die Kosten von KI-Anwendungen sinken, werden einige menschliche Aktivitäten heute so gestaltet, dass sie als Vorhersageproblem behandelt werden können. In der medizinischen Diagnose nutzen Ärzte beispielsweise Daten über die Symptome eines Patienten, um fehlende Informationen über die Ursache dieser Symptome zu gewinnen. Wenn vorliegende Daten verwendet werden, um fehlende Informationen zu gewinnen, ist dies eine Vorhersage. Die Objekterkennung ist ebenfalls ein Vorhersageproblem: Das menschliche Auge nimmt Daten in Form von Lichtsignalen auf und das Gehirn ergänzt diese Daten um Informationen, die zur Einordnung des Konzepts erforderlich sind.

Da Vorhersagen ein wichtiger Input für die Entscheidungsfindung sind und KI die Vorhersagekosten senkt, kann sie vielfältig angewandt werden. Vorhersagen helfen, Entscheidungen zu treffen, und Entscheidungen sind etwas, das unser ganzes Leben bestimmt. Manager treffen wichtige Entscheidungen über Einstellungen, Investitionen und Strategien ebenso wie weniger wichtige Entscheidungen, beispielsweise darüber, an welchen Sitzungen sie teilnehmen und was sie dort sagen sollen. Richter treffen wichtige Entscheidungen über Schuld und Unschuld, Verfahren und Urteile genauso wie weniger wichtige Entscheidungen über einzelne Paragrafen oder Anträge. Menschen treffen ständig Entscheidungen: Große Entscheidungen, z. B. über eine Heirat, ebenso wie tagtägliche, etwa darüber, was sie essen oder welche Musik sie hören möchten. Eine große Herausforderung im Entscheidungsprozess ist der Umgang mit Unsicherheit. Da Vorhersagen die Unsicherheit verringern, können sie Entscheidungen erleichtern und neue Chancen eröffnen.

Menschliche Vorhersagen werden durch maschinelle Vorhersagen ersetzt

Ein weiteres wichtiges wirtschaftliches Konzept ist die Substitution. Wenn der Preis eines Rohstoffs (z. B. Kaffee) fällt, kaufen die Menschen nicht nur mehr davon, sondern sie kaufen u. U. auch weniger Substitutionsgüter (wie Tee). Wenn maschinelle Vorhersagen billiger werden, werden Menschen bei Vorhersageaufgaben zunehmend durch Maschinen ersetzt. Das wird einen Abbau von Arbeitsplätzen in diesem Bereich zur Folge haben.

Ebenso wie die Computer dazu geführt haben, dass heute nur noch wenige Menschen bei ihrer Arbeit Rechenaufgaben lösen müssen, wird KI dazu führen, dass weniger Menschen für Vorhersageaufgaben gebraucht werden. Die Transkription – die Übertragung gesprochener Wörter in Schriftsprache – ist beispielsweise eine Vorhersage, insofern sie fehlende Informationen über die Symbole, die den gesprochenen Wörtern entsprechen, liefert. KI erledigt Transkriptionsaufgaben heute bereits schneller und genauer als viele Menschen, in deren Tätigkeitsbereich dies bislang fiel.

Maschinelle Vorhersagen werden durch Daten, Handlungen und Beurteilungen ergänzt

Wenn der Preis eines Rohstoffs (z. B. Kaffee) sinkt, werden auch mehr Komplementärgüter gekauft (wie etwa Sahne und Zucker). Eine große Herausforderung im Hinblick auf die jüngsten Fortschritte der KI besteht also darin, festzustellen, welche Faktoren zur Ergänzung der Vorhersagen nötig sind. Vorhersagen sind ein wichtiger Input des Entscheidungsprozesses, sie sind jedoch noch keine Entscheidungen. Um zu einer Entscheidung zu gelangen, sind komplementäre Faktoren nötig: Daten, Handlungen und Beurteilungen.

Daten sind Informationen, die in eine Vorhersage einfließen. Viele in letzter Zeit entwickelte KI-Systeme benötigen große Mengen digitaler Daten, um Vorhersagen auf der Basis von Beispielen aus der Vergangenheit zu treffen. Allgemein gilt: Je mehr Beispiele aus der Vergangenheit vorliegen, desto genauer sind die Vorhersagen. KI führt also dazu, dass der Zugang zu großen Datenmengen für Unternehmen und andere Organisationen immer wertvoller wird. Der strategische Wert von Daten ist jedoch nicht eindeutig, weil er davon abhängt, ob die Daten nützlich sind, um Vorhersagen zu treffen, die für die betreffende Organisation wichtig sind. Er hängt auch davon ab, ob die Daten nur rückblickend verfügbar sind oder ob die Organisation im Zeitverlauf kontinuierlich Feedback einholen kann. Die Fähigkeit, kontinuierlich durch neue Daten zu lernen, kann einen dauerhaften Wettbewerbsvorteil schaffen (Agrawal, A., J. Gans und A. Goldfarb, 2018[1]).

Aus den anderen Elementen einer Entscheidung – **Handlungen** und **Beurteilungen** – ergeben sich weitere neue Aufgaben. Einige *Handlungen* sind von Natur aus wertvoller, wenn sie von einem Menschen anstatt von einer Maschine durchgeführt werden (z. B. solche von Berufssportlern, Erziehern oder Verkäufern). Am wichtigsten ist wohl das Konzept der *Beurteilung*, d. h. des Bestimmens des Nutzens einer bestimmten Handlung in einem bestimmten Umfeld. Wenn KI für Vorhersagen genutzt wird, müssen Menschen entscheiden, was vorhergesagt werden soll und wozu die Vorhersage dienen soll.

Die Einführung von KI in einer Organisation erfordert komplementäre Investitionen und Prozessänderungen

Die künstliche Intelligenz kann ebenso wie die elektronische Datenverarbeitung, der elektrische Strom und die Dampfmaschine als Universaltechnologie angesehen werden (Bresnahan, T. und M. Trajtenberg, 1992[2]; Brynjolfsson, E., D. Rock und C. Syverson, 2017[3]). Das bedeutet, dass KI das Potenzial hat, die Produktivität in vielen Sektoren erheblich zu steigern. Zugleich macht KI Investitionen in eine Reihe komplementärer Inputs erforderlich. Das kann dazu führen, dass ein Unternehmen seine Gesamtstrategie ändert.

Organisationen müssen eine Reihe komplementärer Investitionen tätigen, bevor KI einen deutlichen Effekt auf ihre Produktivität haben kann. Dazu gehören Investitionen in die Infrastruktur für die kontinuierliche Erhebung von Daten, in Fachkräfte, die diese Daten nutzen können, sowie in Prozessänderungen, um die sich aus der reduzierten Unsicherheit ergebenden neuen Chancen zu nutzen.

In jeder Organisation gibt es viele Prozesse, die nur dazu dienen, Situationen zu bewältigen, die sich aus Unsicherheit ergeben. Die Unsicherheit selbst wird damit jedoch nicht angegangen. Flughafenlounges dienen beispielsweise dazu, den Kunden die Wartezeit so angenehm wie möglich zu machen. Wenn die Passagiere zutreffende Vorhersagen darüber hätten, wie lange es dauert, zum Flughafen und durch die Sicherheitskontrolle zu kommen, wären diese Lounges u. U. gar nicht nötig.

Bessere Vorhersagen dürften je nach Branche und Unternehmen unterschiedliche Möglichkeiten eröffnen. Google, Baidu und andere große Internetplattformen sind gut aufgestellt, um von hohen KI-Investitionen zu profitieren. Was die Angebotsseite betrifft, verfügen sie schon über die nötigen Systeme zur Datenerhebung. Auf der Nachfrageseite haben sie bereits genug Kunden, um die hohen Fixkosten von Investitionen in Technologien zu rechtfertigen, die sich noch in einem frühen Entwicklungsstadium befinden. Viele andere Unternehmen haben ihre Arbeitsabläufe hingegen noch nicht vollständig digitalisiert und sind noch nicht in der Lage, KI-Tools direkt in die bestehenden Prozesse zu integrieren. Da die Kosten jedoch im Lauf der Zeit sinken werden, dürften auch diese Unternehmen nach und nach erkennen, welche Möglichkeiten eine Reduzierung der Unsicherheit bieten kann. Unter dem Druck der Notwendigkeit werden sie den Branchenführern folgen und in KI investieren.

Private-Equity-Investitionen in KI-Start-ups

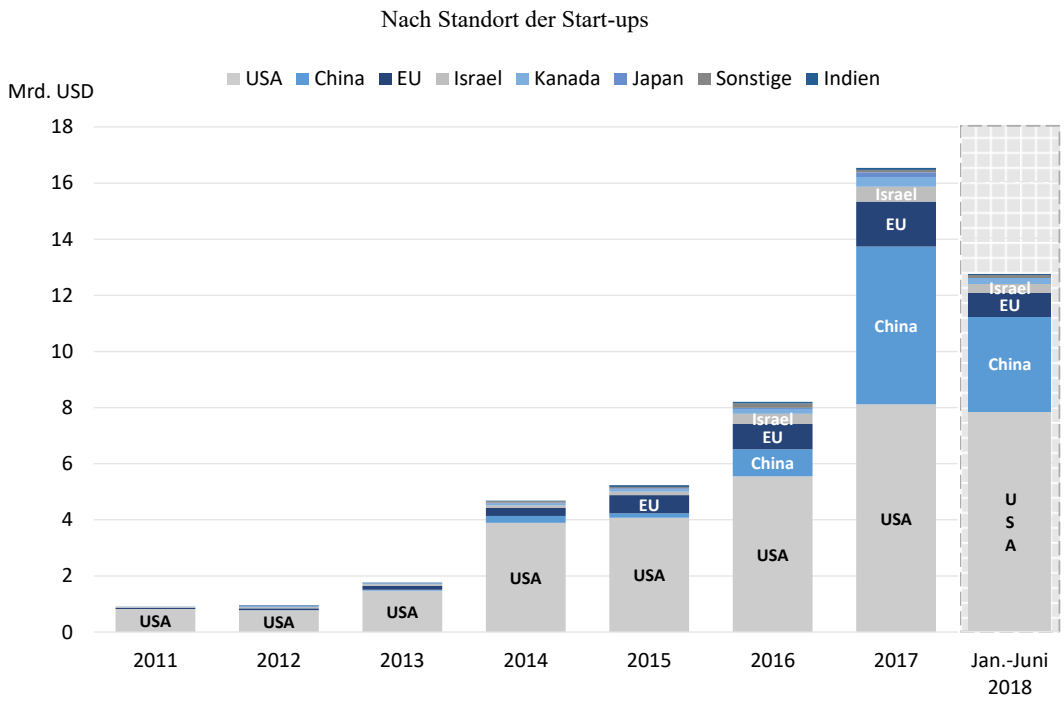
Die KI-Investitionen expandieren insgesamt schnell, und KI hat bereits erhebliche Auswirkungen auf die Unternehmen. Laut Schätzungen von MGI (2017[4]) waren 2016 weltweit 26-39 Mrd. USD in KI investiert. Davon entfielen rd. 70% auf interne Unternehmensinvestitionen, rd. 20% auf Investitionen in KI-Start-ups und rd. 10% auf KI-Übernahmen (Dilda, 2017[5]). Drei Viertel dieser Investitionen wurden von großen Technologieunternehmen getätigt. Außerhalb des Technologiesektors befindet sich die KI noch im Anfangsstadium; nur wenige Unternehmen haben in großem Umfang KI-Lösungen eingeführt. Großunternehmen in anderen digital reifen Sektoren, die über nutzbare Daten verfügen, insbesondere im Finanz- und Automobilsektor, greifen ebenfalls auf KI zurück.

Die großen Technologieunternehmen übernehmen in großem Umfang KI-Start-ups. Laut CBI (2018[6]) gehören Google, Apple, Baidu, Facebook, Amazon, Intel, Microsoft, Twitter und Salesforce zu den Unternehmen, die seit 2010 die meisten KI-Start-ups übernommen haben. 2017 und Anfang 2018 wurden mehrere im Bereich der Cybersicherheit tätige KI-Start-ups übernommen. So hat Amazon beispielsweise Sqrrl aufgekauft und Oracle Zenedge.

KI-Start-ups sind auch Übernahmekandidaten für Unternehmen aus traditionelleren Branchen. Dazu gehören insbesondere Automobilunternehmen, Unternehmen im Gesundheitswesen wie Roche Holding oder Athena Health sowie Versicherungs- und Einzelhandelsunternehmen.

Die Private-Equity-Investitionen in KI-Start-ups haben sich seit 2016 deutlich beschleunigt, nachdem sie in den vorangegangenen fünf Jahren bereits stetig gewachsen waren. Ihr Volumen hat sich zwischen 2016 und 2017 verdoppelt (Abbildung 2.1). Schätzungen zufolge wurden zwischen 2011 und Mitte 2018 mehr als 50 Mrd. USD in KI-Start-ups investiert (Kasten 2.1).

Abbildung 2.1. Gesamtinvestitionen in KI-Start-ups, 2011-2017 und erstes Halbjahr 2018, geschätzte Zahlen



Anmerkung: Die Schätzungen für 2018 sind möglicherweise konservativ, weil Verzögerungen beim Eingang der Daten nicht berücksichtigt sind (vgl. Kasten 2.1. Anmerkung zur Methodik).
Quelle: OECD-Schätzung auf der Basis von Crunchbase (Juli 2018), www.crunchbase.com.

Derzeit entfallen mehr als 12 % der Private-Equity-Investitionen in Start-ups auf KI

Im ersten Halbjahr 2018 mobilisierten im KI-Bereich tätige Start-ups rd. 12 % der gesamten weltweiten Private-Equity-Investitionen. Dies ist ein starker Anstieg gegenüber 2011, als ihr Anteil lediglich 3 % betrug (Abbildung 2.2). Der Anteil der Investitionen in im KI-Bereich tätige Start-ups hat in allen betrachteten Ländern zugenommen. In den Vereinigten Staaten und der Volksrepublik China (im Folgenden „China“) gingen im ersten Halbjahr 2018 rd. 13 % der Start-up-Investitionen an KI-Start-ups. Die stärkste Zunahme war in Israel zu beobachten, dort stieg der Anteil der Investitionen in KI-Start-ups zwischen 2011 und der ersten Hälfte von 2018 von 5 % auf 25 %; 2017 entfielen 50 % der Investitionen auf autonomes Fahren.

Die meisten Investitionen in KI-Start-ups entfallen auf die Vereinigten Staaten und China

Auf Start-ups in den Vereinigten Staaten entfallen weltweit die meisten Beteiligungsinvestitionen in KI-Start-ups. Dies gilt sowohl für die Zahl der Investitionstransaktionen als auch für die investierten Beträge, die zwei Drittel der seit 2011 investierten Gesamtsumme ausmachen (Abbildung 2.1). Dies ist nicht überraschend, wenn man bedenkt, dass technologieübergreifend 70-80 % der weltweiten Wagniskapitalinvestitionen auf die Vereinigten Staaten entfallen (Breschi, S., J. Lassébie und C. Menon, 2018[7]).

Kasten 2.1. Anmerkung zur Methodik

Die in diesem Abschnitt aufgeführten Schätzungen der Private-Equity-Investitionen in KI-Start-ups basieren auf Crunchbase-Daten (Stand Juli 2018). Crunchbase ist eine kommerzielle Datenbank über innovative Unternehmen, die 2007 eingerichtet wurde und Informationen über mehr als 500 000 Unternehmen in 199 Ländern enthält. Breschi, Lassébie und Menon (2018[7]) vergleichen Crunchbase mit anderen aggregierten Datenquellen. Sie stellen übereinstimmende Muster für einen breiten Fächer von Ländern fest, darunter die meisten OECD-Länder (mit Ausnahme Japans und Koreas). Einheitliche Muster wurden außerdem für Brasilien, die Russische Föderation, Indien, die Volksrepublik China (im Folgenden „China") und Südafrika festgestellt. Die Unternehmen werden in Crunchbase einem oder mehreren Technologiebereichen aus einer Liste von 45 großen Kategorien zugeordnet.

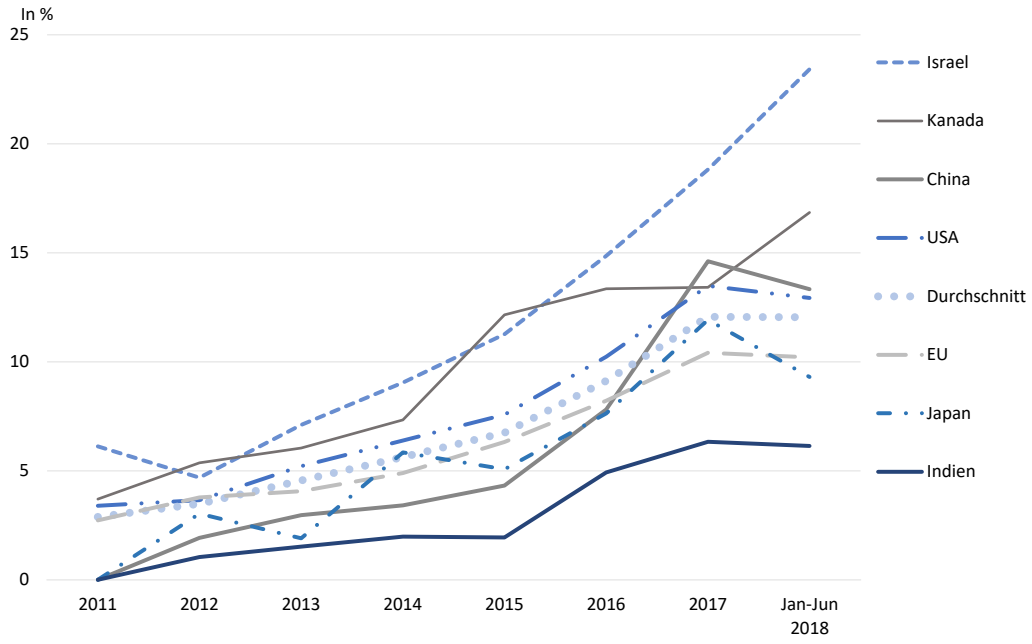
Bei der Nutzung von Crunchbase-Daten müssen einige Punkte bedacht werden, beispielsweise dass der Erfassungsbereich der Datenbank relativ breit abgegrenzt ist, dass Eigenangaben u. U. nur begrenzt zuverlässig sind und dass die Stichprobenauswahl zu Verzerrungen führen kann. Neue Transaktionen dürften zudem nicht sofort in die Datenbank eingehen, wobei die Verzögerungen in den einzelnen Ländern wohl unterschiedlich lang sind. Es ist auch möglich, dass sich Start-ups angesichts des wachsenden Interesses der Investoren an KI inzwischen selbst häufiger als KI-Start-ups bezeichnen.

In diesem Bericht bezieht sich der Begriff „KI-Start-up" auf Unternehmen, die nach 2000 gegründet wurden und bei Crunchbase dem Technologiebereich „künstliche Intelligenz" zugeordnet werden (2 436 Unternehmen). Außerdem werden Unternehmen erfasst, die in der Kurzbeschreibung ihrer Geschäftstätigkeit Schlagworte mit KI-Bezug benutzen (weitere 689 Unternehmen). Dabei werden drei Arten von Schlagwörtern berücksichtigt. Die erste Kategorie umfasst generische Schlagwörter, insbesondere „künstliche Intelligenz", „KI", „maschinelles Lernen" und „maschinelle Intelligenz". Die zweite Kategorie umfasst Schlagwörter, die sich auf KI-Technologien beziehen, insbesondere „neuronales Netz", „Deep Learning" und „bestärkendes Lernen". Die dritte Kategorie bezieht sich auf KI-Anwendungsbereiche, insbesondere „maschinelles Sehen", „Predictive Analytics", „maschinelle Sprachverarbeitung", „autonomes Fahren", „intelligente Systeme" und „virtueller Assistent".

Mehr als ein Viertel (26 %) der Datenbankeinträge über Investitionen in KI-Start-ups enthalten keine Informationen über Wagniskapitalgeber. In dieser Analyse werden die Beträge dieser Transaktionen geschätzt, indem die durchschnittliche Summe der kleineren Investitionen (Beträge unter 10 Mio. USD) im betreffenden Zeitraum und Land herangezogen wird. Größere Transaktionen werden dabei ausgeklammert, weil ihre Beträge mit größerer Wahrscheinlichkeit offengelegt werden. Der Betrag der nicht offengelegten Transaktionen im Zeitraum von 2011 bis Mitte 2018 wird auf rd. 6 % der Gesamtsumme geschätzt, was konservativ veranschlagt sein könnte. Die Zahlen für das erste Halbjahr 2018 sind wahrscheinlich konservativ, weil die Datenerfassung häufig mit Verzögerung erfolgt.

Abbildung 2.2. KI-Anteil an den Private-Equity-Investitionen in Start-ups, 2011 bis 2017 und erstes Halbjahr 2018

In Prozent der Gesamtinvestitionen



Anmerkung: Der Anteil für 2018 umfasst nur die erste Jahreshälfte (vgl. Kasten 2.1. Anmerkung zur Methodik).
Quelle: OECD-Schätzung auf der Basis von Crunchbase (Juli 2018), www.crunchbase.com.

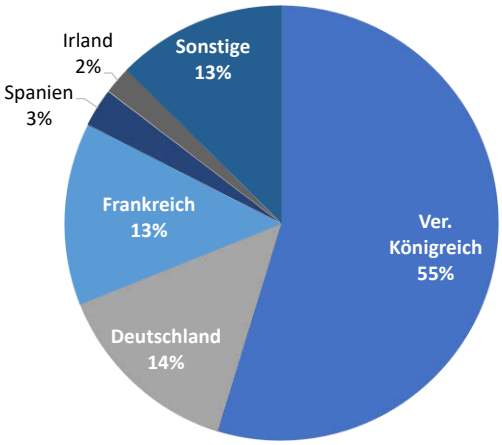
China verzeichnet seit 2016 einen gewaltigen Anstieg der Investitionen in KI-Start-ups. Das Land scheint heute gemessen am Wert der mobilisierten Beteiligungsinvestitionen im KI-Bereich weltweit an zweiter Stelle zu stehen. Gingen 2015 lediglich 3 % der globalen Private-Equity-Investitionen in KI an chinesische Unternehmen, waren es 2017 bereits 36 %. Von 2011 bis Mitte 2018 lag der Durchschnitt bei 21 %.

Auf die Europäische Union entfielen 2017 8 % der globalen Beteiligungsinvestitionen im KI-Bereich. Das bedeutet einen beträchtlichen Anstieg für die Region insgesamt, die 2013 einen Anteil von nur 1 % hatte. Das Investitionsniveau variierte jedoch erheblich zwischen den einzelnen Mitgliedstaaten. Von 2011 bis Mitte 2018 mobilisierten Start-ups im Vereinigten Königreich 55 % der Gesamtinvestitionen der Europäischen Union, gefolgt von Unternehmen in Deutschland (14 %) und Frankreich (13 %). Auf die übrigen 25 Länder entfielen damit weniger als 20 % aller in der Europäischen Union im KI-Bereich mobilisierten Private-Equity-Investitionen (Abbildung 2.3).

Von 2011 bis Mitte 2018 entfielen insgesamt mehr als 93 % aller Private-Equity-Investitionen im KI-Bereich auf die Vereinigten Staaten, China und die Europäische Union. Neben diesen Marktführern spielten Israel (3 %) und Kanada (1,6 %) ebenfalls eine Rolle.

Abbildung 2.3. Private-Equity-Investitionen in KI-Start-ups mit Sitz in der Europäischen Union, 2011 bis Mitte 2018

In Prozent des in diesem Zeitraum in Start-ups mit Sitz in der EU investierten Gesamtbetrags



Anmerkung: Der Anteil für 2018 umfasst nur die erste Jahreshälfte.
Quelle: OECD-Schätzung auf der Basis von Crunchbase (Juli 2018), www.crunchbase.com.

Die Zahl der KI-Transaktionen ist bis 2017 gestiegen – ebenso wie ihr Volumen

Die Zahl der Investitionstransaktionen ist weltweit gestiegen, von weniger als 200 Transaktionen im Jahr 2011 auf mehr als 1 400 im Jahr 2017. Das bedeutet für den Zeitraum von 2011 bis zur ersten Jahreshälfte 2018 eine durchschnittliche jährliche Wachstumsrate von 35 % (Abbildung 2.4). Ein beträchtlicher Anteil aller Investitionstransaktionen entfiel auf Start-ups mit Sitz in den Vereinigten Staaten: Zwischen 2011 und 2017 stieg die Zahl der entsprechenden Transaktionen von 130 auf über 800. Die Europäische Union verzeichnete im gleichen Zeitraum ebenfalls einen Anstieg der Zahl der Transaktionen, von rd. 30 auf über 350.

Auf Start-ups mit Sitz in China entfiel eine geringere Zahl von Transaktionen als auf Unternehmen in den Vereinigten Staaten oder in der Europäischen Union: Bei ihnen war ein Anstieg von null im Jahr 2011 auf rd. 60 im Jahr 2017 zu beobachten. Der hohe Gesamtwert der Investitionen in China deutet jedoch darauf hin, dass das durchschnittliche Volumen dieser Transaktionen deutlich höher war als in der Europäischen Union.

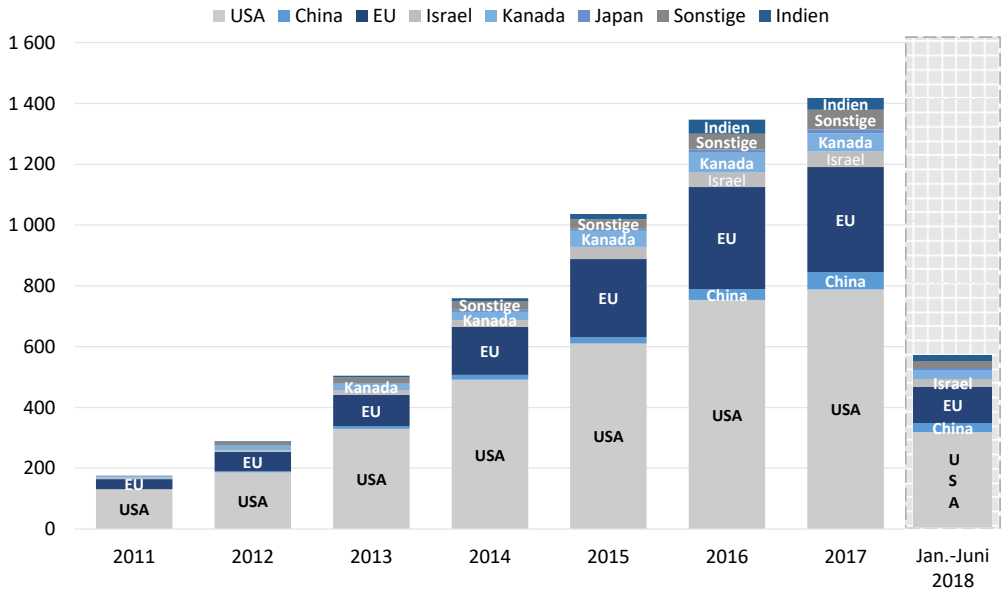
Der hohe durchschnittliche Betrag der Investitionen in China entspricht einem allgemeinen Trend hin zu wertmäßig immer größeren Investitionen. 2012 und 2013 beliefen sich fast 90 % aller verzeichneten Investitionstransaktionen auf weniger als 10 Mio. USD. Nur etwa 10 % der Transaktionen betrugen zwischen 10 Mio. und 100 Mio. USD. Keine Transaktion hatte einen Wert von über 100 Mio. USD. 2017 waren mehr als 20 % aller Transaktionen größer als 10 Mio. USD und fast 3 % lagen über 100 Mio. USD. Dieser Trend verstärkte sich in der ersten Hälfte von 2018, als 40 % der verzeichneten Transaktionen mehr als 10 Mio. USD betrugen und 4,4 % 100 Mio. USD überschritten.

Auf „Mega-Deals“ (Investitionen über 100 Mio. USD) entfielen in der ersten Jahreshälfte 2018 66 % des insgesamt in KI-Start-ups investierten Betrags. Diese Zahlen erklären sich aus dem zunehmenden Reifegrad der KI-Technologien sowie aus Investitionsstrategien, bei denen größere Beträge in eine geringere Zahl von KI-Unternehmen fließen. Das

chinesische Start-up Toutiao mobilisierte 2017 beispielsweise die größte Investition (3 Mrd. USD). Bei diesem Unternehmen handelt es sich um ein KI-gestütztes Empfehlungssystem, das mit Data-Mining arbeitet, um Nutzern in China durch die Analyse sozialer Netzwerke relevante personalisierte Inhalte vorzuschlagen.

Abbildung 2.4. Zahl der Private-Equity-Investitionen in KI-Start-ups, nach Standort

2011-2017 und erstes Halbjahr 2018



Anmerkung: Die Schätzungen für 2018 sind möglicherweise konservativ, weil Verzögerungen beim Eingang der Daten nicht berücksichtigt sind (vgl. Kasten 2.1. Anmerkung zur Methodik).
Quelle: OECD-Schätzung auf der Basis von Crunchbase (Juli 2018), www.crunchbase.com.

Seit 2016 wurden auch in Israel (Voyager Labs), der Schweiz (Mindmaze), Kanada (LeddarTech und Element AI) und im Vereinigten Königreich (Oaknorth und Benevolent AI) Investitionen im Wert von mindestens 100 Mio. USD mobilisiert. Dies zeigt, dass sich der KI-Bereich auch außerhalb der Vereinigten Staaten und Chinas dynamisch entwickelt.

Die Investitionsmuster unterscheiden sich zwischen Ländern und Regionen

Der Gesamtbetrag und die Gesamtzahl der Transaktionen sind seit 2011 erheblich gestiegen, die Investitionsmuster unterscheiden sich jedoch stark zwischen den einzelnen Ländern und Regionen.

Die Merkmale der Investitionen in chinesische Start-ups scheinen sich besonders von denen in der übrigen Welt zu unterscheiden. Die bei Crunchbase erfassten Private-Equity-Investitionen in chinesische KI-Start-ups hatten 2017 und in der ersten Jahreshälfte 2018 einen durchschnittlichen Einzelwert von 150 Mio. USD. In den anderen Ländern belief sich das durchschnittliche Investitionsvolumen 2017 nur auf ein Zehntel dieses Betrags.

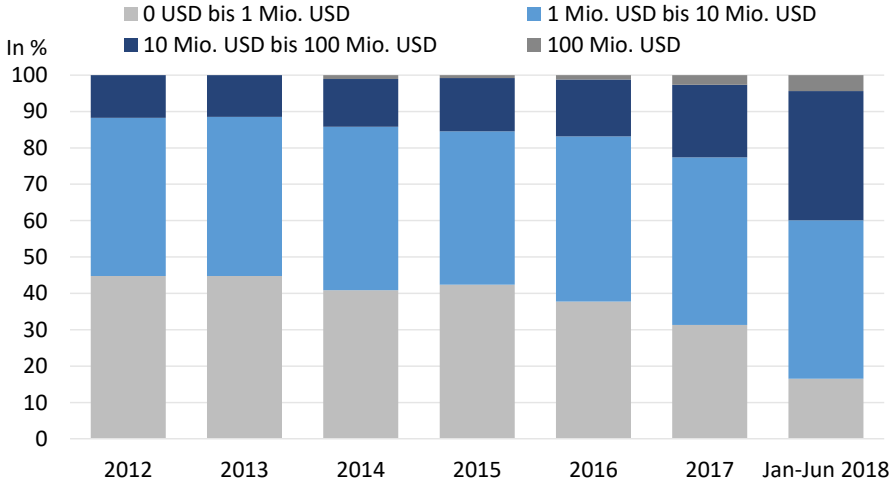
Insgesamt sind drei Muster zu beobachten. 1. Es gibt nur wenige chinesische Start-ups, die allerdings hohe Investitionen mobilisieren. 2. Die Start-ups in der EU mobilisieren eine stetig steigende Anzahl kleinerer Investitionen. Das durchschnittliche Investitionsvolumen stieg von 3,2 Mio. USD im Jahr 2016 auf 5,5 Mio. USD 2017 und 8,5 Mio. USD in der

ersten Jahreshälfte 2018. 3. Die Vereinigten Staaten verzeichnen eine stetig steigende Anzahl größerer Investitionen. Das durchschnittliche Investitionsvolumen erhöhte sich von 9,5 Mio. USD im Jahr 2016 auf 13,2 Mio. USD 2017 und 32 Mio. USD in der ersten Jahreshälfte 2018. Diese Unterschiede bei den Investitionsmerkmalen bleiben auch dann beträchtlich, wenn Transaktionen über 100 Mio. USD aus der Stichprobe ausgeklammert werden (Tabelle 2.1 und Tabelle 2.2).

Die vorstehend beschriebenen Investitionsmuster sind nicht auf KI-Start-ups begrenzt. Sie sind branchenübergreifend zu beobachten. Chinesische Start-ups mobilisierten 2017 branchenübergreifend durchschnittlich 200 Mio. USD pro Investitionsrunde. In den Vereinigten Staaten und in der Europäischen Union belief sich dieser Wert auf lediglich 22 Mio. USD bzw. 10 Mio. USD.

Abbildung 2.5. Umfang der Investitionstransaktionen, 2012-2017 und erstes Halbjahr 2018

In Prozent der Gesamtzahl der Transaktionen



Anmerkung: Die Anteile für 2018 umfassen nur die erste Jahreshälfte.
Quelle: OECD-Schätzung auf der Basis von Crunchbase (Juli 2018), www.crunchbase.com.

Tabelle 2.1. Durchschnittlicher Betrag pro Transaktion, für Investitionen bis 100 Mio. USD

In Mio. US-Dollar

	Kanada	China	EU	Israel	Japan	Vereinigte Staaten
2015	2	12	2	4	4	6
2016	4	20	3	6	5	6
2017	2	26	4	12	14	8

Quelle: OECD-Schätzung auf der Basis von Crunchbase (April 2018), www.crunchbase.com.

Tabelle 2.2. Durchschnittlicher Betrag pro Transaktion, für alle KI-Investitionen

In Mio. US-Dollar

	Kanada	China	EU	Israel	Japan	Vereinigte Staaten
2015	2	12	3	4	4	8
2016	4	73	3	6	5	10
2017	8	147	6	12	14	14

Quelle: OECD-Schätzung auf der Basis von Crunchbase (April 2018), www.crunchbase.com.

Start-ups für autonomes Fahren erhalten erhebliche Finanzmittel

Die Höhe der Private-Equity-Investitionen in KI variiert stark je nach Bereich. Ein immer größerer Teil der Private-Equity-Investitionen in KI-Start-ups entfällt auf autonomes Fahren. Bis 2015 machte autonomes Fahren noch weniger als 5 % der Gesamtinvestitionen in KI-Start-ups aus. 2017 war dieser Anteil auf 23 % gestiegen, Mitte 2018 sogar auf 30 %. Der größte Teil der Wagniskapitalinvestitionen in Start-ups für autonomes Fahren ging an Start-ups mit Sitz in den USA (80 % zwischen 2017 und Mitte 2018), gefolgt von Start-ups in China (15 %), Israel (3 %) und der Europäischen Union (2 %). Dieses Wachstum ist auf einen erheblichen Anstieg der Investitionsvolumen zurückzuführen; die Zahl der Investitionen blieb weitgehend konstant (87 im Jahr 2016 und 95 im Jahr 2017). In den Vereinigten Staaten hat sich der durchschnittliche Betrag pro Investition in diesem Bereich zwischen 2016 und der ersten Jahreshälfte 2018 von 20 Mio. auf fast 200 Mio. USD verzehnfacht. Dies war zum großen Teil auf die Investition von Softbank in Cruise Automation (3,35 Mrd. USD) zurückzuführen. Dieses Unternehmen für selbstfahrende Autos, das sich im Besitz von General Motors befindet, entwickelt Autopilotsysteme für bereits existierende Fahrzeuge. 2017 investierte Ford 1 Mrd. USD in Argo AI.

Allgemeinere Trends in der Entwicklung und Verbreitung von KI

Derzeit wird an empirischen Messgrößen für KI gearbeitet, allerdings gilt es noch einige Schwierigkeiten zu überwinden, u. a. im Hinblick auf Definitionsfragen. Klare Definitionen sind für die Entwicklung genauer und vergleichbarer Messgrößen von entscheidender Bedeutung. Die OECD und das Max-Planck-Institut für Innovation und Wettbewerb (MPI) haben in gemeinsamen experimentellen Arbeiten einen Ansatz entwickelt, der auf der Messung von drei Größen beruht: 1. KI-Entwicklungen in der Wissenschaft (zugrunde gelegt werden hier die wissenschaftlichen Veröffentlichungen), 2. Technologische Entwicklungen im KI-Bereich, gemessen an entsprechenden Patentanmeldungen, sowie 3. Softwareentwicklungen, insbesondere im Bereich der Open-Source-Software. Unter fachlicher Beratung werden bei diesem Ansatz Dokumente (Publikationen, Patente und Software) gesucht, die eindeutig KI-bezogen sind. Diese Dokumente werden dann als Benchmark genutzt, um den KI-Bezug anderer Dokumente zu beurteilen (Baruffaldi et al., erscheint demnächst[8]).

Wissenschaftliche Publikationen werden schon lange als Hilfsindikator zur Messung von Forschungsergebnissen und wissenschaftlichen Fortschritten herangezogen. Die OECD nutzt bibliometrische Daten von Scopus, einer umfangreichen Abstract- und Zitatdatenbank für begutachtete Fachliteratur und Konferenzbeiträge. Konferenzbeiträge sind in neuen Bereichen wie KI besonders wichtig. Sie bieten zeitnah einen Überblick über neue Entwicklungen, die vor der Veröffentlichung auf Fachkonferenzen erörtert werden. Bei dem von OECD und MPI entwickelten Ansatz wird mit einer von KI-Experten validierten Liste KI-bezogener Schlagwörter gearbeitet, anhand der es möglich ist, in verschiedensten wissenschaftlichen Bereichen Dokumente zu finden, die einen KI-Bezug aufweisen.

Bei der auf Patenten basierenden Komponente des Ansatzes von OECD und MPI geht es darum, KI-bezogene Erfindungen und andere technische Entwicklungen, die KI-bezogene Elemente enthalten, zu erkennen und zuzuordnen. Dabei werden mehrere Methoden herangezogen, darunter die Schlagwortsuche in Patentzusammenfassungen oder -anträgen, Analysen der Patentportfolios von KI-Start-ups sowie Analysen von Patenten, in denen KI-bezogene wissenschaftliche Dokumente zitiert werden. Dieser Ansatz wurde unter der Federführung der von der OECD geleiteten Intellectual Property (IP) Statistics Task Force verfeinert.[1]

Um KI-Entwicklungen im Softwarebereich zu erkennen, werden Daten von GitHub – der größten Hosting-Plattform für Open-Source-Software – genutzt. Die KI-Codes werden durch Modellanalysen thematisch in verschiedene Bereiche unterteilt, um die wichtigsten KI-Felder abzubilden. Die allgemeinen Felder umfassen ML (einschließlich Deep Learning), Statistik, Mathematik und wissenschaftliches Rechnen. Zu den Einzelbereichen und -anwendungen gehören Text-Mining, Bilderkennung und Biologie.

Literaturverzeichnis

Agrawal, A., J. Gans und A. Goldfarb (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business School Press, Brighton, MA. [1]

Baruffaldi, S. et al. (erscheint demnächst), "Identifying and measuring developments in artificial intelligence", *OECD Science, Technology and Industry Working Papers*, OECD Publishing, Paris. [8]

Breschi, S., J. Lassébie und C. Menon (2018), "A portrait of innovative start-ups across countries", *OECD Science, Technology and Industry Working Papers*, No. 2018/2, OECD Publishing, Paris, http://dx.doi.org/10.1787/f9ff02f4-en. [7]

Bresnahan, T. und M. Trajtenberg (1992), "General purpose technologies: 'Engines of growth?'", *NBER Working Paper*, No. 4148, http://dx.doi.org/10.3386/w4148. [2]

Brynjolfsson, E., D. Rock und C. Syverson (2017), "Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics", *NBER Working Paper*, No. 24001, http://dx.doi.org/10.3386/w24001. [3]

CBI (2018), "The race for AI: Google, Intel, Apple in a rush to grab artificial intelligence startups", *CB Insights*, 27. Februar, https://www.cbinsights.com/research/top-acquirers-ai-startups-ma-timeline/. [6]

Dilda, V. (2017), *AI: Perspectives and Opportunities*, Präsentation ei der OECD-Konferenz "AI: Intelligent Machines, Smart Policies", Paris, 26.-27. Oktober, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-dilda.pdf. [5]

MGI (2017), "Artificial Intelligence: The Next Digital Frontier?", Discussion Paper, McKinsey Global Institute, June, https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx. [4]

Anmerkungen

[1] Mitgewirkt haben dabei Experten und Patentprüfer des australischen Amts für geistiges Eigentum (Australian IP Office), des kanadischen Amts für geistiges Eigentum (Canadian Intellectual Property Office), des Europäischen Patentamts, des israelischen Patentamts (Israel Patent Office), des italienischen Patent- und Markenamts (Ufficio Italiano Brevetti e Marchi), des chilenischen Instituts für gewerbliches Eigentum (Instituto Nacional de Propiedad Industrial), des britischen Amts für geistiges Eigentum (United Kingdom Intellectual Property Office) und des US-Patent- und Markenamts (United States Patent and Trademark Office).

3. KI-Anwendungen

In diesem Kapitel werden die Möglichkeiten der künstlichen Intelligenz (KI) in verschiedenen Bereichen aufgezeigt, in denen eine rasche Einführung dieser Technologien zu beobachten ist. Dazu gehören Verkehr, Landwirtschaft, Finanzwirtschaft, Marketing und Werbung, Wissenschaft, Gesundheitswesen, Strafverfolgung, Sicherheit, der öffentliche Sektor sowie Anwendungen der erweiterten und virtuellen Realität. In diesen Bereichen können KI-Systeme genutzt werden, um in riesigen Datenmengen Muster zu erkennen und komplexe, interdependente Systeme zu modellieren. Ziel ist es, die Effizienz der Entscheidungsfindung zu steigern, Kosten einzusparen und eine bessere Ressourcenallokation zu ermöglichen. Der Abschnitt über KI im Verkehrswesen stammt vom Internet Policy Research Institute des Massachusetts Institute of Technology. Mehrere andere Abschnitte basieren auf Arbeiten, die in verschiedenen OECD-Gremien durchgeführt wurden, z. B. im Ausschuss für digitale Wirtschaft und in seiner Arbeitsgruppe Sicherheit und Datenschutz, im Ausschuss für Wissenschafts- und Technologiepolitik, in der E-Leaders-Initiative des Ausschusses für öffentliche Governance sowie im Ausschuss für Verbraucherpolitik und in seiner Arbeitsgruppe Sicherheit von Verbraucherprodukten.

KI im Verkehr: autonome Fahrzeuge

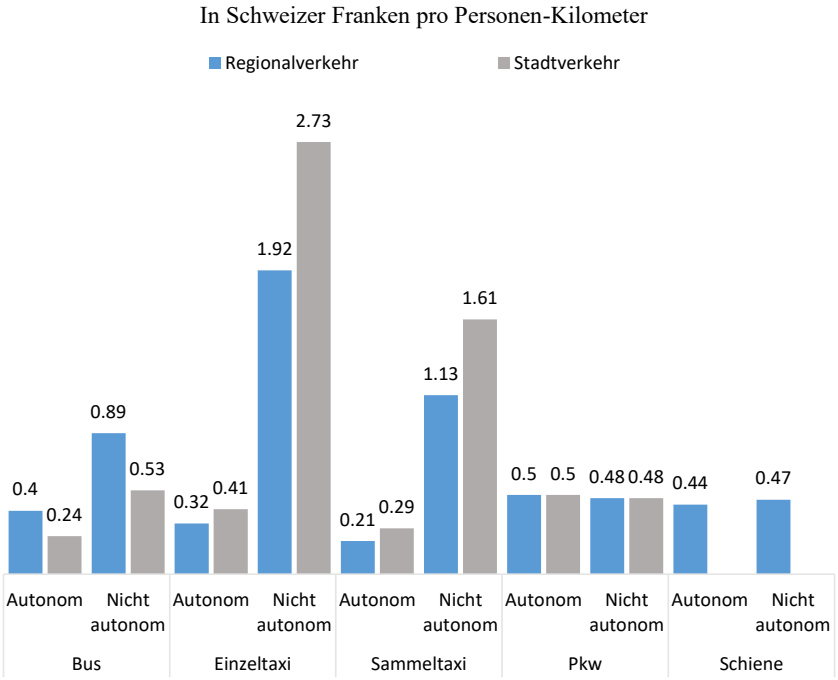
In allen Wirtschaftsbereichen entwickeln sich Systeme der künstlichen Intelligenz (KI). Besonders tiefgreifend sind indessen die Veränderungen im Verkehrssektor, insbesondere mit dem Aufkommen selbstfahrender bzw. autonomer Fahrzeuge.

Wirtschaftliche und soziale Auswirkungen autonomer Fahrzeuge

Der Verkehr ist einer der wichtigsten Wirtschaftssektoren im OECD-Raum. 2016 machte er 5,6 % des Bruttoinlandsprodukts des OECD-Raums aus (OECD, 2018[1]).[1] Dank Einsparungen aufgrund geringerer Unfallzahlen, weniger Verkehrsstauungen und anderer Vorteile könnte das autonome Fahren (AF) erhebliche wirtschaftliche Auswirkungen haben. Schätzungen zufolge könnten in den Vereinigten Staaten durch autonomes Fahren bei einer Adoptionsrate von 10 % jährlich 1 100 Leben gerettet und Einsparungen von 38 Mrd. USD erzielt werden. Bei einer Adoptionsrate von 90 % könnten jährlich sogar 21 700 Leben gerettet und 447 Mrd. USD gespart werden (Fagnant, D. und K. Kockelman, 2015[2]).

Neuere Untersuchungen für die Schweiz haben ergeben, dass die Kosten pro Kilometer für verschiedene Verkehrsmittel mit und ohne Fahrzeugautomatisierung erheblich voneinander abweichen (Bösch et al., 2018[3]). Die Ergebnisse dieser Untersuchungen deuten darauf hin, dass bei Taxis die größten Kosteneinsparungen erzielt würden. Bei Personen mit Privatfahrzeugen ist mit geringeren Kosteneinsparungen zu rechnen (Abbildung 3.1). Die Einsparungen bei Taxis sind erwartungsgemäß großenteils auf den Wegfall der Fahrervergütungen zurückzuführen.

Abbildung 3.1. Vergleich der Kosten verschiedener Verkehrsmittel mit und ohne AF-Technologie



Quelle: Nach Bösch et al. (2018[3]), "Cost-based analysis of autonomous mobility services", https://doi.org/10.1016/j.tranpol.2017.09.005.

Marktentwicklung

Der Verkehrssektor befindet sich aufgrund von drei bedeutenden aktuellen Marktveränderungen im Wandel: der Entwicklung von AF-Systemen, dem Aufkommen von Ridesharing-Diensten und der Einführung von Elektrofahrzeugen. Zwei Trends machen es traditionellen Automobilherstellern dabei schwer, eine klare Strategie festzulegen. Erstens erfreuen sich Ridesharing-Dienste zunehmender Beliebtheit, vor allem bei jüngeren Generationen. Zweitens bestehen Zweifel an der langfristigen Tragfähigkeit des traditionellen Konzepts des eigenen Autos. Premium-Hersteller experimentieren bereits mit neuen Geschäftsmodellen wie Abonnement-Services. Als Beispiele wären „Access by BMW", „Mercedes Collection" und „Porsche Passport" zu nennen. Gegen eine monatliche Pauschalgebühr können die Kunden dieser Programme nach Belieben das Fahrzeug wechseln.

Technologieunternehmen, von großen multinationalen Unternehmen bis hin zu kleinen Start-ups, steigen zunehmend in AF-Systeme, Ridesharing-Dienste oder Elektrofahrzeuge ein oder kombinieren diese Elemente. Laut jüngsten Schätzungen von Morgan Stanley könnte die Alphabet-Tochtergesellschaft Waymo bis zu 175 Mrd. USD wert sein, größtenteils aufgrund ihres Potenzials im Bereich autonome Speditions- und Lieferdienste (Ohnsman, 2018[4]). Das junge Start-up-Unternehmen Zoox, das sich auf KI-Systeme für städtische Ballungsgebiete konzentriert, hat bereits 790 Mio. USD Kapital eingeworben. Damit hat es einen geschätzten Marktwert von 3,2 Mrd. USD[2] erreicht, ohne jemals Umsätze erwirtschaftet zu haben (vgl. auch den Abschnitt „Private-Equity-Investitionen in KI-Start-ups" in Kapitel 2). Neben Technologieunternehmen investieren auch traditionelle Automobilhersteller und Zulieferer in KI-basierte Fahrzeugtechnologien.

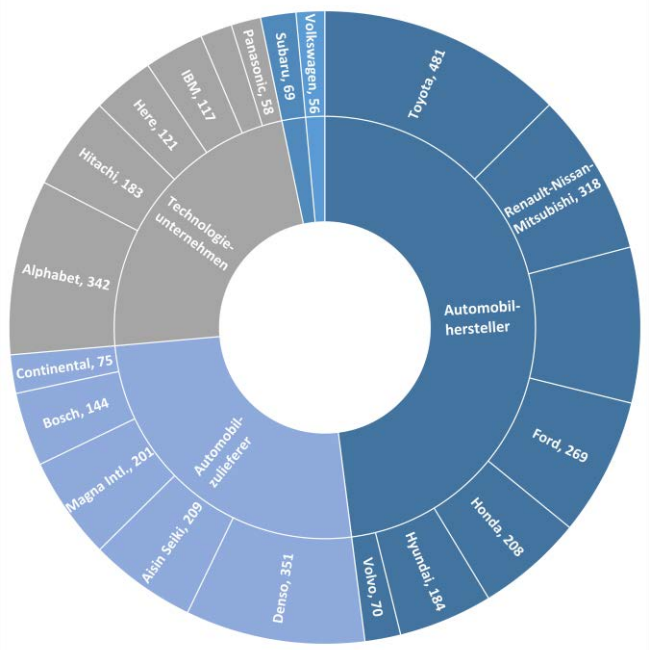
Angesichts der Komplexität von AF-Systemen konzentrieren sich die Unternehmen in der Regel auf ihre jeweilige Fachkompetenz und arbeiten dann mit Firmen zusammen, die sich auf andere Bereiche spezialisiert haben. Waymo ist eines der führenden Unternehmen im Bereich autonomes Fahren, da es sich auf große Datensätze und maschinelles Lernen (ML) spezialisiert hat. Es baut jedoch keine eigenen Fahrzeuge, sondern setzt auf Partner wie General Motors (GM) und Jaguar (Higgins, T. und C. Dawson, 2018[5]).

Auch große Automobilhersteller haben sich mit kleineren Start-ups zusammengeschlossen, um Zugang zu Spitzentechnologie zu erhalten. So kündigte Honda im Oktober 2018 an, 2,75 Mrd. USD in das zu GM gehörende Roboterwagen-Start-up Cruise Automation zu investieren (Carey, N. und P. Lienert, 2018[6]). Ridesharing-Unternehmen wie Uber haben ebenfalls erheblich in autonome Fahrzeuge investiert und sind Partnerschaften mit führenden technischen Universitäten eingegangen (CMU, 2015[7]). Dies hat jedoch Fragen der Haftung bei Unfällen aufgeworfen, insbesondere wenn mehrere Akteure für verschiedene Aspekte zuständig sind.

Die Vielfalt der Marktteilnehmer, die in AF-Kapazitäten investieren, zeigt sich in der Zahl der Patente, die von verschiedenen Unternehmensgruppen in diesem Bereich angemeldet wurden (Abbildung 3.2). Große Automobilhersteller investieren besonders stark in geistiges Eigentum, dicht gefolgt von Automobilzulieferern und Technologieunternehmen.

Abbildung 3.2. Patentanmeldungen im Bereich autonomes Fahren nach Unternehmen, 2011-2016

Unternehmen mit mehr als 50 Patentanmeldungen im Bereich autonomes Fahren



Quelle: Nach Lippert et al. (2018[8]), „Toyota's vision of autonomous cars is not exactly driverless", https://www.bloomberg.com/news/features/2018-09-19/toyota-s-vision-of-autonomous-cars-is-not-exactly-driverless.

Technologieentwicklung

Autonome Fahrzeuge verfügen über neue Sensorsysteme und Verarbeitungskapazitäten, die zu neuen Komplexitäten im Extraktions-, Transformations- und Ladeprozess ihrer Datensysteme führen. Da in allen Schlüsselbereichen des autonomen Fahrens viel investiert wird, floriert die Innovationstätigkeit. So können beispielsweise weniger kostspielige Lidar-Systeme (*light detection and ranging*) die Umgebung abbilden. Neue Technologien des maschinellen Sehens können zudem die Augenbewegungen und die Konzentration des Fahrers verfolgen und so feststellen, wenn er abgelenkt ist. Nach dem Einlesen der Daten und ihrer Verarbeitung kommt dank KI nun ein weiterer Schritt hinzu: In Sekundenschnelle werden operative Entscheidungen getroffen.

Fortschritte bei der Entwicklung autonomer Fahrzeuge werden in der Regel nach einem sechsstufigen Standard gemessen, der von der Society of Automotive Engineers (SAE) herausgegeben wurde (ORAD, 2016[9]). Die Stufen können folgendermaßen zusammengefasst werden:

Stufe 0 (keine Automatisierung): Der menschliche Fahrer kontrolliert alles. Es gibt keine automatische Lenkung, Beschleunigung, Bremsung usw.

Stufe 1 (Fahrerassistenz): Es gibt eine Grundautomatisierung, der Fahrer behält jedoch die Kontrolle über die meisten Funktionen. Der SAE zufolge können auf dieser Stufe Querführung (Lenkung) oder Längsführung (z. B. Beschleunigung) autonom erfolgen, nicht jedoch gleichzeitig.

Stufe 2 (Teilautomatisierung): Sowohl die Quer- als auch die Längsführung werden autonom übernommen, z. B. mit einem Abstandsregelautomaten und Funktionen, die das Fahrzeug in der Spur halten.

Stufe 3 (bedingte Automatisierung): Das Fahrzeug kann eigenständig fahren, bei Bedarf muss jedoch der menschliche Fahrer übernehmen. Der Fahrer ist die Rückfallebene des Systems und muss wachsam und einsatzbereit bleiben.

Stufe 4 (hohe Automatisierung): Das Fahrzeug kann selbst fahren und ist nicht auf einen Menschen angewiesen, der bei Problemen die Kontrolle übernehmen kann. Das System ist jedoch noch nicht in der Lage, in allen Situationen (besondere Anwendungsfälle, Geländemerkmale usw.) autonom zu fahren.

Stufe 5 (volle Automatisierung): Das Fahrzeug kann selbst fahren, ohne dass menschliches Eingreifen notwendig wäre, und in allen Fahrsituationen eingesetzt werden. In Fachkreisen wird lebhaft darüber diskutiert, wie weit der Weg zum vollautonomen Fahren noch ist. Uneinigkeit herrscht auch darüber, welches der richtige Ansatz ist, um autonome Fahrfunktionen zu integrieren.

Bei den Diskussionen geht es vor allem um zwei Punkte, die Rolle des Fahrers und den Einsatzbereich der Technologie:

a) **Die Rolle des Fahrers**

Verzicht auf einen menschlichen Fahrer: Einige Unternehmen, die autonome Fahrzeuge entwickeln, wie Waymo und Tesla, glauben, dass es bald möglich sein wird, auf einen menschlichen Fahrer (Fahrzeughalter oder Sicherheitsfahrer) zu verzichten. Tesla verkauft Fahrzeuge der Autonomiestufe 3. Waymo hatte Pläne, bis Ende 2018 einen vollautonomen Taxidienst ohne Fahrer in Arizona einzuführen (Lee, 2018[10]).

Unterstützung des Fahrers: Andere Systementwickler sind überzeugt, dass der beste Einsatz von AF-Systemen auf kurze Sicht nicht darin bestehen wird, den Fahrer zu ersetzen, sondern Unfälle zu vermeiden. Toyota, der weltgrößte Automobilhersteller gemessen an der Börsenkapitalisierung, setzt sich maßgeblich für die Entwicklung von Fahrzeugen ein, die keinen Unfall verursachen können (Lippert et al., 2018[8]).

b) **Einsatzbereich der Technologie**

Es gibt zwei Ansätze für die Einführung von Fahrautomatisierungssystemen, die von Walker-Smith (2013[11]) und vom Weltverkehrsforum (International Transport Forum – ITF) (2018[12]) beschrieben wurden.

Alles, aber örtlich begrenzt (*everything somewhere*): Bei diesem Ansatz ist eine sehr hohe Autonomie nur in bestimmten Gegenden oder auf bestimmten Straßen möglich, die im Detail kartografiert wurden. Das autonome Fahrassistenzsystem Super Cruise von Cadillac ist beispielsweise nur an bestimmten Orten verfügbar (z. B. funktioniert es nur auf Autobahnen, die kartografisch erfasst wurden).

Überall etwas (*something everywhere*): Bei diesem Ansatz werden Fahrzeuge nur dann mit autonomen Fahrfunktionen ausgestattet, wenn sie auf jeder Straße und in jeder Situation eingesetzt werden können. Dies hat zwar einen begrenzten Funktionsumfang zur Folge, dafür sollte es aber möglich sein, die Funktionen überall zu nutzen. Viele Automobilhersteller scheinen diesen Ansatz zu bevorzugen.

Einige Unternehmen sind indessen optimistischer. Sie scheinen sich zum Ziel gesetzt zu haben, bis 2020 oder 2021 autonome Fahrzeuge der Stufe 4 liefern zu können. Tesla und

Zoox wollen dies z. B. bis 2020 schaffen, Audi/Volkswagen, Baidu und Ford bis 2021. Renault Nissan strebt 2022 an. Auch andere Hersteller investieren hohe Summen in autonomes Fahren. Sie konzentrieren sich jedoch auf die Vermeidung von Unfällen durch menschliche Fahrer. Sie halten die Technologie für nicht ausgereift genug, um in naher Zukunft autonomes Fahren der Stufe 4 zu ermöglichen. Dazu gehören BMW, Toyota, Volvo und Hyundai (Welsch, D. und E. Behrmann, 2018[13]).

Politikfragen

Die Einführung autonomer Fahrzeuge wirft eine Reihe wichtiger rechtlicher und regulatorischer Fragen auf (Inners, M. und A. Kun, 2017[14]). Dabei geht es insbesondere um Sicherheit und Datenschutz (Bose et al., 2016[15]), aber auch ganz allgemein um Auswirkungen auf Wirtschaft und Gesellschaft (Surakitbanharn et al., 2018[16]). Herausforderungen für die Politik bestehen im OECD-Raum vor allem in folgenden Bereichen:

Sicherheit und Regulierung

Die Politik muss nicht nur Sicherheit gewährleisten (vgl. Unterabschnitt „Robustheit und Sicherheit“ in Kapitel 4), sondern sich u. a. auch mit Haftungsfragen, Betriebsmittelvorschriften für Steuerungen und Signalisierungen sowie Fahrervorschriften auseinandersetzen und sicherstellen, dass Straßenverkehrsordnungen und Betriebsvorschriften eingehalten werden (Inners, M. und A. Kun, 2017[14]).

Daten

Wie bei allen KI-Systemen wird der Zugriff auf Daten zum Trainieren und Anpassen der Systeme für den Erfolg des autonomen Fahrens von entscheidender Bedeutung sein. Die Hersteller autonomer Fahrzeuge sammeln bei ihren Testläufen immense Datenmengen. Fridman (2018[17]) schätzt, dass Tesla über Daten aus mehr als 2,4 Milliarden Kilometern verfügt, die mit seinem Autopiloten zurückgelegt wurden. Die Echtzeit-Fahrdaten, die von Entwicklern autonomer Fahrzeuge gesammelt werden, sind eigentumsrechtlich geschützt und werden nicht an andere Unternehmen weitergegeben. Es gibt jedoch Initiativen, z. B. beim Massachusetts Institute of Technology (MIT) (Fridman et al., 2018[18]), um frei zugängliche Datensätze aufzubauen, die das Verständnis des Fahrerverhaltens verbessern sollen. Aufgrund ihrer freien Zugänglichkeit sind diese Datensätze sehr wichtig für Forscher und Entwickler autonomer Fahrzeuge, die die Systeme verbessern möchten. Der Zugang zu den von verschiedenen Systemen gesammelten Daten und die Rolle staatlicher Stellen bei der Finanzierung offener Datensammlungen ist daher ein Thema für Politikdiskussionen.

Sicherheit und Datenschutz

Damit AF-Systeme zuverlässig und sicher funktionieren, bedarf es großer Mengen an Daten über System, Fahrerverhalten und Umgebung. Außerdem werden diese Systeme mit verschiedenen Netzwerken verbunden, um Informationen weiterzugeben. Die von AF-Systemen gesammelten, abgerufenen und verwendeten Daten müssen daher ausreichend gegen unerwünschte Zugriffe gesichert sein. Teilweise handelt es sich um sensible Daten wie Informationen zum Standort und Nutzerverhalten, die verwaltet und geschützt werden müssen (Bose et al., 2016[15]). Aus diesem Grund fordert das Weltverkehrsforum umfassende Cybersicherheitsbestimmungen für das automatisierte Fahren (ITF, 2018[12]). Von neuen kryptografischen Protokollen und Systemen verspricht man sich außerdem einen besseren Schutz der Privatsphäre sowie Datenschutz. Unter Umständen verlangsamen

diese Systeme jedoch die Verarbeitungszeit für erfolgs- und sicherheitskritische Aufgaben. Darüber hinaus befinden sie sich noch im Anfangsstadium und stehen noch nicht in dem Umfang und der Geschwindigkeit zur Verfügung, die für Echtzeitanwendungen im Bereich des autonomen Fahrens erforderlich sind.

Umwälzungen auf dem Arbeitsmarkt

Die Umstellung auf autonome Fahrzeuge könnte erhebliche Auswirkungen auf Fracht-, Taxi- und Lieferdienste sowie andere Dienstleistungstätigkeiten haben. In den Vereinigten Staaten gehen schätzungsweise 2,86 % der Erwerbsbevölkerung Fahrtätigkeiten nach (Surakitbanharn et al., 2018[16]). Bösch et al. (2018[3]) weisen auf die potenziell erheblichen Kosteneinsparungen hin, die in diesen Branchen durch eine Umstellung auf autonome Systeme erzielt werden könnten. Aus gewinnmaximierender Sicht wäre daher ein schneller Übergang zu autonomen Fahrzeugen zu erwarten, sobald die Technologie ausreichend ausgereift ist. Dazu müssen aber auch nichttechnische Hemmnisse, z. B. im Regulierungsbereich, überwunden werden. Der Technologiewandel wird Arbeitsplätze kosten. Die Politik muss ihr Augenmerk daher verstärkt auf die Sicherung von Kompetenzen und Beschäftigungschancen in einer sich wandelnden Arbeitswelt richten (OECD, 2014[19]).

Infrastruktur

Da es mit der Einführung des autonomen Fahrens zu einem Nebeneinander von menschlichen Fahrern und autonomen Fahrzeugen kommen wird, dürften Veränderungen der Infrastruktur erforderlich sein. Zwar dürften autonome Fahrzeuge über die notwendige Ausrüstung verfügen, um in Zukunft miteinander zu kommunizieren, herkömmliche Fahrzeuge mit menschlichen Fahrern würden jedoch ein bedeutender Unsicherheitsfaktor bleiben. Autonome Fahrzeuge müssten ihr Verhalten an Fahrzeuge anpassen, die noch von Menschen gesteuert werden. Derzeit wird über die Möglichkeit spezieller Fahrspuren für autonome Fahrzeuge oder andere Vorrichtungen diskutiert, durch die menschliche Fahrer in Zukunft von autonomen Fahrzeugen getrennt werden könnten (Surakitbanharn et al., 2018[16]). Autonome Fahrzeuge werden daher zunehmend bei der Infrastrukturplanung berücksichtigt werden müssen.

KI in der Landwirtschaft

Durch die wachsende Genauigkeit von Cognitive-Computing-Technologien, wie z. B. der Bilderkennung, erfährt die Landwirtschaft derzeit einen Wandel. Bislang mussten sich die Landwirtinnen und Landwirte auf ihren eigenen geschulten Blick verlassen, um zu entscheiden, wann welche Felder bewässert, gedüngt oder abgeerntet werden mussten. Heute können „Ernteroboter“, die mit KI-Technologien und Daten von Kameras und Sensoren ausgestattet sind, solche Entscheidungen in Echtzeit treffen. Diese Art von Roboter kann zunehmend Aufgaben übernehmen, die bisher menschliche Arbeitsleistung und Wissen erforderten.

Technologie-Start-ups entwickeln innovative Lösungen, um KI erfolgreich in der Landwirtschaft einzusetzen (FAO, 2017[20]). Diese Lösungen lassen sich in folgende Kategorien einteilen (Tabelle 3.1):

Agrarroboter übernehmen wichtige landwirtschaftliche Arbeiten wie z. B. die Ernte. Im Vergleich zu Menschen werden sie immer schneller und produktiver.

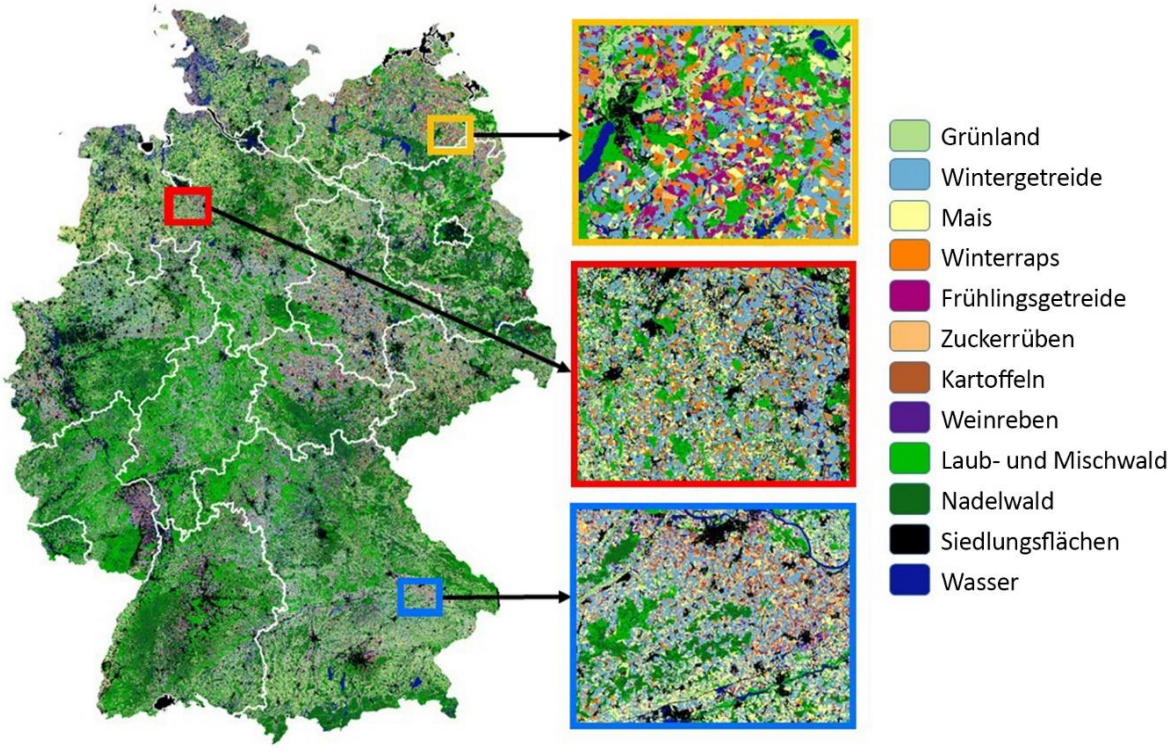
Bei der **Pflanzen- und Bodenüberwachung** werden Computer-Vision- und Deep-Learning-Algorithmen genutzt, um den Zustand von Pflanzen und Böden zu überwachen.

In immer größerer Menge zur Verfügung stehende Satellitendaten erleichtern dies (Abbildung 3.3).

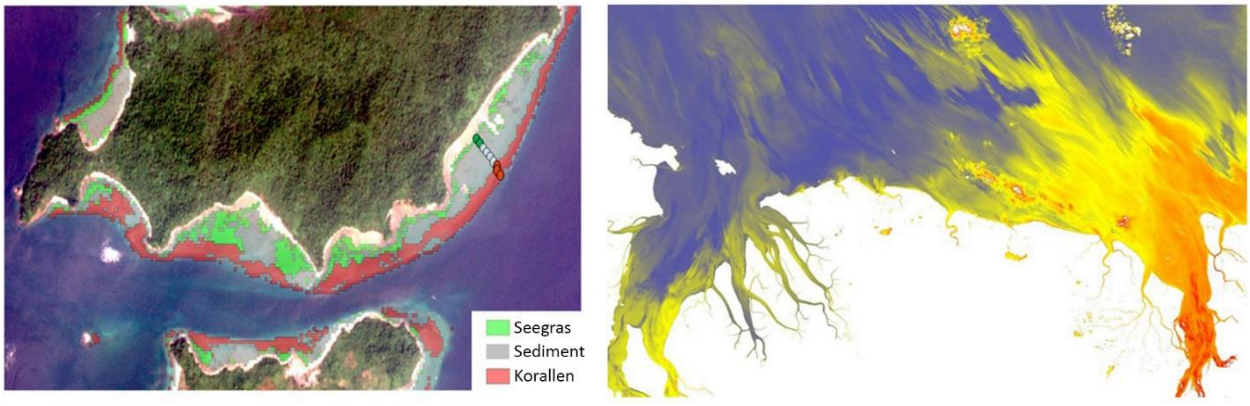
Im Bereich **Predictive Analytics** werden maschinelle Lernmodelle genutzt, um die Auswirkungen von Umweltfaktoren auf den Ernteertrag zu beobachten und vorherzusagen.

Abbildung 3.3. Beispiele für die Nutzung von Satellitendaten für ein besseres Monitoring

A. Eine ML-Anwendung analysiert Zeitreihen von Satellitendaten für die Klassifizierung von Kulturpflanzen in Deutschland



B. Meeres- und Küstenüberwachung in Australien: Meeresboden (links) und Trübung (rechts)



Quelle: Roeland (2017[21]), *EC Perspectives on the Earth Observation*, www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-roeland.pdf; Cooke (2017[22]), *Digital Earth Australia*, www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-cooke.pdf.

Tabelle 3.1. KI-Start-ups in der Landwirtschaft (Auswahl)

Kategorie	Unternehmen	Beschreibung
Agrarroboter	Abundant Robotics	Entwicklung eines Apfel-Pflückroboters, der mit Hilfe von Computer-Vision-Technologie Äpfel mit der gleichen Genauigkeit und Sorgfalt erkennt und pflückt wie ein Mensch. Laut Aussage des Unternehmens entspricht die Arbeitsleistung eines Roboters der von zehn Arbeitskräften.
	Blue River Technology	Entwicklung eines Roboters namens See & Spray, der Pflanzen und Böden überwachen und Unkräuter auf Salat- und Baumwollfeldern mit Herbiziden besprühen kann. Präzises Sprühen kann helfen, Herbizid-resistenzen zu verhindern und die Menge der verwendeten Chemikalien um 80 % zu reduzieren. Im September 2017 wurde Blue River für 305 Mio. USD von John Deere übernommen.
	Harveset CROO Robotics	Entwicklung eines Roboters, der beim Pflücken und Verpacken von Erdbeeren hilft. Pro Tag kann er 3,2 Hektar schaffen und 30 menschliche Arbeitskräfte ersetzen. So kann er Personalmangel in wichtigen Agrar-regionen begegnen und Einnahmeverluste verhindern.
Pflanzen- und Bodenüberwachung	PEAT	Entwicklung einer Deep-Learning-Anwendung, mit der potenzielle Mängel und Nährstoffdefizite im Boden aufgezeigt werden sollen. Anhand von Bildern, die von Landwirten aufgenommen wurden, erkennt diese Anwendung, ob die Pflanzen gesund sind.
	Resson	Entwicklung von Bilderkennungsalgorithmen, die Schädlings- und Krank-heitsbefall von Pflanzen genau erkennen und klassifizieren können. Resson ist eine Partnerschaft mit McCain Foods eingegangen, um Verluste in der Kartoffelproduktionskette zu minimieren.
	SkySquirrel Technologies	Entwicklung eines Systems, das anhand von Bildern die Gesundheit von Weinreben analysiert. Nutzer können von Drohnen aufgenommene Bilder in das Cloud-System des Unternehmens hochladen, welches dann den Zustand der Weinblätter diagnostiziert. Laut Aussage des Unternehmens kann seine Technologie in 24 Minuten 20 Hektar scannen und Daten-analysen mit einer Genauigkeit von 95 % liefern.
Predictive Analytics	aWhere	Entwicklung von ML-Algorithmen auf der Grundlage von Satellitendaten für Wettervorhersagen und die individuelle Beratung von Landwirten, Pflan-zenexperten und Forschern. Darüber hinaus bietet das Unternehmen seinen Nutzern täglich Zugriff auf über eine Milliarde agronomische Datenpunkte.
	FarmShots	Entwicklung eines Systems zur Analyse landwirtschaftlicher Daten, die aus Satelliten- und Drohnenbildern gewonnen werden. Das System kann Krankheiten, Schädlinge und schlecht ernährte Pflanzen erkennen und genau aufzeigen, wo die Felder gedüngt werden müssen. Der Dünge-mittelverbrauch kann so um nahezu 40 % reduziert werden.

Quelle: Websites der Unternehmen.

Herausforderungen bei der Einführung von KI in der Landwirtschaft

Projektionen der Welternährungsorganisation FAO zufolge wird die Weltbevölkerung gegenüber dem heutigen Stand bis 2050 um fast 30 % wachsen – von 7 Milliarden auf 9 Milliarden Menschen. Allerdings wird nur 4 % mehr Land bewirtschaftet werden (FAO, 2009[23]). Die OECD hat Chancen und Herausforderungen des digitalen Wandels in der Landwirtschaft und im Nahrungsmittelsektor untersucht (Jouanjean, 2019[24]). Unter den digitalen Technologien versprechen insbesondere KI-Anwendungen eine Steigerung der landwirtschaftlichen Produktivität. Bevor solche Technologien allgemein eingeführt werden können, müssen jedoch noch eine Reihe von Herausforderungen bewältigt werden (Rakestraw, 2017[25]):

- **Mangel an Infrastruktur**: In vielen ländlichen Gebieten ist der Netzzugang nach wie vor schlecht. Außerdem wären für robuste Anwendungen Data-Warehouse-Systeme erforderlich.

- **Produktion hochwertiger Daten:** Für KI-Anwendungen in der Landwirtschaft werden hochwertige Daten zur Erkennung von Kulturpflanzen bzw. Blattformen benötigt. Die Erfassung dieser Daten kann kostspielig sein, da sie nur während der jährlichen Wachstumssaison möglich ist.
- **Unterschiedliche Denkmuster von Technologie-Start-ups und Landwirten**: Produkte und Dienstleistungen werden von Technologie-Start-ups in der Regel schnell entwickelt und auf den Markt gebracht, Landwirte übernehmen neue Prozesse und Technologien hingegen eher zögerlich. Selbst große landwirtschaftliche Betriebe führen zunächst umfangreiche Feldstudien durch, um sicherzustellen, dass der Einsatz neuer Technologien mit einem klaren Nutzen verbunden ist und die Leistung gleichbleibt.
- **Kosten und insbesondere Transaktionskosten**: Um Agrarbetriebe in Hightech-Betriebe zu verwandeln (die z. B. Agrarroboter einsetzen), sind hohe Investitionen in Sensoren und Automatisierungswerkzeuge erforderlich. In Frankreich z. B. wird derzeit an Maßnahmen zur Förderung von Investitionen in bestimmte landwirtschaftliche KI-Anwendungen gearbeitet. Dadurch könnte die Einführung neuer Technologien auch für Kleinbauern erleichtert werden (OECD, 2017[26]).

Möglichkeiten zur Förderung der Einführung von KI in der Landwirtschaft

Aktuell wird nach Lösungen für die verschiedenen Herausforderungen gesucht, mit denen die künstliche Intelligenz in der Landwirtschaft konfrontiert ist. Wie in anderen Bereichen wird auch hier Open-Source-Software entwickelt. Dies könnte helfen, Kostenprobleme zu lösen. So hat Connectra beispielsweise ausgehend von der Open-Source-Software-Suite TensorFlow von Google einen Bewegungsmelder entwickelt, der am Hals von Kühen befestigt wird und deren Gesundheit überwacht (Webb, 2017[27]). Transferlernen (vgl. Unterabschnitt „Datenzugang und -nutzung“ in Kapitel 4) kann Datenprobleme verringern, da es die Möglichkeit bietet, Algorithmen mit viel kleineren Datensätzen zu trainieren. Forscher haben beispielsweise ein System entwickelt, das Krankheiten von Maniokpflanzen erkennt. Dazu wurde bei einer anderen Pflanzenart erworbenes Wissen genutzt. Mit einem Input von nur 2 756 Bildern von Maniokblättern, die von Pflanzen in Tansania stammten, gelang es den Forschern, die braune Blattfleckenkrankheit an Maniokpflanzen mit einer Genauigkeit von 98 % richtig zu erkennen (Simon, 2017[28]).

KI im Finanzsektor

Im Finanzsektor wurde KI von großen Unternehmen wie JPMorgan, Citibank, State Farm und Liberty Mutual rasch eingeführt. Das Gleiche gilt für Start-ups wie Zest Finance, Insurify, WeCash, CreditVidya und Aire. Finanzdienstleistungsunternehmen kombinieren verschiedene Methoden des maschinellen Lernens. So analysiert das französische Start-up-Unternehmen QuantCube Technology mehrere Milliarden Datenpunkte, die in über 40 Ländern gesammelt wurden. Um KI-Lösungen für die Entscheidungsfindung in Finanzunternehmen zu entwickeln, nutzt es u. a. Sprachverarbeitung, Deep Learning und Graphentheorie.

Durch den Einsatz von KI im Finanzsektor lassen sich erhebliche Nutzeffekte erzielen. So dürfte es z. B. möglich sein, die Kundenzufriedenheit zu erhöhen und Investitionschancen schneller zu erkennen. Vielleicht könnten auch mehr Kredite zu besseren Konditionen vergeben werden. Eine solche Entwicklung wirft jedoch auch Grundsatzfragen auf, beispielsweise in Bezug darauf, wie die Richtigkeit der Entscheidungen sichergestellt und

Diskriminierung vermieden werden kann und welche weitergehenden Auswirkungen die Automatisierung auf die Beschäftigung hat.

Dieser Abschnitt bietet einen Überblick über KI-Anwendungen im Finanzsektor. Er befasst sich mit den Themen Kredit-Scoring, Finanztechnologie (FinTech), Algorithmushandel, Kostensenkungen bei Finanzdienstleistungen, Kundenerlebnis und Compliance.

Kredit-Scoring

In der Finanzdienstleistungsbranche werden seit Langem statistische Ansätze für verschiedene Zwecke genutzt, u. a. zur Berechnung von Anzahlungsbeträgen und zur Abschätzung des Ausfallrisikos. Kredit-Scoring ist eine statistische Analyse, die von Finanzinstituten durchgeführt wird, um die Kreditwürdigkeit eines Kunden zu beurteilen. Anders ausgedrückt wird geprüft, mit welcher Wahrscheinlichkeit ein Kreditnehmer mit seinen Schuldendienstverpflichtungen in Verzug geraten dürfte. In traditionellen Kredit-Scoring-Modellen stellen Analysten Hypothesen über die Attribute auf, die sich auf den Kreditscore auswirken, und erstellen Kundensegmente.

Seit einiger Zeit ermöglichen neuronale Netzwerktechniken die Analyse großer Datenmengen aus Kreditauskünften. Maßgebliche Faktoren und die zwischen ihnen bestehenden Zusammenhänge können so im Detail analysiert werden. Die auf großen Datensätzen beruhenden Algorithmen der KI-Systeme bestimmen automatisch die bestmögliche Konfiguration der neuronalen Netze und davon ausgehend die verschiedenen Kundensegmente und ihre Gewichtung. Laut Kreditauskunfteien in den Vereinigten Staaten können Deep-Learning-Techniken, die Daten auf neue Art und Weise analysieren, die Genauigkeit von Prognosen um bis zu 15 % verbessern (Press, 2017[29]).

Wie in anderen Bereichen auch stellt die schwere Nachvollziehbarkeit der Ergebnisse von ML-basierten Kreditscoring-Algorithmen ein Problem dar. Die Rechtsnormen vieler Länder verlangen ein hohes Maß an Transparenz im Finanzdienstleistungssektor. Aus den Regeln des Fair Credit Reporting Act (1970) und des Equal Credit Opportunity Act (1974) in den Vereinigten Staaten ergibt sich z. B., dass die einzelnen Schritte eines Algorithmus und deren Ergebnisse nachvollziehbar sein müssen. Die Unternehmen versuchen diesen Anforderungen gerecht zu werden. So haben beispielsweise die Auskunftei Equifax und das Datenanalyse-Unternehmen SAS ein interpretierbares Kredit-Scoring-Tool entwickelt, das auf Deep Learning basiert.

Kreditvergabe von Finanztechnologie-Unternehmen

Seit einigen Jahren befinden sich Finanztechnologie-Unternehmen auf einem schnellen Wachstumskurs. FinTech-Kreditplattformen ermöglichen es Verbrauchern, in Sekundenschnelle online Kredite zu finden, zu beantragen und zu erhalten. Kreditgebern liefern sie die üblichen Kreditauskunftsdaten (etwa zum Zahlungsverhalten, zu den geschuldeten Beträgen, zur Länge der Kredithistorie und zur Anzahl der Konten). FinTech-Kreditgeber nutzen aber auch eine Vielzahl alternativer Datenquellen. Dazu gehören Versicherungsansprüche, Aktivitäten in den sozialen Medien, Informationen zu Online-Einkäufen bei Marktplätzen wie Amazon, Versanddaten von Postdiensten, Browsing-Muster und Information zum verwendeten Telefontyp oder Browser (Jagtiani, J. und C. Lemieux, 2019[30]). Untersuchungen zeigen, dass die alternativen Daten, die von FinTech-Unternehmen mit KI verarbeitet werden, den Zugang zu Krediten für Personen erleichtern können, die keine klassische Kredithistorie haben. Sie können auch die Kosten der Kreditvergabe senken, und zwar sowohl für die Verbraucher als auch für die Kreditgeber (FSB, 2017[31]).

In einer jüngeren Untersuchung wurde die Leistung von Algorithmen verglichen, die Kreditausfallwahrscheinlichkeiten auf der Basis des in den Vereinigten Staaten traditionell verwendeten FICO[3]-Scores und auf der Basis alternativer Daten vorhersagen (Berg et al., 2018[32]). Mit dem FICO-Score allein betrug die Genauigkeit 68,3 %. Die Genauigkeit eines auf alternativen Daten basierenden Algorithmus lag bei 69,6 %. Durch die Kombination beider Datentypen stieg die Genauigkeit auf 73,6 %. Diese Ergebnisse legen den Schluss nahe, dass alternative Daten die Informationen von Kreditauskunfteien zwar ergänzen, nicht aber ersetzen können. Kreditgeber können bessere Kreditentscheidungen treffen, wenn sie sowohl Informationen aus herkömmlichen Quellen (FICO) als auch alternative Daten verwenden.

In der Volksrepublik China (nachstehend „China“) hat Ant Financial gezeigt, wie KI das Kreditgeschäft beflügeln kann (Zeng, 2018[33]). Mit Hilfe von Algorithmen verarbeitet das Unternehmen riesige Mengen an Transaktionsdaten, die von kleinen Unternehmen auf seiner Plattform generiert werden. So konnte Ant bereits Kredite in Höhe von mehr als 13,4 Mrd. USD an nahezu 3 Millionen Kleinunternehmen vergeben. Die Algorithmen von Ant analysieren automatisch die Transaktions- und Verhaltensdaten aller Kreditnehmer in Echtzeit. Kleinkredite im Umfang von mehreren Hundert Yuan Renminbi (rd. 50 USD) können in wenigen Minuten bearbeitet werden. Jede auf der Plattform Alibaba ausgeführte Aktion – sei es eine Transaktion, eine Kommunikation zwischen Verkäufer und Käufer oder eine Verbindung zu anderen Diensten – wirkt sich auf den Kreditscore der Unternehmen aus. Die Algorithmen, die die Scores berechnen, entwickeln sich zudem mit der Zeit weiter, sodass die Qualität der Entscheidungsfindung mit jeder Iteration verbessert wird. Im Geschäftsbereich Mikrokredite hat Ant eine Ausfallquote von rd. 1 %. Weltweit lag die durchschnittliche Ausfallquote in diesem Bereich laut Schätzungen der Weltbank 2016 bei 4 %.

Das Kredit-Scoring-Unternehmen Alipay verwendet Verbraucherdatenpunkte, um Kreditscores zu bestimmen (O'Dwyer, 2018[34]). Dazu gehören Kaufhistorie, verwendeter Telefontyp, gespielte Spiele und Daten zu Freunden in den sozialen Medien. Zusätzlich zum herkömmlichen Scoring zur Kreditgewährung können die chinesischen „sozialen“ Kreditscores auch Einfluss auf bestimmte Entscheidungen haben, z. B. darauf, wie hoch die Kaution bei der Anmietung einer Wohnung angesetzt wird oder welche Kandidaten bei einer Online-Partnervermittlung vorgeschlagen werden. Jemand, der tagtäglich stundenlang Videospiele spielt, könnte beispielsweise einen niedrigeren sozialen Kreditscore haben als jemand, der Windeln kauft und von dem deshalb angenommen wird, dass er sich als Vater oder Mutter verantwortungsvoll verhält (Rollet, 2018[35]). 2020 soll in China ein umfassenderes soziales Kreditsystem eingeführt werden, um die „Vertrauenswürdigkeit“ von Einzelpersonen, Unternehmen und Regierungsvertretern zu bewerten.

Alternative Daten bieten die Möglichkeit, den Zugang zu Krediten zu erweitern. Die Verwendung alternativer Daten wirft aber auch Bedenken in Bezug auf Gleichbehandlung, Datenschutz, Sicherheit und Transparenz auf (Gordon, M. und V. Stewart, 2017[36]). Das Consumer Financial Protection Bureau in den Vereinigten Staaten hat daher eine Studie über die Verwendung alternativer Daten beim Kredit-Scoring angestrengt (CFPB, 2017[37]).

Einsatz von KI zur Kostensenkung bei Finanzdienstleistungen

Der Einsatz von KI kommt sowohl den Kunden als auch den Finanzinstituten zugute, im Front-Office (z. B. bei den Kontakten mit den Kunden), im Middle-Office (etwa bei der Unterstützung des Front-Office) und im Back-Office (z. B. bei der Zahlungsabwicklung, im Personalmanagement oder bei der Compliance). Es wird erwartet, dass Finanzinstitute

in den USA durch die Einführung von KI im Front-, Middle- und Back-Office bis 2030 schätzungsweise 1 Bill. USD einsparen können, was Auswirkungen auf 2,5 Millionen Mitarbeiter im Finanzdienstleistungsbereich haben dürfte (Sokolin, L. und M. Low, 2018[38]). Durch zunehmend ausgereifte KI-Anwendungen sind menschliche Eingriffe immer seltener notwendig.

Im Front-Office werden Finanzdaten und Kontobewegungen in KI-gestützte Software-Agenten integriert. Diese Agenten können sich mit Kunden auf Plattformen wie Facebook Messenger oder Slack, die fortgeschrittene Sprachverarbeitung nutzen, unterhalten. KI verbessert nicht nur den traditionellen Kundenservice, sondern wird von vielen Finanzunternehmen auch für sog. Robo-Advisor eingesetzt. Diese algorithmenbasierten Systeme bieten automatisierte Finanzberatungs- und Geldanlageleistungen an (OECD, 2017[39]).

Eine weitere interessante Entwicklung ist der Einsatz der Sentiment-Analyse auf Social-Media-Finanzplattformen. Unternehmen wie Seeking Alpha und StockTwits, die sich auf den Aktienmarkt konzentrieren, ermöglichen es Nutzern, sich zu vernetzen und Fachleute zu konsultieren, um Anlageentscheidungen zu optimieren. Die auf diesen Plattformen produzierten Daten können dann in Entscheidungsprozesse eingebunden werden (Sohangir et al., 2018[40]). KI kommt auch beim Online- und Mobile-Banking zum Einsatz, wenn Nutzer per Fingerabdruck oder Gesichtserkennung über das Smartphone authentifiziert werden. Manche Banken nutzen anstelle von numerischen Zugangscodes auch Verfahren der Spracherkennung (Sokolin, L. und M. Low, 2018[38]).

Im Middle-Office kann KI das Risikomanagement und aufsichtsrechtliche Aufgaben erleichtern. Darüber hinaus hilft KI Portfoliomanagern, effizientere und genauere Investitionsentscheidungen zu treffen. Beim Produktdesign im Back-Office sorgt KI für ein breiteres Datenangebot zur Kreditrisikobewertung, zur Übernahme des versicherungstechnischen Risikos und zur Bewertung von Schadenfällen (z. B. Begutachtung einer gebrochenen Windschutzscheibe mittels maschinellem Sehen).

Compliance

Der Finanzsektor ist für die hohen Kosten bekannt, die mit der Einhaltung von Normen und Berichterstattungspflichten verbunden sind. Neue Vorschriften in den Vereinigten Staaten und der Europäischen Union haben in den letzten zehn Jahren zu einem weiteren Anstieg der Kosten geführt, die Banken durch regulatorische Auflagen entstehen. In den letzten Jahren gaben Banken jährlich schätzungsweise 70 Mrd. USD für Compliance- und Corporate-Governance-Software aus. Diese Ausgaben spiegeln die Kosten wider, die durch die Überprüfung der Rechtskonformität der Geschäftsvorfälle durch Bankanwälte, Juristen und Compliance Officers entstehen. Es wird erwartet, dass die Kosten für solche Tätigkeiten 2020 auf nahezu 120 Mrd. USD ansteigen werden (Chintamaneni, 2017[41]). Durch den Einsatz von KI-Technologien, insbesondere von Sprachverarbeitung, dürfte es möglich sein, die Compliance-Kosten der Banken um rd. 30 % zu verringern. Der Zeitaufwand für die Überprüfung der einzelnen Geschäftsvorfälle wird erheblich sinken. KI kann bei der Interpretation von Regulierungsdokumenten und der Kodifizierung von Compliance-Regeln helfen. So überprüft beispielsweise das von JPMorgan Chase entwickelte Programm Coin Unterlagen auf der Basis von Geschäftsregeln und Methoden der Datenvalidierung. In Sekundenschnelle kann das Programm Dokumente untersuchen, für deren Prüfung ein Mensch 360 000 Stunden benötigen würde (Song, 2017[42]).

Betrugserkennung

Auch bei der Betrugserkennung setzen Finanzunternehmen verstärkt auf KI. Banken haben immer schon Kontobewegungsmuster überwacht. Inzwischen ist es dank Fortschritten im Bereich des maschinellen Lernens zunehmend möglich, Kontobewegungen nahezu in Echtzeit zu verfolgen. Dadurch können Anomalien, die eine Überprüfung erforderlich machen, sofort erkannt werden. Dass KI kontinuierlich neue Verhaltensmuster analysieren und sich automatisch anpassen kann, ist für die Betrugserkennung von besonderer Bedeutung, da sich die Verhaltensmuster rasch ändern. 2016 gründete die Credit Suisse Group AG zusammen mit dem im Sillicon Valley ansässigen Überwachungs- und Sicherheitsunternehmen Palantir Technologies ein KI-Gemeinschaftsunternehmen. Um Banken zu helfen, unerlaubte Geschäfte aufzudecken, haben die beiden Unternehmen eine Lösung entwickelt, mit der unethisch handelnde Mitarbeiter aufgespürt werden sollen, bevor sie der Bank Schaden zufügen können (Voegeli, 2016[43]). Auch im Telekommunikationssektor gewinnt die Betrugserkennung auf Basis ML-gestützter biometrischer Sicherheitssysteme an Bedeutung.

Algorithmushandel

Unter Algorithmushandel ist der Einsatz von Computeralgorithmen zu verstehen, um automatisch Entscheidungen in Bezug auf Transaktionen zu treffen, Aufträge zu erteilen und diese Aufträge anschließend zu verwalten. In den letzten zehn Jahren hat der Algorithmushandel so stark an Beliebtheit gewonnen, dass er heute die Mehrzahl der weltweit an den Börsen getätigten Transaktionen ausmacht. 2017 schätzte JPMorgan, dass es sich nur noch bei 10 % des Aktienhandelsvolumens um auf herkömmliche Weise ausgewählte Aktien (*stock picking*) handelt (Cheng, 2017[44]). Erweiterte Rechenkapazitäten begünstigen zudem den sog. Hochfrequenzhandel, bei dem täglich Millionen von Aufträgen übermittelt und viele Märkte gleichzeitig „gescannt“ werden. Während menschliche Börsenhändler zumeist die gleiche Art von Prädiktoren verwenden, können durch den Einsatz von KI vielfältigere Faktoren berücksichtigt werden.

KI in Marketing und Werbung

KI beeinflusst Marketing und Werbung in vielerlei Hinsicht. Zentral ist dabei, dass KI eine Personalisierung des Online-Erlebnisses ermöglicht. So können Inhalte angezeigt werden, an denen die Verbraucher mit großer Wahrscheinlichkeit interessiert sind. Entwicklungen im Bereich des maschinellen Lernens sowie große Mengen an generierten Daten bieten Werbetreibenden immer mehr Möglichkeiten, ihre Werbekampagnen auf bestimmte Zielgruppen auszurichten. In bislang ungekanntem Umfang können sie personalisierte und dynamische Anzeigen schalten (Chow, 2017[45]). Personalisierte Werbung bietet Unternehmen und Verbrauchern erhebliche Vorteile. Unternehmen gestattet sie, ihre Umsätze zu steigern und Investitionen in Marketingkampagnen rentabler zu machen. Verbraucher können profitieren, da durch Werbeeinnahmen finanzierte Online-Dienste den Endnutzern häufig kostenlos zur Verfügung gestellt werden. Internetsuchen werden so deutlich billiger.

In der folgenden nicht erschöpfenden Liste werden einige Entwicklungen im Bereich der künstlichen Intelligenz beschrieben, die großen Einfluss auf Marketing- und Werbepraktiken auf der ganzen Welt haben könnten:

Sprachverarbeitung: Einer der wichtigsten Teilbereiche der KI, mit dem Werbung und Marketingbotschaften stärker personalisiert werden können, ist die maschinelle Verarbeitung natürlicher Sprache (*natural language processing – NLP*). Sie ermöglicht eine

gezielte Anpassung von Marketingkampagnen ausgehend von sprachlichen Kontextinformationen, z. B. Beiträgen in den sozialen Medien, E-Mails, Kundendienstkontakten und Produktbewertungen. Durch NLP-Algorithmen lernen Maschinen Wörter und identifizieren Wortmuster in der natürlichen Sprache. Dabei verbessern sie ständig ihre Genauigkeit. So können sie Informationen über die Präferenzen und Kaufabsichten eines Kunden gewinnen (Hinds, 2018[46]). NLP kann die Qualität von Online-Suchergebnissen verbessern und dafür sorgen, dass die geschalteten Anzeigen den Interessen des Kunden stärker entsprechen, wodurch sich die Werbeeffizienz erhöht. Wenn Kunden beispielsweise online nach einer bestimmten Schuhmarke suchen, könnte ein KI-basierter Werbealgorithmus gezielt Anzeigen für diese Marke schalten, während die Kunden anderen Aktivitäten im Internet nachgehen. Mit einem solchen Algorithmus können sogar Benachrichtigungen auf das Mobiltelefon der Kunden geschickt werden, wenn sie an einem Schuhgeschäft vorbeikommen, das mit Preisnachlässen lockt.

Strukturierte Datenanalyse: Künstliche Intelligenz wird im Marketing jedoch nicht nur zur Analyse „unstrukturierter Daten" mithilfe von NLP-Modellen genutzt. Dank KI können die heutigen Online-Empfehlungsalgorithmen deutlich mehr leisten als einfache Empfehlungen oder historische Benutzer-Bewertungen. Da sie auf eine sehr breite Datenbasis zugreifen, können sie individuelle Empfehlungen abgeben. So erstellt Netflix beispielsweise eine personalisierte „Watchlist" mit Vorschlägen, bei der berücksichtigt wird, welche Filme der betreffende Kunde gesehen oder wie er sie bewertet hat. Zudem wird analysiert, welche Filme mehrfach angesehen und vor- und zurückgespult wurden (Plummer, 2017[47]).

Ermittlung der Erfolgswahrscheinlichkeit: In der Online-Werbung ist die Klickrate (auch Click-through-Rate – CTR) – d. h. die Anzahl der Personen, die auf eine Anzeige klicken, geteilt durch die Anzahl der Personen, die die Anzeige gesehen haben – eine wichtige Kennzahl zur Beurteilung der Anzeigenleistung. Daher wurden auf der Basis von ML-Algorithmen Systeme entwickelt, die die Anzahl der Klicks vorhersagen, um die Wirkung von gesponserten Anzeigen und Online-Marketingkampagnen zu maximieren. In den meisten Fällen wird mit Algorithmen des bestärkenden Lernens, sog. Reinforced-Learning-Algorithmen, eine Anzeige ausgewählt, deren Merkmale zu einer Maximierung der Klickrate in der Zielgruppe führen dürfte. Eine höhere Klickrate kann die Einnahmen von Unternehmen deutlich steigern: Bereits 1 % mehr Klicks können ausreichen, um den Umsatz und damit auch die Gewinne deutlich zu steigern (Hong, 2017[48]).

Personalisierte Preisgestaltung:[4] KI-Technologien ermöglichen es Unternehmen, ihre Preise laufend an das Verhalten und die Präferenzen der Kunden anzupassen. Gleichzeitig können sie auf die Gesetze von Angebot und Nachfrage, Ertragsanforderungen und äußere Einflüsse reagieren. ML-Algorithmen können den Höchstpreis vorhersagen, den ein Kunde für ein Produkt zu zahlen bereit ist. Die Preise werden dann direkt an dem Ort, an dem der Kundenkontakt stattfindet (z. B. auf einer Online-Plattform), individuell auf die einzelnen Verbraucher zugeschnitten (Waid, 2018[49]). Dies kann zwar für den Kunden von Vorteil sein, wenn solche Praktiken aber der Erzielung überhöhter Preise oder der Wettbewerbsverzerrung bzw. -verdrängung dienen, können sie negative Auswirkungen auf die Verbraucher haben (Brodmerkel, 2017[50]).

KI-gesteuerte erweiterte Realität: Erweiterte Realität (Augmented Reality – AR) ermöglicht die Überlagerung der vom Kunden wahrgenommenen realen Umgebung mit der digitalen Darstellung eines Produkts. In Kombination mit KI kann AR Kunden eine Vorstellung davon vermitteln, wie das Produkt aussehen dürfte, wenn es hergestellt und in den vorgesehenen physischen Kontext gestellt würde. KI-gestützte AR-Systeme können

aus den Kundenpräferenzen lernen. Sie können die computergenerierten Bilder der Produkte dann entsprechend anpassen und so das Kundenerlebnis verbessern und die Kaufwahrscheinlichkeit erhöhen (De Jesus, 2018[51]). AR könnte somit das Wachstum der Onlinemärkte und der Online-Werbeeinnahmen beflügeln.

KI in der Wissenschaft

Die Zahl der globalen Herausforderungen wächst, vom Klimawandel bis zur Antibiotikaresistenz von Bakterien. Viele dieser Herausforderungen lassen sich nur lösen, wenn der Kenntnisstand der Wissenschaft wächst. In einer Zeit, von der einige behaupten, dass es zunehmend schwer wird, neue Ideen zu entwickeln, könnte KI die Produktivität der Wissenschaft erhöhen (Bloom et al., 2017[52]). Eine solche Produktivitätssteigerung wäre auch angesichts knapper öffentlicher Forschungsbudgets zu begrüßen. Um neue Erkenntnisse zu gewinnen, müssen riesige Mengen wissenschaftlicher Daten, die mit neuen wissenschaftlichen Instrumenten generiert werden, analysiert werden. Der Einsatz von KI wird hier unverzichtbar. Da zudem die Menge an wissenschaftlichen Aufsätzen weiter wächst und die Grenze dessen, was die Wissenschaftlerinnen und Wissenschaftler lesen können, bereits erreicht sein dürfte, werden sie zur Auswertung von Forschungsergebnissen immer mehr auf die Unterstützung von KI angewiesen sein.[5]

KI kann in der Wissenschaft auch neue Formen wissenschaftlicher Entdeckung ermöglichen und die Reproduzierbarkeit wissenschaftlicher Forschung verbessern. Die Zahl wissenschaftlicher und industrieller Anwendungen von KI nimmt zu, ebenso wie ihre Bedeutung. Mit KI gelang es beispielsweise, das Verhalten chaotischer Systeme vorherzusagen, komplexe Rechenprobleme in der Genetik zu lösen, die Qualität der Astrofotografie zu verbessern und mehr über die Regeln der chemischen Synthese zu erfahren. Darüber hinaus wird KI u. a. auch zur Analyse großer Datensätze, zur Hypothesengenerierung, zur Analyse wissenschaftlicher Literatur und zur Erleichterung der Datenerfassung, des experimentellen Designs und der Versuchsdurchführung eingesetzt.

Treiber der KI in der Wissenschaft

Verschiedene Formen von KI werden bereits seit Längerem in der Wissenschaft eingesetzt, wenn auch zunächst nur sporadisch. Schon in den 1960er Jahren half das KI-Programm DENDRAL, chemische Strukturen zu erkennen. In den 1970er Jahren unterstützte eine als Automated Mathematician bekannte KI die mathematische Forschung. Seitdem wurden Computerhardware und -software erheblich verbessert, und die Datenverfügbarkeit hat deutlich zugenommen. Auch andere Faktoren begünstigen den Einsatz von KI in der Wissenschaft: KI ist zumindest im gewerblichen Bereich gut finanziert, wissenschaftliche Daten liegen in zunehmend großen Mengen vor, Hochleistungsrechner werden immer besser, und Wissenschaftler haben nun Zugang zu Open-Source-Code für künstliche Intelligenz.

Vielfältige KI-Anwendungen in der Wissenschaft

KI wird in vielen Forschungsbereichen eingesetzt. In der Teilchenphysik wird sie z. B. häufig genutzt, um komplexe räumliche Muster in riesigen Datenströmen zu finden, die von Teilchendetektoren erzeugt wurden. Durch die Verarbeitung von Daten aus den sozialen Medien gibt KI Auskunft über Zusammenhänge zwischen Sprachgebrauch, Psychologie und Gesundheit sowie sozialen und wirtschaftlichen Entwicklungen. KI hilft u. a. auch bei der Lösung komplexer Rechenprobleme in der Genetik, der Verbesserung

der Qualität der Bildgebung in der Astronomie und bei der Entschlüsselung der Regeln der chemischen Synthese (OECD, 2018[53]). KI-Anwendungen dürften in Zukunft immer häufiger und in immer größerem Umfang eingesetzt werden. Je weiter der Prozess des automatisierten maschinellen Lernens fortschreitet, desto leichter kann KI von Wissenschaftlern, Unternehmen und anderen Anwendern eingesetzt werden.

Fortschritte hat es auch in der KI-gestützten Hypothesengenerierung gegeben. Beispielsweise hat IBM mit KnIT ein Prototypsystem entwickelt, mit dem durch Mining-Methoden Informationen aus der wissenschaftlichen Literatur gewonnen werden. Das System stellt diese Informationen explizit in einem abfragbaren Netzwerk dar und zieht aus diesen Daten Schlüsse, um neue, testbare Hypothesen zu generieren. KnIT hat durch Text-Mining veröffentlichte Literatur analysiert, um Informationen über neue Kinasen zu gewinnen – Enzyme, die den Transfer von Phosphatgruppen hochenergetischer, phosphatspendender Moleküle auf bestimmte Substrate katalysieren. Dabei ging es um Kinasen, die ein Tumorsuppressorprotein phosphorylieren (Spangler et al., 2014[54]).

KI erleichtert auch die Überprüfung, das Verständnis und die Analyse wissenschaftlicher Literatur. Durch maschinelle Sprachverarbeitung ist es inzwischen möglich, sowohl Bezüge als auch Kontexte automatisch aus wissenschaftlichen Aufsätzen zu extrahieren. Das System KnIT ermöglicht es z. B., durch Text-Mining der wissenschaftlichen Literatur automatisch Hypothesen zu generieren. Das Start-up-Unternehmen Iris.AI[6] bietet ein kostenloses Tool an, um Schlüsselkonzepte aus Forschungsabstracts zu extrahieren. Dabei werden die Konzepte visuell dargestellt, damit der Benutzer interdisziplinäre Zusammenhänge erkennen kann. Außerdem stellt dieses Tool relevante Arbeiten aus einer Bibliothek mit über 66 Millionen frei zugänglichen Veröffentlichungen zusammen.

KI unterstützt die Erhebung großer Datenmengen. In der Bürgerwissenschaft (Citizen-Science) beispielsweise nutzen Anwendungen KI, um Nutzern bei der Identifizierung unbekannter Tier- und Pflanzenproben zu helfen (Matchar, 2017[55]).

KI kann auch in einem geschlossenen Forschungskreislauf mit Robotersystemen kombiniert werden

Die Konvergenz von KI und Robotertechnik birgt viele potenzielle Vorteile für die Wissenschaft. Laborautomationssysteme können Techniken aus dem KI-Bereich physikalisch nutzen, um wissenschaftliche Experimente durchzuführen. In einem Labor der Universität Aberystwyth in Wales z. B. verwendet der Roboter Adam KI-Techniken, um Zyklen wissenschaftlicher Experimente automatisch durchzuführen. Er gilt als die erste Maschine, die selbstständig neue wissenschaftliche Erkenntnisse gewinnen kann. So entdeckte er die Verbindung Triclosan, die gegen die wildtypischen, arzneimittelresistenten Parasiten Plasmodium falciparum und Plasmodium vivax wirkt (King et al., 2004[56]). Eine vollautomatisierte Forschung könnte verschiedene Vorteile haben (OECD, 2018[57]):

- **Schnellere Gewinnung wissenschaftlicher Erkenntnisse**: Automatisierte Systeme können Tausende von Hypothesen parallel generieren und überprüfen. Die kognitiven Fähigkeiten des Menschen reichen hingegen nur aus, um ein paar Hypothesen gleichzeitig zu untersuchen (King et al., 2004[56]).
- **Kostengünstigeres Experimentieren**: KI-Systeme können Experimente auswählen, deren Durchführung weniger kostenintensiv ist (Williams et al., 2015[58]). Dank ihrer Leistungsfähigkeit können sie unbekannte Versuchslandschaften effi-

zient erforschen und nutzen. Dadurch könnte es z. B. möglich sein, neue Medikamente (Segler, M., M. Preuss und M. Waller, 2018[59]), Materialien (Butler et al., 2018[60]) und Geräte zu entwickeln (Kim et al., 2017[61]).

- **Einfachere Ausbildung**: Die Ausbildung eines menschlichen Wissenschaftlers kann über zwanzig Jahre dauern und erfordert enorme Ressourcen. Menschen können sich Wissen nur langsam durch Unterricht und Erfahrung aneignen. Roboter hingegen können Wissen direkt voneinander übernehmen.
- **Verbesserter Wissens- und Datenaustausch und wissenschaftliche Reproduzierbarkeit**: Eine der wichtigsten Fragen in der Biologie – und anderen wissenschaftlichen Bereichen – ist die Reproduzierbarkeit. Roboter besitzen die dem Menschen versagte Fähigkeit, experimentelle Handlungen und deren Ergebnisse einschließlich der entsprechenden Metadaten und Verfahren automatisch, vollständig und nach anerkannten Standards aufzuzeichnen, ohne dass dadurch zusätzliche Kosten entstehen. Von Menschen durchgeführte Experimente werden hingegen um bis zu 15 % teurer, wenn dabei Daten, Metadaten und Verfahren erfasst werden.

Für die meisten wissenschaftlichen und technischen Bereiche ist Laborautomation von entscheidender Bedeutung. Aufgrund der geringen Zahl verkaufter Geräte und der mangelnden Marktreife ist sie jedoch teuer und schwierig anzuwenden. Folglich ist der Einsatz von Laborautomation an großen zentralen Standorten am wirtschaftlichsten. Deshalb wird Laborautomation in Unternehmen und Universitäten zunehmend gebündelt. Das beste Beispiel für diesen Trend ist die Cloud-Automatisierung, bei der eine große Zahl an Geräten an einem einzigen Standort bereitgestellt wird. Biologen senden dann z. B. ihre Proben ein und nutzen eine Anwendung, die ihnen beim Aufbau ihrer Experimente hilft.

Politische Grundsatzfragen

Der zunehmende Einsatz von KI-Systemen in der Wissenschaft könnte sich auch auf soziologische, institutionelle und sonstige Aspekte auswirken, z. B. auf den Wissenstransfer, auf die Systeme für die Anerkennung wissenschaftlicher Entdeckungen, auf das Peer-Review-Prinzip und auf die Systeme zum Schutz von geistigem Eigentum. Je umfassender KI in der Welt der Wissenschaft genutzt wird, umso wichtiger werden Regelungen in Bezug auf Datenzugang und Hochleistungsrechnen. Überdies wirft die zunehmend wichtige Rolle, die KI bei Neuentdeckungen spielt, neue, noch unbeantwortete Fragen auf: Sollten Maschinen in akademischen Veröffentlichungen zitiert werden? Müssen die Systeme zum Schutz von geistigem Eigentum in einer Welt, in der Maschinen erfinden können, umgestaltet werden? Wie steht es um die grundsätzliche Frage der Aus- und Weiterbildung (OECD, 2018[57])?

KI im Gesundheitswesen

Hintergrund

KI-Anwendungen in der Pharma- und Gesundheitsbranche können helfen, Erkrankungen frühzeitig zu erkennen, Präventionsleistungen zu erbringen, die klinische Entscheidungsfindung zu optimieren und neue Behandlungsmethoden und Medikamente zu entdecken. Dank KI-gestützter Instrumente, Anwendungen und Trackern zur Beobachtung von Körperdaten können sie den Weg für eine personalisierte Gesundheitsversorgung und Präzisionsmedizin ebnen. Durch KI könnten sich Vorteile sowohl in Bezug auf die Qualität als auch auf die Kosten der Versorgung ergeben. Der Einsatz von KI wirft jedoch einige

Grundsatzfragen auf, insbesondere im Hinblick auf den Zugang zu (Gesundheits-)Daten und den Datenschutz (vgl. Unterabschnitt „Schutz personenbezogener Daten“ in Kapitel 4). Dieser Abschnitt befasst sich hauptsächlich mit den speziellen Auswirkungen der KI auf das Gesundheitswesen.

In gewisser Weise ist der Gesundheitssektor eine ideale Plattform für KI-Systeme und ein hervorragendes Beispiel für ihre möglichen Auswirkungen. Als wissensintensiver Sektor ist er auf Daten und Analysen angewiesen, um Therapien und Praktiken zu verbessern. Die Bandbreite der gesammelten Informationen, darunter klinische, genetische, verhaltens- und umweltbezogene Daten, hat stark zugenommen. Täglich produzieren Gesundheitsfachkräfte, biomedizinische Forscher und Patienten mit einer Vielzahl von Geräten eine gewaltige Menge an Daten, z. B. durch elektronische Patientenakten (ePA), Genomsequenzierer, hochauflösende medizinische Bildgebung, Smartphone-Anwendungen und Sensoren sowie IoT-Geräte zur Überwachung des Gesundheitszustands der Patienten (OECD, 2015[62]).

Positive Effekte der KI auf die Gesundheitsversorgung

Die Auswertung der durch KI generierten Daten könnte für die medizinische Versorgung und Forschung von großem Nutzen sein. Durch die Nutzung der Chancen, die die Informations- und Kommunikationstechnologien bieten, erfährt der Gesundheitssektor derzeit länderübergreifend einen tiefgreifenden Wandel. Hauptziel ist dabei die Verbesserung der Effizienz, der Produktivität und der Qualität der Versorgung (OECD, 2017[26]).

Konkrete Beispiele

Verbesserung der Patientenversorgung: Durch die sekundäre Nutzung von Gesundheitsdaten kann die Qualität und Wirksamkeit der Patientenversorgung sowohl in der klinischen als auch in der häuslichen Pflege verbessert werden. So können KI-Systeme beispielsweise das administrative oder klinische Personal warnen, wenn Messgrößen für Qualität und Patientensicherheit außerhalb des Normbereichs liegen. Außerdem können sie auf Faktoren hinweisen, die möglicherweise für diese Abweichungen verantwortlich sind (Canadian Institute for Health Information, 2013[63]). Ein besonderer Aspekt der Verbesserung der Patientenversorgung ist die **Präzisionsmedizin**, die auf der schnellen Verarbeitung vieler komplexer Datensätze beruht, z. B. Patientenakten, physiologischer Reaktionen und genomischer Daten. Ein weiterer Aspekt ist die **mobile Gesundheit**: Mobile Technologien bieten hilfreiches Echtzeit-Feedback über das gesamte Versorgungsspektrum – von der Prävention über die Diagnose und Behandlung bis hin zur Überwachung von Körperdaten. In Verbindung mit anderen personenbezogenen Daten wie Standortinformationen und Präferenzen können KI-gestützte Technologien risikoreiche Verhaltensweisen erkennen oder vorteilhafte Verhaltensweisen fördern. So können sie personalisiert ein gesünderes Verhalten fördern (z. B. Treppen steigen statt mit dem Fahrstuhl fahren, Wasser trinken oder mehr zu Fuß gehen). Ebenso wie sensorbasierte Monitoringsysteme ermöglichen diese Technologien ein kontinuierliches und direktes Monitoring sowie auf individuelle Bedürfnisse zugeschnittene Maßnahmen. Sie können daher besonders nützlich sein, um die Qualität der Altenpflege und der Betreuung von Menschen mit Behinderungen zu verbessern (OECD, 2015[62]).

Management von Gesundheitssystemen: Gesundheitsdaten können in Entscheidungen über Programme, Maßnahmen und Finanzierungen einfließen. Auf diese Weise können sie das Management von Gesundheitssystemen erleichtern und zur Verbesserung ihrer

Effektivität und Effizienz beitragen. Beispielsweise können KI-Systeme erkennen, wenn Behandlungsmaßnahmen ineffektiv sind, Chancen verpasst oder Leistungen doppelt erbracht werden und so Kosten senken. Sie können auf vier Arten den Zugang zur Gesundheitsversorgung verbessern und Wartezeiten verkürzen: 1. KI-Systeme sind in der Lage zu analysieren, welche Versorgungsetappen der Patient durchlaufen muss. 2. Sie sorgen dafür, dass die Patienten die für ihre Bedürfnisse am besten geeigneten Leistungen erhalten. 3. Sie können zukünftige Gesundheitsbedürfnisse der Bevölkerung vorhersagen. 4. Sie optimieren die Ressourcenallokation im gesamten System (Canadian Institute for Health Information, 2013[63]). Vor dem Hintergrund des zunehmenden Monitorings von Therapien und Ereignissen im Zusammenhang mit Arzneimitteln und Medizinprodukten (OECD, 2015[62]) können Länder KI nutzen, um Muster besser zu erkennen, z. B. systemische Erfolge oder Misserfolge. Darüber hinaus fördern datengesteuerte Innovationen die Vision eines „lernenden Gesundheitssystems", das in der Lage ist, kontinuierlich neue Daten von Forschern, Leistungserbringern und Patienten einzubeziehen. Dadurch können umfangreiche klinische Algorithmen verbessert werden, die an bestimmten Entscheidungsknoten der klinischen Versorgung helfen zu bestimmen, welches die optimale Vorgehensweise ist (OECD, 2015[62]).

Analyse und Lösung von Fragen der öffentlichen Gesundheit: Daten können nicht nur für eine zeitnahe Überwachung von Gesundheitskrisen wie Grippe-Epidemien und anderen Viruserkrankungen verwendet werden, sondern auch, um unerwartete Nebenwirkungen und Kontraindikationen neuer Arzneimittel zu erkennen (Canadian Institute for Health Information, 2013[63]). KI-Technologien ermöglichen es, Krankheitsausbrüche frühzeitig zu erkennen und ihre Ausbreitung zu überwachen. Über die sozialen Medien ist es beispielsweise möglich, Informationen, die die öffentliche Gesundheit betreffen, zu erhalten und zu verbreiten. KI kann Instrumente der maschinellen Sprachverarbeitung (NLP) nutzen, um Beiträge in sozialen Medien zu analysieren und ihnen Informationen über potenzielle Nebenwirkungen zu entnehmen (Comfort et al., 2018[64]; Patton, 2018[65]).

Förderung der Gesundheitsforschung: Gesundheitsdaten können die klinische Forschung unterstützen und die Entdeckung neuer Therapien beschleunigen. Durch Big-Data-Analysen ergeben sich neue, leistungsfähigere Optionen zur Beobachtung von Krankheitsverlauf und Gesundheitszustand. Dies kann eine verbesserte Diagnose und Versorgung ermöglichen und auch die translationale und klinische Forschung unterstützen, z. B. zur Entwicklung neuer Arzneimittel. 2015 arbeitete das Pharmaunternehmen Atomwise beispielsweise mit Forschern der Universität Toronto und IBM zusammen, um KI-Technologie bei der Suche nach einem Mittel gegen das Ebola-Virus einzusetzen.[7] Auch in der medizinischen Diagnostik wird zunehmend mit KI experimentiert. Ein Meilenstein war hier die Entscheidung der Food and Drug Administration der Vereinigten Staaten, die Vermarktung des ersten medizinischen Geräts zuzulassen, das KI einsetzt, um bei diabeteskranken Erwachsenen eine über die milde Form hinausgehende diabetische Retinopathie zu erkennen (FDA, 2018[66]). Techniken des maschinellen Lernens können zudem herangezogen werden, um Modelle zur Klassifizierung von Bildern des Auges zu trainieren. So könnten z. B. Kataraktdetektoren entwickelt werden, die in Smartphones integriert und damit auch in entlegenen Gebieten eingesetzt werden könnten (Lee, 2017[67]; Patton, 2018[65]). In einer neueren Studie wurden einem Deep-Learning-Algorithmus mehr als 100 000 Bilder von bösartigen Melanomen und gutartigen Muttermalen zugeführt. Am Ende war dieser Algorithmus erfolgreicher bei der Erkennung von Hautkrebs als eine Gruppe von 58 internationalen Dermatologen (Mar, 2018[68]).

KI im Gesundheitswesen – Erfolgs- und Risikofaktoren

Es bedarf einer ausreichenden Infrastruktur und Mechanismen der Risikominderung, um die Möglichkeiten der KI im Gesundheitssektor voll auszuschöpfen.

In den verschiedenen Ländern werden zunehmend Systeme für elektronische Patientenakten eingerichtet und mobile Gesundheitstechnologien (*m-health*) eingeführt (OECD, 2017[69]). Eine Studie lieferte stichhaltige Belege dafür, dass elektronische Patientenakten zur Reduzierung von Medikationsfehlern und zu einer besseren Koordination der Gesundheitsversorgung beitragen können (OECD, 2017[26]). Die gleiche Studie zeigte aber auch, dass es bislang nur wenigen Ländern gelungen ist, solche Technologien umfassend zu integrieren und die Möglichkeiten der Datenextraktion aus elektronischen Patientenakten für Forschung, Statistik und andere sekundäre Anwendungen zu nutzen. Silostrukturen mit getrennter Datenerfassung und -analyse sind im Gesundheitssystem immer noch die Regel. Deshalb sind Standards und Interoperabilität zentrale Herausforderungen, die angegangen werden müssen, um das Potenzial elektronischer Patientenakten voll auszuschöpfen (OECD, 2017[26]).

Ein weiterer kritischer Faktor für den Einsatz von KI im Gesundheitssektor ist die Minimierung der **Datenschutzrisiken für die Betroffenen**. Die Risiken, die bei einer verstärkten Erfassung und Verarbeitung personenbezogener Daten bestehen, werden ausführlich im Unterabschnitt „Schutz personenbezogener Daten“ in Kapitel 4 beschrieben. Dieser Unterabschnitt befasst sich mit der hohen Sensibilität gesundheitsbezogener Daten. Bias-bedingte Fehlleistungen eines Algorithmus, der eine bestimmte Behandlung empfiehlt, könnten in einigen Gruppen zu echten Gesundheitsrisiken führen. Andere Datenschutzrisiken sind Besonderheiten des Gesundheitssektors. Beispielsweise stellt sich die Frage, ob Daten aus implantierbaren Medizinprodukten wie Herzschrittmachern als Beweismittel vor Gericht verwendet werden können.[8] Da solche Medizinprodukte zudem immer komplexer werden, können die an sie geknüpften Sicherheitsrisiken steigen, wie z. B. das Risiko eines böswilligen Eingriffs in das System, um einen schädlichen Vorgang auszuführen. Die Verwendung biologischer Proben (z. B. Gewebe) für maschinelles Lernen ist ein weiteres Beispiel, das komplexe Einwilligungs- und Eigentumsfragen aufwirft (OECD, 2015[62]; Ornstein, C. und K. Thomas, 2018[70]).[9]

Aufgrund solcher Bedenken verweisen viele OECD-Länder auf rechtliche Hindernisse bei der Verwendung personenbezogener Gesundheitsdaten. Zu diesen Hindernissen gehören die Deaktivierung von Datenverknüpfungen und Hemmnisse für die Entwicklung von Datenbanken aus elektronischen Patientenakten. Die OECD-Ratsempfehlung zur Governance von Gesundheitsdaten aus dem Jahr 2016 ist ein wichtiger Schritt auf dem Weg zu einem kohärenteren Ansatz für die Verwaltung und Nutzung von Gesundheitsdaten (OECD, 2016[71]). Sie zielt in erster Linie darauf ab, die Einrichtung und Umsetzung eines nationalen Governance-Rahmens für Gesundheitsdaten zu fördern. Ein solcher Rahmen würde die Verfügbarkeit und Verwendung personenbezogener Gesundheitsdaten im öffentlichen Interesse begünstigen und zugleich den Schutz von Persönlichkeitsrechten und personenbezogenen Gesundheitsdaten sowie die Datensicherheit stärken. Ein kohärenter Ansatz in der Datenverwaltung könnte dazu beitragen, den Zielkonflikt zwischen Datennutzung und Datensicherheit zu überwinden.

Die Einbeziehung aller relevanten Akteure trägt maßgeblich dazu bei, Vertrauen zu schaffen und die öffentliche Akzeptanz des Einsatzes von KI und der Erfassung von Daten für Gesundheitszwecke zu erhöhen. Die zuständigen staatlichen Stellen könnten Schulungsprogramme für speziell mit Gesundheitsdaten befasste Datenwissenschaftler konzipieren oder Datenwissenschaftler mit Gesundheitsfachkräften zusammenbringen. Auf

diese Weise könnten sie zu einem besseren Verständnis der Chancen und Risiken in diesem neuen Bereich beitragen (OECD, 2015[62]). Die Einbeziehung medizinischer Fachkräfte in das Design und die Entwicklung von KI-Gesundheitssystemen könnte sich als unerlässlich erweisen, damit Patienten und Leistungserbringer KI-basierten Gesundheitsprodukten und -dienstleistungen Vertrauen schenken.

KI in der Strafverfolgung

KI und prädiktive Algorithmen für die Justiz

Künstliche Intelligenz kann den Zugang zur Justiz verbessern und eine effektive und unparteiische Rechtsprechung fördern. Bedenken bestehen jedoch wegen der Fragen, die KI-Systeme im Hinblick auf Bürgerbeteiligung, Transparenz, Würde, Persönlichkeitsrechte und Freiheit aufwerfen könnten. Dieser Abschnitt befasst sich hauptsächlich mit KI-Fortschritten im Bereich der Strafverfolgung, geht jedoch auch auf Entwicklungen in anderen Rechtsgebieten ein.

KI wird zunehmend in verschiedenen strafrechtlichen Abläufen eingesetzt, u. a. um vorherzusagen, wo Verbrechen begangen werden könnten oder wie ein bestimmtes Strafverfahren ausgehen wird, das von einem Angeklagten ausgehende Risiko zu bewerten oder zu einer effizienteren Prozessführung beizutragen. Viele KI-Anwendungen sind zwar noch in der Erprobung, einige ausgereifte Prognoseprodukte werden jedoch bereits in der Justizverwaltung und im Gesetzesvollzug eingesetzt. Mit KI lassen sich u. U. besser Verbindungen herstellen, Muster erkennen sowie Verbrechen verhindern und aufklären (Wyllie, 2013[72]). Der zunehmende Einsatz solcher Instrumente folgt dem generellen Trend, sich auf faktenbasierte Methoden zu stützen, die eine effizientere, rationellere und kostengünstigere Nutzung der knappen Ressourcen für den Gesetzesvollzug gewährleisten (Horgan, 2008[73]).

Die Strafjustiz ist ein kritischer Kontaktpunkt zwischen Staat und Bürgern, an dem Macht- und Informationsgefälle besonders deutlich zum Ausdruck kommen. Ohne hinreichende Sicherheitsmechanismen kann es hier zu überproportional negativen Effekten kommen, können systemische Benachteiligungen verstärkt werden oder möglicherweise sogar neue Benachteiligungen entstehen (Barocas, S. und A. Selbst, 2016[74]).

Vorausschauende Polizeiarbeit

Beim Predictive Policing („vorausschauende Polizeiarbeit“) wird KI zur Mustererkennung eingesetzt, um statistische Vorhersagen über potenzielle kriminelle Tätigkeiten zu treffen (Ferguson, 2014[75]). Bereits vor der Einführung von KI wurden in der Polizeiarbeit prädiktive Verfahren eingesetzt. So wurde beispielsweise gesammeltes Datenmaterial analysiert, um Städte in Viertel mit hohem und niedrigem Kriminalitätsrisiko einzuteilen (Brayne, S., A. Rosenblat und D. Boyd, 2015[76]). Dank KI können nun mehrere Datensätze miteinander verknüpft sowie komplexe und feinkörnigere Analysen durchgeführt werden, die präzisere Vorhersagen ermöglichen. Automatische Nummernschilderfassung, ein engmaschiges Kameranetz, eine zunehmend kostengünstige Datenspeicherung und erweiterte Rechenkapazitäten können es der Polizei gestatten, wesentliche Daten über eine große Zahl von Personen zu gewinnen. Anhand dieser Daten kann die Polizei dann Muster – d. h. auch kriminelle Verhaltensmuster – erkennen (Joh, 2017[77]).

Im Predictive Policing werden zwei Herangehensweisen unterschieden. Bei der **ortsbezogenen Vorhersage** werden retrospektive Kriminalitätsdaten genutzt, um vorherzusagen, wann und wo Verbrechen verübt werden könnten. Zu den berücksichtigten Orten können

Spirituosengeschäfte, Bars und Parks gehören, in denen in der Vergangenheit bestimmte Verbrechen begangen wurden. Die Polizei kann dann einen Beamten entsenden, der an einem bestimmten Wochentag bzw. zu einer bestimmten Uhrzeit in diesen Bereichen auf Streife geht, um Verbrechen vorzubeugen. Bei der **personenbezogenen Vorhersage** ziehen die Strafverfolgungsbehörden Kriminalstatistiken heran, um vorherzusagen, welche Personen oder Gruppen mit besonders großer Wahrscheinlichkeit mit Verbrechen in Berührung kommen, sei es als Opfer oder als Täter.

KI-gestützte Initiativen für die vorausschauende Polizeiarbeit werden in Städten auf der ganzen Welt erprobt, so z. B. in Manchester, Durham, Bogota, London, Madrid, Kopenhagen und Singapur. Im Vereinigten Königreich hat die Greater Manchester Police 2012 ein prädiktives System zur Verbrechenskartierung entwickelt. In der Grafschaft Kent setzt die Polizei seit 2013 das System PredPol ein. Diese beiden Systeme schätzen die Wahrscheinlichkeit von an bestimmten Orten während eines bestimmten Zeitfensters verübten Verbrechen. Sie beruhen auf einem Algorithmus, der ursprünglich zur Vorhersage von Erdbeben entwickelt wurde.

In Bogota (Kolumbien) nutzt die Data-Pop Alliance Kriminalitäts- und Verkehrsdaten, um Kriminalitätsbrennpunkte zu ermitteln. Die Polizeikräfte werden dann gezielt an den Orten und zu den Uhrzeiten eingesetzt, an denen das Kriminalitätsrisiko besonders hoch ist.

Viele Polizeibehörden stützen sich zudem zu unterschiedlichen Zwecken auf die sozialen Medien, beispielsweise zur Aufdeckung krimineller Tätigkeiten, zur Erlangung von Durchsuchungsbefehlen auf der Basis eines konkretisierten Anfangsverdachts oder zur Sammlung von Beweismitteln für Gerichtsverhandlungen. Sie nutzen die sozialen Medien auch zur Ermittlung der Aufenthaltsorte von Straftätern, für den Umgang mit unberechenbaren Situationen, zur Identifizierung von Zeugen, zur Verbreitung von Informationen und für Bitten um Hinweise aus der Öffentlichkeit (Mateescu et al., 2015[78]).

Allerdings wirft der Einsatz künstlicher Intelligenz auch Fragen in Bezug auf die Verwendung personenbezogener Daten (vgl. Unterabschnitt „Schutz personenbezogener Daten“ in Kapitel 4) und Bias-Risiken (vgl. Unterabschnitt „Fairness und Ethik“ in Kapitel 4) auf. Insbesondere ihre z. T. mangelnde Transparenz und ihre nicht immer nachvollziehbare Funktionsweise ist ein heikler Punkt, der Anlass zu Besorgnis gibt. Ein Ansatz zur Erhöhung der Transparenz von Algorithmen ist das im Vereinigten Königreich angewandte Tool ALGO-CARE. Mit ihm soll sichergestellt werden, dass die Polizei bei der Nutzung algorithmischer Instrumente zur Risikobewertung rechtliche und praxisrelevante Grundregeln berücksichtigt (Burgess, 2018[79]). Grundsätze des öffentlichen Rechts und der Menschenrechte werden dabei in praktische Begriffe und Leitlinien für Polizeibehörden umgesetzt.

Nutzung von KI durch die Justiz

In vielen Staaten wird KI von der Justiz hauptsächlich zur Risikobewertung eingesetzt. Risikobewertungen fließen in eine Reihe von strafrechtlichen Entscheidungen ein, z. B. bei der Festlegung der Kautionshöhe oder anderer Bedingungen für Haftentlassungen und Strafaussetzungen zur Bewährung (Kehl, D., P. Guo und S. Kessler, 2017[80]). Richter nutzen seit Langem mathematisch-statistische Instrumente zur Risikobewertung, der Übergang zur KI war daher eine logische Entwicklung (Christin, A., A. Rosenblat und D. Boyd, 2015[81]). Forscher des Berkman Klein Center an der Harvard-Universität arbeiten gegenwärtig an einer Datenbank, in der alle Instrumente zur Risikobewertung erfasst werden sollen, die in den Strafjustizsystemen der Vereinigten Staaten zur Unterstützung des Entscheidungsprozesses eingesetzt werden (Bavitz, C. und K. Hessekiel, 2018[82]).

Algorithmen zur Risikobewertung prognostizieren das Risikoniveau auf der Basis einer kleinen Anzahl von Faktoren, die in der Regel in zwei Gruppen unterteilt sind. Dies sind zum einen Strafregistereinträge (z. B. frühere Verhaftungen und Verurteilungen oder Vorladungen vor Gericht, denen nicht Folge geleistet wurde) und zum anderen soziodemografische Merkmale (z. B. Alter, Geschlecht, Beschäftigung und Aufenthaltsstatus). Prädiktive Algorithmen fassen die relevanten Informationen zusammen, um ausgehend davon Entscheidungen zu treffen. Dabei sollen sie effizienter sein als das menschliche Gehirn, da sie größere Datenmengen rascher verarbeiten können und weniger durch Vorurteile beeinflusst sein dürften (Christin, A., A. Rosenblat und D. Boyd, 2015[81]).

KI-basierte Instrumente zur Risikobewertung, die von Privatunternehmen entwickelt werden, wecken jedoch Bedenken im Hinblick auf ihre Transparenz und Nachvollziehbarkeit. Dies liegt daran, dass Geheimhaltungsvereinbarungen häufig den Zugang zu proprietärem Code verhindern, um geistiges Eigentum zu schützen oder böswilligen Handlungen vorzubeugen (Joh, 2017[77]). Ohne Zugriff auf den Code ist es nur begrenzt möglich, die Gültigkeit und Zuverlässigkeit der Instrumente zu überprüfen.

Ein interessantes Beispiel hierfür ist das in einigen US-Bundesstaaten eingesetzte proprietäre Instrument COMPAS, das von der gemeinnützigen Nachrichtenredaktion ProPublica einem Test unterzogen wurde. ProPublica fand heraus, dass die Vorhersagen von COMPAS für alle Formen der Kriminalität zusammengenommen eine Genauigkeit von durchschnittlich 60 % aufwiesen. Die Vorhersagegenauigkeit für Gewaltverbrechen betrug jedoch nur 20 %. Darüber hinaus traten bei der Studie Unterschiede je nach Hautfarbe der Betroffenen zutage: Schwarze wurden von dem Algorithmus doppelt so häufig fälschlicherweise als zukünftige Straftäter eingestuft wie Weiße (Angwin et al., 2016[83]). Die Studie erregte die Aufmerksamkeit der Medien, und ihre Ergebnisse wurden auf der Basis statistischer Fehler infrage gestellt (Flores, A., K. Bechtel und C. Lowenkamp, 2016[84]). COMPAS ist ein sog. „Black-Box"-Algorithmus. Das bedeutet, dass niemand Zugriff auf den Quellcode hat – auch nicht seine Anwender.

Die Nutzung von COMPAS wurde gerichtlich angefochten. Seinen Gegnern zufolge verletze sein proprietärer Charakter das Recht des Angeklagten auf ein faires Gerichtsverfahren. Der Oberste Gerichtshof von Wisconsin hat den Einsatz von COMPAS bei der Strafzumessung gebilligt. Es muss jedoch ein unterstützendes Instrument bleiben, und der Richter muss nach wie vor nach eigenem Ermessen bestimmen können, welche weiteren Faktoren zu berücksichtigen sind und welches Gewicht ihnen beizumessen ist.[10] Der Oberste Gerichtshof der Vereinigten Staaten lehnte einen Antrag auf Anhörung des Falls ab.[11]

Kleinberg et al. (2017[85]) entwickelten im Rahmen einer anderen Studie, die sich mit den Auswirkungen von KI auf die Strafjustiz befasst, einen ML-Algorithmus. Dieser Algorithmus sollte vorhersagen, welche Angeklagten im Zeitraum bis zum Gerichtsprozess eine weitere Straftat begehen würden oder versuchen würden, sich der Justiz zu entziehen. Die Eingangsvariablen waren bekannt. Der Algorithmus sollte die relevanten Unterkategorien und ihr jeweiliges Gewicht bestimmen. Für die Altersvariable beispielsweise bestimmte der Algorithmus die Altersstufen mit der größten statistischen Signifikanz, nämlich 18-25 und 25-30 Jahre. Die Autoren stellten fest, dass dieser Algorithmus die Inhaftierungsquoten und die Risiken einer Ungleichbehandlung von Personen unterschiedlicher Hautfarbe erheblich reduzieren könnte. Der Algorithmus verringerte nach Ansicht der Autoren auch die Verzerrungen durch menschliche Vorurteile. Sie kamen zu dem Schluss, dass alle Informationen, die über die für die Vorhersage notwendigen Elemente hinausgehen, die Richter ablenken und das Risiko verzerrter Urteile erhöhen könnten.

Moderne KI-gestützte Instrumente zur Risikobewertung werden auch im Vereinigten Königreich eingesetzt. Bei der Polizei von Durham wurde das sog. Harm Assessment Risk Tool entwickelt, um das Rückfallrisiko von verurteilten Straftätern zu bewerten. Das Instrument basiert auf den vergangenen Gesetzesverstößen der betreffenden Personen, ihrem Alter, ihrer Postleitzahl und anderen Hintergrundmerkmalen. Anhand dieser Indikatoren wird das mit ihnen verbundene Risiko vom Algorithmus als gering, mittel oder hoch eingestuft.

KI zur Vorhersage des Ausgangs gerichtlicher Verfahren

Unter Verwendung fortschrittlicher Sprachverarbeitungstechniken und Datenanalysefunktionen haben mehrere Forscher Algorithmen entwickelt, um den Ausgang gerichtlicher Verfahren mit hoher Genauigkeit vorherzusagen. Forscher des University College London und der Universitäten Sheffield und Pennsylvania haben beispielsweise einen ML-Algorithmus entwickelt, der das Ergebnis von vor dem Europäischen Gerichtshof für Menschenrechte verhandelten Verfahren mit einer Genauigkeit von 79 % vorhersagen kann (Aletras et al., 2016[86]). In einer weiteren Studie wurde von Forschern des Illinois Institute of Technology in Chicago ein Algorithmus entwickelt, der das Ergebnis von Rechtssachen, die vor dem Obersten Gerichtshof der Vereinigten Staaten entschieden werden, mit einer Genauigkeit von 79 % vorhersagen kann (Hutson, 2017[87]). Solche Algorithmen könnten den Parteien helfen, die Wahrscheinlichkeit eines Erfolgs im Straf- oder Rechtsmittelverfahren (auf der Basis ähnlicher früherer Fälle) zu beurteilen. Anwälte könnten mit ihnen eventuell auch leichter herausfinden, welche Punkte sie hervorheben müssen, um ihre Gewinnchancen zu erhöhen.

Andere Einsatzmöglichkeiten künstlicher Intelligenz in Rechtsverfahren

In zivilrechtlichen Verfahren wird KI vielfältig genutzt. Anwälte setzen künstliche Intelligenz für die Vertragsgestaltung, die Analyse von Dokumenten und die Extraktion von darin enthaltenen Informationen bei der Sachverhaltserforschung und bei Due-Diligence-Prüfungen ein (Marr, 2018[88]). Der Einsatz von KI könnte auf ähnliche Bereiche des Strafjustizsystems ausgedehnt werden, z. B. auf Verständigungen im Strafverfahren und Ermittlungen. Da sich die Gestaltung und Nutzung der Algorithmen auf die Ergebnisse auswirken könnte, müssen die Konsequenzen des Einsatzes der künstlichen Intelligenz allerdings sorgfältig geprüft werden.

KI im Sicherheitsbereich

Künstliche Intelligenz verspricht, zur Bewältigung komplexer digitaler und physischer Herausforderungen im Sicherheitsbereich beizutragen. Für 2018 wurde ein Anstieg der weltweiten Verteidigungsausgaben auf 1,67 Bill. USD erwartet, was im Vorjahresvergleich einer Zunahme um 3,3 % entspricht (IHS, 2017[89]). Die Sicherheit ist jedoch nicht nur eine Domäne des öffentlichen Sektors. Im privaten Sektor wurden 2018 Schätzungen zufolge Ausgaben in Höhe von 96 Mrd. USD getätigt, um Sicherheitsrisiken zu begegnen. Das entspricht einem Anstieg um 8 % gegenüber 2017 (Gartner, 2017[90]). Einige große Cyberangriffe haben in jüngster Zeit das Bewusstsein der Öffentlichkeit für digitale Sicherheit geschärft. Sie haben gezeigt, dass Datenschutzverletzungen weitreichende Folgen für Wirtschaft, Gesellschaft und nationale Sicherheit haben können. Vor diesem Hintergrund nutzen öffentliche und private Akteure zunehmend KI-Technologien, um sich an die sich weltweit verändernde Sicherheitslandschaft anzupassen. In diesem Abschnitt werden zwei sicherheitsrelevante Bereiche beschrieben, die ein besonders rasches Wachstum verzeichnen: digitale Sicherheit und Überwachung.[12,13]

KI und digitale Sicherheit

KI wird bereits umfassend in digitalen Sicherheitsanwendungen eingesetzt, z. B. in den Bereichen Netzwerksicherheit, Anomalieerkennung, Automatisierung von Sicherheitsoperationen und Gefahrenerkennung (OECD, 2017[26]). Gleichzeitig dürfte der Missbrauch von KI jedoch zunehmen, beispielsweise wenn Software-Schwachstellen aufgespürt und ausgenutzt werden, um die Verfügbarkeit, Integrität oder Vertraulichkeit von Systemen, Netzwerken und Daten zu verletzen. Dies wird sich auf Art und Gesamtumfang des digitalen Sicherheitsrisikos auswirken.

Zwei Trends bewirken, dass die Sicherheitsrelevanz von KI-Systemen steigt: die wachsende Zahl von Cyberangriffen und der Fachkräftemangel im Bereich digitale Sicherheit (ISACA, 2016[91]). Infolge dieser Trends werden ML-Anwendungen und KI-Systeme immer wichtiger, um die Erkennung und Abwehr von Bedrohungen zu automatisieren (MIT, 2018[92]). Da sich Malware ständig weiterentwickelt, ist maschinelles Lernen (ML) unerlässlich geworden, um Angriffe wie polymorphe Viren, Denial-of-Service-Attacken und Phishing zu bekämpfen.[14] Tatsächlich wird maschinelles Lernen von führenden E-Mail-Anbietern wie Gmail und Outlook seit mehr als einem Jahrzehnt mit unterschiedlichem Erfolg eingesetzt, um unerwünschte oder schädliche E-Mail-Nachrichten zu filtern. Kasten 3.1 verdeutlicht, wie KI zum Schutz von Unternehmen vor böswilligen Bedrohungen eingesetzt werden kann.

Kasten 3.1. KI zur Steuerung digitaler Sicherheitsrisiken in Geschäftsumgebungen

Unternehmen wie Darktrace, Vectra und viele andere setzen maschinelles Lernen und künstliche Intelligenz ein, um Cyberangriffe in Echtzeit zu erkennen und abzuwehren. Darktrace setzt auf die Technologie Enterprise Immune System, die potenzielle Gefahren auch erkennen kann, ohne zuvor schon einmal eine ähnliche Bedrohung erlebt zu haben. KI-Algorithmen machen sich durch iteratives Lernen mit dem einzigartigen „Lebensmuster“ eines Netzwerks vertraut, um Bedrohungen zu erkennen, die andernfalls unbemerkt bleiben würden. Die Funktionsweise von Darktrace ist dem menschlichen Immunsystem nachempfunden, das lernt, was für den Körper normal ist, und das dieser Normalität nicht entsprechende Situationen automatisch erkennen und beseitigen kann.

Vectra macht mit seiner rund um die Uhr laufenden, automatisierten und ständig lernenden „Cognito-Plattform“ Jagd auf Angreifer in Cloud-Umgebungen. Diese Plattform macht das Verhalten von Angreifern vollkommen transparent, auf Cloud- und Rechenzentren-Workloads ebenso wie Benutzer- und IoT-Geräten. Auf diese Weise wird es für Angreifer immer schwieriger, unsichtbar zu agieren.

Quelle: www.darktrace.com/; https://vectra.ai/.

Computercode ist anfällig für menschliche Fehler. Schätzungen zufolge sind neun von zehn Cyberangriffen auf Fehler im Softwarecode zurückzuführen. Solche Fehler entstehen trotz der enorm hohen Entwicklungszeit, die für Tests aufgewendet wird – zwischen 50 % und 75 % (Oliver, 2018[93]). Angesichts der Milliarden von Zeilen Code, die jedes Jahr geschrieben werden, und der dafür genutzten proprietären Bibliotheken von Drittanbietern ist das Erkennen und Korrigieren von Fehlern im Softwarecode eine für den Menschen kaum noch zu bewältigende Aufgabe. Länder wie die Vereinigten Staaten und China finanzieren Forschungsprojekte, um KI-Systeme zu entwickeln, die Sicherheitslücken in Software erkennen können. Unternehmen wie der Videospielehersteller Ubisoft haben

begonnen KI einzusetzen, die Fehler im Code kennzeichnen soll, bevor er implementiert wird. Dadurch lässt sich die Testzeit um 20 % reduzieren (Oliver, 2018[93]). In der Praxis funktionieren KI-Technologien zur Prüfung von Software wie die Rechtschreibprüfung in Textverarbeitungsprogrammen, die Tipp- und Syntaxfehler finden soll. KI-Technologien lernen jedoch dazu und werden mit der Zeit immer effizienter (Oliver, 2018[93]).

KI in der Überwachung

In vielen Städten wird die digitale Infrastruktur ausgebaut. Dies gilt insbesondere für den Überwachungsbereich, wo verschiedene KI-gestützte Instrumente und Geräte zur Erhöhung der öffentlichen Sicherheit eingesetzt werden. So können Smart-Kameras beispielsweise Schlägereien erkennen. Schussdetektoren können Schüsse automatisch melden und dabei genau angeben, wo sie gefallen sind. In diesem Abschnitt wird untersucht, wie künstliche Intelligenz die Welt der Sicherheit und Überwachung revolutioniert. Videoüberwachung wird mittlerweile immer häufiger eingesetzt, um die Sicherheit im öffentlichen Raum zu erhöhen. Laut Schätzungen einer im Vereinigten Königreich durchgeführten jüngeren Studie lieferten Sicherheitsaufnahmen verwertbare Beweise für 65 % der Verbrechen, die zwischen 2011 und 2015 im britischen Eisenbahnnetz begangen wurden und für die Bildmaterial zur Verfügung stand (Ashby, 2017[94]). Die hohe Zahl von Überwachungskameras – 2014 waren es weltweit 245 Millionen – hat zur Folge, dass immer mehr Daten erzeugt werden. Die an einem einzigen Tag produzierte Datenmenge erhöhte sich zwischen 2013 und 2017 von 413 Petabyte (PB) auf geschätzte 860 PB (Jenkins, 2015[95]), (Civardi, 2017[96]). Menschen sind nur bedingt in der Lage, so große Datenmengen zu verarbeiten. Die Nutzung von KI-Technologien zur Verarbeitung großer Datenmengen und Automatisierung mechanischer Erkennungs- und Kontrollprozesse bietet sich daher an. Sicherheitssysteme können zudem dank KI Verbrechen in Echtzeit erkennen und darauf reagieren (Kasten 3.2).

Kasten 3.2. Überwachung mit „intelligenten“ Kameras

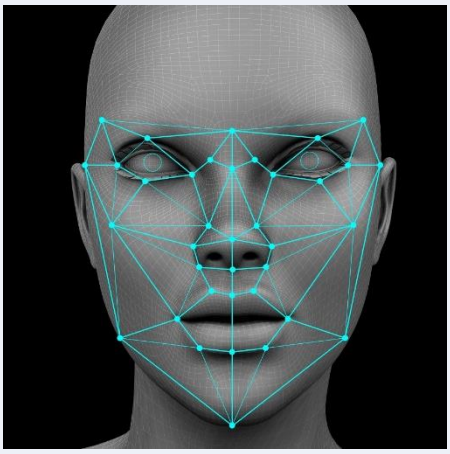
In Zusammenarbeit mit Thales setzt die französische Kernenergiebehörde (Commissariat à l’énergie atomique et aux énergies alternatives – CEA) Deep Learning ein, um Videos für Sicherheitsanwendungen automatisch zu analysieren und auszuwerten. Ein Modul zur Erkennung von Gewaltakten erkennt z. B. automatisch Schlägereien oder tätliche Angriffe, die von Überwachungskameras aufgenommen wurden, und alarmiert die Betreiber in Echtzeit. Ein weiteres Modul hilft, die Täter im Kameranetzwerk zu lokalisieren. Diese Anwendungen werden derzeit von den französischen öffentlichen Verkehrsbetrieben RATP und SNCF in den Stationen Châtelet-Les-Halles und Gare du Nord getestet, zwei der verkehrsreichsten Bahn- und Metro-Stationen in Paris. Die französische Stadt Toulouse setzt ebenfalls Smart-Kameras ein, um verdächtiges Verhalten und herrenlose Gepäckstücke zu melden. Ähnliche Projekte laufen in Berlin, Rotterdam und Shanghai.

Quelle: Der OECD von CEA Tech und Thales 2018 zur Verfügung gestellte Informationen. Weitere Informationen (auf Französisch) unter: www.gouvernement.fr/sites/default/files/contenu/piece-jointe/2015/11/projet_voie_videoprotection_ouverte_et_integree_appel_a_projets.pdf.

Kasten 3.3. Gesichtserkennung als Instrument der Überwachung

Gesichtserkennungstechnologien werden zunehmend von privaten oder öffentlichen Akteuren zur wirksamen Überwachung eingesetzt (Abbildung 3.4). KI verbessert traditionelle Gesichtserkennungssysteme, da sie eine schnellere und genauere Identifizierung ermöglicht, wo herkömmliche Systeme versagen, z. B. bei schlechten Lichtverhältnissen oder wenn die Zielperson durch etwas verdeckt ist. Unternehmen wie FaceFirst kombinieren Gesichtserkennungstools mit KI, um Lösungen zur Diebstahls-, Betrugs- und Gewaltprävention anzubieten. In ihrem Design wurden besondere Überlegungen berücksichtigt, damit sie höchsten Datenschutz- und Sicherheitsstandards gerecht werden, z. B. durch Anti-Profiling-Funktionen zum Schutz vor Diskriminierung, Verschlüsselung von Bilddaten und strikte Zeitvorgaben für die Löschung der Daten. Diese Überwachungsinstrumente werden in verschiedensten Bereichen eingesetzt, vom Einzelhandel (z. B. zur Bekämpfung von Ladendiebstahl) über den Bankensektor (beispielsweise zur Verhinderung von Identitätsbetrug) und den Gesetzesvollzug (z. B. zur Grenzsicherung) bis hin zum Eventmanagement (etwa zur Erkennung von gesperrten Fans) und Casinos (z. B. zur Erkennung wichtiger Personen).

Abbildung 3.4. Veranschaulichung einer Gesichtserkennungssoftware



Quelle: www.facefirst.com.

Aufgrund der Dualität, die KI eigen ist, könnten KI-gestützte Überwachungsinstrumente entgegen den in Kapitel 4 beschriebenen Grundsätzen zu illegalen Zwecken genutzt werden. Legitim ist der Einsatz von KI u. a. im Gesetzesvollzug zur Vereinfachung strafrechtlicher Ermittlungen, zur frühestmöglichen Aufdeckung und Bekämpfung von Straftaten und zur Terrorismusbekämpfung. Gesichtserkennungstechnologien haben sich hier als zielführend erwiesen (Kasten 3.3). Die Auswirkungen von KI gehen in der Überwachung allerdings über die Gesichtserkennung hinaus. Auch bei der Verbesserung von Technologien der gesichtslosen Erkennung spielt KI eine zunehmend wichtige Rolle. Dies sind Technologien, die zur Identifikation alternative Informationen wie Größe, Kleidung, Körperbau, Körperhaltung usw. verwenden. Darüber hinaus hat sich KI in Kombination mit Bildschärfungstechnologien bewährt: Neuronale Netze werden mit großen Bilddatensätzen trainiert, damit sie typische Eigenschaften physischer Elemente wie Haut, Haare oder sogar Mauersteine erkennen. Sie können diese Merkmale dann in

neuen Bildern wiedererkennen und das zuvor erworbene Wissen nutzen, um ihnen zusätzliche Details und Texturen hinzuzufügen. So lassen sich Unzulänglichkeiten von Bildern mit schlechter Auflösung beheben und kann die Wirksamkeit von Überwachungssystemen verbessert werden (Medhi, S. M., B. Scholkopf und M. Hirsch, 2017[97]).

KI im öffentlichen Sektor

Künstliche Intelligenz eröffnet in der öffentlichen Verwaltung vielfältige Möglichkeiten. Die Entwicklung von KI-Technologien hat bereits Auswirkungen auf die Arbeit des öffentlichen Sektors und die Ausgestaltung von Politikmaßnahmen für Bürger und Unternehmen. Dies gilt besonders für Bereiche wie Gesundheitsversorgung, Verkehr sowie Sicherheit.[15]

Verschiedene staatliche Stellen im OECD-Raum experimentieren mit KI, um dem Bedarf der Nutzer öffentlicher Dienstleistungen besser gerecht zu werden. Außerdem möchten sie eine bessere Ressourcennutzung gewährleisten (z. B. um den Zeitaufwand öffentlich Bediensteter für Kundenbetreuung und Verwaltungsaufgaben zu verringern). KI-Instrumente könnten die Effizienz und Qualität vieler Verfahren des öffentlichen Sektors verbessern. Zum Beispiel könnten sie den Bürgern die Möglichkeit bieten, sich bereits im Vorfeld in die Dienstleistungsgestaltung einzubringen und flexibler, effektiver und persönlicher mit dem Staat zu interagieren. Bei richtiger Planung und Implementierung könnten KI-Technologien in den gesamten politischen Entscheidungsprozess eingebunden werden, Reformen im öffentlichen Sektor unterstützen und die Produktivität des öffentlichen Sektors steigern.

Einige Staaten haben bereits KI-Systeme eingesetzt, um Sozialprogramme zu stärken. KI könnte beispielsweise dazu beitragen, die Lagerhaltung in Gesundheits- und Sozialeinrichtungen zu optimieren. Dies würde über Technologien des maschinellen Lernens erfolgen, die Transaktionsdaten analysieren und zunehmend genaue Vorhersagen für die Bedarfs- und Nachschubplanung treffen. Dies wiederum würde die Erstellung von Prognosen und die Politikgestaltung erleichtern. Im Vereinigten Königreich setzen staatliche Stellen KI-Algorithmen auch ein, um Sozialleistungsbetrug aufzudecken (Marr, 2018[98]).

KI-Anwendungen in Verbindung mit erweiterter und virtueller Realität

Unternehmen nutzen KI-Technologie und visuelle Erkennungsfunktionen wie Bildklassifizierung und Objekterkennung, um Hardware und Software für erweiterte und virtuelle Realität (AR und VR) zu entwickeln. Diese neuen Lösungen können u. a. für immersive Erlebnisse, Aus- und Weiterbildung, zur Unterstützung von Menschen mit Behinderungen und für Unterhaltungsangebote genutzt werden. VR und AR haben sich seit dem ersten VR-Headset-Prototyp, den Ivan Sutherland 1968 zur Darstellung von 3D-Bildern entwickelte, in bemerkenswerter Weise weiterentwickelt. Das Headset von Sutherland war noch zu schwer zum Tragen und musste deshalb an der Decke montiert werden (Günger, C. und K. Zengin, 2017[99]). Heute bieten VR-Unternehmen 360-Grad-Video-Streaming-Erlebnisse mit viel leichteren Headsets an. Durch Pokemon Go entdeckte 2016 auch die breite Öffentlichkeit die AR, und die Erwartungen sind nach wie vor hoch. Anwendungen mit integrierter künstlicher Intelligenz sind bereits im Handel erhältlich. IKEA bietet eine mobile App, mit der Kunden mit einer Genauigkeit von bis zu 1 Millimeter sehen können, wie ein Möbelstück aussehen und in einen bestimmten Raum passen würde (Jesus, 2018[100]). Einige Technologieunternehmen entwickeln auch Anwendungen für Menschen mit Sehbehinderung.[16]

KI für interaktive AR/VR-Anwendungen

Die Weiterentwicklung von AR/VR geht mit dem Einsatz von KI einher, die mehr Interaktivität sowie attraktivere und intuitivere Inhalte ermöglicht. Dank KI-Technologien können AR/VR-Anwendungen Bewegungen der Nutzer, z. B. Augen- und Handbewegungen, erkennen und mit hoher Genauigkeit interpretieren. Dies ermöglicht es, Inhalte entsprechend der Reaktion des Nutzers in Echtzeit anzupassen (Lindell, 2017[101]). Durch die Kombination von KI und VR kann man beispielsweise erkennen, wann ein Nutzer einen bestimmten Bereich betrachtet, und nur dann Inhalte in hoher Auflösung anzeigen. So lässt sich der Bedarf an Systemressourcen reduzieren, Verzögerungen können verkürzt und Bildverluste vermieden werden (Hall, 2017[102]). In Bereichen wie Marktforschung, Trainingssimulationen und Bildung wird eine symbiotische Entwicklung von AR/VR- und KI-Technologien erwartet (Kilpatrick, 2018[103]); (Stanford, 2016[104]).

VR zum Trainieren von KI-Systemen

Einige KI-Systeme benötigen große Mengen an Trainingsdaten. Der Mangel an Daten ist daher noch ein Problem. So müssen beispielsweise KI-Systeme in selbstfahrenden Fahrzeugen für den Umgang mit kritischen Situationen trainiert werden. Allerdings gibt es z. B. kaum reale Daten über Kinder, die auf die Straße laufen. Eine Alternative wäre die Entwicklung einer digitalen Realität. In diesem Fall würde das KI-System in einer von Computern simulierten Umgebung trainiert, die relevante Merkmale der realen Welt wirklichkeitsgetreu nachbildet. Eine solche simulierte Umgebung kann auch verwendet werden, um die Leistung von KI-Systemen zu prüfen (z. B. als „Führerscheinprüfung“ für selbstfahrende Fahrzeuge) (Slusallek, 2018[105]).

Selbstfahrende Fahrzeuge sind jedoch nicht der einzige Anwendungsbereich. Mit dem Household Multimodal Environment (HoME) wurde eine Plattform entwickelt, die eine simulierte Umgebung zum Trainieren von Haushaltsrobotern bietet. HoME verfügt über eine Datenbank mit über 45 000 verschiedenen Grundrissen von Häusern in 3D. Dank dieser Datenbank kann sie eine realistische Umgebung schaffen, in der künstliche Agenten durch Sehen, Hören, Semantik, Physik und Interaktion mit Objekten und anderen Agenten lernen können (Brodeur et al., 2017[106]).

Cloud-basierte VR-Simulationen, in denen KI-Systeme durch Ausprobieren (d. h. durch Versuch und Irrtum) lernen können, wären für das Training ideal, insbesondere für den Umgang mit kritischen Situationen. Durch die kontinuierliche Weiterentwicklung der Cloud-Technologie dürfte es möglich sein, solche Umgebungen zu verwirklichen. So kündigte NVIDIA im Oktober 2017 einen cloud-basierten VR-Simulator an, der physikalische Gesetzmäßigkeiten in realen Umgebungen präzise nachbilden kann. Für die nächsten Jahre wird mit der Entwicklung eines neuen Trainingsgebiets für KI-Systeme gerechnet (Solotko, 2017[107]).

Literaturverzeichnis

Aletras, N. et al. (2016), "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective", *PeerJ Computer Science*, Vol. 2, S. e93, http://dx.doi.org/10.7717/peerj-cs.93. [86]

Angwin, J. et al. (2016), "Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks", *ProPublica*, 23. Mai, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. [83]

Ashby, M. (2017), "The value of CCTV surveillance cameras as an investigative tool: An empirical analysis", *European Journal on Criminal Policy and Research*, Vol. 23/3, S. 441-459, http://dx.doi.org/10.1007/s10610-017-9341-6. [94]

Barocas, S. und A. Selbst (2016), "Big data's disparate impact", *California Law Review*, Vol. 104, S. 671-729, http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf. [74]

Bavitz, C. und K. Hessekiel (2018), "Algorithms and Justice: Examining the Role of the State in the Development and Deployment of Algorithmic Technologies", Berkman Klein Center for Internet and Society, https://cyber.harvard.edu/story/2018-07/algorithms-and-justice. [82]

Berg, T. et al. (2018), "On the rise of FinTechs – Credit scoring using digital footprints", *Michael J. Brennan Irish Finance Working Paper Series Research Paper*, No. 18-12, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3163781. [32]

Bloom, N. et al. (2017), "Are ideas getting harder to find?", *NBER Working Paper*, No. 23782, http://dx.doi.org/10.3386/w23782. [52]

Bösch, P. et al. (2018), "Cost-based analysis of autonomous mobility services", *Transport Policy*, Vol. 64, S. 76-91, https://doi.org/10.1016/j.tranpol.2017.09.005. [3]

Bose, A. et al. (2016), "The VEICL Act: Safety and security for modern vehicles", *Willamette Law Review*, Vol. 53, S. 137. [15]

Brayne, S., A. Rosenblat und D. Boyd (2015), "Predictive policing, data & civil rights: A new era of policing and justice", *Pennsylvania Law Review*, Vol. 163/327, http://www.datacivilrights.org/pubs/2015-1027/Predictive_Policing.pdf. [76]

Brodeur, S. et al. (2017), "HoME: A household multimodal environment", arXiv:1711.11017, https://arxiv.org/abs/1711.11017. [106]

Brodmerkel, S. (2017), "Dynamic pricing: Retailers using artificial intelligence to predict top price you'll pay", *ABC News,* 27. Juni, http://www.abc.net.au/news/2017-06-27/dynamic-pricing-retailers-using-artificial-intelligence/8638340. [50]

Brundage, M. et al. (2018), *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Centre for a New American Security, Electronic Frontier Foundation and Open AI, https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf. [108]

Burgess, M. (2018), "UK police are using AI to make custodial decisions but it could be discriminating against the poor", *WIRED,* 1. März, http://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit. [79]

Butler, K. et al. (2018), "Machine learning for molecular and materials science", *Nature*, Vol. 559/7715, S. 547-555, http://dx.doi.org/10.1038/s41586-018-0337-2. [60]

Canadian Institute for Health Information (2013), "Better information for improved health: A vision for health system use of data in Canada", in Zusammenarbeit mit Canada Health Infoway, http://www.cihi.ca/cihi-ext-portal/pdf/internet/hsu_vision_report_en. [63]

Carey, N. und P. Lienert (2018), "Honda to invest $2.75 billion in GM's self-driving car unit" Reuters, 3. Oktober, https://www.reuters.com/article/us-gm-autonomous/honda-buys-in-to-gm-cruise-self-driving-unit-idUSKCN1MD1GW. [6]

CFPB (2017), "CFPB explores impact of alternative data on credit access for consumers who are credit invisible", Consumer Financial Protection Bureau, https://www.consumerfinance.gov/about-us/newsroom/cfpb-explores-impact-alternative-data-credit-access-consumers-who-are-credit-invisible/. [37]

Cheng, E. (2017), "Just 10% of trading is regular stock picking, JPMorgan estimates", *CNBC*, 13. Juni, https://www.cnbc.com/2017/06/13/death-of-the-human-investor-just-10-percent-of-trading-is-regular-stock-picking-jpmorgan-estimates.html. [44]

Chintamaneni, P. (2017), "How banks can use AI to reduce regulatory compliance burdens", digitally.cognizant blog, 26. Juni, https://digitally.cognizant.com/how-banks-can-use-ai-to-reduce-regulatory-compliance-burdens-codex2710/. [41]

Chow, M. (2017), "AI and machine learning get us one step closer to relevance at scale", Google, https://www.thinkwithgoogle.com/marketing-resources/ai-personalized-marketing/. [45]

Christin, A., A. Rosenblat und D. Boyd (2015), *Courts and Predictive Algorithms*, Beitrag zur Konferenz "Data & Civil Rights, A New Era of Policing and Justice" conference, Washington, 27. Oktober, http://www.law.nyu.edu/sites/default/files/upload_documents/Angele%20Christin.pdf. [81]

Civardi, C. (2017), *Video Surveillance and Artificial Intelligence: Can A.I. Fill the Growing Gap Between Video Surveillance Usage and Human Resources Availability?*, Balzano Informatik, http://dx.doi.org/10.13140/RG.2.2.13330.66248. [96]

CMU (2015), "Uber, Carnegie Mellon announce strategic partnership and creation of advanced technologies center in Pittsburgh", *Carnegie Mellon University News,* 2. Februar, https://www.cmu.edu/news/stories/archives/2015/february/uber-partnership.html. [7]

Comfort, S. et al. (2018), "Sorting through the safety data haystack: Using machine learning to identify individual case safety reports in social-digital media", *Drug Safety*, Vol. 41/6, S. 579-590, https://www.ncbi.nlm.nih.gov/pubmed/29446035. [64]

Cooke, A. (2017), *Digital Earth Australia*, Präsentation bei der OECD-Konferenz "AI: Intelligent Machines, Smart Policies" conference, Paris, 26.-27. Oktober, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-cooke.pdf. [22]

De Jesus, A. (2018), "Augmented reality shopping and artificial intelligence – Near-term applications", *Emerj,* 18. Dezember, https://www.techemergence.com/augmented-reality-shopping-and-artificial-intelligence/. [51]

Fagnant, D. und K. Kockelman (2015), "Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations", *Transportation Research A: Policy and Practice*, Vol. 77, S. 167-181, https://doi.org/10.1016/j.tra.2015.04.003. [2]

FAO (2017), "Can artificial intelligence help improve agricultural productivity?", *e-agriculture* 19. Dezember, Ernährungs- und Landwirtschaftsorganisation der Vereinten Nationen, Rom, http://www.fao.org/e-agriculture/news/can-artificial-intelligence-help-improve-agricultural-productivity. [20]

FAO (2009), *How to Feed the World in 2050*, Ernährungs- und Landwirtschaftsorganisation der Vereinten Nationen, Rom, http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf. [23]

FDA (2018), "FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems", FDA News Release, 11. April, Food and Drug Administration, https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm. [66]

Ferguson, A. (2014), "Big Data and Predictive Reasonable Suspicion", *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.2394683. [75]

Flores, A., K. Bechtel und C. Lowenkamp (2016), "False positives, false negatives, and false analyses: A rejoinder to 'Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks'", *Federal Probation Journal*, Vol. 80/2, https://www.uscourts.gov/sites/default/files/80_2_6_0.pdf. [84]

Fridman, L. (2018), "Tesla autopilot miles", MIT Human-Centered AI blog, 8. Oktober, https://hcai.mit.edu/tesla-autopilot-miles/. [17]

Fridman, L. et al. (2018), "MIT autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation", arXiv:1711.06976, https://arxiv.org/pdf/1711.06976.pdf. [18]

FSB (2017), *Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications*, Finanzstabilitätsrat, Basel, https://www.fsb.org/wp-content/uploads/P011117.pdf. [31]

Gartner (2017), "Gartner's worldwide security spending forecast", Pressemitteilung, 7. Dezember, Gartner, https://www.gartner.com/en/newsroom/press-releases/2017-12-07-gartner-forecasts-worldwide-security-spending-will-reach-96-billion-in-2018. [90]

Gordon, M. und V. Stewart (2017), "CFPB insights on alternative data use on credit scoring" Law 360, 3. Mai, https://www.law360.com/articles/919094/cfpb-insights-on-alternative-data-use-in-credit-scoring. [36]

Günger, C. und K. Zengin (2017), "A Survey on Augmented Reality Applications using Deep Learning", *GE-International Journal of Engineering Research*, Associated Asia Research Foundation, https://www.researchgate.net/publication/322332639_A_Survey_On_Augmented_Reality_Applications_Using_Deep_Learning. [99]

Hall, N. (2017), "8 ways AI makes virtual & augmented reality even more real", Topbots, 13. Mai, https://www.topbots.com/8-ways-ai-enables-realistic-virtual-augmented-reality-vr-ar/. [102]

Higgins, T. und C. Dawson (2018), "Waymo orders up to 20,000 Jaguar SUVs for driverless fleet", *The Wall Street Journal*, 27. März, https://www.wsj.com/articles/waymo-orders-up-to-20-000-jaguar-suvs-for-driverless-fleet-1522159944. [5]

Hinds, R. (2018), "How Natural Language Processing is shaping the Future of Communication", MarTechSeries – Marketing Technology Insights, 5. Februar, https://martechseries.com/mts-insights/guest-authors/how-natural-language-processing-is-shaping-the-future-of-communication/. [46]

Hong, P. (2017), "Using machine learning to boost click-through rate for your ads", LinkedIn Blog, 27. August, https://www.linkedin.com/pulse/using-machine-learning-boost-click-through-rate-your-ads-tay/. [48]

Horgan, J. (2008), "Against prediction: Profiling, policing, and punishing in an actuarial age – by Bernard E. Harcourt", *Review of Policy Research*, Vol. 25/3, S. 281-282, http://dx.doi.org/10.1111/j.1541-1338.2008.00328.x. [73]

Hutson, M. (2017), "Artificial intelligence prevails at predicting Supreme Court decisions", *Science Magazine* 2. Mai, http://www.sciencemag.org/news/2017/05/artificial-intelligence-prevails-predicting-supreme-court-decisions. [87]

Hu, X. (ed.) (2017), "Human-in-the-loop Bayesian optimization of wearable device parameters", *PLOS ONE*, Vol. 12/9, http://dx.doi.org/10.1371/journal.pone.0184054. [61]

IHS (2017), "Global defence spending to hit post-Cold War high in 2018", IHS Markit, 18. Dezember, https://ihsmarkit.com/research-analysis/global-defence-spending-to-hit-post-cold-war-high-in-2018.html. [89]

Inners, M. und A. Kun (2017), "Beyond Liability: Legal Issues of Human-Machine Interaction for Automated Vehicles", *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* 24.-27. September, Oldenburg, S. 245-253, http://dx.doi.org/10.1145/3122986.3123005. [14]

ISACA (2016), *The State of Cybersecurity: Implications for 2016*, An ISACA and RSA Conference Survey, Cybersecurity Nexus, https://www.isaca.org/cyber/Documents/state-of-cybersecurity_res_eng_0316.pdf. [91]

ITF (2018), "Safer Roads with Automated Vehicles?", *International Transport Forum Policy Papers* No. 55, OECD Publishing, Paris, https://doi.org/10.1787/b2881ccb-en. [12]

Jagtiani, J. und C. Lemieux (2019), "The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform", Working Paper, No. 18-15, Federal Reserve Bank of Philadelphia, http://dx.doi.org/10.21799/frbp.wp.2018.15. [30]

Jenkins, N. (2015), "245 million video surveillance cameras installed globally in 2014", IHS Markit, Market Insight, 11. Juni, https://technology.ihs.com/532501/245-million-video-surveillance-cameras-installed-globally-in-2014. [95]

Jesus, A. (2018), "Augmented Reality Shopping and Artificial Intelligence – Near-Term Applications", Emerj, 12. Dezember, https://www.techemergence.com/augmented-reality-shopping-and-artificial-intelligence/. [100]

Joh, E. (2017), "The Undue Influence of Surveillance Technology Companies on Policing", *New York University Law Review*, Vol. 91/101, http://dx.doi.org/10.2139/ssrn.2924620. [77]

Jouanjean, M. (2019), "Digital opportunities for trade in the agriculture and food sectors", *OECD Food, Agriculture and Fisheries Papers*, No. 122, OECD Publishing, Paris, https://doi.org/10.1787/91c40e07-en. [24]

Kehl, D., P. Guo und S. Kessler (2017), *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessment in Sentencing*, Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041. [80]

Kilpatrick, S. (2018), "The rising force of deep learning in VR and AR" Logikk, 28. März, https://www.logikk.com/articles/deep-learning-in-vr-and-ar/. [103]

King, R. et al. (2004), "Functional genomic hypothesis generation and experimentation by a robot scientist", *Nature*, Vol. 427/6971, S. 247-252, http://dx.doi.org/10.1038/nature02236. [56]

Kleinberg, J. et al. (2017), "Human decisions and machine predictions", *NBER Working Paper*, No. 23180, https://www.nber.org/papers/w23180.pdf. [85]

Lee, C. (2017), "Deep learning is effective for classifying normal versus age-related macular degeneration OCT images", *Opthamology Retina*, Vol. 1/4, S. 322-327, https://doi.org/10.1016/j.oret.2016.12.009. [67]

Lee, T. (2018), "Fully driverless Waymo taxis are due out this year, alarming critics", *Ars Technica*, 1. Oktober, https://arstechnica.com/cars/2018/10/waymo-wont-have-to-prove-its-driverless-taxis-are-safe-before-2018-launch/. [10]

Lindell, T. (2017), "Augmented reality needs AI in order to be effective", AI Business, 6. November, https://aibusiness.com/holographic-interfaces-augmented-reality/. [101]

Lippert, J. et al. (2018), "Toyota's vision of autonomous cars is not exactly driverless", *Bloomberg Business Week,* 19. September, https://www.bloomberg.com/news/features/2018-09-19/toyota-s-vision-of-autonomous-cars-is-not-exactly-driverless. [8]

Marr, B. (2018), "How AI and machine learning are transforming law firms and the legal sector", *Forbes*, 23. Mai, https://www.forbes.com/sites/bernardmarr/2018/05/23/how-ai-and-machine-learning-are-transforming-law-firms-and-the-legal-sector/#7587475832c3. [88]

Marr, B. (2018), "How the UK government uses artificial intelligence to identify welfare and state benefits fraud", *Forbes*, 29. Oktober, https://www.forbes.com/sites/bernardmarr/2018/10/29/how-the-uk-government-uses-artificial-intelligence-to-identify-welfare-and-state-benefits-fraud/#f5283c940cb9. [98]

Mar, V. (2018), "Artificial intelligence for melanoma diagnosis: How can we deliver on the promise?", *Annals of Oncology*, Vol. 29/8, S. 1625-1628, http://dx.doi.org/10.1093/annonc/mdy193. [68]

Matchar, E. (2017), "AI plant and animal identification helps us all be citizen scientists", *Smithsonian.com,* 7. Juni, https://www.smithsonianmag.com/innovation/ai-plant-and-animal-identification-helps-us-all-be-citizen-scientists-180963525/. [55]

Mateescu, A. et al. (2015), "Social Media Surveillance and Law Enforcement" Beitrag zur Konferenz "Data & Civil Rights: A New Era of Criminal Justice and Policing", http://www.datacivilrights.org/pubs/2015-1027/Social_Media_Surveillance_and_Law_Enforcement.pdf. [78]

Medhi, S. M., B. Scholkopf und M. Hirsch (2017), "EnhanceNet: Single image super-resolution through automated texture synthesis", *arXiv* arXiv:1612.07919, https://arxiv.org/abs/1612.07919. [97]

MIT (2018), "Cybersecurity's insidious new threat: workforce stress", *MIT Technology Review,* 7. August, https://www.technologyreview.com/s/611727/cybersecuritys-insidious-new-threat-workforce-stress/. [92]

O'Dwyer, R. (2018), *Algorithms are making the same mistakes assessing credit scores that humans did a century ago*, Quartz, 14. Mai, https://qz.com/1276781/algorithms-are-making-the-same-mistakes-assessing-credit-scores-that-humans-did-a-century-ago/. [34]

OECD (2018), "Artificial intelligence and machine learning in science", in *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, OECD Publishing, Paris, http://dx.doi.org/10.1787/sti_in_outlook-2018-10-en. [57]

OECD (2018), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, OECD Publishing, Paris, https://dx.doi.org/10.1787/sti_in_outlook-2018-en. [53]

OECD (2018), "Personalised Pricing in the Digital Era – Note by the United Kingdom", Key Paper für das gemeinsame Treffen der OECD-Ausschüsse für Verbraucherschutz und Wettbewerb am 28. November, OECD, Paris, http://www.oecd.org/daf/competition/personalised-pricing-in-the-digital-era.htm. [109]

OECD (2018), *Structural Analysis Database (STAN)*, Rev. 4, Abteilungen 49 bis 53, http://www.oecd.org/industry/ind/stanstructuralanalysisdatabase.htm (Abruf: 31. Januar 2018). [1]

OECD (2017), *New Health Technologies: Managing Access, Value and Sustainability*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264266438-en. [69]

OECD (2017), *OECD Digital Economy Outlook 2017*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264276284-en. [26]

OECD (2017), *Technology and Innovation in the Insurance Sector*, OECD, Paris, https://www.oecd.org/finance/Technology-and-innovation-in-the-insurance-sector.pdf. [39]

OECD (2016), "Recommendation of the Council on Health Data Governance" OECD, Paris, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0433. [71]

OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264229358-en. [62]

OECD (2014), "Skills and Jobs in the Internet Economy", *OECD Digital Economy Papers*, No. 242, OECD Publishing, Paris, https://dx.doi.org/10.1787/5jxvbrjm9bns-en. [19]

Ohnsman, A. (2018), "Waymo Dramatically Expanding Autonomous Taxi fleet, Eyes Sales To Individuals", *Forbes,* 31. Mai, https://www.forbes.com/sites/alanohnsman/2018/05/31/waymo-adding-up-to-62000-minivans-to-robot-fleet-may-supply-tech-for-fca-models. [4]

Oliver, J. (2018), "US and China back AI bug-detecting projects", *Financial Times, Cyber Security and Artificial Intelligence,* 26. September, https://www.ft.com/content/64fef986-89d0-11e8-affd-da9960227309. [93]

ORAD (2016), "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles", On-Road Automated Driving (ORAD) Committee, SAE International, http://dx.doi.org/10.4271/j3016_201609. [9]

Ornstein, C. und K. Thomas (2018), "Sloan Kettering's Cozy Deal With Start-Up Ignites a New Uproar", *New York Times*, 20. September, https://www.nytimes.com/2018/09/20/health/memorial-sloan-kettering-cancer-paige-ai.html. [70]

Patton, E. (2018), *Integrating Artificial Intelligence for Scaling Internet of Things in Health Care*, OECD-GCOA-Cornell-Tech Expert Consultation on Growing and Shaping the Internet of Things Wellness and Care Ecosystem, 4.-5. Oktober, New York. [65]

Plummer, L. (2017), "This is how Netflix's top-secret recommendation system works", *WIRED,* 22. August, https://www.wired.co.uk/article/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like. [47]

Press, G. (2017), "Equifax And SAS Leverage AI And Deep Learning To Improve Consumer Access To Credit", *Forbes,* 20. Februar, https://www.forbes.com/sites/gilpress/2017/02/20/equifax-and-sas-leverage-ai-and-deep-learning-to-improve-consumer-access-to-credit/2/#2ea15ddd7f69. [29]

Rakestraw, R. (2017), "Can Artificial Intelligence Help Feed The World?", *Forbes,* 6. September, https://www.forbes.com/sites/themixingbowl/2017/09/05/can-artificial-intelligence-help-feed-the-world/#16bb973646db. [25]

Roeland, C. (2017), "EC Perspectives on the Earth Observation" Präsentation bei der OECD-Konferenz "AI: Intelligent Machines, Smart Policies", Paris, 26.-27. Oktober, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-roeland.pdf. [21]

Rollet, C. (2018), "The odd reality of life under China's all-seeing credit score system", *WIRED,* 5. Juni, https://www.wired.co.uk/article/china-blacklist. [35]

Segler, M., M. Preuss und M. Waller (2018), "Planning chemical syntheses with deep neural networks and symbolic AI", *Nature*, Vol. 555, S. 604-610, http://dx.doi.org/10.1038/nature25978. [59]

Simon, M. (2017), "Phone-powered AI spots sick plants with remarkable accuracy", *WIRED,* 2. Februar, https://www.wired.com/story/plant-ai/. [28]

Slusallek, P. (2018), *Artificial Intelligence and Digital Reality: Do We Need a CERN for AI?*, The Forum Network, OECD, Paris, https://www.oecd-forum.org/channels/722-digitalisation/posts/28452-artificial-intelligence-and-digital-reality-do-we-need-a-cern-for-ai. [105]

Smith, B. (2013), "Automated Vehicles Are Probably Legal in the United States", *Texas A&M Law Review*, Vol. 1/3, S. 411-521. [11]

Sohangir, S. et al. (2018), "Big data: Deep learning for financial sentiment analysis", *Journal of Big Data*, Vol. 5/3, http://dx.doi.org/10.1186/s40537-017-0111-6. [40]

Sokolin, L. und M. Low (2018), *Machine Intelligence and Augmented Finance: How Artificial Intelligence Creates $1 Trillion Dollar of Change in the Front, Middle and Back Office*, Autonomous Research LLP, https://next.autonomous.com/augmented-finance-machine-intelligence. [38]

Solotko, S. (2017), "Virtual Reality is the Next Training Ground for Artificial Intelligence", *Forbes,* 11. Oktober, https://www.forbes.com/sites/tiriasresearch/2017/10/11/virtual-reality-is-the-next-training-ground-for-artificial-intelligence/#6e0c59cc57a5. [107]

Song, H. (2017), "JPMorgan Software Does in Seconds What Took Lawyers 360,000 Hours" Bloomberg, 28. Februar, https://www.bloomberg.com/news/articles/2017-02-28/jpmorgan-marshals-an-army-of-developers-to-automate-high-finance. [42]

Spangler, S. et al. (2014), *Automated Hypothesis Generation based on Mining Scientific Literature*, ACM Press, New York, http://dx.doi.org/10.1145/2623330.2623667. [54]

Stanford (2016), *Artificial Intelligence and Life in 2030*, AI100 Standing Committee and Study Panel, Stanford University, https://ai100.stanford.edu/2016-report. [104]

Surakitbanharn, C. et al. (2018), *Preliminary Ethical, Legal and Social Implications of Connected and Autonomous Transportation Vehicles (CATV)*, Purdue University, https://www.purdue.edu/discoverypark/ppri/docs/Literature%20Review_CATV.pdf. [16]

Voegeli, V. (2016), "Credit Suisse, CIA-Funded Palantir to Target Rogue Bankers" Bloomberg, 22. März, https://www.bloomberg.com/news/articles/2016-03-22/credit-suisse-cia-funded-palantir-build-joint-compliance-firm. [43]

Waid, B. (2018), "AI-Enabled Personalization: The New Frontier In Dynamic Pricing", *Forbes,* 9. Juli, https://www.forbes.com/sites/forbestechcouncil/2018/07/09/ai-enabled-personalization-the-new-frontier-in-dynamic-pricing/#71e470b86c1b. [49]

Webb, L. (2017), "Machine learning in action" Präsentation bei der OECD-Konferenz "AI: Intelligent Machines, Smart Policies", Paris, 26.-27. Oktober, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-webb.pdf. [27]

Welsch, D. und E. Behrmann (2018), "Who's Winning the Self-Driving Car Race?", *Bloomberg,* 7. Mai, https://www.bloomberg.com/news/features/2018-05-07/who-s-winning-the-self-driving-car-race. [13]

Williams, K. et al. (2015), "Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases", *Journal of The Royal Society Interface*, Vol. 12/104, http://dx.doi.org/10.1098/rsif.2014.1289. [58]

Wyllie, D. (2013), "How 'Big Data' is helping law enforcement", PoliceOne.Com, 20. August, https://www.policeone.com/police-products/software/Data-Information-Sharing-Software/articles/6396543-How-Big-Data-is-helping-law-enforcement/. [72]

Zeng, M. (2018), "Alibaba and the future of business", *Harvard Business Review,* September-Oktober, https://hbr.org/2018/09/alibaba-and-the-future-of-business. [33]

Anmerkungen

[1] STAN Industrial Analysis, 2018, Wertschöpfung der Wirtschaftszweige „Verkehr und Lagerei", ISIC Rev. 4, Abteilungen 49 bis 53, als Anteil an der Gesamtwertschöpfung, ungewichteter OECD-Durchschnitt 2016. Der gewichtete OECD-Durchschnitt betrug 2016 4,3 %.

[2] Aus https://www.crunchbase.com/.

[3] Der FICO-Kreditscore wurde 1989 von Fair, Isaac and Company (FICO) eingeführt. Er wird nach wie vor von der Mehrheit der Banken und Kreditgeber verwendet.

[4] Der OECD-Ausschuss für Verbraucherschutz hat die Definition der personalisierten Preisgestaltung des Office of Fair Trading des Vereinigten Königreichs übernommen: Personalisierte Preisgestaltung kann als eine Form der Preisdiskriminierung definiert werden, bei der Unternehmen durch Beobachtungen gewonnene, freiwillig bereitgestellte, abgeleitete oder gesammelte Informationen über das Verhalten oder die Merkmale einzelner Personen verwenden, um für verschiedene Verbraucher (bzw. Verbrauchergruppen) unterschiedliche Preise festzulegen, und zwar auf der Basis dessen, was sie ihrer Ansicht nach zu zahlen bereit sind (OECD, 2018[109]). Wenn sie von Verkäufern genutzt wird, kann die personalisierte Preisgestaltung dazu führen, dass einige Verbraucher für eine bestimmte Ware oder Dienstleistung weniger, andere hingegen mehr bezahlen, als sie gezahlt hätten, wenn allen Kunden der gleiche Preis angeboten worden wäre.

[5] Dieser Abschnitt stützt sich auf die Arbeit des OECD-Ausschusses für Wissenschafts- und Technologiepolitik und insbesondere auf Kapitel 3 „Artificial Intelligence and Machine Learning in Science" in OECD (2018[53]). Die Hauptautoren dieses Kapitels waren Professor Stephen Roberts von der Universität Oxford und Professor Ross King von der Universität Manchester.

[6] Vgl. https://iris.ai/.

[7] Vgl. https://www.atomwise.com/2015/03/24/new-ebola-treatment-using-artificial-intelligence/.

[8] Vgl. https://www.bbc.com/news/technology-40592520.

[9] Vgl. https://www.nytimes.com/2018/09/20/health/memorial-sloan-kettering-cancer-paige-ai.html.

[10] *State of Wisconsin v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

[11] *Loomis v. Wisconsin*, 137 S.Ct. 2290 (2017).

[12] Trotz der anerkannten Bedeutung öffentlicher Ausgaben für KI-Technologien zu Verteidigungszwecken wird dieser Bereich hier nicht behandelt.

[13] Sofern nicht anders angegeben, bezieht sich „digitale Sicherheit" in dieser Publikation auf das Management wirtschaftlicher und sozialer Risiken, die sich aus Verletzungen der Verfügbarkeit, Integrität und Vertraulichkeit von Informations- und Kommunikationstechnologien und Daten ergeben.

[14] Ein Phishing-Angriff ist ein betrügerischer Versuch, durch elektronische Nachrichten sensible Informationen abzufangen. Die Empfänger der Nachrichten werden dabei durch eine scheinbar vertrauenswürdige Fassade getäuscht. Mehr Arbeit erfordert Spear-Phishing, das auf bestimmte Personen oder Organisationen als Zielgruppe zugeschnitten ist. Dabei werden sensible Informationen wie Name, Geschlecht, Zugehörigkeit usw. gesammelt und verwendet (Brundage et al., 2018[108]). Spear-Phishing ist der häufigste Infektionsvektor: 2017 begannen 71 % der Cyberangriffe mit Spear-Phishing-E-Mails.

[15] Die E-Leaders-Initiative des OECD-Ausschusses für öffentliche Governance bestätigt dies. Ihre Arbeitsgruppe zu neuen Technologien, die aus Vertretern aus 16 Ländern besteht, befasst sich hauptsächlich mit den Themen KI und Blockchain.

[16] BlindTool (https://play.google.com/store/apps/details?id=the.blindtool&hl=en) und Seeing AI (https://www.microsoft.com/en-us/seeing-ai) sind Beispiele für solche Anwendungen.

4. Überlegungen zur Politikgestaltung

In diesem Kapitel werden Überlegungen zur Politikgestaltung erörtert, die es zu berücksichtigen gilt, damit Systeme der künstlichen Intelligenz (KI) vertrauenswürdig und menschenzentriert sind. Es behandelt Bedenken in Bezug auf Ethik und Fairness, die Achtung der Menschenrechte und der demokratischen Werte, darunter auch den Schutz der Privatsphäre. Außerdem geht es um die Gefahren, die mit der Übertragung bestehender Voreingenommenheiten bzw. Verzerrungen, sog. Biases, aus der analogen in die digitale Welt verbunden sind, insbesondere im Hinblick auf Geschlecht oder ethnische Herkunft. Betont wird die Notwendigkeit, Fortschritte auf dem Weg zu robusteren, sichereren und transparenteren KI-Systemen mit klaren Rechenschaftsmechanismen zu erzielen.

Die Förderung vertrauenswürdiger KI-Systeme erfolgt insbesondere über Politikmaßnahmen, die Anreize für Investitionen in verantwortungsvolle KI-Forschung und -Entwicklung setzen, die ein digitales Ökosystem ermöglichen, in dem der Schutz der Privatsphäre nicht durch einen breiteren Datenzugang infrage gestellt wird, die kleinen und mittleren Unternehmen Erfolgschancen eröffnen, die den Wettbewerb unterstützen und gleichzeitig geistiges Eigentum schützen und die Arbeitskräftemobilität bei sich verändernden Arbeitsmarktverhältnissen erleichtern.

Menschenzentrierte KI

Der Einfluss der künstlichen Intelligenz (KI) wächst zusehends. Je stärker sich diese Technologien durchsetzen, umso größere Auswirkungen haben ihre Vorhersagen, Empfehlungen oder Entscheidungen auf das Leben der Menschen. In Fachwelt, Wirtschaft und Politik wird aktiv untersucht, wie sich eine menschenzentrierte und vertrauenswürdige KI am besten gewährleisten lässt, wie ihr Nutzen maximiert und ihre Risiken minimiert werden können und wie ihre gesellschaftliche Akzeptanz erhöht werden kann.

Kasten 4.1. „Black-Box"-KI-Systeme lassen neue Herausforderungen entstehen

Neuronale Netze werden häufig als „Black Box" beschrieben. Obwohl das Verhalten solcher Systeme beobachtet werden kann, unterscheiden sie sich im Hinblick auf ihre Überwachbarkeit deutlich von bisherigen Technologien. Daher ist dieser Begriff durchaus passend. Neuronale Netze iterieren über die Daten, mit denen sie trainiert werden. Sie finden komplexe, probabilistische Korrelationen mit mehreren Variablen, die Teil des Modells werden, das sie aufbauen. Sie liefern jedoch keine Hinweise dazu, wie die Daten zusammenhängen könnten (Weinberger, 2018[1]). Die Daten sind viel zu komplex, als dass Menschen sie analysieren könnten. KI unterscheidet sich von früheren technologischen Entwicklungen u. a. durch folgende Merkmale, die sich auf Transparenz und Verantwortlichkeit auswirken:

- **Erkennbarkeit**: Regelbasierte Algorithmen können Regel für Regel gelesen und geprüft werden, sodass bestimmte Fehlerarten vergleichsweise einfach zu finden sind. Dagegen sind bestimmte Arten von Systemen des maschinellen Lernens (ML), insbesondere neuronale Netze, nur abstrakte mathematische Beziehungen zwischen Faktoren. Diese können äußerst komplex und schwer nachvollziehbar sein, selbst für diejenigen, die sie programmieren und trainieren (OECD, 2016).
- **Evolutiver Charakter**: Einige ML-Systeme sind iterativ und entwickeln sich im Lauf der Zeit weiter; sie können sogar ihr eigenes Verhalten auf unvorhergesehene Weise ändern.
- **Geringe Reproduzierbarkeit**: Es kann sein, dass ein ML-System nur dann eine bestimmte Prognose aufstellt oder eine bestimmte Entscheidung trifft, wenn bestimmte Bedingungen oder Daten vorliegen, welche nicht unbedingt reproduzierbar sind.
- **Zunehmende Zielkonflikte beim Schutz personenbezogener und sensibler Daten:**
  - **Inferenz**: Selbst wenn ihnen keine geschützten oder sensiblen Daten vorliegen, können KI-Systeme solche Informationen und Korrelationen aus Ersatzvariablen ableiten, die nicht personenbezogen oder sensibel sind, wie z. B. dem Einkaufsverhalten oder Standortdaten (Kosinski, M., D. Stillwell und T. Graepel, 2013[2]).
  - **Unerwünschte Proxy-Variablen**: Politische und technische Konzepte zum Schutz der Privatsphäre und zur Nichtdiskriminierung führen in der Regel dazu, dass die Datensammlung auf ein Minimum reduziert, die Verwendung bestimmter Daten verboten oder Daten gelöscht wurden, um ihre Verwendung zu verhindern. Ein KI-System könnte eine Prognose jedoch auf Ersatzvariablen

basieren, die zu den verbotenen und nicht gesammelten Daten in enger Verbindung stehen. Solche Proxys lassen sich nur erkennen, indem auch sensible oder personenbezogene Daten, beispielsweise zur ethnischen Zugehörigkeit, erhoben werden. Wenn solche Daten erfasst werden, muss sichergestellt werden, dass sie stets in angemessener Weise verwendet werden.

- **Das Datenschutz-Paradoxon:** Bei vielen KI-Systemen können mehr Trainingsdaten die Genauigkeit von KI-Prognosen verbessern und dazu beitragen, das Bias-Risiko aufgrund verzerrter Stichproben zu reduzieren. Je mehr Daten jedoch gesammelt werden, desto größer sind die Datenschutzrisiken für die betroffenen Personen.

Einige Arten von KI – die häufig als „Black Boxes“ beschrieben werden – bringen im Vergleich zu früheren technologischen Fortschritten neue Herausforderungen mit sich (Kasten 4.1). Daher hat die OECD – auf der Basis der Arbeit ihrer Sachverständigengruppe für KI (AIGO) – die wichtigsten Prioritäten für eine menschenzentrierte KI festgelegt: Erstens sollte sie zu inklusivem und nachhaltigem Wachstum und zur Lebensqualität beitragen. Zweitens sollte sie menschenzentrierte Werte und Fairness respektieren. Drittens sollten die Nutzung und die Funktionsweise von KI-Systemen transparent sein. Viertens sollten KI-Systeme robust und sicher sein. Fünftens sollte es eine Rechenschaftspflicht für die Ergebnisse von KI-Prognosen und die daraus resultierenden Entscheidungen geben. Bei Vorhersagen, bei denen besonders viel auf dem Spiel steht, werden diese Maßnahmen als entscheidend angesehen. Sie sind aber auch für betriebswirtschaftliche Empfehlungen oder KI-Anwendungsformen wichtig, die weniger starke Auswirkungen haben.

Inklusives und nachhaltiges Wachstum und Lebensqualität

KI verfügt über ein erhebliches Potenzial zur Förderung der Ziele für nachhaltige Entwicklung

KI kann zum Gemeinwohl und zur Erreichung der Ziele der Vereinten Nationen für nachhaltige Entwicklung (SDG) eingesetzt werden, u. a. in Bereichen wie Bildung, Gesundheit, Verkehr, Landwirtschaft und nachhaltige Städte. Viele öffentliche und private Organisationen, darunter die Weltbank, eine Reihe von Organisationen der Vereinten Nationen und die OECD, arbeiten daran, mithilfe von KI die Verwirklichung dieser Ziele voranzubringen.

Die Entwicklung gerechter und inklusiver künstlicher Intelligenz hat zunehmend Priorität

Die Entwicklung gerechter und inklusiver künstlicher Intelligenz hat zunehmend Priorität. Dies gilt insbesondere angesichts von Bedenken, dass KI Ungleichheit verstärken oder Unterschiede innerhalb und zwischen Industrie- und Entwicklungsländern vergrößern könnte. Solche Unterschiede sind auf die Konzentration von KI-Ressourcen – KI-Technologien, Kompetenzen, Datensätze und Rechenleistung – in einigen wenigen Unternehmen und Ländern zurückzuführen. Außerdem besteht die Sorge, dass KI Verzerrungen verfestigen könnte (Talbot et al., 2017[3]). So wird beispielsweise befürchtet, dass KI besondere Auswirkungen auf benachteiligte und unterrepräsentierte Bevölkerungsgruppen haben könnte, z. B. auf Menschen mit niedrigem Bildungsniveau, Geringqualifizierte, Frauen und ältere Menschen, insbesondere in Ländern der unteren und der mittleren Einkommensgruppe (Smith, M. und S. Neupane, 2018[4]). Das International

Development Research Centre in Kanada empfahl kürzlich die Einrichtung eines globalen Fonds für entwicklungsorientierte KI. Damit könnten in Ländern der unteren und der mittleren Einkommensgruppe Exzellenzzentren für künstliche Intelligenz eingerichtet werden, die die Gestaltung und Umsetzung evidenzbasierter, inklusiver Politikmaßnahmen unterstützen (Smith, M. und S. Neupane, 2018[4]). Ziel ist es sicherzustellen, dass durch KI entstehende Nutzeffekte gleichmäßig verteilt werden und zu gerechteren Gesellschaftsstrukturen führen. Mit inklusiven KI-Initiativen soll erreicht werden, dass die wirtschaftlichen Vorteile der KI möglichst breiten Bevölkerungskreisen zugutekommen und niemand den Anschluss verliert.

Inklusive und nachhaltige KI ist ein Bereich, dem Länder wie Indien (Indien, 2018[5]), Unternehmen wie Microsoft[1] und Hochschulinitiativen wie das Berkman Klein Center in Harvard besondere Aufmerksamkeit zuteilwerden lassen. So hat Microsoft beispielsweise Projekte wie Seeing AI lanciert, eine mobile Anwendung, die Menschen mit Sehbehinderung helfen soll. Diese Anwendung scannt und erkennt alle Elemente in der Umgebung der Person und stellt ihr eine Audiobeschreibung zur Verfügung. Microsoft investiert außerdem 2 Mio. USD in Initiativen, bei denen KI zur Sicherung der Nachhaltigkeit genutzt werden soll, z. B. zum Erhalt der Biodiversität oder im Kampf gegen den Klimawandel (Heiner, D. und C. Nguyen, 2018[6]).

Menschenzentrierte Werte und Fairness

Menschenrechte und Ethikrichtlinien

Ethische Normen sind in den internationalen Menschenrechtsbestimmungen verankert

Ethische Normen sind in den internationalen Menschenrechtsbestimmungen verankert. KI kann zur Gewährleistung der Menschenrechte beitragen, aber auch neue Risiken vorsätzlicher oder unbeabsichtigter Menschenrechtsverletzungen schaffen. Die Menschenrechtsbestimmungen und die an sie geknüpften rechtlichen und sonstigen institutionellen Strukturen könnten ein Instrument sein, um sicherzustellen, dass KI menschenzentriert ist (Kasten 4.2).

Kasten 4.2. Menschenrechte und KI

Bei den internationalen Menschenrechtsbestimmungen handelt es sich um verschiedene internationale Rechtsakte, etwa die Internationale Menschenrechtscharta,[1] sowie regionale Systeme zum Schutz der Menschenrechte, die in den letzten siebzig Jahren auf der ganzen Welt entwickelt wurden. Sie geben eine Reihe universeller Mindeststandards vor, die u. a. auf Werten wie Menschenwürde, Autonomie und Gleichheit beruhen und mit dem Prinzip der Rechtsstaatlichkeit im Einklang stehen. Aufgrund dieser Standards und der an sie geknüpften Rechtsgrundlagen sind Länder rechtlich verpflichtet, die Menschenrechte zu achten, zu schützen und zu gewährleisten. Sie verlangen zudem, dass Personen, denen Rechte verwehrt wurden oder deren Rechte verletzt wurden, Anspruch auf effektive Rechtsmittel haben.

Zu den Menschenrechten gehören das Recht auf Gleichheit, das Recht auf Nichtdiskriminierung, das Recht auf Vereinigungsfreiheit und das Vereinigungsrecht, das Recht auf Schutz der Privatsphäre und wirtschaftliche, soziale und kulturelle Rechte wie das Recht auf Bildung oder das Recht auf Gesundheit.

Neuere zwischenstaatliche Instrumente wie die „Leitprinzipien für Wirtschaft und Menschenrechte“ der Vereinten Nationen (OHCHR, 2011[7]) befassen sich auch mit der Rolle privater Akteure im Bereich der Menschenrechte. Schließlich sehen sie diese in der Verantwortung, die Menschenrechte zu achten. Darüber hinaus enthält die 2011 aktualisierte Version der *OECD-Leitsätze für multinationale Unternehmen* (OECD, 2011[8]) ein Kapitel über Menschenrechte.

Die Menschenrechte überschneiden sich mit weiterreichenden ethischen Fragen und anderen für die künstliche Intelligenz relevanten Regulierungsbereichen, wie etwa dem Schutz personenbezogener Daten oder dem Produktsicherheitsrecht. Allerdings haben diese anderen Anliegen und Fragen häufig eine andere Tragweite.

1. Die Internationale Menschenrechtscharta besteht aus der Allgemeinen Erklärung der Menschenrechte, dem Internationalen Pakt über bürgerliche und politische Rechte und dem Internationalen Pakt über wirtschaftliche, soziale und kulturelle Rechte.

KI könnte zur Förderung der Menschenrechte beitragen

Angesichts ihres potenziell breiten Anwendungs- und Nutzungsspektrums könnte KI den Schutz und die Gewährleistung der Menschenrechte verbessern. Zum Beispiel könnte KI eingesetzt werden, um Muster der Nahrungsmittelknappheit zur Hungerbekämpfung zu analysieren, medizinische Diagnosen und Behandlungen zu verbessern, das Angebot an Gesundheitsleistungen sowie den Zugang zur Gesundheitsversorgung auszuweiten und Diskriminierungen ans Licht zu bringen.

KI könnte Menschenrechte aber auch gefährden

Auf dem Gebiet der Menschenrechte kann KI aber auch eine Reihe von Herausforderungen mit sich bringen, was in Diskussionen über KI und Ethik häufig angesprochen wird. Bestimmte KI-Systeme könnten auf unbeabsichtigte Weise oder vorsätzlich Menschenrechte verletzen oder für Menschenrechtsverletzungen benutzt werden. Besonders über unbeabsichtigte Auswirkungen wird viel diskutiert. ML-Algorithmen, die Rückfallwahrscheinlichkeiten vorhersagen, können z. B. ein Bias aufweisen, das unerkannt bleibt. KI-Technologien können aber auch mit vorsätzlichen Menschenrechtsverletzungen in Verbindung gebracht werden. Beispiele hierfür sind der Einsatz von KI-Technologien, um politische Dissidenten ausfindig zu machen oder das Recht des Einzelnen auf freie Meinungsäußerung und politische Partizipation einzuschränken. In solchen Fällen ist der Verstoß selbst in der Regel nicht – oder nicht nur – auf die Nutzung von KI zurückzuführen. Allerdings könnte er durch die technische Raffinesse und Effizienz von KI verschärft werden.

Auch in Situationen, in denen die Auswirkungen auf die Menschenrechte unbeabsichtigt oder schwer zu erkennen sind, kann der Einsatz von KI besondere Herausforderungen darstellen. Dies kann auf die Verwendung qualitativ unzureichender Trainingsdaten, das Systemdesign oder komplexe Interaktionen zwischen dem KI-System und seiner Umgebung zurückzuführen sein. Ein Beispiel ist die algorithmusbedingte Verstärkung von Verhetzung oder Anstiftung zu Gewalt im Internet. Ein weiteres Beispiel ist die unbeabsichtigte Verbreitung sogenannte Fake News, die das Recht auf Teilhabe am politischen und öffentlichen Geschehen beeinträchtigen könnte. Welches Ausmaß und welche Auswirkungen die dadurch verursachten Schäden haben, hängt von der Tragweite der Entscheidungen des betreffenden KI-Systems ab. Zum Beispiel haben die Entscheidungen

eines KI-Systems, das Nachrichten empfiehlt, potenziell geringere Auswirkungen als Entscheidungen eines Algorithmus, der das Rückfallrisiko von zur Bewährung Verurteilten vorhersagt.

Menschenrechtsbestimmungen können durch KI-Ethikrichtlinien ergänzt werden

Mit Ethikrichtlinien kann dem Risiko begegnet werden, dass KI nicht menschenzentriert ist oder nicht mit menschlichen Werten im Einklang steht. Sowohl private Unternehmen als auch staatliche Stellen haben zahlreiche Ethikrichtlinien verabschiedet, die sich mit künstlicher Intelligenz befassen.

So hat beispielsweise das zu Google gehörende Unternehmen DeepMind im Oktober 2017 eine Ethik-Abteilung (DeepMind Ethics & Society) gegründet.[2] Diese Abteilung soll Technologieexperten helfen, die ethischen Implikationen ihrer Arbeit zu verstehen, und dazu beitragen, dass die Gesellschaft insgesamt besser entscheiden kann, wie KI für sie nutzbringend sein kann. Sie wird auch externe Forschung zu Themen wie algorithmische Verzerrungen, Zukunft der Arbeit und letale autonome Waffensysteme finanzieren. Google selbst hat eine Reihe von Ethikgrundsätzen aufgestellt, die seine Forschung, Produktentwicklung und Geschäftsentscheidungen leiten sollen.[3] Das Unternehmen hat ein Weißbuch zur KI-Governance veröffentlicht, in dem Fragen aufgezeigt werden, die in Zusammenarbeit mit Staat und Zivilgesellschaft geklärt werden müssen.[4] Microsoft verfolgt in Bezug auf KI die Vision, „die menschliche Genialität mit intelligenter Technologie zu verstärken“ (Heiner, D. und C. Nguyen, 2018[6]). Das Unternehmen hat Projekte ins Leben gerufen, die eine inklusive und nachhaltige Entwicklung gewährleisten sollen.

Die Menschenrechtsbestimmungen geben mit ihren institutionellen Mechanismen und ihrer globalen Architektur die Richtung vor und schaffen die Basis für eine ethische und menschenzentrierte Entwicklung und Nutzung von KI in der Gesellschaft.

Menschenrechtsbestimmungen sind im KI-Kontext ein wichtiges Instrument

Im KI-Kontext ist die Nutzung von Menschenrechtsbestimmungen insbesondere deshalb von Vorteil, weil sie es ermöglicht, sich auf etablierte Institutionen, eine bestehende Rechtsprechung und eine universelle Sprache zu stützen, und weil diese Bestimmungen große internationale Akzeptanz genießen:

- **Etablierte Institutionen**: Auf dem Gebiet der Menschenrechte ist im Lauf der Zeit eine umfassende internationale, regionale und nationale Infrastruktur entwickelt worden. Sie besteht aus zwischenstaatlichen Organisationen, Gerichten, Nichtregierungsorganisationen, wissenschaftlichen Einrichtungen und anderen Institutionen und Organen, in denen Menschenrechte geltend gemacht und Rechtsmittel eingelegt werden können.
- **Rechtsprechung**: Als Rechtsnormen werden die durch die Menschenrechte geschützten Werte in konkreten Situationen durch die Rechtsprechung und die Auslegungsarbeit internationaler, regionaler und nationaler Institutionen umgesetzt und rechtsverbindlich gemacht.
- **Universelle Sprache**: Die Menschenrechte bieten eine universelle Sprache für ein globales Thema. Zusammen mit der allgemeinen Menschenrechtsinfrastruktur kann sie einer größeren Zahl von Akteuren die Möglichkeit geben, an der Debatte über den Platz von KI in der Gesellschaft teilzuhaben. So kann diese Debatte über den Kreis der Akteure, die direkt an der KI-Entwicklung beteiligt sind, hinaus ausgedehnt werden.

- **Internationale Akzeptanz**: Die Menschenrechte genießen hohe internationale Akzeptanz und Legitimität. Bereits wenn der Eindruck entsteht, dass ein bestimmter Akteur Menschenrechte verletzt, kann dies für ihn erhebliche Folgen haben: Die Kosten des Reputationsverlusts sind hoch.

Ein an den Menschenrechten ausgerichteter KI-Ansatz kann dazu beitragen, Risiken, Prioritäten und benachteiligte Gruppen zu identifizieren und Abhilfe zu schaffen

- **Risikoerkennung**: Menschenrechtsbestimmungen können helfen, Schadensrisiken zu erkennen. Insbesondere bieten sie eine Grundlage, um menschenrechtlichen Sorgfaltspflichten (Due Diligence) nachzukommen, z. B. im Rahmen von Menschenrechtsverträglichkeitsprüfungen (Kasten 4.3).
- **Kernanforderungen**: Als Mindeststandards legen die Menschenrechte unantastbare Kernanforderungen fest. Zum Beispiel hilft die Menschenrechtsprechung, Regeln für die Meinungsäußerung in sozialen Netzwerken festzulegen: z. B. indem Verhetzung klar als eine Grenze gekennzeichnet wird, die nicht überschritten werden darf.
- **Ermittlung von Risikobereichen**: Menschenrechte können ein nützliches Instrument sein, um Bereiche oder Aktivitäten mit hohem Risiko zu erkennen. In diesen Bereichen ist dann erhöhte Vorsicht geboten oder muss u. U. auf die Nutzung von KI verzichtet werden.
- **Ermittlung benachteiligter Gruppen oder Gemeinden**: Menschenrechte können helfen, in Bezug auf KI benachteiligte oder gefährdete Gruppen oder Gemeinden zu ermitteln. Manche Personen oder Gruppen können z. B. aufgrund fehlender Möglichkeiten zur Nutzung von Smartphones unterrepräsentiert sein.
- **Abhilfe**: Als Rechtsnormen mit daraus erwachsenden Verpflichtungen können die Menschenrechte für diejenigen Abhilfe schaffen, deren Rechte verletzt werden. Beispiele für solche Abhilfemaßnahmen sind die Einstellung der Tätigkeit, die Entwicklung neuer Verfahren oder Maßnahmen, eine Entschuldigung oder eine Schadenersatzzahlung.

Kasten 4.3. Menschenrechtsverträglichkeitsprüfungen

Menschenrechtsverträglichkeitsprüfungen (Human Rights Impact Assessments – HRIA) können helfen, Risiken zu erkennen, die die am KI-Lebenszyklus beteiligten Akteure sonst möglicherweise übersehen würden. Bei solchen Prüfungen richtet sich der Blick stärker auf unbeabsichtigte Auswirkungen auf den Menschen als auf die Optimierung der Technologie oder ihrer Ergebnisse. Mit solchen Prüfungen oder ähnlichen Verfahren kann sichergestellt werden, dass die Menschenrechte während des gesamten Lebenszyklus der Technologien geachtet werden.

Bei Menschenrechtsverträglichkeitsprüfungen werden Technologien umfassend in Bezug auf ein breites Spektrum an Auswirkungen bewertet, die sie auf die Menschenrechte haben könnten, was recht ressourcenintensiv ist. Dabei kann es einfacher sein, mit dem KI-System selbst zu beginnen und sich nach außen vorzuarbeiten. So kann eine begrenzte Anzahl von Bereichen untersucht werden, in denen rechtliche Herausforderungen am

wahrscheinlichsten sind. Branchenverbände können bei der Durchführung von Menschenrechtsverträglichkeitsprüfungen in kleinen und mittleren Unternehmen (KMU) oder Nicht-Technologieunternehmen helfen, die KI-Systeme zwar einsetzen, sich mit solchen Technologien aber möglicherweise nicht genügend auskennen. In Bezug auf die Meinungsfreiheit und den Schutz der Privatsphäre ist die Global Network Initiative ein gutes Beispiel für solche Organisationen. Sie hilft Unternehmen, vorausschauend zu planen und Menschenrechtsverträglichkeitsprüfungen in ihre Pläne für neue Produkte einzubeziehen (https://globalnetworkinitiative.org/).

Menschenrechtsverträglichkeitsprüfungen haben den Nachteil, dass sie im Allgemeinen auf der Ebene der einzelnen Unternehmen durchgeführt werden. An KI-Systemen können jedoch viele Akteure beteiligt sein. Es kann daher ineffektiv sein, nur einen Teil dieser Akteure zu prüfen. Microsoft war das erste große Technologieunternehmen, das 2018 eine Menschenrechtsverträglichkeitsprüfung zu KI durchführte.

Auch die Umsetzung eines an den Menschenrechten ausgerichteten KI-Ansatzes ist mit erheblichen Herausforderungen verbunden. Dabei geht es z. B. darum, dass sich die Menschenrechtsbestimmungen an staatliche Instanzen richten, dass ihre Durchsetzung an eine Gebietshoheit geknüpft ist, dass sie eher geeignet sind, bei schweren Beeinträchtigungen einer geringen Zahl von Personen Abhilfe zu schaffen, und dass sie für Unternehmen kostspielig sein können:

- **Menschenrechtsbestimmungen richten sich an staatliche Instanzen, nicht an private Akteure.** Akteure des privaten Sektors spielen jedoch eine Schlüsselrolle bei der Erforschung, Entwicklung und Einführung von KI. Dies ist ein Problem, das nicht nur im KI-Bereich besteht. In mehreren zwischenstaatlichen Initiativen wird versucht, den Graben zwischen öffentlichem und privatem Sektor zu schließen. Außerdem setzt sich zunehmend die Erkenntnis durch, dass eine gute Menschenrechtsbilanz auch gut fürs Geschäft ist.[5]
- **Die Durchsetzung der Menschenrechte ist an die Gebietshoheit geknüpft**. Im Allgemeinen müssen Kläger nachweisen, dass sie in einem bestimmten Staat Klagebefugnis haben. In Fällen, in denen es um große multinationale Unternehmen und KI-Systeme geht, die in mehreren Staaten eingesetzt werden, ist dies möglicherweise schwierig.
- **Menschenrechtsbestimmungen eignen sich besser, um bei schweren Beeinträchtigungen kleiner Gruppen von Personen Abhilfe zu schaffen,** als wenn einer großen Gruppe von Menschen ein weniger erheblicher Schaden entstanden ist. Zudem können Menschenrechtsbestimmungen und ihre Strukturen auf Außenstehende undurchsichtig wirken.
- **Menschenrechtsbestimmungen stehen teilweise im Ruf, kostspielig für die Unternehmen zu sein.** Daher dürften Ansätze, die Ethik, Verbraucherschutz oder verantwortungsvolles unternehmerisches Handeln sowie wirtschaftliche Argumente für die Einhaltung der Menschenrechte in den Mittelpunkt rücken, besonders vielversprechend sein.

Einige der allgemeineren Herausforderungen, die im Zusammenhang mit KI bestehen, z. B. die Frage der Transparenz und der Nachvollziehbarkeit, wirken sich auch im Menschenrechtsbereich aus (vgl. Abschnitt „Transparenz und Nachvollziehbarkeit“). Ohne Transparenz ist es schwierig, festzustellen, wann Menschenrechte verletzt wurden, oder eine Beschwerde über eine Menschenrechtsverletzung zu begründen. Dasselbe gilt für die

Einlegung von Rechtsmitteln, die Bestimmung des ursächlichen Zusammenhangs und die Rechenschaftslegung.

Schutz personenbezogener Daten

KI stellt das Konzept der „personenbezogenen Daten" und der Einwilligung infrage

KI ist zunehmend in der Lage, verschiedene Datensätze miteinander zu verknüpfen und verschiedene Arten von Daten abzugleichen, was tiefgreifende Konsequenzen hat. Separat gespeicherte Daten (bzw. Daten, die von persönlichen Identifikatoren befreit, d. h. „de-identifiziert" wurden) galten früher nicht als personenbezogen. Mit KI können jedoch nicht personenbezogene Daten mit anderen Daten korreliert und bestimmten Personen zugeordnet werden, wodurch sie zu personenbezogenen Daten (bzw. „re-identifiziert") werden. Durch die algorithmische Korrelation verschwimmt die Grenze zwischen personenbezogenen und nicht personenbezogenen Daten. Nicht personenbezogene Daten können zunehmend verwendet werden, um Personen zu re-identifizieren oder über sie sensible Informationen abzuleiten, die über das hinausgehen, was diese Personen ursprünglich wissentlich preisgegeben hatten (Cellarius, 2017[9]). 2007 hatten Forscher z. B. bereits angeblich anonyme Daten verwendet, um die Liste der auf Netflix ausgeliehenen Filme mit den in der Internet Movie Database (IMDb) veröffentlichten Bewertungen zu verknüpfen. Auf diese Weise konnten sie die Personen identifizieren, die Filme ausgeliehen haben, und auf ihre vollständige Ausleihhistorie zugreifen. Da die Menge der erfassten Daten zunimmt und die Technologien immer ausgereifter werden, wird es zunehmend möglich sein, solche Verknüpfungen herzustellen. Damit wird es schwierig zu beurteilen, welche Daten tatsächlich als nicht personenbezogen betrachtet werden können und dies auch bleiben werden.

Die Unterscheidung zwischen sensiblen und nicht sensiblen Daten wird immer schwieriger, wie beispielsweise die Datenschutz-Grundverordnung (DSGVO) der Europäischen Union zeigt. Einige Algorithmen können sensible Daten aus „nicht sensiblen" Daten ableiten, z. B. kann der emotionale Zustand von Personen anhand der Art, wie sie auf ihrer Tastatur tippen, bestimmt werden (Privacy International and ARTICLE 19, 2018[10]). Die Verwendung von KI zur Identifikation oder Re-Identifikation von Daten, die ursprünglich nicht personenbezogen bzw. de-identifiziert waren, stellt auch ein rechtliches Problem dar. Verschiedene Texte befassen sich mit dem Schutz personenbezogener Daten, so z. B. die Empfehlung des Rats der OECD über Leitlinien für den Schutz des Persönlichkeitsbereichs und den grenzüberschreitenden Verkehr personenbezogener Daten (*Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, „Leitlinien zum Datenschutz") (Kasten 4.4). Es ist jedoch nicht klar, ob oder ab wann sie für Daten gelten, die unter bestimmten Umständen mit bestimmten Personen verknüpft werden oder verknüpft werden könnten (OVIC, 2018[11]). Eine extreme Auslegung könnte dazu führen, dass der Umfang des Datenschutzes erheblich ausgeweitet würde, was seine Anwendung erschweren würde.

Kasten 4.4. Die OECD-Leitlinien zum Datenschutz

Die Empfehlung des Rats über Leitlinien für den Schutz des Persönlichkeitsbereichs und den grenzüberschreitenden Verkehr personenbezogener Daten („Leitlinien zum Datenschutz“) wurde 1980 angenommen und 2013 aktualisiert (OECD, 2013[12]). Sie enthält Definitionen einschlägiger Begriffe. Insbesondere werden „personenbezogene Daten“ als „alle Informationen, die sich auf eine bestimmte oder bestimmbare Person (Datensubjekt) beziehen“ definiert. Außerdem legt sie Grundsätze fest, die bei der Verarbeitung personenbezogener Daten zu beachten sind. Dabei handelt es sich um die Grundsätze der begrenzten Datenerhebung (die gegebenenfalls nur mit Einwilligung als Mittel zur Gewährleistung dieses Grundsatzes erfolgen darf), der Datenqualität, der Zweckbestimmung, der Nutzungsbegrenzung, der Sicherung, der Offenheit, des Mitspracherechts und der Rechenschaftspflicht. Sie sehen auch vor, dass Mitgliedstaaten bei der Umsetzung der Leitlinien zum Datenschutz sicherzustellen haben, dass Datensubjekte nicht in unfairer Weise diskriminiert werden. Die Umsetzung der Leitlinien zum Datenschutz sollte 2019 überprüft werden, um u. a. den jüngsten Entwicklungen, auch im KI-Bereich, Rechnung zu tragen.

KI stellt auch die Datenschutzgrundsätze der begrenzten Datenerhebung, der Nutzungsbegrenzung und der Zweckbestimmung infrage

Zum Trainieren und Optimieren von KI-Systemen benötigen ML-Algorithmen große Datenmengen. Dadurch entstehen Anreize, die Datenerhebung zu maximieren, anstatt sie auf ein Minimum zu reduzieren. Mit der zunehmenden Nutzung von KI-basierten Geräten und dem Internet der Dinge (Internet of Things – IoT) werden immer mehr Daten erhoben. Die Datenerhebung erfolgt immer häufiger und ist insgesamt einfacher. Die erhobenen Daten werden zudem mit anderen Daten verknüpft, z. T. ohne Wissen oder Einwilligung der betroffenen Personen.

Die identifizierten Muster und die Abläufe des „Lernprozesses“ lassen sich nur schwer antizipieren. Daher kann der Umfang der Datenerhebung und -nutzung über das hinausgehen, was den Datensubjekten ursprünglich bekannt war, was sie selbst an Daten preisgegeben und wozu sie ihre Einwilligung gegeben hatten (Privacy International and ARTICLE 19, 2018[10]). Dies könnte mit den in den Leitlinien zum Datenschutz festgelegten Grundsätzen der begrenzten Datenerhebung, der Nutzungsbegrenzung und der Zweckbestimmung unvereinbar sein (Cellarius, 2017[9]). Die ersten beiden dieser Grundsätze beruhen z. T. auf der Einwilligung des Datensubjekts (wobei anerkannt wird, dass eine Einwilligung in einigen Fällen eventuell nicht möglich ist). Die Einwilligung ist die Basis für die Erhebung personenbezogener Daten oder ihre Nutzung zu anderen als den ursprünglich angegebenen Zwecken. KI-Technologien wie z. B. Deep Learning, deren Funktionsweise schwer nachvollziehbar und überwachbar ist, lassen sich den betroffenen Personen auch nur schwer erklären. Dies ist eine Herausforderung für die Unternehmen. Aus Unternehmenskreisen wurde verlautet, dass sich die Datenschutzgrundsätze angesichts des rasanten Tempos, mit dem KI Zugang zu Daten erhält, diese analysiert und nutzt, zunehmend schwer umsetzen lassen (OECD, 2018[13]).

Diese Schwierigkeiten verstärken sich durch die Verquickung von KI-Technologien und Entwicklungen im IoT-Bereich, d. h. der Anbindung einer immer größeren Zahl von Geräten und Gegenständen an das Internet. Die immer stärkere Verflechtung von KI- und IoT-Technologien (z. B. IoT-Geräte, die mit KI oder KI-Algorithmen zur Analyse von IoT-Daten ausgestattet sind) führt dazu, dass ständig mehr Daten, auch personenbezogene,

erhoben werden. Diese können zunehmend miteinander verknüpft und dann zusammen analysiert werden. Während die Zahl der Geräte wächst, die Daten sammeln (z. B. Überwachungskameras oder autonome Fahrzeuge), verbessert sich zugleich die KI-Technologie (z. B. Gesichtserkennung). Die Kombination dieser beiden Trends birgt die Gefahr eines stärkeren Eingriffs in die Privatsphäre als jeder dieser beiden Faktoren für sich genommen (OVIC, 2018[11]).

KI kann die Prinzipien der Beteiligung und Zustimmung des Einzelnen aber auch stärken

KI könnte den Datenschutz aber auch verbessern. Beispielsweise laufen in einer Reihe von technischen Normungsorganisationen Initiativen zum Aufbau von KI-Systemen nach den Grundsätzen *Privacy by design* (Datenschutz durch Technikgestaltung) und *Privacy by default* (Datenschutz durch datenschutzfreundliche Voreinstellungen). Die meisten dieser Organisationen verwenden und adaptieren Datenschutzrichtlinien, darunter auch die OECD-Leitlinien zum Datenschutz. Darüber hinaus wird KI eingesetzt, um den Nutzern personalisierte, auf ihre Bedürfnisse zugeschnittene Dienste anzubieten, die auf ihren im Lauf der Zeit erfassten persönlichen Datenschutzpräferenzen basieren (OVIC, 2018[11]). Diese Dienste können dem Einzelnen helfen, sich im Dickicht der Maßnahmen zur Verarbeitung personenbezogener Daten zurechtzufinden und sicherzustellen, dass seine Präferenzen insgesamt berücksichtigt werden. Auf diese Weise stärkt KI die informierte Zustimmung und Beteiligung des Einzelnen. Ein Forscherteam entwickelte z. B. Polisis, ein automatisiertes Framework, das mithilfe von Klassifikatoren neuronaler Netze Datenschutzrichtlinien analysiert (Harkous, 2018[14]).

Fairness und Ethik

ML-Algorithmen können implizit in ihren Trainingsdaten enthaltene Verzerrungen reproduzieren

Politikinitiativen im KI-Bereich befassen sich aktuell intensiv mit Fragen der Ethik, Fairness und/oder Gerechtigkeit. Es gibt erhebliche Bedenken, dass ML-Algorithmen in ihren Trainingsdaten implizit enthaltene Voreingenommenheiten, z. B. gegenüber bestimmten ethnischen Gruppen, und stereotype Vorstellungen übernehmen und reproduzieren. Da technologische Artefakte häufig gesellschaftliche Werte verkörpern, muss in der Debatte um Fairness deutlich gemacht werden, welcher Art von Gesellschaft die Technologien dienen sollen, wer Schutz genießen soll und welche Grundwerte dabei zu beachten sind (Flanagan, M., D. Howe und H. Nissenbaum, 2008[15]). Disziplinen wie Philosophie, Recht und Wirtschaft setzen sich seit Jahrzehnten aus unterschiedlichen Perspektiven mit unterschiedlichen Gerechtigkeitskonzepten auseinander. Sie verdeutlichen das breite Spektrum möglicher Interpretationen des Prinzips der Fairness und der daraus erwachsenden Implikationen für die Politik.

Philosophie, Recht und Informatik haben unterschiedliche Auffassungen von Fairness und ethischer KI

Die Philosophie befasst sich mit Konzepten von richtigem und falschem Verhalten, von Gut und Böse und von Moral. Im Kontext ethischer KI sind drei große philosophische Theorien relevant (Abrams et al., 2017[16]):

- Das **Konzept der menschlichen Grundrechte**, bei dem auf Immanuel Kant Bezug genommen wird, stellt auf die formalen Grundsätze der Ethik ab. Dabei geht es um

konkrete Rechte wie das Recht auf Schutz der Privatsphäre oder auf Freiheit. Diese Grundsätze werden durch Vorschriften geschützt, die KI-Systeme einhalten sollten.

- Der **utilitaristische Ansatz**, der von Jeremy Bentham und John Stuart Mill propagiert wurde, beruht auf dem Prinzip, dass staatliches Handeln das menschliche Wohlergehen auf der Grundlage wirtschaftlicher Kosten-Nutzen-Analysen maximieren soll. Für KI wirft der utilitaristische Ansatz die Frage auf, *wessen* Wohlergehen maximiert werden soll (das Wohlergehen des Einzelnen, der Familie, der Gesellschaft oder des Staats). Die Antwort auf diese Frage kann Auswirkungen auf das Design der Algorithmen haben.
- Der **Ansatz der Tugendethik**, der sich auf Aristoteles gründet, richtet den Blick auf die Werte und ethischen Normen, die eine Gesellschaft braucht, um die Menschen in ihren täglichen Bemühungen um ein lebenswertes Leben zu unterstützen. Dies wirft die Frage auf, welche Werte und welche ethischen Normen schützenswert sind.

In der Rechtswissenschaft werden Konzepte der Fairness häufig mit den Begriffen „Gleichheit" und „Gerechtigkeit" umrissen. Zwei zentrale Konzepte sind hier die individuelle Fairness und die Gruppenfairness.

- Die **Individuelle Fairness** bzw. Gerechtigkeit für den Einzelnen entspricht dem Konzept der Gleichheit vor dem Gesetz. Sie impliziert, dass alle Personen gleichbehandelt und nicht aufgrund besonderer Merkmale diskriminiert werden sollten. Gleichheit ist als internationales Menschenrecht anerkannt.
- Bei der **Gruppenfairness** geht es um die Fairness der Ergebnisse. Es gilt sicherzustellen, dass sich die Situation für Personen, die unterschiedlichen, durch bestimmte Merkmale (z. B. ethnische Herkunft oder Geschlecht) gekennzeichneten Gruppen angehören, im Ergebnis nicht systematisch anders darstellt. Dem liegt der Gedanke zugrunde, dass bestimmte Unterschiede und historische Umstände dazu führen können, dass verschiedene Gruppen auf bestimmte Situationen unterschiedlich reagieren. Um Gruppenfairness zu gewährleisten, werden unterschiedliche Ansätze verfolgt. Einer davon ist die positive Diskriminierung.

Entwickler von KI-Systemen haben darüber nachgedacht, wie man Fairness in KI-Systemen gewährleisten kann. Unterschiedliche Definitionen von Fairness schlagen sich dabei in verschiedenen Herangehensweisen nieder (Narayanan, 2018[17]):

- Das **Konzept der Unkenntnis**, bei dem ein KI-System keine identifizierbaren Faktoren kennen sollte, beruht auf dem rechtlichen Prinzip der individuellen Fairness. Das KI-System darf in diesem Fall keine Daten über sensible Attribute wie Geschlecht, ethnische Herkunft und sexuelle Orientierung berücksichtigen (Yona, 2017[18]). Allerdings können zahlreiche andere Faktoren mit dem geschützten Attribut (z. B. Geschlecht) korreliert sein. Zudem könnte sich die Entfernung dieser Daten negativ auf die Genauigkeit des KI-Systems auswirken.
- Beim Konzept der **kenntnisbasierten Fairness** wird Unterschieden zwischen bestimmten Gruppen Rechnung getragen. Ziel ist es, vergleichbare Personen gleich zu behandeln. Die Herausforderung besteht jedoch darin festzulegen, wer mit wem gleichbehandelt werden soll. Um zu verstehen, wer für eine bestimmte Aufgabe als vergleichbar anzusehen ist, bedarf es der Kenntnis sensibler Merkmale.

- Mit **Konzepten der Gruppenfairness** soll sichergestellt werden, dass die Ergebnisse für Personen, die verschiedenen Gruppen angehören, sich nicht systematisch unterscheiden. Es besteht die Befürchtung, dass KI-Systeme nicht fair sind und traditionelle Voreingenommenheiten verfestigen oder verstärken, da sie sich oft auf Datensätze aus der Vergangenheit stützen.

Unterschiedliche Auffassungen von Fairness führen für unterschiedliche gesellschaftliche Gruppen und unterschiedliche Akteure zu unterschiedlichen Ergebnissen. Nicht alle Ziele können gleichzeitig erreicht werden. Daher sollten bei technologischen Designentscheidungen, die sich nachteilig auf bestimmte Gruppen auswirken könnten, auch politische Erwägungen bzw. gegebenenfalls Entscheidungen berücksichtigt werden.

Im Personalwesen zeigen sich die Chancen und Risiken von KI besonders deutlich

Im Personalwesen können durch künstliche Intelligenz Verzerrungen bei der Einstellung von Mitarbeitern entweder verfestigt oder im Gegenteil aufgedeckt und verringert werden. Eine von Carnegie Mellon durchgeführte Studie, in der die Muster von Online-Stellenausschreibungen untersucht wurden, ergab z. B., dass eine Annonce für eine hoch bezahlte Führungsposition Männern 1 816 Mal, Frauen hingegen nur 311 Mal angezeigt wurde (Simonite, 2018[19]). Die Zusammenarbeit zwischen Mensch und KI ist daher sinnvoll, um sicherzustellen, dass KI-Anwendungen für Personaleinstellungen und -bewertungen transparent sind. Es muss gewährleistet sein, dass Verzerrungen nicht in den Code eingehen können, dass es also z. B. nicht möglich ist, dass Kandidaten aus bestimmten Kulturkreisen bei der Besetzung von Posten, die ihnen in der Vergangenheit nicht offenstanden, automatisch ausgeschlossen werden (OECD, 2017[20]).

Diskriminierung in KI-Systemen kann auf verschiedene Weise verringert werden

Um das Diskriminierungsrisiko in KI-Systemen zu verringern, wurden bereits verschiedene Ansätze vorgeschlagen, so z. B. Sensibilisierungsmaßnahmen, diversitätsfördernde organisatorische Maßnahmen, Normen, technische Lösungen zur Erkennung und Korrektur algorithmischer Verzerrungen sowie Selbstregulierungs- bzw. Regulierungsansätze. Im Bereich des Predictive Policing wurde beispielsweise die Einführung von Algorithmus-Folgenabschätzungen oder entsprechenden Erklärungen empfohlen. Dazu müsste die Polizei die Wirksamkeit, den Nutzen und die möglichen diskriminierenden Auswirkungen der verschiedenen technologischen Optionen, die ihr für die vorausschauende Polizeiarbeit zur Verfügung stehen, bewerten (Selbst, 2017[21]). Rechenschaftspflicht und Transparenz sind wichtig, um Fairness zu erreichen. Aber selbst zusammengenommen sind sie noch keine Garantie für Fairness (Weinberger, 2018[22]); (Narayanan, 2018[17]).

Anstrengungen zur Gewährleistung von Fairness in KI-Systemen können Zielkonflikte mit sich bringen

KI-Systeme sollen „fair“ sein. Ihre Vorhersagen sollen z. B. gewährleisten, dass nur Angeklagte, bei denen ein hohes Risiko besteht, nicht auf Kaution freigelassen werden oder dass einem Kreditkunden das je nach seiner Rückzahlungsfähigkeit günstigste Finanzierungsangebot gemacht wird. Bei **falsch-positiven Fehlern** wird irrtümlicherweise ein negativer Ausgang vorausgesagt. Zum Beispiel können KI-Systeme fälschlicherweise vorhersagen, dass ein Angeklagter erneut straffällig wird, der in Wirklichkeit keine neuen Straftaten begehen wird. Sie können auch fälschlicherweise eine Erkrankung vorhersagen, an der die untersuchte Person gar nicht leidet. **Falsch-negative Fehler** sind hingegen Fälle,

bei denen ein KI-System fälschlicherweise einen positiven Ausgang vorhersagt, beispielsweise dass ein Angeklagter nicht erneut straffällig wird oder dass ein Patient eine bestimmte Krankheit nicht hat, an der er in Wirklichkeit leidet.

Konzepte der Gruppenfairness versuchen der Tatsache Rechnung zu tragen, dass die Ausgangssituation nicht für alle Gruppen die gleiche ist. Sie berücksichtigen solche Unterschiede mathematisch, indem sie gewährleisten, dass die Richtigkeitsquote bzw. die Fehlerquote für alle Gruppen gleich hoch ist. Dies bedeutet z. B., dass der Anteil der fälschlicherweise als Wiederholungstäter eingestuften Personen unter den Männern gleich hoch sein muss wie unter den Frauen (oder dass die falsch-positiven und die falsch-negativen Ergebnisse einander ausgleichen).

Der Ausgleich von falsch-positiven und falsch-negativen Ergebnissen stellt eine Herausforderung dar. Falsch-negative Ergebnisse werden häufig als weniger wünschenswert und risikoreicher angesehen als falsch-positive Ergebnisse, weil sie kostspieliger sind (Berk, R. und J. Hyatt, 2015[23]). Zum Beispiel sind die Kosten, die einer Bank entstehen, wenn sie einer Person einen Kredit gewährt, die den Prognosen eines KI-Systems zufolge den Kredit zurückzahlen wird, ihren Zahlungsverpflichtungen dann aber nicht nachkommt, größer als der Gewinn aus diesem Kredit. Eine Person, die der Diagnose zufolge frei von einer bestimmten Krankheit ist, diese Krankheit aber hat, wird möglicherweise schwer leiden. Der Ausgleich von echt-positiven und echt-negativen Ergebnissen kann ebenfalls zu unerwünschten Ergebnissen führen: So könnte es z. B. geschehen, dass Frauen inhaftiert bleiben, die keine Sicherheitsbedrohung darstellen, weil gewährleistet werden soll, dass ein gleicher Anteil an Männern und Frauen aus der Haft entlassen wird (Berk, R. und J. Hyatt, 2015[23]). Einige Ansätze zielen darauf ab, sowohl falsch-positive als auch falsch-negative Ergebnisse auszugleichen. Allerdings ist es schwierig, gleichzeitig verschiedenen Konzepten von Fairness gerecht zu werden (Chouldechova, 2016[24]).

Politikverantwortliche sollten über den angemessenen Umgang mit sensiblen Daten im KI-Kontext nachdenken

Es könnte angebracht sein, erneut darüber nachzudenken, wie mit sensiblen Daten umzugehen ist. In einigen Fällen müssen Unternehmen möglicherweise sensible Daten speichern und verwenden, um sicherzustellen, dass ihre Algorithmen diese Daten nicht unbeabsichtigt rekonstruieren. Eine weitere Priorität der Politik ist die Überwachung unbeabsichtigter Rückkopplungseffekte. Wenn sich die Polizei beispielsweise in Viertel begibt, die von Algorithmen als Orte mit hohem Kriminalitätsrisiko eingestuft wurden, könnte dies zu einer verzerrten Datenerhebung führen und in der Folge die Voreingenommenheit des Algorithmus – und der Gesellschaft – gegenüber diesen Vierteln verstärken (O'Neil, 2016[25]).

Transparenz und Nachvollziehbarkeit

Transparenz ist beim Einsatz von KI und bei der Funktionsweise von KI-Systemen von entscheidender Bedeutung

Der Begriff „Transparenz" hat technisch und politisch gesehen nicht die gleiche Bedeutung. In der Politik bezieht sich Transparenz traditionell darauf, wie eine Entscheidung getroffen wird, wer an diesem Prozess beteiligt ist und welche Faktoren in die Entscheidung einfließen (Kosack, S. und A. Fung, 2014[26]). Transparenzmaßnahmen können hier darin

bestehen, offenzulegen, wie KI für Prognosen, Empfehlungen oder Entscheidungen eingesetzt wird. Gegebenenfalls könnte der Nutzer auch darauf aufmerksam gemacht werden, wenn er mit einem KI-System interagiert.

Für die Technologieexperten geht es bei der Transparenz eines KI-Systems in erster Linie um prozessbezogene Fragen. Menschen sollen verstehen können, wie ein KI-System entwickelt, trainiert und eingesetzt wird. Das kann auch bedeuten, dass Einblick in die Faktoren gegeben wird, die eine bestimmte Prognose oder Entscheidung beeinflussen. Um eine gemeinsame Nutzung von bestimmtem Code oder bestimmten Datensätzen geht es dabei in der Regel nicht. In vielen Fällen sind die Systeme zu komplex, als dass dies wirklich Transparenz schaffen könnte (Wachter, S., B. Mittelstadt und C. Russell, 2017[27]). Darüber hinaus könnten durch die gemeinsame Nutzung von Code oder Datensätzen Geschäftsgeheimnisse oder sensible Nutzerdaten preisgegeben werden.

Generell wird es als wichtig erachtet, das Bewusstsein für die in der KI verwendeten Schlussfolgerungsprozesse zu schärfen und deren Verständnis zu fördern, damit diese Technologien allgemein akzeptiert werden und allen Nutzen bringen.

Transparenzansätze in KI-Systemen

Experten der Working Group on Explanation and the Law des Berkman Klein Center der Harvard-Universität haben drei Ansätze zur Erhöhung der Transparenz von KI-Systemen aufgezeigt und festgestellt, dass es bei jedem dieser Ansätze zu Zielkonflikten kommen kann (Doshi-Velez et al., 2017[28]). Zusätzlich zu diesen drei Ansätzen gibt es noch das Konzept der Optimierungstransparenz, d. h. der Transparenz in Bezug auf die Ziele eines KI-Systems und die in Verbindung mit diesen Zielen erhaltenen Ergebnisse. Die drei in Harvard untersuchten Ansätze sind: a) theoretische Garantien, b) empirische Evidenz und c) Erklärung (Tabelle 4.1).

Tabelle 4.1. Ansätze zur Erhöhung der Transparenz und Rechenschaftspflicht von KI-Systemen

Ansatz	Beschreibung	Gut geeignete Kontexte	Schlecht geeignete Kontexte
Theoretische Garantien	In einigen Situationen ist es möglich, theoretische Garantien für ein KI-System zu geben, die durch Beweise gestützt sind.	Die Umgebung ist vollständig beobachtbar (ein Beispiel ist das Go-Spiel) und sowohl Problem als auch Lösung können formalisiert werden.	Die Situation kann nicht eindeutig beschrieben werden (wie die meisten realen Situationen).
Statistische Evidenz/ Wahrscheinlichkeit	Anhand von empirischer Evidenz wird die Gesamtleistung eines Systems gemessen und der durch dieses System entstehende Nutzen oder Schaden aufgezeigt, ohne bestimmte Entscheidungen zu erklären.	Ergebnisse können vollständig formalisiert werden; es ist möglich, negative Ergebnisse abzuwarten, um sie zu messen; Probleme sind u. U. nur aggregiert sichtbar.	Das Ziel kann nicht vollständig formalisiert werden; es ist möglich, für eine bestimmte Entscheidung die Verantwortlichkeiten festzulegen (Schuld oder Unschuld).
Erklärung/ Nachvollziehbarkeit	Menschen können Informationen über die Logik interpretieren, nach der ein System mit einem bestimmten Satz von Eingangsdaten zu einer bestimmten Schlussfolgerung gelangt ist.	Probleme sind nicht vollständig spezifiziert, Ziele sind nicht klar und Eingangsdaten könnten falsch sein.	Andere Formen der Rechenschaftslegung sind möglich.

Quelle: Nach Doshi-Velez et al. (2017[28]), „Accountability of AI under the law: The role of explanation“, https://arxiv.org/pdf/1711.01134.pdf.

Einige Systeme bieten theoretische Garantien für die Einhaltung bestimmter Grenzen

In einigen Fällen ist es möglich, **theoretische Garantien** zu geben, dass ein System nachweislich innerhalb enger Grenzen (*constraints*) operiert. Dies ist in Situationen möglich, in denen die Umgebung vollständig beobachtbar ist und sowohl das Problem als auch die Lösung vollständig formalisiert werden können, wie z. B. beim Go-Spiel. In solchen Fällen können bestimmte Ergebnisse nicht eintreten, selbst wenn ein KI-System neue Arten von Daten verarbeitet. Beispielsweise könnte ein System entwickelt werden, das vereinbarte Prozesse für Abstimmungen und Stimmauszählungen nachweislich einhält. In diesem Fall ist möglicherweise keine Erklärung oder Evidenz erforderlich: Das System muss nicht erklären, wie es zu einem bestimmten Ergebnis gekommen ist, weil die Arten von Ergebnissen, die Anlass zur Sorge geben, mathematisch unmöglich sind. Es kann bereits frühzeitig geprüft werden, ob die festgelegten einschränkenden Bedingungen ausreichend sind.

In einigen Fällen kann statistische Evidenz für die Gesamtleistung eines Systems erbracht werden

In einigen Fällen kann es ausreichend sein, sich auf **statistische Evidenz** für die Gesamtleistung eines Systems zu verlassen. Evidenz dafür, dass ein KI-System einen bestimmten Nutzen oder Schaden für die Gesellschaft insgesamt oder für den Einzelnen messbar erhöht, kann zur Erfüllung der Rechenschaftspflicht ausreichen. Zum Beispiel könnte statistisch belegt werden, dass es mit einem autonomen Flugzeuglandesystem zu weniger sicherheitsrelevanten Zwischenfällen kommt als mit menschlichen Piloten oder dass ein klinisches Instrument zur Diagnoseunterstützung die Sterblichkeit verringern kann. Statistische Evidenz könnte ein geeigneter Rechenschaftsmechanismus für viele KI-Systeme sein, weil bei ihrer Verwendung Geschäftsgeheimnisse geschützt werden können und weil sie es gestattet, häufig auftretende, aber wenig risikoträchtige Schäden zu erkennen, die erst aggregiert sichtbar werden (Barocas, S. und A. Selbst, 2016[29]; Crawford, 2016[30]). Bias- oder Diskriminierungsfragen können statistisch festgestellt werden: Ein Kreditgenehmigungssystem würde z. B. ein Bias aufweisen, falls es (bei Ausklammerung anderer Faktoren) mehr Kredite für Männer als für Frauen genehmigt. Die zulässige Fehlerrate und die tolerierte Unsicherheit sind je nach Anwendung unterschiedlich. Die für eine Übersetzungssoftware als akzeptabel erachtete Fehlerquote könnte für autonomes Fahren oder medizinische Untersuchungen z. B. nicht akzeptabel sein.

Optimierungstransparenz ist die Transparenz in Bezug auf die Ziele und Ergebnisse eines Systems

Ein anderer Ansatz zur Erhöhung der Transparenz von KI-Systemen besteht darin, den Fokus von den Instrumenten des Systems auf seine Ziele zu verlagern: Es geht dann nicht mehr darum, die Nachvollziehbarkeit der Funktionsweise eines Systems zu fordern, sondern seine Ergebnisse zu messen – d. h. das, wofür das System „optimiert" wurde. Dies setzt voraus, dass erklärt wird, wofür das betreffende KI-System optimiert wurde, wobei zu berücksichtigen ist, dass Optimierungen unvollkommen sind, mit Zielkonflikten verbunden sind und „kritischen Erfordernissen" wie dem der Sicherheit und der Fairness unterliegen sollten. Nach diesem Ansatz sollen KI-Systeme dafür eingesetzt werden, wofür sie optimiert wurden. Dabei wird gegebenenfalls auf bestehende Ethik- und Rechtsvorschriften sowie auf gesellschaftliche Diskussionen und politische Prozesse Bezug genommen, um klarzustellen, wofür die KI-Systeme optimiert werden sollen (Weinberger, 2018[1]).

Das Konzept der Nachvollziehbarkeit bezieht sich auf ein bestimmtes Ergebnis eines KI-Systems

Nachvollziehbarkeit ist in Situationen unerlässlich, in denen konkret die Verantwortung für ein bestimmtes Ereignis zugewiesen werden muss – Situationen also, wie sie immer häufiger auftreten dürften, wenn KI-Systeme eingesetzt werden, um Empfehlungen abzugeben oder Entscheidungen zu treffen, die derzeit noch im menschlichen Ermessen liegen (Burgess, 2016[31]). Die DSGVO schreibt vor, dass die betroffenen Personen aussagekräftige Informationen über die zugrunde liegende Logik, die Bedeutung und voraussichtlichen Konsequenzen automatisierter Entscheidungssysteme erhalten. In der Regel muss die Nachvollziehbarkeit nicht für den gesamten Entscheidungsprozess des Systems gewährleistet sein. Meistens reicht die Beantwortung einer der folgenden Fragen aus (Doshi-Velez et al., 2017[28]):

1. **Hauptfaktoren für eine Entscheidung**: Bei vielen Arten von Entscheidungen, z. B. bei Sorgerechtsstreitigkeiten, Kreditprüfungen und Untersuchungshaftentlassungen, müssen eine Reihe von Faktoren berücksichtigt werden (oder ist es im Gegenteil ausdrücklich verboten, bestimmte Faktoren zu berücksichtigen). Mit einer Auflistung der Faktoren, die für eine KI-Prognose wichtig waren – idealerweise in der Reihenfolge ihrer Bedeutung –, kann sichergestellt werden, dass die richtigen Faktoren berücksichtigt wurden.
2. **Maßgebliche Faktoren, d. h. Faktoren, die das Ergebnis entscheidend beeinflussen**: Manchmal ist es wichtig zu wissen, ob ein bestimmter Faktor das Ergebnis beeinflusst hat. Durch die Änderung einer bestimmten Variablen, z. B. der ethnischen Herkunft bei der Vergabe von Studienplätzen, kann aufgezeigt werden, ob der jeweilige Faktor richtig eingesetzt wurde.
3. **Warum kam es in zwei scheinbar ähnlichen Fällen zu unterschiedlichen Ergebnissen oder umgekehrt warum kam es in zwei unterschiedlichen Fällen zu einem gleichen Ergebnis?** Es ist möglich, die Konsistenz und Integrität von KI-basierten Prognosen zu bewerten. Beispielsweise sollte das Einkommen bei der Entscheidung über eine Darlehensgewährung berücksichtigt werden, aber es sollte in ansonsten ähnlichen Fällen nicht mal entscheidend und mal unerheblich sein.

Im Bereich der Nachvollziehbarkeit wird aktiv geforscht, dies ist aber mit Kosten und möglichen Zielkonflikten verbunden

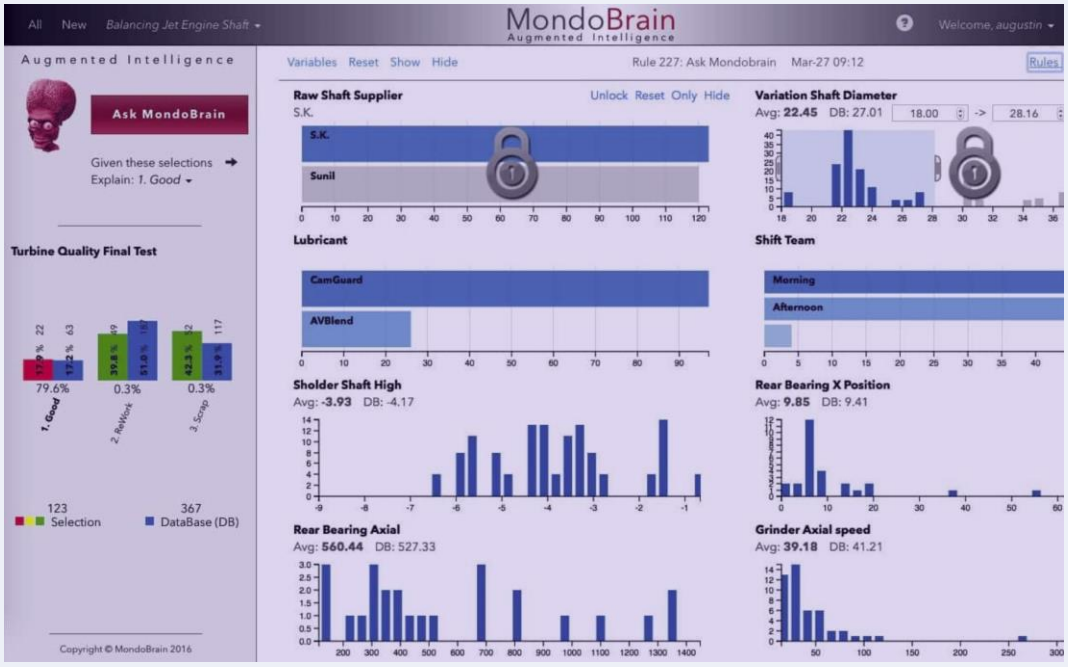
Einzelne Unternehmen, Normungsgremien, gemeinnützige Organisationen und öffentliche Einrichtungen betreiben technische Forschung, um KI-Systeme zu schaffen, die ihre Prognosen erklären können. Unternehmen in stark regulierten Bereichen wie der Finanzbranche, dem Gesundheitswesen und dem Personalwesen sind besonders aktiv bemüht, potenzielle finanzielle, rechtliche und Reputationsrisiken anzugehen, die sich aus den Prognosen von KI-Systemen ergeben. Zum Beispiel hat die US-amerikanische Bank Capital One 2016 ein Forschungsteam gegründet, um Wege zu finden, die Nachvollziehbarkeit von KI-Techniken zu verbessern (Knight, 2017[32]). Unternehmen wie MondoBrain haben Benutzeroberflächen entworfen, die helfen, wichtige Faktoren zu erklären (Kasten 4.5). Gemeinnützige Organisationen wie OpenAI erforschen Ansätze zur Entwicklung erklärbarer KI und zur Prüfung von KI-Entscheidungen. Es wird auch öffentlich finanzierte Forschung betrieben. Die Defense Advanced Research Projects Agency (DARPA) finanziert beispielsweise 13 verschiedene Forschungsgruppen, die an einer Reihe von Ansätzen arbeiten, um die Erklärbarkeit von KI zu verbessern.

Kasten 4.5. Nachvollziehbarkeitsprobleme durch besser gestaltete Benutzeroberflächen angehen

Einige Unternehmen haben begonnen, Nachvollziehbarkeit in ihre Lösungen einzubauen, damit Benutzer die im Hintergrund ablaufenden KI-Prozesse besser verstehen. Eines dieser Unternehmen ist MondoBrain. Das in Frankreich ansässige Unternehmen kombiniert menschliche, kollektive und künstliche Intelligenz, um Unternehmen eine Augmented-Reality-Lösung anzubieten. Mit interaktiven Dashboards zur Datenvisualisierung werden alle in einem Unternehmen vorhandenen Daten ausgewertet (z. B. Daten aus der Enterprise-Resource-Planning-, Business-Programme-Management- oder Customer-Relationship-Management-Software) und auf der Basis von Kundenanfragen präskriptive Empfehlungen abgegeben (Abbildung 4.1). Mit einem besonderen ML-Algorithmus werden dabei betriebswirtschaftliche Variablen ausgeklammert, die für die Abfrage uninteressant sind, und die Variablen mit dem größten Effekt extrahiert.

Eine einfache Ampelsymbolik führt den Benutzer durch alle Schritte der Abfrage und erleichtert ihm so das Verständnis des Entscheidungsprozesses. Jede einzelne Entscheidung wird automatisch dokumentiert, wodurch sie überprüfbar und nachvollziehbar wird. Damit werden alle Schritte, die zur endgültigen Geschäftsempfehlung führten, vollständig, aber einfach dargestellt.

Abbildung 4.1. Datenvisualisierungsinstrumente zur Verbesserung der Erklärbarkeit



Quelle: www.mondobrain.com.

In vielen Fällen ist es möglich, eine oder mehrere Erklärungen zu den Ergebnissen von KI-Systemen zu erstellen. Solche Erklärungen sind jedoch mit Kosten verbunden. Die Konzeption eines Systems, dessen Entscheidungen nachvollziehbar sind, kann komplex und teuer sein. Für alle Systeme Nachvollziehbarkeit zu fordern, ist angesichts des unterschiedlichen Zwecks der verschiedenen Systeme möglicherweise nicht angemessen und

kann insbesondere KMU benachteiligen. KI-Systeme müssen oft ex ante konzipiert werden, um eine bestimmte Art von Erklärung zu liefern. Die nachträgliche Suche nach Erklärungen erfordert in der Regel zusätzliche Arbeit; möglicherweise muss das gesamte Entscheidungssystem neu gestaltet werden. So kann ein KI-System beispielsweise nicht alle wichtigen Faktoren erklären, die sich auf ein Ergebnis ausgewirkt haben, wenn es aufgrund seines Designs nur auf die Erklärung eines dieser Faktoren ausgelegt ist. Ein KI-System zur Erkennung von Herzerkrankungen kann z. B. nicht nach dem Einfluss des Geschlechts auf eine Diagnose abgefragt werden, wenn es nicht mit Geschlechtsdaten trainiert wurde. Dies ist selbst dann der Fall, wenn das KI-System das Geschlecht tatsächlich über Ersatzvariablen berücksichtigt, wie z. B. andere Erkrankungen, die bei Frauen häufiger auftreten als bei Männern.

In einigen Fällen gibt es einen Zielkonflikt zwischen Nachvollziehbarkeit und Genauigkeit. Damit sie erklärbar sind, müssen die Lösungsvariablen möglicherweise auf eine hinreichend kleine Zahl reduziert werden, um von Menschen verstanden zu werden. Dies könnte bei komplexen, hochdimensionalen Problemen suboptimal sein. Einige ML-Modelle, die in der medizinischen Diagnostik verwendet werden, können beispielsweise die Wahrscheinlichkeit einer Krankheit genau vorhersagen, sind aber zu komplex, als dass sie von Menschen verstanden werden könnten. In solchen Fällen sollte der Schaden, der durch ein weniger präzises System entstehen könnte, das klare Erklärungen bietet, gegen den Schaden abgewogen werden, den ein präziseres System verursachen würde, bei dem Fehler schwieriger zu erkennen sind. Zum Beispiel kann die Rückfallprognose einfache, erklärbare Modelle erfordern, bei denen Fehler erkennbar sind (Dressel, J. und H. Farid, 2018[33]). In Bereichen wie der Klimavorhersage hingegen werden komplexere Modelle, die zwar bessere Prognosen liefern, aber weniger erklärbar sind, u. U. eher akzeptiert. Dies ist insbesondere dann der Fall, wenn es andere Mechanismen zur Rechenschaftslegung gibt, z. B. statistische Daten zur Erkennung möglicher Verzerrungen oder Fehler.

Robustheit und Sicherheit

Was unter Robustheit und Sicherheit zu verstehen ist

Robustheit kann als die Fähigkeit verstanden werden, widrigen Bedingungen, einschließlich digitaler Sicherheitsrisiken, zu widerstehen oder sie zu überwinden (OECD, 2019[34]). Als sichere KI-Systeme gelten Systeme, die bei normalem oder vorhersehbarem Gebrauch oder Missbrauch während ihres gesamten Lebenszyklus keine unvertretbaren Sicherheitsrisiken darstellen (OECD, 2019[35]). Die Fragen der Robustheit und Sicherheit von KI-Systemen sind eng miteinander verknüpft. Die digitale Sicherheit kann sich beispielsweise auf die Produktsicherheit auswirken, wenn vernetzte Produkte wie selbstfahrende Fahrzeuge oder KI-fähige Haushaltsgeräte nicht ausreichend sicher sind; Hacker könnten die Kontrolle über sie übernehmen und Einstellungen aus der Ferne ändern.

Risikomanagement in KI-Systemen

Die Bestimmung des erforderlichen Sicherheitsniveaus sollte auf einer Nutzen-Risiko-Abwägung beruhen

Der Schaden, der durch ein KI-System verursacht werden könnte, sollte gegen die Kosten für die Integration von Transparenz und Rechenschaftspflicht in KI-Systeme abgewogen werden. Dabei geht es beispielsweise um Risiken in Bezug auf Menschenrechte, Datenschutz, Fairness und Robustheit. Allerdings birgt nicht jede Nutzung von KI die gleichen

Risiken, und die Forderung nach Nachvollziehbarkeit ist z. B. auch mit Kosten verbunden. Im Risikomanagement scheint ein breiter Konsens darüber zu bestehen, dass Kontexte, in denen viel auf dem Spiel steht, ein höheres Maß an Transparenz und Verantwortlichkeit erfordern, insbesondere wenn es um das Leben oder die Freiheit des Einzelnen geht.

Anwendung von Risikomanagementkonzepten während des gesamten Lebenszyklus von KI-Systemen

Unternehmen nutzen Konzepte des Risikomanagements, um potenzielle Risiken, die das Verhalten und die Ergebnisse eines Systems negativ beeinflussen können, zu erkennen, zu bewerten, zu priorisieren und anzugehen. Solche Konzepte können auch genutzt werden, um Risiken für verschiedene Akteure zu bestimmen und um festzulegen, wie diesen Risiken während des gesamten Lebenszyklus des KI-Systems begegnet werden kann (vgl. Abschnitt „Lebenszyklus eines KI-Systems“ in Kapitel 1).

KI-Akteure – d. h. Personen, die eine aktive Rolle im Lebenszyklus eines KI-Systems spielen – bewerten und mindern Risiken in diesem System als Ganzem sowie in jeder Phase seines Lebenszyklus. Das Risikomanagement in KI-Systemen untergliedert sich in folgende Schritte, deren Bedeutung je nach Lebenszyklusphase des KI-Systems variiert:

1. **Ziele**: Definition von Zielen, Funktionen oder Eigenschaften des KI-Systems im Kontext. Diese Funktionen und Eigenschaften können sich je nach Phase des KI-Lebenszyklus ändern.
2. **Beteiligte bzw. betroffene Akteure**: Bestimmung der Akteure, die in den verschiedenen Phasen des Lebenszyklus des Systems direkt oder indirekt von dessen Funktionen oder Eigenschaften betroffen sind.
3. **Risikobewertung**: Bewertung der potenziellen Auswirkungen (Vorteile und Risiken) für die beteiligten bzw. betroffenen Akteure. Diese hängen von den Akteuren selbst sowie von der Lebenszyklusphase ab, die das KI-System erreicht hat.
4. **Risikominderung**: Bestimmung von Strategien zur Risikominderung, die dem Risiko angemessen sind und im Verhältnis zu ihm stehen. Dabei sollten Faktoren wie die Ziele des Unternehmens, die beteiligten bzw. betroffenen Akteure, die Eintrittswahrscheinlichkeit der Risiken und der potenzielle Nutzen berücksichtigt werden.
5. **Umsetzung**: Umsetzung von Strategien zur Risikominderung.
6. **Überwachung, Bewertung und Rückmeldung**: Überwachung, Bewertung und Rückmeldung zu den Ergebnissen der Umsetzung.

Der Einsatz von Risikomanagementkonzepten im Lebenszyklus von KI-Systemen und die Dokumentation der Entscheidungen in jeder Lebenszyklusphase können dazu beitragen, die Transparenz von KI-Systemen und die Rechenschaftspflicht der Unternehmen für diese Systeme zu verbessern.

Das Ausmaß des Gesamtschadens und der unmittelbare Risikokontext sollten nebeneinander betrachtet werden

Isoliert betrachtet stellen einige Anwendungen von KI-Systemen ein geringes Risiko dar. Aufgrund ihrer gesellschaftlichen Auswirkungen können sie jedoch ein höheres Maß an Robustheit erfordern. Ein System, durch dessen Betrieb einer großen Zahl von Menschen ein leichter Schaden entsteht, könnte insgesamt einen erheblichen Schaden verursachen.

Dies könnte beispielsweise geschehen, wenn eine kleine Zahl von KI-Tools in mehrere Dienste integriert ist und in verschiedenen Branchen genutzt wird, etwa für Kreditanträge, den Abschluss von Versicherungsverträgen oder Hintergrundprüfungen. Ein einziger Fehler oder eine einzige Verzerrung in einem System könnte dann zu einer ganzen Kaskade von Ablehnungen führen (Citron, D. und F. Pasquale, 2014[36]). Für sich genommen dürften die einzelnen Ablehnungen kaum von Bedeutung sein. Zusammengenommen könnten sie jedoch einen disruptiven Effekt haben. In Politikdiskussionen sollte daher neben dem unmittelbaren Risikokontext auch das Ausmaß des Gesamtschadens berücksichtigt werden.

Robustheit gegenüber den mit KI verbundenen digitalen Sicherheitsrisiken

KI ermöglicht raffiniertere Angriffe potenziell größeren Ausmaßes

Je kostengünstiger und je leichter einsetzbar KI wird, umso mehr dürfte neben ihrer Nutzung zur Verbesserung der digitalen Sicherheit auch ihre Nutzung zu böswilligen Zwecken zunehmen (vgl. Unterabschnitt „KI und digitale Sicherheit“ in Kapitel 3). Cyberkriminelle arbeiten daran, ihre KI-Fähigkeiten auszubauen. Schnellere und raffiniertere Angriffe stellen eine zunehmende Gefahr für die digitale Sicherheit dar.[6] Vor diesem Hintergrund weiten sich bestehende Bedrohungen aus, während neue Bedrohungen entstehen und sich die Art der Bedrohungen selbst verändert.

Heutige KI-Systeme weisen eine Reihe von Schwachstellen auf. Böswillige Akteure können z. B. die Daten manipulieren, mit denen ein KI-System trainiert wird („Datenvergiftung“). Sie können auch die Eigenschaften identifizieren, die in einem digitalen Sicherheitsmodell zur Erkennung von Malware verwendet werden. Mit diesen Informationen können sie nicht erkennbaren bösartigen Code entwickeln oder Informationen absichtlich falsch klassifizieren (Adversarial Examples bzw. „feindliche Beispiele“) (Kasten 4.6) (Brundage et al., 2018[37]). Mit der zunehmenden Verfügbarkeit von KI-Technologien können immer mehr Menschen KI nutzen, um raffiniertere Angriffe potenziell größeren Ausmaßes durchzuführen. Die Häufigkeit und die Effizienz arbeitsintensiver digitaler Sicherheitsangriffe wie z. B. gezieltes Spear-Phishing könnten mit deren Automatisierung durch Algorithmen des maschinellen Lernens zunehmen.

Kasten 4.6. Die Gefahr von Adversarial Examples für maschinelles Lernen

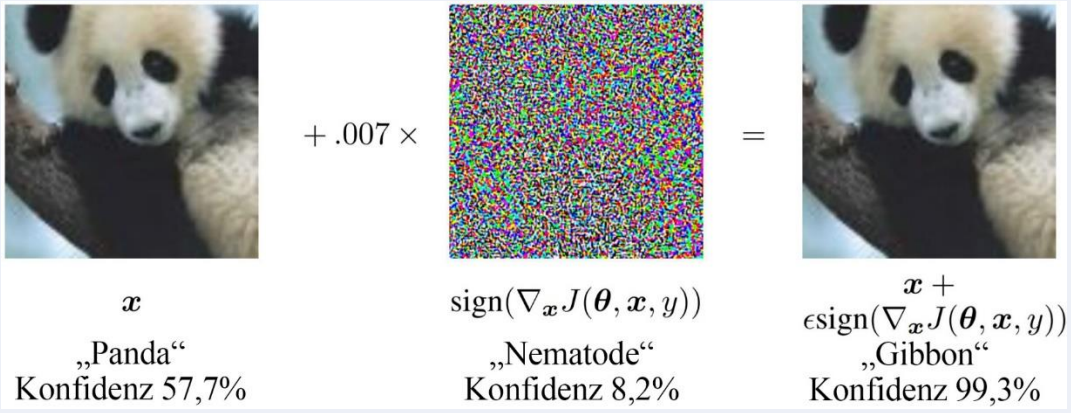
Adversarial Examples bzw. „feindliche Beispiele“ sind Informationen, die von Angreifern absichtlich in ML-Modelle eingegeben werden, damit das Modell einen Fehler macht, dabei aber zugleich ein hohes Konfidenzniveau anzeigt. Feindliche Beispiele sind ein echtes Problem für die Robustheit und Sicherheit von KI-Systemen, da verschiedene ML-Modelle, darunter auch dem neuesten Stand der Technik entsprechende neuronale Netze, für sie anfällig sind.

Diese feindlichen Beispiele können sehr subtil funktionieren. In Abbildung 4.2 wurde dem Bild eines Pandas eine unmerklich kleine Veränderung – ein „feindlicher Input“ – hinzugefügt. Dieser Input wurde speziell entwickelt, um das Bildklassifikationsmodell zu täuschen. Der Algorithmus klassifiziert den Panda daraufhin mit einem Konfidenzniveau von nahezu 100 % als Gibbon.

Neuere Forschungsarbeiten haben zudem gezeigt, dass sich feindliche Beispiele erstellen lassen, indem ein Bild auf Normalpapier gedruckt und mit einem Smartphone mit Standardauflösung abfotografiert wird. Solche Bilder könnten gefährlich sein: Ein gezielt

aufgebrachter Aufkleber auf einem Stoppschild könnte dazu führen, dass ein selbstfahrendes Auto es als Vorfahrts- oder anderes Schild interpretiert.

Abbildung 4.2. Durch eine kleine Veränderung wird ein Algorithmus so getäuscht, dass er einen Panda als Gibbon klassifiziert



Quelle: Goodfellow, Shlens und Szegedy (2015[38]), „Explaining and harnessing adversarial examples“, https://arxiv.org/pdf/1412.6572.pdf; Kurakin, Goodfellow und Bengio (2017[39]), „Adversarial examples in the physical world“, https://arxiv.org/abs/1607.02533.

Sicherheit

Lernende und autonome KI-Systeme haben sicherheitsrechtliche Konsequenzen

Das Spektrum von Produkten mit integrierter künstlicher Intelligenz nimmt rasant zu – von der Robotik und selbstfahrenden Fahrzeugen bis hin zu gängigen Konsumgütern und verbraucherorientierten Dienstleistungen, wie z. B. intelligenten Haushaltsgeräten und Haussicherheitssystemen. KI-Produkte bieten erhebliche Sicherheitsvorteile, lassen im Bereich der Produktsicherheitsbestimmungen aber auch neue praktische und rechtliche Herausforderungen entstehen (OECD, 2017[20]). Sicherheitsbestimmungen sind in der Regel eher auf „fertige“ Hardware-Produkte als auf Software ausgerichtet, zudem lernen viele KI-Software-Produkte während ihres gesamten Lebenszyklus und entwickeln sich weiter.[7] KI-Produkte können auch „autonom“ oder „halbautonom“ sein, d. h. Entscheidungen ohne oder mit wenig menschlichem Input treffen und ausführen.

Verschiedene Arten von KI-Anwendungen dürften unterschiedliche politische Maßnahmen erfordern (Freeman, 2017[40]). Ganz allgemein sind für KI-Systeme vier Erwägungen von Bedeutung. Erstens muss geklärt werden, wie am besten gewährleistet werden kann, dass Produkte sicher sind bzw. – anders ausgedrückt – während ihres gesamten Lebenszyklus bei normalem oder vorhersehbarem Gebrauch oder Missbrauch kein unvertretbares Sicherheitsrisiko darstellen. Dies gilt auch für Fälle, in denen nur wenige Daten vorliegen, mit denen das System trainiert werden kann (Kasten 4.7). Zweitens stellt sich die Frage, wer in welchem Umfang für Schäden, die durch ein KI-System verursacht wurden, haftbar gemacht werden sollte. Gleichzeitig ist zu überlegen, welche Akteure zur Sicherheit von autonomen Maschinen beitragen können. Dabei könnte es sich um Nutzer, Produkt- und Sensorhersteller, Softwarehersteller, Designer, Infrastrukturanbieter und Datenanalysefirmen handeln. Drittens ist es wichtig, über die Art(en) der Haftung nachzudenken – ob es sich um eine verschuldensunabhängige oder verschuldensabhängige Haftung handeln soll

und welche Rolle die Versicherung spielen soll. Die mangelnde Transparenz einiger KI-Systeme verschärft das Problem der Haftung. Viertens müssen politische Entscheidungsträger überlegen, wie geltendes Recht durchgesetzt werden kann, was ein „Fehler" in einem KI-Produkt ist, wie die Beweislast verteilt wird und welche Rechtsmittel zur Verfügung stehen.

Kasten 4.7. Synthetische Daten für eine sicherere und genauere KI: autonome Fahrzeuge

Im Bereich des maschinellen Lernens werden zunehmend synthetische Daten verwendet, da sie die Simulation von Szenarien ermöglichen, die sich unter realen Bedingungen nur schwer beobachten oder reproduzieren lassen. Philipp Slusallek, wissenschaftlicher Direktor am Deutschen Forschungszentrum für Künstliche Intelligenz, nennt als Beispiel hierfür den Fall eines Kindes, das über die Straße rennt: Es muss sichergestellt werden, dass autonome Fahrzeuge in einer solchen Situation richtig reagieren.

Eine „digitale Realität" – d. h. eine simulierte Umgebung, die die relevanten Merkmale der realen Welt nachbildet – könnte vier Dinge ermöglichen. Erstens könnte sie synthetische Eingabedaten generieren, aus denen KI-Systeme lernen können, mit komplexen Situationen umzugehen. Zweitens könnte sie Ergebnisse validieren und synthetische Daten gegenüber realen Daten neu kalibrieren. Drittens könnte sie für die Durchführung von Tests, z. B. eine Führerscheinprüfung für autonome Fahrzeuge, verwendet werden. Viertens bietet sie die Möglichkeit, den Entscheidungsprozess des Systems und die möglichen Effekte alternativer Entscheidungen zu untersuchen. Dank eines solchen Ansatzes konnte Google z. B. seine selbstfahrenden Fahrzeuge mit mehr als 4,8 Millionen simulierten Kilometern pro Tag trainieren (dies entspricht mehr als 500 Hin- und Rückfahrten zwischen New York City und Los Angeles).

Quelle: Golson (2016[41]), „Google's self-driving cars rack up 3 million simulated miles every day", https://www.theverge.com/2016/2/1/10892020/google-self-driving-simulator-3-million-miles; Slusallek (2018[42]), *Artificial Intelligence and Digital Reality: Do We Need a CERN for AI?*, https://www.oecd-forum.org/channels/722-digitalisation/posts/28452-artificial-intelligence-and-digital-reality-do-we-need-a-cern-for-ai.

Die Produkthaftungsrichtlinie der Europäischen Union (Richtlinie 85/374/EWG) von 1985 legt das Prinzip der „verschuldensunabhängigen Haftung" (Gefährdungshaftung) fest. Wenn ein fehlerhaftes Produkt einem Verbraucher Schaden zufügt, haftet der Hersteller nach diesem Prinzip auch, wenn keine Fahrlässigkeit oder Verschulden vorliegt. Diese Richtlinie wird gegenwärtig von der Europäischen Kommission überarbeitet. Den ersten Schlussfolgerungen zufolge ist das derzeitige Modell im Großen und Ganzen angemessen (Ingels, 2017[43]). Aktuelle und in Zukunft zu erwartende KI-Technologien stellen allerdings die Konzepte von „Produkt", „Sicherheit", „Fehler" und „Schaden" infrage, was die Last der Beweisführung schwieriger macht.

Im Bereich des autonomen Fahrens ist Sicherheit ein sehr wichtiges Thema für die Politik. Beispielsweise muss festgelegt werden, wie autonome Fahrzeuge getestet werden können, damit gewährleistet ist, dass ihr Einsatz sicher ist. Dies beinhaltet z. B. Zulassungsregelungen, die die Möglichkeit der Vorerprobung von AF-Systemen vorsehen, oder Anforderungen an die Systeme, die die Wachsamkeit menschlicher Fahrer überwachen müssen, wenn diese als Rückfallebene dienen. In einigen Fällen stellt die Zulassung für Firmen, die Fahrzeuge testen möchten, ein echtes Problem dar. Zudem stehen staatliche Stellen solchen Tests mehr oder weniger offen gegenüber. Von verschiedener Seite wurden Forderungen nach einer verschuldensunabhängigen Haftung von Herstellern autonomer Fahrzeuge laut. Eine solche Haftung würde auf der Kontrollierbarkeit des Risikos beruhen.

Damit würde beispielsweise anerkannt, dass einen Mitfahrer in einem selbstfahrenden Fahrzeug kein Verschulden treffen kann oder er nicht gegen eine Sorgfaltspflicht verstoßen kann. Juristen zufolge könnte selbst das Konzept des Fahrzeughalters hier nicht greifen, weil der Halter in der Lage sein muss, das Risiko zu kontrollieren (Borges, 2017[44]). Einige schlagen vor, dass Versicherungen das Risiko von durch autonome Fahrzeuge verursachten Schäden übernehmen könnten. Grundlage dafür könnte eine auf Risikobewertungen beruhende Klassifikation der zugelassenen autonomen Fahrzeuge sein.

Möglicherweise müssen die Standards für sichere Arbeitsbedingungen aktualisiert werden

Zu den direkten Auswirkungen von KI auf die Arbeitsbedingungen kann auch die Notwendigkeit neuer Sicherheitsprotokolle gehören. So wird immer deutlicher, dass neue bzw. überarbeitete Branchenstandards und Technologievereinbarungen zwischen Geschäftsleitung und Beschäftigten erforderlich sind, um zuverlässige, sichere und produktive Arbeitsplätze zu gewährleisten. Der Europäische Wirtschafts- und Sozialausschuss (EWSA) riet den „Interessenträgern, sich gemeinsam für komplementäre KI-Systeme und ihre Ko-Kreation am Arbeitsplatz einzusetzen“ (EWSA, 2017[45]).

Rechenschaftspflicht

Die zunehmende Nutzung von KI verlangt nach einer Rechenschaftspflicht für die Funktionsweise von KI-Systemen

Rechenschaftspflicht bedeutet im Wesentlichen, dass die Beweislast für die ordnungsgemäße Funktionsweise eines KI-Systems auf die entsprechenden Organisationen oder Personen übertragen werden kann. Zu den Verantwortlichkeitskriterien, auf denen sie beruht, gehören die Achtung menschlicher Werte sowie Fairness, Transparenz, Robustheit und Sicherheit. Die Rechenschaftspflicht hängt von der Rolle der einzelnen KI-Akteure, dem Kontext und dem Stand der Technik ab. Für politische Entscheidungsträger ist sie an Mechanismen geknüpft, die verschiedene Funktionen erfüllen. Diese Mechanismen bestimmen, welche Akteure für eine bestimmte Empfehlung oder Entscheidung verantwortlich sind. Sie korrigieren die Empfehlung oder Entscheidung, bevor sie umgesetzt wird. Sie könnten die Entscheidung auch nachträglich anfechten und sie könnten sogar das für die Entscheidung verantwortliche System anfechten (Helgason, 1997[46]).

In der Praxis hängt die Verantwortlichkeit von KI-Systemen häufig davon ab, wie gut ein bestimmtes System nach Indikatoren der Genauigkeit oder Effizienz abschneidet. Zunehmend werden dabei auch Indikatoren für die Ziele Fairness, Sicherheit und Robustheit herangezogen. Allerdings werden solche Indikatoren immer noch seltener eingesetzt als Messgrößen der Effizienz oder Genauigkeit. Wie bei allen Messgrößen können Monitoring und Evaluierung kostspielig sein. Daher müssen Art und Häufigkeit der Erhebungen in einem angemessenen Verhältnis zu den potenziellen Risiken und Vorteilen stehen.

Das erforderliche Maß an Rechenschaftslegung hängt vom Risikokontext ab

Welcher Politikansatz angemessen ist, hängt vom Kontext und vom Anwendungsfall ab. Beispielsweise können die Rechenschaftserwartungen bei der Nutzung von KI im öffentlichen Sektor höher sein. Dies gilt insbesondere für die Ausübung hoheitlicher Aufgaben, wie Sicherheit und Rechtsvollzug, bei denen ein erhebliches Schadenspotenzial besteht. Formale Rechenschaftsmechanismen sind häufig auch für privatwirtschaftliche Anwendungen in den stark regulierten Bereichen Verkehr, Finanzen und Gesundheitswesen

erforderlich. In Bereichen des privaten Sektors, die weniger stark reguliert sind, unterliegt der Einsatz von KI weniger formalen Rechenschaftsmechanismen. Technische Ansätze für Transparenz und Rechenschaftspflicht spielen in diesem Fall eine noch wichtigere Rolle. Sie müssen sicherstellen, dass die von privatwirtschaftlichen Akteuren konzipierten und betriebenen Systeme gesellschaftlichen Normen und rechtlichen Auflagen gerecht werden.

Einige Anwendungen oder Entscheidungen können das Eingreifen eines Menschen erfordern, der den sozialen Kontext und mögliche unbeabsichtigte Konsequenzen berücksichtigt. Wenn Entscheidungen erhebliche Auswirkungen auf das Leben von Menschen haben, besteht weitgehend Einigkeit darüber, dass sie nicht allein aufgrund von KI-basierten Ergebnissen (z. B. einem Score) getroffen werden sollten. Die DSGVO beispielsweise befürwortet in solchen Fällen ein menschliches Eingreifen. Menschen müssen z. B. informiert werden, wenn KI genutzt wird, um in einem strafrechtlichen Verfahren ein Urteil zu fällen, um Kreditvergabeentscheidungen zu treffen, Plätze in bestimmten Bildungsgängen zu vergeben oder Kandidaten für eine Stelle auszuwählen. Wenn viel auf dem Spiel steht, sind meist formale Rechenschaftsmechanismen erforderlich. Zum Beispiel ist ein Richter, der KI bei der Urteilsfällung in Strafverfahren nutzt, jemand, der direkt eingreifen kann („human-in-the-loop“). Andere Rechenschaftsmechanismen – darunter auch traditionelle gerichtliche Berufungsverfahren – helfen jedoch sicherzustellen, dass KI-Empfehlungen für Richter nur ein Kriterium unter anderen sind, die es zu berücksichtigen gilt (Wachter, S., B. Mittelstadt und L. Floridi, 2017[47]). In risikoarmen Kontexten, z. B. bei Restaurantempfehlungen, kann man sich möglicherweise ganz auf Maschinen verlassen. Ein mehrere Ebenen umfassender Ansatz, der unnötige Kosten nach sich ziehen kann, ist hier vermutlich nicht nötig.

Politikumfeld für künstliche Intelligenz

Es bedarf Maßnahmen auf nationaler Ebene zur Förderung vertrauenswürdiger KI-Systeme. Solche Maßnahmen können nutzbringende und faire Ergebnisse für die Menschen und den Planeten begünstigen, insbesondere in vielversprechenden Bereichen, in die nicht genügend marktorientierte Investitionen fließen. Zur Schaffung eines für vertrauenswürdige KI günstigen Politikumfelds gehört u. a., dass öffentliche und private Investitionen in KI-Forschung und -Entwicklung erleichtert und dass die Menschen mit den notwendigen Kompetenzen ausgestattet werden, um in einer sich wandelnden Arbeitswelt erfolgreich zu sein. In den folgenden Unterabschnitten werden verschiedene Politikbereiche erörtert, die für die Förderung und Entwicklung vertrauenswürdiger KI von entscheidender Bedeutung sind.

In KI-Forschung und -Entwicklung investieren

Langfristige Investitionen in öffentliche Forschung können maßgeblichen Einfluss auf die Innovationstätigkeit im KI-Bereich haben

Die OECD befasst sich in ihren Arbeiten auch mit dem Einfluss innovationspolitischer Maßnahmen auf den digitalen Wandel und die Einführung von KI (OECD, 2018[48]). Dazu untersucht sie u. a. die Rolle der öffentlichen Forschungspolitik, des Wissenstransfers und der Ko-Kreation von Maßnahmen zur Förderung der Entwicklung von Forschungsinstrumenten und Forschungsinfrastruktur für KI. Die Politikverantwortlichen müssen prüfen, welches Niveau staatlichen Engagements in der KI-Forschung angemessen ist, um gesellschaftliche Herausforderungen zu bewältigen (OECD, 2018[13]). Außerdem werden For-

schungseinrichtungen in allen Bereichen leistungsfähige KI-Systeme benötigen, um wettbewerbsfähig zu bleiben, insbesondere in der Biomedizin und den Biowissenschaften. Neue Instrumente wie Plattformen zur gemeinsamen Nutzung von Daten und Superrechner können sich auf die Forschung im KI-Bereich förderlich auswirken und möglicherweise neue Investitionen erforderlich machen. Japan investiert z. B. jährlich mehr als 120 Mio. USD in den Aufbau einer Hochleistungsrecheninfrastruktur für Universitäten und öffentliche Forschungszentren.

KI gilt als Universaltechnologie mit potenziellen Auswirkungen auf eine große Zahl von Wirtschaftszweigen (Agrawal, A., J. Gans und A. Goldfarb, 2018[49]) (Brynjolfsson, E., D. Rock und C. Syverson, 2017[50]). Deshalb wurde KI auch schon als „Erfindung einer Methode der Erfindung" bezeichnet (Cockburn, I., R. Henderson und S. Stern, 2018[51]). Als solche wird sie von Wissenschaftlern und Erfindern bereits umfassend genutzt, um Innovationen zu erleichtern. Zudem könnten dank der bahnbrechenden technologischen Entwicklungen, die KI ermöglicht, ganz neue Wirtschaftszweige entstehen. Das zeigt, wie wichtig Grundlagenforschung und eine langfristig ausgerichtete Forschungspolitik sind (OECD, 2018[52]).

Ein digitales Ökosystem für KI fördern

KI-Technologien und -Infrastruktur

In den letzten Jahren wurden bei KI-Technologien erhebliche Fortschritte erzielt. Dies ist auf die zunehmende Reife statistischer Modellierungstechniken, z. B. neuronaler Netze und insbesondere tiefer neuronaler Netze (bekannt als Deep Learning), zurückzuführen. Viele der Tools zur Verwaltung und Nutzung von KI sind gemeinfreie Open-Source-Ressourcen. Dies erleichtert ihre Einführung und ermöglicht die Behebung von Softwarefehlern durch Crowdsourcing. Beispiele solcher Tools sind TensorFlow (Google), Michelangelo (Uber) und Cognitive Toolkit (Microsoft). Einige Unternehmen und Forscher geben auch kuratierte Trainingsdatensätze und Trainingswerkzeuge öffentlich weiter, um die Verbreitung von KI-Technologien zu unterstützen.

Die jüngsten Fortschritte im KI-Bereich sind z. T. dem exponentiellen Anstieg der Rechenleistung zu verdanken, wobei sich auch der Effekt des Mooreschen Gesetzes bemerkbar macht (das besagt, dass sich die Anzahl der Transistoren in einem integrierten Schaltkreis etwa alle zwei Jahre verdoppelt). Zusammen führen diese beiden Entwicklungen dazu, dass KI-Algorithmen enorme Datenmengen schnell verarbeiten können. Wenn KI-Projekte die Konzeptphase verlassen und aus ihnen kommerzielle Anwendungen werden, steigt der Bedarf an spezialisierten und teuren Cloud-Computing- und Grafikprozessor-Ressourcen. Die bei KI-Systemen zu beobachtenden Trends gehen mit einem gewaltigen Anstieg der erforderlichen Rechenleistung einher. Einer Schätzung zufolge wurde für das größte Experiment der letzten Zeit, AlphaGo Zero, 300 000 Mal mehr Rechenleistung benötigt als für das größte Experiment sechs Jahre zuvor (OpenAI, 2018[53]). Die Erfolge von AlphaGo Zero im Go- und Schachspiel wurden mit einer Rechenleistung erzielt, die die Leistung der zehn leistungsstärksten Supercomputer der Welt zusammengenommen übertreffen dürfte (OECD, 2018[52]).

Datenzugang und -nutzung

Datenzugang und Datenaustausch können Fortschritte im KI-Bereich beschleunigen oder umgekehrt bremsen

Zum Training und zur Weiterentwicklung aktueller ML-Technologien werden kuratierte und genaue Daten benötigt. Der Zugang zu qualitativ hochwertigen Datensätzen ist für die KI-Entwicklung daher von entscheidender Bedeutung. Folgende Faktoren sind für Datenzugang und Datenaustausch relevant und können Fortschritte im KI-Bereich beschleunigen oder umgekehrt bremsen (OECD, erscheint demnächst[54]):

- **Standards**: Standards bzw. Normen sind erforderlich, um Interoperabilität sowie die Weiterverwendung von Daten in verschiedenen Anwendungen zu ermöglichen, den Zugang zu fördern und sicherzustellen, dass Daten auffindbar, katalogisiert und/oder durchsuchbar und effektiv weiterverwendbar sind.
- **Risiken**: „Data-Sharing", d. h. der Austausch bzw. die Weitergabe oder gemeinsame Nutzung von Daten, kann für Privatpersonen, Organisationen und Staaten mit verschiedenen Risiken verbunden sein, darunter Vertraulichkeits- und Datenschutzverletzungen, Risiken in Bezug auf Rechte geistigen Eigentums und kommerzielle Interessen sowie potenzielle Risiken für die nationale und die digitale Sicherheit.
- **Datenkosten**: Die Datengewinnung, der Datenzugang, die gemeinsame Nutzung oder Weiterverwendung von Daten erfordern Vorab- und Folgeinvestitionen. Neben den Investitionen für die Datenerfassung sind zusätzliche Investitionen für die Datenbereinigung, die Datenkuratierung, die Pflege der Metadaten, die Datenspeicherung und -verarbeitung sowie eine sichere IT-Infrastruktur erforderlich.
- **Anreize**: Marktbasierte Ansätze können Anreize schaffen, den Zugang zu Datenmärkten und -plattformen, die Daten kommerzialisieren und Mehrwertdienstleistungen wie Zahlungs- und Datenaustauschinfrastrukturen anbieten, zu ermöglichen und Daten mit ihnen auszutauschen.
- **Unsicherheiten bezüglich des Dateneigentums**: Verschiedene rechtliche Bestimmungen – Regelungen zum Schutz geistigen Eigentums, strafrechtliche Regelungen (auch in Bezug auf Cyberkriminalität), Wettbewerbsrecht und Datenschutzgesetze – haben in Verbindung mit der Vielzahl der an der Erzeugung von Daten beteiligten Akteuren zu Unsicherheiten in der Frage des Dateneigentums geführt.
- **Nutzerbefähigung, einschließlich KI-Agenten**: Nutzerbefähigung und eine leichtere Übertragbarkeit der Daten – sowie die Einführung wirksamer Einwilligungsmechanismen und Wahlmöglichkeiten für die betroffenen Personen – können Einzelne und Unternehmen dazu bewegen, personenbezogene oder geschäftliche Daten weiterzugeben. Teilweise wird auch die Ansicht vertreten, dass KI-Agenten, die die Präferenzen einer Person kennen, ihr helfen könnten, den komplexen Datenaustausch mit anderen KI-Systemen auszuhandeln (Neppel, 2017[55]).
- **Vertrauenswürdige Dritte:** Dritte können Vertrauen schaffen und den Austausch und die Weiterverwendung von Daten unter allen Akteuren erleichtern. Datenmittler können als Zertifizierungsstellen fungieren. Vertrauenswürdige Datenaustauschplattformen, wie z. B. Datentrusts, liefern qualitativ hochwertige Daten.

Institutionelle Prüfungskommissionen stellen zudem sicher, dass die legitimen Interessen Dritter respektiert werden.

- **Repräsentativität der Daten**: Die Prognosen von KI-Systemen beruhen auf Mustern, die in Trainingsdatensätzen festgestellt wurden. Daher müssen Trainingsdatensätze sowohl aus Gründen der Genauigkeit als auch der Fairness inklusiv, vielfältig und repräsentativ sein, damit bestimmte Gruppen nicht unter- oder falsch repräsentiert werden.

Bestimmte Politikmaßnahmen können den Datenzugang und -austausch für die KI-Entwicklung verbessern

Zur Verbesserung des Datenzugangs und Datenaustauschs bieten sich u. a. folgende Maßnahmen an (OECD, erscheint demnächst[54]):

- **Zugang zu Daten des öffentlichen Sektors gewähren**, einschließlich Open-Government-Daten, Geodaten (z. B. Karten) und Verkehrsdaten.
- **Datenaustausch im privaten Sektor erleichtern**, in der Regel auf freiwilliger Basis oder, im Fall verbindlicher Maßnahmen, beschränkt auf vertrauenswürdige Nutzer. Besondere Schwerpunktbereiche sind „Daten von öffentlichem Interesse", Daten in netzgebundenen Sektoren wie Verkehr und Energie (zur Sicherung der Interoperabilität der Dienste) sowie die Übertragbarkeit personenbezogener Daten.
- **Statistische/Datenanalyse-Kapazitäten entwickeln**, indem Technologiezentren eingerichtet werden, die Unterstützung und Orientierungshilfen bei der Nutzung und Analyse von Daten bieten.
- **Nationale Datenstrategien entwickeln**, um die Kohärenz nationaler Data-Governance-Rahmen und ihre Vereinbarkeit mit nationalen KI-Strategien zu gewährleisten.

Es werden technische Lösungen für Datenengpässe entwickelt

Die Leistungsfähigkeit einiger ML-Algorithmen, z. B. solcher zur Bilderkennung, übersteigt bereits die durchschnittlichen menschlichen Fähigkeiten. Um zu diesem Punkt zu gelangen, mussten sie jedoch mit großen Datenbanken mit Millionen von beschrifteten Bildern trainiert werden. Daher wird aktiv an der Entwicklung maschineller Lernverfahren gearbeitet, durch die sich der Datenbedarf zum Training von KI-Systemen verringert. Dazu bieten sich mehrere Methoden an:

- **Tiefes bestärkendes Lernen (*deep reinforcement learning*)** ist eine ML-Technik, die tiefe neuronale Netze mit bestärkendem Lernen kombiniert (vgl. Unterabschnitt „Cluster 2: ML-Techniken" in Kapitel 1). Die Netze lernen dabei, bestimmte Verhaltensweisen zu bevorzugen, die zum gewünschten Ergebnis führen (Mousave, S., M. Schukat und E. Howley, 2018[56]). Verschiedene KI-Agenten konkurrieren in einer komplexen Umgebung miteinander, in der sie für ihre Handlungen entweder belohnt oder bestraft werden, je nachdem, ob diese zum gewünschten Ergebnis führen oder nicht. Die Agenten passen ihr Handeln dann entsprechend diesem „Feedback" an.[8]
- Beim **Transferlernen oder Vortraining** (Pan, S. und Q. Yang, 2010[57]) werden Modelle wiederverwendet, die zur Ausführung verschiedener Aufgaben im selben Bereich trainiert wurden. Zum Beispiel könnten einige Schichten eines Modells,

das darauf trainiert ist, Bilder von Katzen zu erkennen, wiederverwendet werden, um Bilder von blauen Kleidern zu erkennen. Gelingt dies, wären wesentlich kleinere Bilderdatensammlungen erforderlich als für traditionelle AL-Algorithmen (Jain, 2017[58]).

- Durch **erweitertes Datenlernen (*augmented data learning*)** oder „Datensynthetisierung" können Daten durch Simulationen oder Interpolationen aus vorhandenen Daten künstlich erzeugt werden. Dadurch erhöht sich das Datenvolumen und wird der Lernprozess effizienter. Diese Methode bietet sich vor allem an, wenn die Datennutzung durch Datenschutzbestimmungen eingeschränkt ist oder wenn Szenarien simuliert werden sollen, die in der Realität nur selten auftreten (Kasten 4.7).[9]
- **Hybride Lernmodelle** können Unsicherheit modellieren, indem sie verschiedene Arten tiefer neuronaler Netze mit probabilistischen oder Bayesschen Ansätzen kombinieren. Die Modellierung von Unsicherheit dient dazu, die Leistung und Nachvollziehbarkeit zu verbessern und die Wahrscheinlichkeit fehlerhafter Prognosen zu verringern (Kendall, 2017[59]).

Aufgrund von Datenschutz-, Vertraulichkeits- und Sicherheitsbedenken könnten Datenzugang und Datenaustausch eingeschränkt werden. Dies könnte zu einem Missverhältnis zwischen der Geschwindigkeit, mit der KI-Systeme lernen können, und den für ihr Training zur Verfügung stehenden Datensätzen führen. Jüngste Fortschritte in der Kryptografie, z. B. bei der sicheren Mehrparteienberechnung (*Multi-party computation – MPC*) und der homomorphen Verschlüsselung, könnten Datenanalysen ermöglichen, bei denen die Datenschutzrechte gewahrt werden. So könnte insbesondere erreicht werden, dass KI-Systeme operieren können, ohne sensible Daten erfassen oder auf sensible Daten zugreifen zu müssen (Kasten 4.8). KI-Modelle sind zunehmend in der Lage, mit verschlüsselten Daten zu arbeiten.[10] Da diese Lösungen jedoch rechenintensiv sind, lassen sie sich möglicherweise nur schwer skalieren (Brundage et al., 2018[37]).

Kasten 4.8. Neue kryptografische Werkzeuge ermöglichen datenschutzgerechte Berechnungen

Die im Bereich der Verschlüsselung erzielten Fortschritte ebnen den Weg für vielversprechende KI-Anwendungen. So könnten ML-Modelle beispielsweise mit kombinierten Daten verschiedener Organisationen trainiert werden. Dabei würden die Daten aller Beteiligten vertraulich bleiben. Dies könnte dazu beitragen, Hindernisse im Zusammenhang mit Datenschutz- oder Vertraulichkeitsbedenken zu überwinden. Die Verschlüsselungstechniken, die solche Berechnungen ermöglichen, sind nicht neu: homomorphe Verschlüsselung und sichere MPC wurden bereits vor Jahren bzw. Jahrzehnten entdeckt. Für die praktische Anwendung waren sie jedoch bislang zu ineffizient. Dank der jüngsten Fortschritte bei den Algorithmen und bei der Umsetzung entwickeln sie sich nun zunehmend zu praktischen Werkzeugen, die Datensätze aus der realen Welt produktiv analysieren können.

- Die **Homomorphe Verschlüsselung** ist eine Technik, mit der Berechnungen an verschlüsselten Daten durchgeführt werden können, ohne dass die unverschlüsselten Daten angezeigt werden müssen.
- Die **sichere MPC** ist die Berechnung einer Funktion von Daten, die aus vielen verschiedenen Quellen stammen, ohne dass Informationen über die Daten einer Quelle an eine andere Quelle weitergegeben werden. Sichere MPC-Protokolle

ermöglichen es mehreren Beteiligten, Algorithmen gemeinsam zu berechnen, wobei die Daten, die die einzelnen Beteiligten in den Algorithmus eingegeben haben, vertraulich bleiben.

Quelle: Brundage et al. (2018[37]), *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf; Dowlin (2016[60]), *CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy*, https://www.microsoft.com/en-us/research/wp-content/uploads/2016/04/CryptonetsTechReport.pdf.

KI-Modelle könnten auch Blockchain-Technologien nutzen, die ebenfalls kryptografische Werkzeuge zur sicheren Datenspeicherung einsetzen (Kasten 4.9). Lösungen, die KI- und Blockchain-Technologien kombinieren, könnten die Datenverfügbarkeit erhöhen. Gleichzeitig könnten sie die mit einer unverschlüsselten Datenverarbeitung verbundenen Datenschutz- und Sicherheitsrisiken auf ein Minimum reduzieren.

Kasten 4.9. KI-basierte datenschutzgerechte Identitätsprüfung dank Blockchain-Technologie

Kairos, ein Unternehmen, das Gesichtserkennungslösungen anbietet, hat in sein Portfolio Blockchain-Technologien aufgenommen. Durch die Kombination von Gesichtsbiometrie und Blockchain-Technologie ermöglicht es Nutzern einen besseren Schutz der Privatsphäre. Ein Algorithmus vergleicht das Bild einer Person mit charakteristischen Gesichtsmerkmalen (oder Identifizierungsmerkmalen), bis eine eindeutige Übereinstimmung hergestellt ist. Diese Übereinstimmung wird dann in eine eindeutige, zufällige Zahlenfolge umgewandelt. Das Originalbild kann danach verworfen werden. Diese „biometrische Blockchain“ wird unter der Prämisse aufgebaut, dass Unternehmen oder staatliche Stellen nicht die Identität einer Person kennen müssen, um zu überprüfen, dass sie es tatsächlich mit dieser Person zu tun haben.

Quelle: https://kairos.com/.

Wettbewerb

Die OECD hat sich mit der Frage der Auswirkungen des digitalen Wandels auf den Wettbewerb und die Wettbewerbspolitik auseinandergesetzt (OECD, 2020[61]). Einige dieser Auswirkungen hängen besonders mit KI zusammen. Sie sind Gegenstand dieses Unterabschnitts. Es wird allgemein anerkannt, dass KI einen wettbewerbsfördernden Effekt hat, da sie den Marktzutritt neuer Anbieter erleichtert. Ein Großteil der Aufmerksamkeit, die die Wettbewerbspolitik großen KI-Akteuren schenkt, ist darauf zurückzuführen, dass sie Betreiber von Online-Plattformen und Inhaber großer Datenmengen sind. Sie hängt nicht mit der Nutzung von KI als solcher zusammen.

Speziell im Zusammenhang mit KI stellt sich jedoch die Frage, ob es datengetriebene Netzwerkeffekte gibt. Von einem Netzwerkeffekt spricht man, wenn der einzelne Nutzer umso größeren Nutzen aus einer bestimmten Plattform zieht, je mehr andere Personen diese Plattform ebenfalls nutzen. Mit jedem neuen Nutzer erhält die Plattform z. B. mehr Informationen, mit deren Hilfe ihre Algorithmen lernen, wie sie die Nutzer besser bedienen können (OECD, 2019[62]). Dem steht die Annahme gegenüber, dass im Fall von Daten von abnehmenden Skalenerträgen auszugehen ist: Ist ein bestimmtes Datenvolumen erreicht, führt eine weitere Zunahme der Daten zu immer geringeren Verbesserungen der Prognosen. Daher ist nicht klar, ob KI langfristig Wettbewerbsprobleme hervorrufen könnte (Bajari et al., 2018[63]; OECD, 2016[64]; Varian, 2018[65]).

Im Hinblick auf den geschäftlichen Wert zusätzlicher Daten sind Skaleneffekte möglich. Wenn ein Unternehmen aufgrund seiner im Vergleich zur Konkurrenz etwas besseren Datenqualität deutlich mehr Kunden gewinnen kann, könnte dies zu einer positiven Feedbackschleife führen. Mehr Kunden bedeuten mehr Daten, wodurch sich dieser Kreislauf beschleunigt und im Lauf der Zeit eine immer größere Marktbeherrschung erzielt werden könnte. Zu Skaleneffekten könnte es auch bei dem für den Aufbau effektiver KI-Systeme erforderlichen Fachwissen kommen.

Zudem besteht die Befürchtung, dass Algorithmen durch die Überwachung von Marktbedingungen, Preisen und Reaktionen der Konkurrenz auf Preisänderungen Absprachen zwischen verschiedenen Anbietern erleichtern könnten. Dadurch könnten Unternehmen über neue und verbesserte Instrumente zur Koordinierung von Strategien, zur Festsetzung von Preisen und zur Durchsetzung von Kartellabsprachen verfügen. Etwas spekulativer ist die Sorge, dass im Fall ausgereifterer Deep-Learning-Algorithmen nicht einmal tatsächliche Absprachen zwischen den Unternehmen nötig wären, um eine kartellartige Situation herbeizuführen. Eine solche Situation könnte vielmehr ohne menschliches Eingreifen erreicht werden. Dies würde große Herausforderungen in Bezug auf die Rechtsdurchsetzung mit sich bringen. Nach den geltenden wettbewerbsrechtlichen Bestimmungen müssen Beweise für Absprachen oder „korrespondierende Willenserklärungen“ vorliegen, bevor ein Kartellverstoß festgestellt und geahndet werden kann (OECD, 2017[66]).

Geistiges Eigentum

In diesem Unterabschnitt werden einige mögliche Auswirkungen von KI auf geistiges Eigentum erörtert. Es handelt sich hier um einen Bereich, in dem viel Bewegung herrscht und man gerade erst beginnt, sich auf evidenzbasierte Analysen zu stützen. Regelungen zum Schutz des geistigen Eigentums im KI-Bereich erhöhen im Allgemeinen den Umfang und das Tempo von Neuentdeckungen, Erfindungen und Technologieverbreitung. Damit ähneln sie den Regelungen für andere Technologien, die durch Rechte des geistigen Eigentums geschützt sind. Regelungen zum Schutz des geistigen Eigentums sollen Erfinder, Autoren, Künstler und Markeninhaber belohnen. Die Politik in diesem Bereich muss aber auch das Potenzial von KI als Input für weitere Innovationen berücksichtigen.

Wenn zum Schutz von geistigem Eigentum im KI-Bereich andere Instrumente als Geschäftsgeheimnisse eingesetzt werden, stellen sich neue Fragen in Bezug darauf, wie man Innovatoren dazu anregen kann, KI-Innovationen offenzulegen, einschließlich Algorithmen und deren Training. Das Büro des Europäischen Parlaments hat auf einer Konferenz drei mögliche Arten der KI-Patentierung diskutiert (EPO, 2018[67]). Der erste Typ (*Core AI*) bezieht sich häufig auf Algorithmen, die als mathematische Methoden nicht patentierbar sind. Beim zweiten Typ – trainierte Modelle/ML – könnten Ansprüche auf Variationen und Parameterbereiche problematisch sein. Im dritten Fall könnte KI als ein Werkzeug in einem Anwendungsbereich patentiert werden, der über technische Effekte definiert ist. Andere internationale Organisationen und OECD-Länder untersuchen ebenfalls die Auswirkungen von KI im Bereich des geistigen Eigentums.[11]

Im Zusammenhang mit der Verbreitung von KI stellt sich zudem die Frage, ob die Systeme zum Schutz der geistigen Eigentumsrechte in einer Welt, in der KI-Systeme selbst Erfindungen machen können, angepasst werden müssen (OECD, 2017[68]). Bestimmte KI-Systeme können bereits patentierbare Erfindungen hervorbringen, insbesondere in der Chemie, der Pharmazie und der Biotechnologie. In diesen Bereichen bestehen viele Erfindungen darin, originale Molekülkombinationen für neue Verbindungen zu kreieren oder neue Eigenschaften bestehender Moleküle zu bestimmen. Zum Beispiel gelang es

KnIT, einem von IBM entwickelten ML-Tool, besondere Kinasen zu entdecken (Kinasen sind Enzyme, die als Katalysator für die Übertragung von Phosphatgruppen auf bestimmte Substrate wirken). Diese Kinasen wiesen in einer Gruppe bekannter Kinasen spezifische Eigenschaften auf, die experimentell getestet wurden. Die spezifischen Eigenschaften dieser Moleküle wurden von einer Software entdeckt, und die entsprechenden Erfindungen wurden zum Patent angemeldet. Diese und andere Aspekte von KI und geistigen Eigentumsrechten werden derzeit von den in diesem Bereich kompetenten Organisationen des OECD-Raums untersucht, wie dem Europäischen Patentamt, dem Patent- und Markenamt der Vereinigten Staaten und der Weltorganisation für geistiges Eigentum (WIPO). Fragen im Zusammenhang mit dem Urheberrechtsschutz von mittels KI verarbeiteten Daten könnten dabei ebenfalls geprüft werden.

Kleine und mittlere Unternehmen

Maßnahmen und Programme, die KMU bei der Einführung von KI unterstützen sollen, gewinnen zunehmend an Priorität. Dies ist ein sich rasch entwickelnder Bereich, der allmählich zum Gegenstand evidenzbasierter Analysen wird. Digitale Ökosysteme, die die Einführung und Nutzung von KI in KMU erleichtern, können u. a. durch folgende Arten von Maßnahmen gefördert werden:

- Weiterqualifizierung ist eine entscheidende Komponente, da es KMU oft schwerfällt, sich im Wettbewerb um KI-Fachkräfte zu behaupten.
- Förderung gezielter Investitionen in ausgewählte vertikale Branchen. In Frankreich laufen z. B. Maßnahmen zur Förderung von Investitionen in spezifische KI-Anwendungen in der Landwirtschaft. Solche Maßnahmen können helfen, wenn einzelne KMU es sich nicht leisten können, allein zu investieren (OECD, 2018[13]).
- Unterstützung von KMU beim Datenzugang, u. a. durch die Schaffung von Plattformen für den Datenaustausch.
- Förderung des Zugangs von KMU zu KI-Technologien – auch durch Technologietransfer von öffentlichen Forschungsinstituten – sowie zu Rechenkapazitäten und Cloud-Plattformen (Deutschland, 2018[69]).
- Verbesserung der Finanzierungsmechanismen, um auf KI spezialisierte KMU bei der Ausweitung ihrer Aktivitäten zu unterstützen, z. B. durch neue öffentliche Investitionsfonds sowie mehr Flexibilität und höhere Finanzierungsobergrenzen bei Programmen zur Unterstützung von Investitionen in wissensintensive Unternehmen (Vereinigtes Königreich, 2017[70]). Die Europäische Kommission engagiert sich insbesondere mit dem AI4EU-Projekt, einer On-Demand-KI-Plattform, für europäische KMU.

Günstige Rahmenbedingungen für KI-Innovationen schaffen

Die OECD analysiert derzeit, welche Veränderungen aufgrund von KI und anderen digitalen Transformationen in der Innovationspolitik und ähnlichen KI-relevanten Politikbereichen erforderlich sind (OECD, 2018[48]). Dabei wird geprüft, wie die Anpassungsfähigkeit, die Reaktionsfähigkeit und die Flexibilität der politischen Instrumente und Experimente verbessert werden können. Staatliche Stellen können Experimente nutzen, um kontrollierte Umgebungen zum Testen von KI-Systemen bereitzustellen. Solche Umgebungen sind z. B. Regulatory Sandboxes („regulatorische Sandkästen“) bzw. Reallabore oder Innovationszentren. Die Experimente können dort im „Start-up-Modus“ erfolgen. In

diesem Fall werden sie durchgeführt, ausgewertet und modifiziert und dann entweder ausgeweitet oder umgekehrt zurückgefahren oder schnell wieder aufgegeben.

Eine weitere Möglichkeit, schnellere und effektivere Entscheidungen zu treffen, besteht im Einsatz digitaler Werkzeuge zur Gestaltung von Politikmaßnahmen (einschließlich innovationsfördernder Maßnahmen) und zum Monitoring von Politikzielen. Einige Staaten greifen beispielsweise auf eine „agentenbasierte Modellierung" zurück, um die Auswirkungen verschiedener Politikoptionen auf verschiedene Arten von Unternehmen vorherzusehen.

Die zuständigen staatlichen Stellen können KI-Akteure ermutigen, Selbstregulierungsmechanismen wie Verhaltenskodizes, freiwillige Standards und Best Practices zu entwickeln. Diese Instrumente können den KI-Akteuren dann als Orientierungshilfen während des gesamten KI-Lebenszyklus dienen, d. h. auch bei Monitoring, Berichterstattung und Bewertung sowie bei der Verhinderung nachteiliger Auswirkungen oder des Missbrauchs von KI-Systemen.

Außerdem können die zuständigen staatlichen Stellen gegebenenfalls Kontrollmechanismen für KI-Systeme im öffentlichen und im privaten Sektor einrichten und fördern. Dazu könnten Compliance-Prüfungen, Audits, Konformitätsbewertungen und Zertifizierungsprogramme gehören. Solche Mechanismen könnten auch nützlich sein, um die besonderen Bedürfnisse von KMU und die Sachzwänge, mit denen sie konfrontiert sind, zu berücksichtigen.

Wandel der Arbeitswelt und Kompetenzentwicklung

Beschäftigung

KI dürfte die Menschen bei einigen Aufgaben ergänzen und bei anderen ersetzen und neue Arbeitsformen schaffen

Die OECD hat sich eingehend mit den Auswirkungen des digitalen Wandels auf die Beschäftigung und den daraus erwachsenden Konsequenzen für die Politik befasst (OECD, 2020[61]). Die Entwicklungen im Bereich KI schreiten rasch voran und die evidenzbasierte Analyse steht erst am Anfang. Sicher ist jedoch bereits, dass es mit der zunehmenden Verbreitung von KI zu erheblichen Veränderungen in der Arbeitswelt kommen wird. Künstliche Intelligenz dürfte die Menschen bei einigen Aufgaben ergänzen und bei anderen ersetzen und neue Arbeitsformen schaffen. In diesem Abschnitt werden einige der Veränderungen erörtert, die KI auf den Arbeitsmärkten auslösen dürfte. Dabei geht es auch um die Frage, wie die Politik den Übergang zu einer KI-Wirtschaft begleiten kann.

KI dürfte zu Produktivitätssteigerungen führen

KI wird die Produktivität voraussichtlich auf zweierlei Weise erhöhen. Erstens werden einige Tätigkeiten, die bisher von Menschen ausgeführt wurden, automatisiert werden. Zweitens werden Systeme dank Maschinenautonomie mit reduzierter oder ohne menschliche Kontrolle funktionieren und sich an die jeweiligen Umstände anpassen (OECD, 2017[68]; Autor, D. und A. Salomons, 2018[71]). Eine in zwölf Industrieländern durchgeführte Untersuchung ergab, dass KI die Arbeitsproduktivität bis 2035 um bis zu 40 % gegenüber dem erwarteten Basisniveau steigern könnte (Purdy, M. und P. Daugherty, 2016[72]). Konkrete Beispiele gibt es viele. Watson von IBM etwa unterstützt Kundenberater der französischen Bank Crédit Mutuel dabei, Kundenanfragen um 60 % schneller zu beantworten.[12] Der Chatbot von Alibaba bearbeitete im Schlussverkauf 2017 mehr als

95 % der Kundenanfragen. Dadurch konnten sich die menschlichen Kundenbetreuer um die komplizierteren oder individuelleren Anfragen kümmern (Zeng, 2018[73]). Eine höhere Arbeitsproduktivität führt theoretisch zu höheren Löhnen, da die Wertschöpfung je Beschäftigten steigt.

Teams, in denen Menschen und KI zusammenwirken, können Fehler begrenzen helfen und menschlichen Arbeitskräften neue Möglichkeiten eröffnen. Es hat sich gezeigt, dass solche Teams produktiver sind als KI oder menschliche Arbeitskräfte allein (Daugherty, P. und H. Wilson, 2018[74]). Bei BMW führte die Bildung solcher gemischten Teams in der Fertigung beispielsweise zu einer Steigerung der Produktivität um 85 % im Vergleich zu nicht integrierten Teams. Bei Walmart verwalten Roboter den Lagerbestand, sodass sich das Ladenpersonal ganz auf die Kundenbetreuung konzentrieren kann. Und wenn Radiologen mit Hilfe von KI-Modellen Röntgenaufnahmen auf Tuberkulose prüfen, liegt die Genauigkeit der Diagnose bei 100 % – und ist damit höher, als wenn nur auf KI oder nur auf menschliche Urteilskraft vertraut wird (Lakhani, P. und B. Sundaram, 2017[75]).

KI kann auch helfen, bereits automatisierte Arbeitsschritte zu verbessern und zu beschleunigen. Dadurch können Unternehmen ihr Produktionsvolumen erhöhen und gleichzeitig ihre Kosten senken. Werden die niedrigeren Kosten an andere Unternehmen oder Privatpersonen weitergegeben, ist mit einer steigenden Nachfrage nach den produzierten Waren zu rechnen. Dies erhöht wiederum die Nachfrage nach Arbeitskräften sowohl innerhalb der betreffenden Unternehmen – z. B. in der Produktion – als auch in anderen Unternehmen, die Vorleistungen erbringen.

KI dürfte die Art der Aufgaben verändern, die automatisiert werden können, und diesen Prozess vielleicht beschleunigen

Automatisierung ist kein neues Phänomen, KI dürfte jedoch die Art der Aufgaben verändern, die automatisiert werden können, und diesen Prozess vielleicht insgesamt beschleunigen. Im Gegensatz zu Computern sind KI-Technologien nicht streng vorprogrammiert und regelbasiert. Computer haben in Routineberufen mit mittleren Qualifikationsanforderungen zu einem tendenziellen Rückgang der Beschäftigung geführt. Neue KI-gestützte Anwendungen sind zunehmend in der Lage, relativ komplexe Aufgaben auszuführen, bei denen Vorhersagen angestellt werden müssen (vgl. Kapitel 3). Zu diesen Aufgaben gehören Transkription, Übersetzung, das Führen von Fahrzeugen, die Diagnose von Krankheiten und die Beantwortung von Kundenanfragen (Graetz, G. und G. Michaels, 2018[76]; Michaels, G., A. Natraj und J. Van Reenen, 2014[77]; Goos, M., A. Manning und A. Salomons, 2014[78]).[13]

In einer von der OECD durchgeführten explorativen Studie wurde geschätzt, inwieweit Technologien in der Lage sind, die Fragen der OECD-Erhebung über die Kompetenzen Erwachsener (PIAAC) in den Bereichen Lesekompetenz und alltagsmathematische Kompetenz zu beantworten (Elliott, 2017[79]). Dabei zeigte sich, dass das Leistungsniveau von KI-Systemen im Bereich Lesekompetenz 2017 dem von 89 % der Erwachsenen im OECD-Raum entsprach. Im Umkehrschluss bedeutet dies, dass nur 11 % der Erwachsenen der KI überlegen waren. Der Studie zufolge ist daher damit zu rechnen, dass sich der wirtschaftliche Druck erhöhen wird, für bestimmte lesekompetenzbasierte und alltagsmathematische Aufgaben Rechnerkapazitäten einzusetzen. Dies dürfte dazu führen, dass die Nachfrage nach menschlichen Arbeitskräften für Aufgaben, die eine geringe bis mittlere Lesekompetenz erfordern, entgegen den Trends der jüngsten Vergangenheit zurückgehen wird. Die Studie wies auch auf die Schwierigkeit hin, bildungspolitische Maßnahmen für Erwachsene zu gestalten, mit denen das Kompetenzniveau dieser

Arbeitskräfte über das aktuelle Leistungsniveau der Computer angehoben werden kann. Neben neuen Instrumenten und Anreizen zur Kompetenzentwicklung von Erwachsenen wurde dabei auch vorgeschlagen, kompetenzfördernde Maßnahmen mit anderen Initiativen zu kombinieren, u. a. im Bereich der sozialen Sicherung und des sozialen Dialogs (OECD, 2018[13]).

Der Beschäftigungseffekt der KI wird davon abhängen, wie schnell sie sich in den verschiedenen Sektoren durchsetzen wird

Der Beschäftigungseffekt der künstlichen Intelligenz wird auch davon abhängen, wie schnell sich KI-Technologien in den kommenden Jahrzehnten in den verschiedenen Sektoren entwickeln und verbreiten werden. Allgemein wird erwartet, dass autonome Fahrzeuge einen disruptiven Effekt auf Arbeitsplätze in der Personen- und Warenbeförderung haben werden. Etablierte Lkw-Hersteller wie Volvo und Daimler konkurrieren beispielsweise mit Start-ups wie Kodiak und Einride bei der Entwicklung und Erprobung fahrerloser Lastkraftwagen (Stewart, 2018[80]). Dem Weltverkehrsforum zufolge könnten fahrerlose Lastkraftwagen in den nächsten zehn Jahren auf vielen Straßen immer häufiger anzutreffen sein. Etwa 50 %-70 % der 6,4 Millionen Arbeitsplätze für Berufskraftfahrer in den Vereinigten Staaten und Europa könnten bis 2030 wegfallen (ITF, 2017[81]). Allerdings werden parallel dazu neue Arbeitsplätze geschaffen werden, um die gestiegene Zahl der fahrerlosen Lastkraftwagen zu unterstützen. Fahrerlose Lastkraftwagen könnten die Betriebskosten im Straßengüterverkehr um rd. 30 % senken, vor allem durch Einsparungen bei den Arbeitskosten. Dies könnte traditionelle Speditionsunternehmen aus dem Geschäft drängen und in der Folge zu einem noch schnelleren Rückgang der Arbeitsplätze im Speditionsgewerbe führen.

KI-Technologien werden sich wahrscheinlich auf Tätigkeiten auswirken, die traditionell höhere Qualifikationen voraussetzen

KI-Technologien übernehmen auch Prognoseaufgaben, die traditionell von höher qualifizierten Arbeitskräften – von Juristen bis hin zu medizinischen Fachkräften – ausgeführt werden. So hat beispielsweise ein Roboter-Anwalt Autofahrern dabei geholfen, Strafzettel in Höhe von insgesamt 12 Mio. USD anzufechten (Dormehl, 2018[82]). Im Jahr 2016 übertrafen die Systeme Watson von IBM und DeepMind Health menschliche Ärzte bei der Diagnose seltener Krebsarten (Frey, C. und M. Osborne, 2017[83]). Bei der Vorhersage von Börsenschwankungen erwies sich KI besser als Finanzexperten (Mims, 2010[84]).

KI kann Menschen ergänzen und neue Arbeitsformen schaffen

KI ergänzt Menschen und wird wahrscheinlich auch neue Beschäftigungsmöglichkeiten für menschliche Arbeitskräfte schaffen, insbesondere in Bereichen, in denen komplementäre Kompetenzen zur Prognose und insbesondere menschliche Kompetenzen wie kritisches Denken, Kreativität und Einfühlungsvermögen erforderlich sind (Vereinigte Staaten, 2016[85]; OECD, 2017[20]).

- **Tätigkeiten von Datenwissenschaftlern und ML-Experten**: Zur Erzeugung und Bereinigung von Daten sowie zur Programmierung und Entwicklung von KI-Anwendungen werden Spezialisten benötigt. Doch selbst wenn Daten und maschinelles Lernen neue Aufgaben für Menschen entstehen lassen, ist nicht mit einem allzu großen Zuwachs zu rechnen.

- **Menschenzentrierte Tätigkeiten**: Einige Handlungen sind von Natur aus wertvoller, wenn sie von einem Menschen anstatt von einer Maschine durchgeführt werden (z. B. solche von Berufssportlern, Erziehern oder Verkäufern). Viele halten es für wahrscheinlich, dass sich Menschen zunehmend auf Aufgaben konzentrieren werden, die das Leben anderer Menschen verbessern, z. B. Kinderbetreuung, Sport-Coaching und Pflege von unheilbar Kranken.
- **Urteilstätigkeiten – Bestimmung des Prognosegegenstands**: Am wichtigsten ist vielleicht das Konzept der Beurteilung, d. h. der Entscheidung über den Nutzen einer bestimmten Handlung in einem bestimmten Umfeld. Wenn KI für Prognosen genutzt wird, muss ein Mensch entscheiden, was vorhergesagt werden soll und wozu die Vorhersage dienen soll. Es bedarf Menschen mit Eigenschaften wie Urteilsvermögen und Fairness, um Dilemmas zu formulieren, Situationen zu interpretieren oder einem Text den richtigen Sinn zu entnehmen (OECD, 2018[13]). In der Wissenschaft kann KI beispielsweise Menschen ergänzen. Deren Fähigkeit zum konzeptuellen Denken ist jedoch unerlässlich, etwa um den Forschungsrahmen abzustecken und den Kontext für bestimmte Experimente festzulegen.
- **Urteilstätigkeiten – Bestimmung der optimalen Vorgehensweise**: Eine Entscheidung kann nicht allein auf der Grundlage einer Prognose getroffen werden. Dies zeigt sich schon an einer alltäglichen Entscheidung wie der, ob man auf einen Spaziergang einen Regenschirm mitnimmt oder nicht. Für diese Entscheidung wird eine Prognose über die Regenwahrscheinlichkeit angestellt. Die eigentliche Entscheidung wird aber großenteils von persönlichen Vorlieben abhängen, z. B. davon, was einem unangenehmer ist: nass zu werden oder einen Regenschirm mit sich herumzutragen. So verhält es sich mit vielen Entscheidungen. In der Cybersicherheit muss bei einer Prognose über die potenziell feindliche Natur einer neuen Anfrage das Risiko, eine freundliche Anfrage abzulehnen, gegen die Gefahr abgewogen werden, einer feindlichen Anfrage unautorisierte Informationen zugänglich zu machen.

Die Prognosen über den Nettoeffekt von KI auf das Beschäftigungsvolumen gehen stark auseinander

In den letzten 5 Jahren wurden unterschiedliche Schätzungen über die Gesamtauswirkungen der Automatisierung auf das Beschäftigungsvolumen angestellt (Winick, 2018[86]; MGI, 2017[87]; Frey, C. und M. Osborne, 2017[83]). Frey und Osborne sagten z. B. voraus, dass 47 % der Arbeitsplätze in den Vereinigten Staaten in den nächsten 10-15 Jahren verschwinden könnten. Das McKinsey Global Institute schätzte 2017, dass bei 60 % der Arbeitsplätze rd. ein Drittel der erledigten Aufgaben automatisiert werden könnte. Die Automatisierung ist in diesen Arbeitsplätzen jedoch nicht nur auf die Entwicklung und den Einsatz von KI zurückzuführen, sondern auch auf andere technologische Entwicklungen.

Allerdings ist nicht nur der Umfang der Beschäftigungsverluste schwer vorauszusehen, sondern auch das Volumen der Beschäftigungsschaffung in neuen Bereichen. Einer Studie zufolge könnte KI bis 2025 zu einem Nettozuwachs an 2 Millionen Arbeitsplätzen führen (Gartner, 2017[88]). Dies ist jedoch nicht nur durch die Entstehung neuer Tätigkeitsfelder bedingt, sondern auch durch indirekte Effekte. Beispielsweise wird KI wahrscheinlich die Kosten für die Produktion von Waren und Dienstleistungen senken und deren Qualität erhöhen. Dies dürfte zu einem Anstieg der Nachfrage und damit auch der Beschäftigung führen.

Die jüngsten Schätzungen der OECD tragen der Aufgabenheterogenität in eng definierten Berufen Rechnung, wobei Daten der Internationalen Vergleichsstudie der Kompetenzen Erwachsener (PIAAC) verwendet werden. Beim aktuellen Stand der technologischen Entwicklung sind 14 % der Arbeitsplätze in den OECD-Ländern einem hohen Automatisierungsrisiko ausgesetzt. 32% der Arbeitskräfte werden sich wahrscheinlich mit erheblichen Veränderungen ihrer Arbeitsplätze konfrontiert sehen (Nedelkoska, L. und G. Quintini, 2018[89]). Jugendliche und ältere Arbeitskräfte sind vom Automatisierungsrisiko am stärksten bedroht. Neuere Analysen der OECD zeigen, dass die Beschäftigung in Berufen, die als „stark automatisierbar" eingestuft werden, in 16 europäischen Ländern in 82 % der Regionen zurückgeht. Gleichzeitig wird in 60 % der Regionen eine größere Zunahme der Zahl der Arbeitsplätze mit „geringem Automatisierungsgrad" festgestellt, die diesen Arbeitsplatzverlust ausgleicht. Diese Untersuchung bestätigt die Annahme, dass die Automatisierung das Verhältnis zwischen den verschiedenen Beschäftigungskategorien verschieben könnte, ohne die Gesamtbeschäftigung zu verringern (OECD, 2018[90]).

KI wird die Arbeitswelt verändern

Die Einführung von KI dürfte die Arbeitswelt verändern. KI kann dazu beitragen, Arbeit interessanter zu machen, indem sie Routineaufgaben automatisiert, flexiblere Arbeitsformen ermöglicht und u. U. auch die Vereinbarkeit von Berufs- und Privatleben verbessert. Menschliche Kreativität und Genialität können mit immer leistungsfähigeren Rechen-, Daten- und Algorithmusressourcen kombiniert werden, wodurch neue Aufgaben und Tätigkeiten entstehen können (Kasparov, 2018[91]).

Ganz allgemein könnte KI effizienzsteigernd wirken und so den Wandel an den Arbeitsmärkten beschleunigen. KI-Techniken in Verbindung mit Big Data können Unternehmen heute potenziell bei der Definition der Rollen unterstützen, die Arbeitskräfte übernehmen sollen – und ihnen helfen, für ihre Beschäftigten die jeweils optimalsten Einsatzmöglichkeiten zu finden und ihre Stellen mit den jeweils geeigneten Personen zu besetzen. IBM beispielsweise nutzt KI zur Optimierung innerbetrieblicher Weiterbildungsmaßnahmen und empfiehlt seinen Mitarbeitern auf der Grundlage ihrer bisherigen Leistung, ihrer Karriereziele und des Kompetenzbedarfs des Unternehmens passende Schulungsmodule. Unternehmen wie KeenCorp und Vibe haben Textanalyse-Techniken entwickelt, um Unternehmen bei der Analyse der Mitarbeiterkommunikation zu unterstützen und so die Bewertung von Messgrößen wie Arbeitsmoral, Arbeitskräfteproduktivität und Netzwerkeffekte zu erleichtern (Deloitte, 2017[92]). Mit diesen Informationen könnte KI Unternehmen helfen, die Produktivität ihrer Arbeitskräfte zu optimieren.

Es müssen Parameter für organisatorische Veränderungen festgelegt werden

Es wird zunehmend deutlich, dass neue bzw. überarbeitete Branchenstandards und Technologievereinbarungen zwischen Geschäftsleitung und Beschäftigten erforderlich sind, um zuverlässige, sichere und produktive Arbeitsplätze zu gewährleisten. Der Europäische Wirtschafts- und Sozialausschuss (EWSA) rät den „Interessenträgern, sich gemeinsam für komplementäre KI-Systeme und ihre Ko-Kreation am Arbeitsplatz einzusetzen" (EWSA, 2017[45]). Darüber hinaus ist es wichtig, die Beschäftigungsflexibilität zu erhöhen, ohne die Autonomie der Arbeitskräfte und die Beschäftigungsqualität zu beeinträchtigen, auch in Bezug auf die Gewinnbeteiligung. Der jüngste Tarifvertrag zwischen der deutschen IG Metall und dem Arbeitgeberverband Gesamtmetall verdeutlicht die wirtschaftliche Machbarkeit variabler Arbeitszeiten. Er zeigt, dass Arbeitgeber und Gewerkschaften heute Arbeitsbedingungen aushandeln können, die den organisatorischen Anforderungen der Unternehmen ebenso gerecht werden wie den persönlichen Bedürfnissen der Beschäftigten

(Kindererziehung, Pflege kranker Angehöriger usw.), und zwar ohne dass dazu der gesetzliche Beschäftigungsschutz überarbeitet werden müsste (Byhovskaya, 2018[93]).

Die Nutzung von KI zur Unterstützung der Arbeitsmarktfunktionen ist ebenfalls vielversprechend

Durch KI gelingt es bereits, den Prozess der Stellenvermittlung ebenso wie die Weiterbildung effizienter zu gestalten. KI kann Arbeitsuchenden und insbesondere freigesetzten Arbeitskräften helfen, passende Weiterbildungsprogramme zu finden, um die für neue, expandierende Berufe erforderlichen Kompetenzen zu erwerben. In vielen OECD-Ländern nutzen Arbeitgeber und öffentliche Arbeitsverwaltungen bereits Online-Plattformen zur Stellenbesetzung (OECD, 2018[90]). KI und andere digitale Technologien könnten künftig helfen, innovative und personalisierte Konzepte für Arbeitsuche und Stellenbesetzung zu verbessern und das Matching von Arbeitsnachfrage und Arbeitsangebot effizienter zu gestalten. Die Plattform LinkedIn nutzt KI, um Personalvermittlern zu helfen, die richtigen Kandidaten zu finden, und Kandidaten auf passende Stellen hinzuweisen. Dabei stützt sie sich auf die Profil- und Aktivitätsdaten ihrer 470 Millionen registrierten Nutzer (Wong, 2017[94]).

KI-Technologien, die große Datenmengen nutzen, können auch dazu beitragen, staatliche Stellen, Arbeitgeber und Arbeitskräfte über die lokalen Arbeitsmarktbedingungen zu informieren. Mit diesen Informationen ist es möglich, den aktuellen Kompetenzbedarf zu ermitteln und den künftigen Kompetenzbedarf vorherzusagen, Weiterbildungsressourcen zu steuern und Arbeitsuchende auf freie Stellen aufmerksam zu machen. Projekte zur Entwicklung solcher Arbeitsmarktinformationen laufen z. B. bereits in Finnland, der Tschechischen Republik und Lettland (OECD, 2018[90]).

Regeln für die Nutzung von Beschäftigtendaten festlegen

KI ist nur mit großen Datensätzen produktiv, bei der Nutzung personenbezogener Beschäftigtendaten bestehen jedoch Risiken. Dies gilt insbesondere, wenn die KI-Systeme, die die Daten analysieren, nicht transparent sind. Zur Personal- und Produktivitätsplanung wird zunehmend auf Beschäftigtendaten und Algorithmen zurückgegriffen. Politikverantwortliche und sonstige Akteure sollten daher untersuchen, wie sich Datenerfassung und -verarbeitung auf Beschäftigungsaussichten und -bedingungen auswirken. Über Anwendungen, Fingerabdrücke, Wearables und Sensoren können in Echtzeit Daten erfasst werden, aus denen sich der Standort und Arbeitsplatz der betreffenden Personen ablesen lässt. Im Kundenservice wird z. B. mit KI-Software die Freundlichkeit des Tonfalls von Mitarbeitern analysiert. Laut Aussage betroffener Mitarbeiter werden dabei allerdings keine Sprachmuster berücksichtigt, und es ist offenbar schwierig, die Auswertungsergebnisse anzufechten (UNI, 2018[95]).

In einigen Ländern wird derzeit an Vereinbarungen über Arbeitnehmerdaten und das „Recht auf Abschalten“ (*right to disconnect*) gearbeitet. Das französische Telekommunikationsunternehmen Orange France Telecom gehörte zusammen mit fünf Gewerkschaftsverbänden zu den ersten, die sich auf Verpflichtungen zum Schutz von Arbeitnehmerdaten geeinigt haben. Vereinbart wurden dabei insbesondere Regelungen für eine transparente Datennutzung, Schulungen und die Einführung neuer Geräte. Um Regelungslücken in Bezug auf Arbeitnehmerdaten zu schließen, kann es auch sinnvoll sein, Data-Governance-Organe in den Unternehmen einzurichten, Rechenschaftspflichten für die Verwendung von (personenbezogenen) Daten einzuführen sowie einen Anspruch auf Datenübertragbarkeit, Erklärung und Datenlöschung zu schaffen (UNI, 2018[95]).

Den Wandel begleiten

Die Politik muss den Übergang zu einer KI-basierten Wirtschaft begleiten, insbesondere mit Maßnahmen im Bereich des Sozialschutzes

Wenn der organisatorische Wandel nicht mit der technologischen Entwicklung Schritt hält, kann es an den Arbeitsmärkten zu Verwerfungen kommen (OECD, 2018[13]). Dass langfristig Grund zu Optimismus besteht, heißt nicht, dass der Übergang zu einer zunehmend KI-basierten Wirtschaft reibungslos verlaufen wird: Einige Wirtschaftszweige werden wahrscheinlich wachsen, andere werden schrumpfen. Manche Arbeitsplätze werden verschwinden, andere werden entstehen. Daher wird die wichtigste beschäftigungspolitische Herausforderung in Bezug auf KI darin bestehen, diese Veränderungen unterstützend zu begleiten. Dabei helfen können soziale Sicherheitsnetze, Krankenversicherungsregelungen, eine progressive Besteuerung von Arbeit und Kapital sowie Bildungsmaßnahmen. OECD-Analysen zeigen, dass zudem Maßnahmen in anderen Bereichen, z. B. im Wettbewerbsrecht, berücksichtigt werden müssen, die sich auf Marktkonzentration, Marktmacht und Einkommensverteilung auswirken können (OECD, 2020[61]).

Zur Nutzung von KI erforderliche Kompetenzen

Mit dem Wandel der Arbeitswelt verändern sich auch die Kompetenzanforderungen

Mit dem Wandel der Arbeitswelt verändern sich auch die Kompetenzanforderungen, die an die Beschäftigten gerichtet werden (OECD, 2017[96]; Acemoglu, D. und P. Restrepo, 2018[97]; Brynjolfsson, E. und T. Mitchell, 2017[98]). In diesem Unterabschnitt wird erörtert, welche Auswirkungen künstliche Intelligenz auf die von den Arbeitskräften benötigten Kompetenzen haben könnte. Auch hier verändert sich die Situation rasch, und die evidenzbasierte Analyse steht noch ganz am Anfang. In der Bildungspolitik dürften Anpassungen erforderlich sein, um die Möglichkeiten für lebenslanges Lernen, Fort- und Weiterbildung und Kompetenzentwicklung zu verbessern. Wie andere Technologien auch dürfte KI voraussichtlich eine zusätzliche Nachfrage nach drei Arten von Kompetenzen erzeugen: Erstens werden **Fachkompetenzen** zur Programmierung und Entwicklung von KI-Anwendungen benötigt werden. Dazu könnten Kompetenzen für KI-bezogene Grundlagenforschung, Technik und Anwendungen sowie Datenwissenschaft und rechnergestütztes Denken gehören. Zweitens werden **allgemeine Kompetenzen** zur Nutzung von KI benötigt werden, z. B. um gemischte Teams aus Menschen und KI im Fertigungsbereich und in der Qualitätskontrolle zu ermöglichen. Drittens wird KI **komplementäre Kompetenzen** erfordern, so z. B. kritisches Denken, Kreativität, Innovationsgeist und unternehmerische Initiative sowie Einfühlungsvermögen (Vereinigte Staaten, 2016[85]; OECD, 2017[20]).

Es bedarf Initiativen zum Aufbau und zur Weiterentwicklung von KI-Kompetenzen, um Kompetenzengpässen entgegenzuwirken

Es wird erwartet, dass die Kompetenzengpässe im KI-Bereich zunehmen und infolge der wachsenden Nachfrage nach Spezialisten in Bereichen wie maschinellem Lernen immer deutlicher zutage treten werden. KMU, Hochschulen und öffentliche Forschungszentren konkurrieren bereits mit den großen marktführenden Unternehmen um Talente. Im öffentlichen, im privaten und im Hochschulsektor werden Initiativen zum Aufbau und zur Weiterentwicklung von KI-Kompetenzen gestartet. Zum Beispiel hat die Regierung von Singapur an der Singapore Management University ein fünfjähriges Forschungsprogramm zur Governance von KI und zur Datennutzung eingerichtet. Das dortige Centre for AI &

Data Governance betreibt industrierelevante Forschung zu den Themen KI und Industrie, Gesellschaft und Vermarktung. Das Massachusetts Institute of Technology (MIT) hat 1 Mrd. USD für die Gründung des Schwarzman College of Computing bereitgestellt. Ziel ist es, Studierende und Forscher aller Fachrichtungen Mittel an die Hand zu geben, um ihre jeweiligen Fachrichtungen mithilfe von Informatik und KI voranzubringen und so umgekehrt auch die KI voranzubringen.

Die Kompetenzengpässe im KI-Bereich haben einige Länder auch dazu veranlasst, die Zuwanderungsverfahren für hochqualifizierte Fachkräfte zu vereinfachen. So hat das Vereinigte Königreich beispielsweise die Zahl der Tier-1-Visa (für Hochqualifizierte) auf 2 000 pro Jahr verdoppelt und das Verfahren zur Erlangung einer Arbeitsgenehmigung für Spitzenstudenten und -forscher beschleunigt (Vereinigtes Königreich, 2017[99]). Kanada hat die Bearbeitungsfrist für Visumanträge von Hochqualifizierten auf zwei Wochen verkürzt und Visumbefreiungen für kurzfristige Forschungsaufträge eingeführt. Dies war Teil der Global Skills Strategy 2017, um hochqualifizierte Arbeitskräfte und Forscher aus dem Ausland anzuziehen (Kanada, 2017[100]).

Zur Nutzung von KI erforderliche allgemeine Kompetenzen

Alle OECD-Länder führen Kompetenzbeurteilungen durch und stellen Prognosen zum kurz-, mittel- oder langfristigen Kompetenzbedarf an. In Finnland wurde ein Programm für künstliche Intelligenz vorgeschlagen, das ein Kompetenzkonto bzw. Gutscheinsystem für lebenslanges Lernen umfasst, das die Nachfrage nach allgemeiner und beruflicher Bildung erhöhen soll (Finnland, 2017[101]). Das Vereinigte Königreich engagiert sich für eine vielfältige Erwerbsbevölkerung mit KI-Fachkompetenzen und investiert rd. 406 Mio. GBP (530 Mio. USD) in die Kompetenzentwicklung. Der Schwerpunkt liegt dabei auf Naturwissenschaften, Technologie, Ingenieurwesen und Mathematik sowie auf der Ausbildung von Informatiklehrern (Vereinigtes Königreich, 2017[99]).

Zunehmend wichtig im Arbeitsleben wird die „branchenspezifische Zweisprachigkeit“: Es reicht dann nicht mehr, sich nur im eigenen Fachbereich auszukennen, z. B. in Wirtschaft, Biologie oder Recht, erforderlich sind zudem Kompetenzen in KI-Techniken wie maschinellem Lernen. Daher kündigte das MIT im Oktober 2018 die bedeutendste Änderung seiner Organisationsstruktur seit fünfzig Jahren an – die Schaffung eines neuen Informatikinstituts, das von den Ingenieurwissenschaften unabhängig und mit allen anderen Fachbereichen verflochten sein wird. Dort sollen solche „zweisprachigen“ Studierenden ausgebildet werden, die KI und ML auf die Herausforderungen ihrer jeweiligen Fachrichtung anwenden können. Die Art und Weise, wie Informatik am MIT gelehrt wird, wandelt sich damit grundlegend. Das MIT hat 1 Mrd. USD für die Gründung dieser neuen Fakultät zur Verfügung gestellt (MIT, 2018[102]).

Komplementäre Kompetenzen

Soziale Kompetenzen („Soft Skills“) werden immer wichtiger. Nach dem aktuellen Stand der Forschung geht es dabei insbesondere um menschliches Urteilsvermögen, Analysefähigkeiten und zwischenmenschliche Kommunikation (Agrawal, A., J. Gans und A. Goldfarb, 2018[103]; Deming, 2017[104]; Trajtenberg, 2018[105]). Die OECD wird ihre internationale Schulleistungsstudie PISA 2021 um ein Modul ergänzen, mit dem kreative und kritische Denkfähigkeiten getestet werden sollen. Die Ergebnisse sollen einen Vergleich des Kreativitätspotenzials der verschiedenen Länder ermöglichen, der dann als Grundlage für von der Politik und den Sozialpartnern zu ergreifenden Maßnahmen dienen kann.

Messung

Die Umsetzung einer menschenzentrierten und vertrauenswürdigen KI hängt vom Kontext ab. Ein wichtiger Teil des Engagements der Politik für eine menschenzentrierte KI wird allerdings darin bestehen, Ziele und Messgrößen zur Bewertung der Leistung von KI-Systemen festzulegen, insbesondere im Hinblick auf die Aspekte Genauigkeit, Effizienz, gesellschaftlicher Wohlstand, Fairness und Robustheit.

Literaturverzeichnis

Abrams, M. et al. (2017), *Artificial Intelligence, Ethics and Enhanced Data Stewardship*, The Information Accountability Foundation, Plano, Texas, http://informationaccountability.org/wp-content/uploads/Artificial-Intelligence-Ethics-and-Enhanced-Data-Stewardship.pdf. [16]

Acemoglu, D. und P. Restrepo (2018), "Artificial Intelligence, Automation and Work", *NBER Working Paper* No. 24196. [97]

Agrawal, A., J. Gans und A. Goldfarb (2018), "Economic policy for artificial intelligence", *NBER Working Paper*, No. 24690, http://dx.doi.org/10.3386/w24690. [49]

Agrawal, A., J. Gans und A. Goldfarb (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business School Press, Brighton, MA. [103]

Autor, D. und A. Salomons (2018), "Is automation labor-displacing? Productivity growth, employment, and the labor share", *NBER Working Paper*, No. 24871, http://dx.doi.org/10.3386/w24871. [71]

Bajari, P. et al. (2018), "The impact of big data on firm performance: An empirical investigation", *NBER Working Paper*, No. 24334, http://dx.doi.org/10.3386/w24334. [63]

Barocas, S. und A. Selbst (2016), "Big Data's Disparate Impact", *California Law Review*, Vol. 104, S. 671-729, http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf. [29]

Berk, R. und J. Hyatt (2015), "Machine Learning Forecasts of Risk to Inform Sentencing Decisions", *Federal Sentencing Reporter*, Vol. 27/4, S. 222-228, http://dx.doi.org/10.1525/fsr.2015.27.4.222. [23]

Borges, G. (2017), *Liability for Machine-Made Decisions: Gaps and Potential Solutions*, Präsentation bei der OECD-Konferenz "AI: Intelligent Machines, Smart Policies", Paris, 26.-27. Oktober, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-borges.pdf. [44]

Brundage, M. et al. (2018), *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Centre for a New American Security, Electronic Frontier Foundation and Open AI, arXiv:1802.07228, https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf. [37]

Brynjolfsson, E. und T. Mitchell (2017), "What can machine learning do? Workforce implications", *Science*, Vol. 358/6370, S. 1530-1534, http://dx.doi.org/10.1126/science.aap8062. [98]

Brynjolfsson, E., D. Rock und C. Syverson (2017), "Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics", *NBER Working Paper*, No. 24001, http://dx.doi.org/10.3386/w24001. [50]

Burgess, M. (2016), "Holding AI to account: Will algorithms ever be free of bias if they are created by humans?", *WIRED,* 11. Januar, https://www.wired.co.uk/article/creating-transparent-ai-algorithms-machine-learning. [31]

Byhovskaya, A. (2018), *Overview of the national strategies on work 4.0: a coherent analysis of the role of the social partners*, Europäischer Wirtschafts- und Sozialausschuss, Brüssel, https://www.eesc.europa.eu/sites/default/files/files/qe-02-18-923-en-n.pdf. [93]

Cellarius, M. (2017), *Artificial Intelligence and the Right to Informational Self-determination*, The Forum Network, OECD, Paris, https://www.oecd-forum.org/users/75927-mathias-cellarius/posts/28608-artificial-intelligence-and-the-right-to-informational-self-determination. [9]

Chouldechova, A. (2016), "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments", arXiv:1610.07524, https://arxiv.org/abs/1610.07524. [24]

Citron, D. und F. Pasquale (2014), "The Scored Society: Due Process for Automated Predictions", *Washington Law Review*, Vol. 89, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376209. [36]

Cockburn, I., R. Henderson und S. Stern (2018), "The impact of artificial intelligence on innovation"*, NBER Working Paper*, No. 24449, http://dx.doi.org/10.3386/w24449. [51]

Crawford, K. (2016), "Artificial Intelligence's White Guy Problem", *The New York Times,* 26. Juni, https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html?_r=0. [30]

Daugherty, P. und H. Wilson (2018), *Human Machine: Reimagining Work in the Age of AI*, Harvard Business Review Press, Cambridge, MA. [74]

Deloitte (2017), *HR Technology Disruptions for 2018: Productivity, Design and Intelligence Reign*, Deloitte, http://marketing.bersin.com/rs/976-LMP-699/images/HRTechDisruptions2018-Report-100517.pdf. [92]

Deming, D. (2017), "The Growing Importance of Social Skills in the Labor Market", *The Quarterly Journal of Economics*, Vol. 132/4, S. 1593-1640, http://dx.doi.org/10.1093/qje/qjx022. [104]

Deutschland (2018), "Eckpunkte der Bundesregierung für eine Strategie Künstliche Intelligenz", Gemeinsame Pressemitteilung der Bundesregierung und des BMWi, 18. Juli, Bundesministerium für Wirtschaft und Energie, https://www.bmwi.de/Redaktion/EN/Pressemitteilungen/2018/20180718-key-points-for-federal-government-strategy-on-artificial-intelligence.html. [69]

Dormehl, L. (2018), "Meet the British whiz kid who fights for justice with robo-lawyer sidekick", *Digital Trends,* 25. März, https://www.digitaltrends.com/cool-tech/robot-lawyer-free-acess-justice/. [82]

Doshi-Velez, F. et al. (2017), "Accountability of AI under the law: The role of explanation", *arXiv* arXiv:1711.01134v2, https://arxiv.org/pdf/1711.01134v2.pdf. [28]

Dowlin, N. (2016), "CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy", *MSR-TR-2016-3* Microsoft Research, https://www.microsoft.com/en-us/research/wp-content/uploads/2016/04/CryptonetsTechReport.pdf. [60]

Dressel, J. und H. Farid (2018), "The accuracy, fairness and limits of predicting recidivism", *Science Advances*, Vol. 4/1, http://advances.sciencemag.org/content/4/1/eaao5580. [33]

Elliott, S. (2017), *Computers and the Future of Skill Demand*, Educational Research and Innovation, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264284395-en. [79]

EPO (2018), *Patenting Artificial Intelligence – Conference Summary*, Europäisches Patentamt, München, 30. Mai, http://documents.epo.org/projects/babylon/acad.nsf/0/D9F20464038C0753C125829E0031B814/$FILE/summary_conference_artificial_intelligence_en.pdf. [67]

EWSA (2017), *Künstliche Intelligenz – die Auswirkungen der künstlichen Intelligenz auf den (digitalen) Binnenmarkt sowie Produktion, Verbrauch, Beschäftigung und Gesellschaft*, Europäischer Wirtschafts- und Sozialausschuss, Brüssel, https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence. [45]

Finnland (2017), "Artificial intelligence programme" Webseite, Ministry of Economic Affairs and Employment, https://tem.fi/en/artificial-intelligence-programme. [101]

Flanagan, M., D. Howe und H. Nissenbaum (2008), "Embodying values in technology: Theory and practice", in van den Hoven, Jeroen und J. Weckert (Hrsg.), S. 322-353, http://dx.doi.org/10.1017/cbo9780511498725.017. [15]

Freeman, R. (2017), *Evolution or Revolution? The Future of Regulation and Liability for AI*, Präsentation bei der OECD-Konferenz "AI: Intelligent Machines, Smart Policies", Paris, 26.-27. Oktober, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-freeman.pdf. [40]

Frey, C. und M. Osborne (2017), "The future of employment: How susceptible are Jobs to computerisation?", *Technological Forecasting and Social Change*, Vol. 114, S. 254-280, http://dx.doi.org/10.1016/j.techfore.2016.08.019. [83]

Gartner (2017), "Gartner says by 2020, artificial intelligence will create more jobs than it eliminates", Pressemitteilung, 13. Dezember, Gartner, https://www.gartner.com/en/newsroom/press-releases/2017-12-13-gartner-says-by-2020-artificial-intelligence-will-create-more-jobs-than-it-eliminates. [88]

Golson, J. (2016), "Google's self-driving cars rack up 3 million simulated miles every day", *The Verge,* 1. Februar, https://www.theverge.com/2016/2/1/10892020/google-self-driving-simulator-3-million-miles. [41]

Goodfellow, I., J. Shlens und C. Szegedy (2015), "Explaining and harnessing adversarial examples", arXiv:1412.6572, https://arxiv.org/pdf/1412.6572.pdf. [38]

Goos, M., A. Manning und A. Salomons (2014), "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring", *American Economic Review*, Vol. 104/8, S. 2509-2526, http://dx.doi.org/10.1257/aer.104.8.2509. [78]

Graetz, G. und G. Michaels (2018), "Robots at Work", *Review of Economics and Statistics*, Vol. 100/5, S. 753-768, http://dx.doi.org/10.1162/rest_a_00754. [76]

Harkous, H. (2018), "Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning", *arXiv* arXiv:1802.02561v2, https://arxiv.org/pdf/1802.02561.pdf. [14]

Heiner, D. und C. Nguyen (2018), "Amplify Human Ingenuity with Intelligent Technology", *Shaping Human-Centered Artificial Intelligence, A.Ideas Series*, The Forum Network, OECD, Paris, https://www.oecd-forum.org/users/86008-david-heiner-and-carolyn-nguyen/posts/30653-shaping-human-centered-artificial-intelligence. [6]

Helgason, S. (1997), *Towards Performance-Based Accountability: Issues for Discussion*, Public Management Service, OECD, Paris, http://www.oecd.org/governance/budgeting/1902720.pdf. [46]

Indien (2018), *National Strategy for Artificial Intelligence #AIforall*, Discussion Paper, NITI Aayog, Juni, http://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf. [5]

Ingels, H. (2017), *Artificial Intelligence and EU Product Liability Law*, Präsentation bei der OECD-Konferenz "AI: Intelligent Machines, Smart Policies", Paris, 26.-27. Oktober, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-ingels.pdf. [43]

ITF (2017), "Driverless Trucks: New Report Maps Out Global Action on Driver Jobs and Legal Issues", Weltverkehrsforum, Paris, https://www.itf-oecd.org/driverless-trucks-new-report-maps-out-global-action-driver-jobs-and-legal-issues. [81]

Jain, S. (2017), "NanoNets: How to use Deep Learning when you have Limited Data, Part 2: Building Object Detection Models with Almost no Hardware" Medium, 30. Januar, https://medium.com/nanonets/nanonets-how-to-use-deep-learning-when-you-have-limited-data-f68c0b512cab. [58]

Kanada (2017), "Government of Canada launches the Global Skills Strategy", Pressemitteilung, Immigration, Refugees and Citizenship Canada, 12. Juni, https://www.canada.ca/en/immigration-refugees-citizenship/news/2017/06/government_of_canadalaunchestheglobalskillsstrategy.html. [100]

Kasparov, G. (2018), *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*, Public Affairs, New York. [91]

Kendall, A. (2017), "Deep Learning Is Not Good Enough, We Need Bayesian Deep Learning for Safe AI", Alex Kendall blog, 23. Mai, https://alexgkendall.com/computer_vision/bayesian_deep_learning_for_safe_ai/. [59]

Knight, W. (2017), "The Financial World Wants to Open AI's Black Boxes", *MIT Technology Review*, 13. April, https://www.technologyreview.com/s/604122/the-financial-world-wants-to-open-ais-black-boxes/. [32]

Kosack, S. und A. Fung (2014), "Does Transparency Improve Governance", *Annual Review of Political Science*, Vol. 17, S. 65-87, https://www.annualreviews.org/doi/pdf/10.1146/annurev-polisci-032210-144356. [26]

Kosinski, M., D. Stillwell und T. Graepel (2013), "Private traits and attributes are predictable from digital records of human behavior", *PNAS,* 11. März, http://www.pnas.org/content/pnas/early/2013/03/06/1218772110.full.pdf. [2]

Kurakin, A., I. Goodfellow und S. Bengio (2017), "Adversarial examples in the physical world", *arXiv* arXiv:1607.02533v4, https://arxiv.org/abs/1607.02533. [39]

Lakhani, P. und B. Sundaram (2017), "Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks", *Radiology*, Vol. 284/2, S. 574-582, http://dx.doi.org/10.1148/radiol.2017162326. [75]

Matheson, R. (2018), *Artificial intelligence model "learns" from patient data to make cancer treatment less toxic*, 9. August, http://news.mit.edu/2018/artificial-intelligence-model-learns-patient-data-cancer-treatment-less-toxic-0810. [106]

MGI (2017), *Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation*, McKinsey Global Institute, New York, https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages. [87]

Michaels, G., A. Natraj und J. Van Reenen (2014), "Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years", *Review of Economics and Statistics*, Vol. 96/1, S. 60-77, http://dx.doi.org/10.1162/rest_a_00366. [77]

Mims, C. (2010), "AI That Picks Stocks Better Than the Pros", *MIT Technology Review,* 10. Juni, https://www.technologyreview.com/s/419341/ai-that-picks-stocks-better-than-the-pros/. [84]

MIT (2018), "Cybersecurity's insidious new threat: workforce stress", *MIT Technology Review* 7. August, https://www.technologyreview.com/s/611727/cybersecuritys-insidious-new-threat-workforce-stress/. [102]

Mousave, S., M. Schukat und E. Howley (2018), "Deep Reinforcement Learning: An Overview", arXiv:1806.08894v1, https://arxiv.org/abs/1806.08894. [56]

Narayanan, A. (2018), "Tutorial: 21 fairness definitions and their politics", https://www.youtube.com/watch?v=jIXIuYdnyyk. [17]

Nedelkoska, L. und G. Quintini (2018), "Automation, skills use and training", *OECD Social, Employment and Migration Working Papers*, No. 202, OECD Publishing, Paris, https://dx.doi.org/10.1787/2e2f4eea-en. [89]

Neppel, C. (2017), *AI: Intelligent Machines, Smart Policies*, Präsentation bei der OECD-Konferenz "AI: Intelligent Machines, Smart Policies", Paris, 26.-27. Oktober, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-neppel.pdf. [55]

OECD (erscheint demnächst), *Enhanced Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use across Societies*, OECD Publishing, Paris. [54]

OECD (2020), *Going Digital: Den digitalen Wandel gestalten, das Leben verbessern*, OECD Publishing, Paris, https://doi.org/10.1787/e78eb379-de. [61]

OECD (2019), *An Introduction to Online Platforms and Their Role in the Digital Transformation*, OECD Publishing, Paris, https://dx.doi.org/10.1787/53e5f593-en. [62]

OECD (2019), *Empfehlung des Rats zu künstlicher Intelligenz*, OECD, Paris, http://www.oecd.org/berlin/presse/Empfehlung-des-Rats-zu-kuenstlicher-Intelligenz.pdf. [35]

OECD (2019), *Scoping Principles to Foster Trust in and Adoption of AI – Proposal by the Expert Group on Artificial Intelligence at the OECD (AIGO)*, OECD, Paris, http://oe.cd/ai. [34]

OECD (2018), "AI: Intelligent machines, smart policies: Conference summary", *OECD Digital Economy Papers*, No. 270, OECD Publishing, Paris, http://dx.doi.org/10.1787/f1a650d9-en. [13]

OECD (2018), *Job Creation and Local Economic Development 2018: Preparing for the Future of Work*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264305342-en. [90]

OECD (2018), *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, OECD Publishing, Paris, https://dx.doi.org/10.1787/sti_in_outlook-2018-en. [52]

OECD (2018), "Perspectives on innovation policies in the digital age", in *OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, OECD Publishing, Paris, https://dx.doi.org/10.1787/sti_in_outlook-2018-8-en. [48]

OECD (2017), *Algorithms and Collusion: Competition Policy in the Digital Age*, OECD Publishing, Paris, http://www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.html. [66]

OECD (2017), *Getting Skills Right: Skills for Jobs Indicators*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264277878-en. [96]

OECD (2017), *OECD Digital Economy Outlook 2017*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264276284-en. [20]

OECD (2017), *The Next Production Revolution: Implications for Governments and Business*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264271036-en. [68]

OECD (2016), "Big Data: Bringing Competition Policy to the Digital Era – Executive Summary", OECD-Dokument DAF/COMP/M(2016)2/ANN4/FINAL, Direktion Finanz- und Unternehmensfragen (DAF), Wettbewerbsaussschuss, OECD, Paris, https://one.oecd.org/document/DAF/COMP/M(2016)2/ANN4/FINAL/en/pdf. [64]

OECD (2013), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, OECD, Paris, http://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf. [12]

OECD (2011), *OECD-Leitsätze für multinationale Unternehmen*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264122352-de. [8]

OHCHR (2011), *Guiding Principles on Business and Human Rights*, Hohes Kommissariat der Vereinten Nationen für Menschenrechte (OHCHR), https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf. [7]

O'Neil, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Broadway Books, New York. [25]

OpenAI (2018), "AI and compute", OpenAI blog, 16. Mai, https://blog.openai.com/ai-and-compute/. [53]

Pan, S. und Q. Yang (2010), "A Survey on Transfer Learning", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22/10, S. 1345-1359, http://dx.doi.org/10.1109/TKDE.2009.191. [57]

Paper, I. (ed.) (2018), "Artificial intelligence and privacy", Issues Paper, Juni, Office of the Victorian Information Commissioner, https://ovic.vic.gov.au/wp-content/uploads/2018/08/AI-Issues-Paper-V1.1.pdf. [11]

Patki, N., R. Wedge und K. Veeramachaneni (2016), "The Synthetic Data Vault", in *IEEEE Proceedings – 3rd IEEE International Conference on Data Science and Advanced Analytics (DSAA 2016)*, S. 399-410, http://dx.doi.org/10.1109/dsaa.2016.49. [107]

Privacy International and ARTICLE 19 (2018), "Privacy and Freedom of Expression in the Age of Artificial Intelligence" Scoping Paper, https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf. [10]

Purdy, M. und P. Daugherty (2016), "Artificial Intelligence Poised to Double Annual Economic Growth Rate in 12 Developed Economies and Boost Labor Productivity by up to 40 Percent by 2035, According to New Research by Accenture", Pressemitteilung, Accenture, 28. September, http://www.accenture.com/news/artificial-intelligence-poised-to-double-annual-economic-growth-rate-in-12-developed-economies-and-boost-labor-productivity-by-up-to-40-percent-by-2035-according-to-new-research-by-accenture.htm. [72]

Selbst, A. (2017), "Disparate impact in big data policing", *Georgia Law Review*, Vol. 52/109, S. 109-195, http://dx.doi.org/10.2139/ssrn.2819182. [21]

Simonite, T. (2018), "Probing the dark side of Google's ad-targeting system", *MIT Technology Review,* 6. Juli, https://www.technologyreview.com/s/539021/probing-the-dark-side-of-googles-ad-targeting-system/. [19]

Slusallek, P. (2018), *Artificial Intelligence and Digital Reality: Do We Need a CERN for AI?*, The Forum Network, OECD, Paris, https://www.oecd-forum.org/channels/722-digitalisation/posts/28452-artificial-intelligence-and-digital-reality-do-we-need-a-cern-for-ai. [42]

Smith, M. und S. Neupane (2018), *Artificial intelligence and human development: toward a research agenda*, International Development Research Centre, Ottawa, https://idl-bnc-idrc.dspacedirect.org/handle/10625/56949. [4]

Stewart, J. (2018), "As Uber Gives up on Self-Driving Trucks, Another Startup Jumps In", *WIRED*, 8. Juli, https://www.wired.com/story/kodiak-self-driving-semi-trucks/. [80]

Talbot, D. et al. (2017), "Charting a Roadmap to Ensure Artificial Intelligence (AI) Benefits All" Medium, 30. November, https://medium.com/berkman-klein-center/charting-a-roadmap-to-ensure-artificial-intelligence-ai-benefits-all-e322f23f8b59. [3]

Trajtenberg, M. (2018), "AI as the next GPT: A political-economy perspective"*, NBER Working Paper*, No. 24245, http://dx.doi.org/10.3386/w24245. [105]

UNI (2018), *Die 10 wichtigsten Grundsätze für Arbeitnehmerdatenschutz und -sicherheit*, UNI Global Union, http://www.thefutureworldofwork.org/media/35483/uni-global-union_-arbeitnehmerdatenschutz-und-sicherheit.pdf. [95]

Varian, H. (2018), "Artificial intelligence, economics and industrial organization", *NBER Working Paper*, Vol. 24839, http://dx.doi.org/10.3386/w24839. [65]

Vereinigtes Königreich (2017), *UK Digital Strategy*, Government of the United Kingdom, Department for Digital, Culture, Media & Sport, https://www.gov.uk/government/publications/uk-digital-strategy/uk-digital-strategy. [70]

Vereinigtes Königreich (2017), *UK Industrial Strategy: A Leading Destination to Invest and Grow*, Great Britain & Northern Ireland, Department for Business, Energy & Industrial Strategy, http://https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/668161/uk-industrial-strategy-international-brochure.pdf. [99]

Vereinigte Staaten (2016), *Artificial Intelligence, Automation and the Economy*, Executive Office of the President, Government of the United States, https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED AI Economy Report.pdf. [85]

Wachter, S., B. Mittelstadt und C. Russell (2017), "Counterfactual explanations without opening the black box: Automated decisions and the GDPR", arXiv:1711.00399, https://arxiv.org/pdf/1711.00399.pdf. [27]

Wachter, S., B. Mittelstadt und L. Floridi (2017), "Transparent, explainable and accountable AI for robotics", *Science Robotics,*, Vol. 2/6, 31. Mai, http://robotics.sciencemag.org/content/2/6/eaan6080. [47]

Weinberger, D. (2018), "Optimization over explanation – Maximizing the benefits of machine learning without sacrificing its intelligence", Medium, 28. Januar, https://medium.com/@dweinberger/optimization-over-explanation-maximizing-the-benefits-we-want-from-machine-learning-without-347ccd9f3a66. [1]

Weinberger, D. (2018), "Playing with AI Fairness", Google PAIR, 17. September, https://pair-code.github.io/what-if-tool/ai-fairness.html. [22]

Winick, E. (2018), "Every study we could find on what automation will do to jobs, in one chart", *MIT Technology Review,* 25. Januar, https://www.technologyreview.com/s/610005/every-study-we-could-find-on-what-automation-will-do-to-jobs-in-one-chart/. [86]

Wong, Q. (2017), "At LinkedIn, artificial intelligence is like 'oxygen'", *Mercury News,* 1. Juni, http://www.mercurynews.com/2017/01/06/at-linkedin-artificial-intelligence-is-like-oxygen. [94]

Yona, G. (2017), "A Gentle Introduction to the Discussion on Algorithmic Fairness", Towards Data Science", 5. Oktober, https://towardsdatascience.com/a-gentle-introduction-to-the-discussion-on-algorithmic-fairness-740bbb469b6. [18]

Zeng, M. (2018), "Alibaba and the future of business", *Harvard Business Review,* September-Oktober, https://hbr.org/2018/09/alibaba-and-the-future-of-business. [73]

Anmerkungen

[1] Weitere Informationen unter: https://www.microsoft.com/en-us/ai/ai-for-good.

[2] Vgl. https://deepmind.com/applied/deepmind-ethics-society/.

[3] Vgl. https://www.blog.google/technology/ai/ai-principles/.

[4] Vgl. https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf.

[5] Als Beispiele sind die Internationale Arbeitsorganisation, die OECD-Leitsätze für multinationale Unternehmen oder die Leitprinzipien für Wirtschaft und Menschenrechte der Vereinten Nationen zu nennen.

[6] Wegen weiterer Studien zu diesem Thema vgl. https://www.dudumimran.com/2018/05/speaking-about-ai-and-cyber-security-at-the-oecd-forum-2018.html und https://maliciousaireport.com/.

[7] Pierre Chalançon, Vorsitzender der Taskforce Verbraucherpolitik des Beratenden Ausschusses der Wirtschaft bei der OECD (BIAC) und Vice President Regulatory Affairs, Vorwerk & Co KG, Vertretung bei der Europäischen Union – *Science-Fiction is not a Sound Basis for Legislation.*

[8] Diese Technik wird u. a. eingesetzt, um autonome Fahrzeuge in der Ausführung komplexer Manöver zu trainieren, das AlphaGo-Programm zu trainieren und Krebspatienten zu behandeln (z. B. indem die kleinste Dosierung und Verabreichungshäufigkeit bestimmt wird, die zur Reduzierung von Hirntumoren noch wirksam ist (Matheson, 2018[106])).

[9] Aus den Ergebnissen einer neueren Studie geht hervor, dass reale Daten vielfach erfolgreich durch synthetische Daten ersetzt werden können, was für Wissenschaftler gerade im Fall datenschutzrechtlicher Beschränkungen sehr hilfreich sein kann (Patki, N., R. Wedge und K. Veeramachaneni, 2016[107]). Die Autoren zeigen, dass sich die Ergebnisse, die mit synthetischen Daten erzeugt wurden, in 70 % der Fälle nicht signifikant von den Ergebnissen unterschieden, die mit realen Daten erzielt wurden.

[10] Lösungen, die z. B. eine vollhomomorphe Verschlüsselung mit neuronalen Netzen kombinieren, wurden in dieser Hinsicht erfolgreich getestet und eingesetzt (Dowlin, 2016[60]).

[11] Vgl. https://www.wipo.int/about-ip/en/artificial_intelligence/ und https://www.uspto.gov/about-us/events/artificial-intelligence-intellectual-property-policy-considerations.

[12] Vgl. https://www.ibm.com/watson/stories/creditmutuel/.

[13] Alibaba beschäftigt z. B. keine Zeitarbeiter mehr, um an Tagen, an denen Hochbetrieb herrscht oder Sonderaktionen angeboten werden, Kundenanfragen zu bearbeiten. An dem Tag, an dem Alibaba 2017 den höchsten Umsatz erzielte, bearbeitete der Chatbot des Unternehmens mehr als 95 % der Kundenanfragen und antwortete rd. 3,5 Millionen Kunden (Zeng, 2018[73]). Wenn Chatbots immer mehr Kundendienstfunktionen übernehmen, verändern sich die Aufgaben der menschlichen Kundenberater, die sich dann auf komplexere oder individuellere Anfragen konzentrieren müssen.

5. KI-Maßnahmen und -Initiativen

Maßnahmen und Initiativen im Bereich der künstlichen Intelligenz gewinnen für staatliche Stellen, Unternehmen, Fachorganisationen, Zivilgesellschaft und Gewerkschaften zunehmend an Bedeutung. Auf zwischenstaatlicher Ebene entstehen ebenfalls KI-Initiativen. Dieses Kapitel befasst sich mit KI-Maßnahmen, -Initiativen und -Strategien verschiedener Akteure auf nationaler und internationaler Ebene weltweit. Es kommt zu dem Ergebnis, dass nationale staatliche Initiativen im Allgemeinen darauf ausgerichtet sind, KI zur Steigerung von Produktivität und Wettbewerbsfähigkeit zu nutzen. Dabei kommen Aktionspläne zum Einsatz, die die folgenden Elemente stärken sollen: 1. Faktorbedingungen, z. B. KI-Forschungskapazitäten, 2. Nachfragebedingungen, 3. verwandte und unterstützende Branchen, 4. Firmenstrategie, Struktur und Wettbewerb sowie 5. Governance und Koordinierung auf nationaler Ebene. Zu den internationalen Initiativen gehört die Empfehlung des Rats der OECD zu künstlicher Intelligenz, mit der erstmals zwischenstaatliche Leitlinien für KI formuliert und Grundsätze und Prioritäten für eine verantwortungsvolle Steuerung vertrauenswürdiger KI festgelegt wurden.

Künstliche Intelligenz für wirtschaftliche Wettbewerbsfähigkeit: Strategien und Aktionspläne

Künstliche Intelligenz (KI) nimmt in der Politikagenda staatlicher Institutionen sowohl auf nationaler als auch auf internationaler Ebene eine immer wichtigere Position ein. Derzeit sind viele nationale staatliche Initiativen darauf ausgerichtet, KI zur Stärkung der Produktivität und der Wettbewerbsfähigkeit zu nutzen. Die in den nationalen KI-Aktionsplänen aufgeführten Prioritäten können in fünf große Themenbereiche aufgeteilt werden, die sich teilweise mit Porters Determinanten der wirtschaftlichen Wettbewerbsfähigkeit decken. Zu diesen Prioritäten gehören: 1. Faktorbedingungen, z. B. KI-Forschungskapazitäten, einschließlich Kompetenzen, 2. Nachfragebedingungen, 3. verwandte und unterstützende Branchen, 4. Firmenstrategie, Struktur und Wettbewerb sowie 5. Governance und Koordinierung auf nationaler Ebene (Kasten 5.1). Außerdem gewinnen KI-Fragen wie Transparenz, Menschenrechte und Ethik in der Politik zunehmend an Bedeutung.

Unter den Mitgliedsländern und Partnervolkswirtschaften der OECD haben die Volksrepublik China (im Folgenden „China“), Deutschland, Frankreich, Indien, Kanada, Schweden, das Vereinigte Königreich und die Vereinigten Staaten gezielte KI-Strategien aufgestellt. Einige Länder wie Dänemark, Japan und Korea führen KI-bezogene Maßnahmen im Rahmen breiter gefasster Aktionspläne durch. Viele andere Länder – darunter Australien, Estland, Finnland, Israel, Italien und Spanien – entwickeln derzeit entsprechende Strategien. Alle diese Strategien zielen darauf ab, die Zahl der Forscher und Hochschulabsolventen im KI-Bereich zu erhöhen, die nationalen KI-Forschungskapazitäten zu stärken und KI-Forschungsergebnisse im öffentlichen und privaten Sektor anzuwenden. Die wirtschaftlichen, sozialen, ethischen, politischen und rechtlichen Auswirkungen von Fortschritten im KI-Bereich werden in den nationalen Initiativen unterschiedlich berücksichtigt. Dies erklärt sich u. a. aus den unterschiedlichen kulturellen Rahmenbedingungen und Rechtssystemen der einzelnen Länder, deren unterschiedlicher Größe sowie dem unterschiedlichen Grad der Verbreitung von KI. Dabei ist allerdings anzumerken, dass sich die Umsetzung noch im Anfangsstadium befindet. Außerdem befasst sich das Kapitel mit den jüngsten rechtlichen und politischen Entwicklungen im KI-Bereich. Inwieweit die Ziele und Teilziele der nationalen Initiativen tatsächlich verwirklicht werden oder wie erfolgreich die verschiedenen Ansätze sind, wird jedoch nicht analysiert oder beurteilt.

KI ist auch Gegenstand von Diskussionen internationaler Gremien wie der Gruppe der 7 (G7), der Gruppe der 20 (G20), der OECD, der Europäischen Union und den Vereinten Nationen. Die Europäische Kommission richtet ihr Augenmerk dabei besonders auf Effizienz- und Flexibilitätssteigerungen durch KI, Interaktion und Kooperation, Produktivität, Wettbewerbsfähigkeit und Wachstum sowie die Lebensqualität ihrer Bürger. Nachdem dieses Thema bereits auf der Tagung der für Informations- und Kommunikationstechnologien (IKT) zuständigen Minister der G7 im April 2016 in Japan behandelt worden war, sprachen sich die IKT- und Industrieminister der G7 auf ihrer Tagung in Turin im September 2017 gemeinsam für die Vision einer „menschenzentrierten“ KI aus. Sie beschlossen, die internationale Zusammenarbeit und den Multi-Stakeholder-Dialog über KI zu fördern und mit Unterstützung der OECD gemeinsam auf ein besseres Verständnis von KI hinzuarbeiten. Die G20 beschäftigt sich ebenfalls zunehmend mit KI, und Japan erklärte KI zu einem der Schwerpunkte seines G20-Vorsitzes im Jahr 2019 (G20, 2018[1]).

Grundsätze für KI in der Gesellschaft

Mehrere Gruppen von Akteuren beschäftigen sich aktiv mit der Frage, wie die Entwicklung und Einführung von KI so gesteuert werden kann, dass sie der gesamten Gesellschaft zugutekommt. Das Institute for Electrical and Electronics Engineers (IEEE) lancierte beispielsweise im April 2016 die Global Initiative on Ethics of Autonomous and Intelligent Systems. Im Dezember 2017 veröffentlichte es die zweite Fassung seiner Leitlinien für ethisches Design. Die endgültige Version war für Anfang 2019 geplant. Die im September 2016 mit einem umfassenden Programm gegründete Partnership on Artificial Intelligence to Benefit People and Society arbeitet an Grundsätzen zu spezifischen Fragen wie etwa Sicherheit. Die KI-Leitsätze von Asilomar befassen sich mit Forschung, ethischen Fragen und Werten für eine kurz- und langfristig sichere und sozial verträgliche Entwicklung von KI. Die AI Initiative bringt Experten, Fachleute aus der Praxis und Bürger weltweit zusammen, um ein gemeinsames Verständnis von Konzepten wie der Erklärbarkeit von KI-Systemen zu entwickeln.

Im Rahmen mehrerer Initiativen (Tabelle 5.1) wurden wertvolle Grundsätze für die Steuerung der KI-Entwicklung erarbeitet. Besonderes Augenmerk richtet sich dabei auf die Einrichtungen und Akteure, die Forschung und Entwicklung (FuE) in Bezug auf KI-Systeme durchführen. Viele dieser Grundsätze wurden in Multi-Stakeholder-Prozessen entwickelt. Die an diesen Prozessen beteiligten Akteure lassen sich in fünf große Gruppen einteilen: Fachwelt, Privatsektor, Staat, Wissenschaft und Arbeitnehmerorganisationen. Zu den Akteuren aus der Fachwelt gehören das Future of Life Institute, das IEEE, die Japanese Society for Artificial Intelligence, die Initiative „Fairness, Accountability and Transparency in Machine Learning“ sowie die Association for Computing Machinery. Unter den Akteuren aus der Privatwirtschaft sind z. B. die Partnership on AI, der Information Industry Technology Council sowie Satya Nadella, Chief Executive Officer von Microsoft, zu nennen. Zu den staatlichen Akteuren gehören das Ministerium für Innere Angelegenheiten und Kommunikation in Japan, die World Commission on the Ethics of Science and Technology der Vereinten Nationen sowie der Engineering and Physical Sciences Research Council im Vereinigten Königreich. Erwähnenswerte Akteure aus der Wissenschaft sind die Université de Montréal und Nicolas Economou, Chief Executive Officer von H5 und Sonderberater der KI-Initiative der Future Society an der Harvard Kennedy School. Die Arbeitnehmerinteressen werden durch Organisationen wie UNI Global Union vertreten.

Den verschiedenen Initiativen sind einige Themen gemeinsam. Die aufgestellten Leitlinien befassen sich z. B. mit Themen wie menschliche Werte und Menschenrechte, Nichtdiskriminierung, Sensibilisierung und Kontrollierbarkeit, Datenzugang, Datenschutz und Datenkontrolle, Sicherheit, Kompetenzen, Transparenz und Nachvollziehbarkeit, Rechenschaft und Verantwortung, gesamtgesellschaftlicher Dialog und Messmethoden.

Im Mai 2018 gründete der OECD-Ausschuss für digitale Wirtschaft die OECD-Sachverständigengruppe für KI (AI Group of Experts at the OECD – AIGO). Ziel war dabei, Grundsätze für die Politik und die internationale Zusammenarbeit festzulegen, die das Vertrauen in die künstliche Intelligenz und deren Verbreitung fördern können (OECD, 2019[2]). Ihre Arbeiten bildeten die Grundlage für die Empfehlung des Rats der OECD zu künstlicher Intelligenz (OECD, 2019[3]), der am 22. Mai 2019 42 Staaten beigetreten sind.

Tabelle 5.1. Von verschiedenen Gremien und Akteuren aufgestellte KI-Leitlinien (nicht erschöpfende Auswahl)

Bezeichnung	KI-Leitlinien
ACM	ACM (2017), "2018 ACM Code of Ethics and Professional Conduct: Draft 3", Association for Computing Machinery Committee on Professional Ethics, https://ethics.acm.org/2018-code-draft-3/ USACM (2017), "Statement on Algorithmic Transparency and Accountability", Association for Computing Machinery US Public Policy Council, www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf
AI Safety	Amodei, D. et al. (2016), "Concrete Problems in AI Safety", 25. Juli, https://arxiv.org/pdf/1606.06565.pdf
Asilomar	FLI (2017), "Asilomar AI Principles", Future of Life Institute, https://futureoflife.org/ai-principles/
COMEST	COMEST (2017), "Report of COMEST on Robotics Ethics", World Commission on the Ethics of Scientific Knowledge and Technology, http://unesdoc.unesco.org/images/0025/002539/253952E.pdf
Economou	Economou, N. (2017) "A 'principled' artificial intelligence could improve justice", 3. Oktober, *Aba Journal* www.abajournal.com/legalrebels/article/a_principled_artificial_intelligence_could_improve_justice
EGE	EGE (2018), "Statement on Artificial Intelligence, Robotics and Autonomous Systems", European Group on Ethics in Science and New Technologies, http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf
EPSRC	EPSRC (2010), "Principles of Robotics" , Engineering and Physical Sciences Research Council, https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/
FATML	FATML (2016), "Principles for Accountable Algorithms and a Social Impact Statement for Algorithms", Fairness, Accountability, and Transparency in Machine Learning, www.fatml.org/resources/principles-for-accountable-algorithms
FPF	FPF (2018), "Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models", The Future of Privacy Forum, https://fpf.org/wp-content/uploads/2018/06/Beyond-Explainability.pdf
Google	Google (2018), "AI at Google: Our Principles", https://www.blog.google/technology/ai/ai-principles/
IEEE	IEEE (2017), *Ethically Aligned Design Version 2 – A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, Institute of Electrical and Electronics Engineers, http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf
Intel	Intel (2017), "AI – The Public Policy Opportunity", https://blogs.intel.com/policy/files/2017/10/Intel-Artificial-Intelligence-Public-Policy-White-Paper-2017.pdf
ITI	ITI (2017), "AI Policy Principles", Information Technology Industry Council, www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf
JSAI	JSAI (2017), "The Japanese Society for Artificial Intelligence Ethical Guidelines", The Japanese Society for Artificial Intelligence, http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf
MIC	Japan (2017), "Draft AI R&D Guidelines for International Discussions", Ministry of Internal Affairs and Communications, http://www.soumu.go.jp/main_content/000507517.pdf
MIC	MIC (2018), "Draft AI Utilization Principles", Japanese Ministry of Internal Affairs and Communication, www.soumu.go.jp/main_content/000581310.pdf
Montreal	UoM (2017), "The Montreal Declaration for a Responsible Development of Artificial Intelligence", University of Montreal, www.montrealdeclaration-responsibleai.com/
Nadella	Nadella, S. (2016) "The Partnership of the Future", 28. Juni, Slate, www.slate.com/articles/technology/future_tense/2016/06/microsoft_ceo_satya_nadella_humans_and_a_i_can_work_together_to_solve_society.html
PAI	PAI (2016), "TENETS", Partnership on AI, www.partnershiponai.org/tenets/
Polonski	Polonski, V. (2018) "The Hard Problem of AI Ethics – Three Guidelines for Building Morality Into Machines" , 28. Februar, Forum Network on Digitalisation and Trust, www.oecd-forum.org/users/80891-dr-vyacheslav-polonski/posts/30743-the-hard-problem-of-ai-ethics-three-guidelines-for-building-morality-into-machines
Taddeo und Floridi	Taddeo, M. und L. Floridi (2018), "How AI can be a force for good", *Science*, 24. August, Vol. 61/6404, S. 751-752, http://science.sciencemag.org/content/361/6404/751
The Public Voice Coalition	UGAI (2018), "Universal Guidelines on Artificial Intelligence", The Public Voice Coalition, https://thepublicvoice.org/ai-universal-guidelines/
Tokyo Statement	Next Generation Artificial Intelligence Research Center (2017), "The Tokyo Statement – Co-operation for Beneficial AI", www.ai.u-tokyo.ac.jp/tokyo-statement.html
Twomey	Twomey, P. (2018), "Toward a G20 Framework for Artificial Intelligence in the Workplace", *CIGI Papers,* No. 178, Centre for International Governance Innovation, www.cigionline.org/sites/default/files/documents/Paper%20No.178.pdf
UNI	UNI Global Union (2017), "Top 10 Principles for Ethical Artificial Intelligence", www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf

Nationale Initiativen

Überblick über nationale KI-Maßnahmen

Viele Länder haben nationale KI-Strategien und -Politikinitiativen angekündigt. Im Allgemeinen geht es darum, eine führende Stellung im KI-Bereich zu erlangen. In diesen Strategien und Initiativen werden Ziele und Teilziele festgelegt, die konzertierte Maßnahmen aller betroffenen Akteure erfordern. Die Regierungen übernehmen dabei häufig die Rolle eines Mittlers und Moderators. Kasten 5.1 beschreibt Elemente, die in Strategien und Maßnahmen zur Förderung der nationalen Wettbewerbsfähigkeit im KI-Bereich häufig enthalten sind. Darüber hinaus haben einige Länder spezielle Behörden für Ethikfragen im KI- und Datenbereich eingerichtet oder bereits bestehende Stellen mit solchen Aufgaben betraut.

Kasten 5.1. Wie versuchen die Länder, einen Wettbewerbsvorteil im KI-Bereich zu erreichen?

Porter nennt vier Determinanten für die Erzielung eines nationalen Wettbewerbsvorteils in einer bestimmten Branche: 1. Faktorbedingungen, 2. Nachfragebedingungen, 3. verwandte und unterstützende Branchen sowie 4. Firmenstrategie, Struktur und Wettbewerb. Die Wettbewerbsvorteile in den einzelnen Branchen werden laut Porter durch die Unternehmen geschaffen. Dem Staat kommt jedoch eine zentrale Rolle dabei zu, die vier Determinanten der nationalen industriellen Entwicklung zu unterstützen und zu fördern.

- **Faktorbedingungen**: Diese Determinante hängt von den geografischen Merkmalen des Landes, dem Fachkräfteangebot, dem Bildungsniveau der Bevölkerung und den Forschungskapazitäten ab. Die Länder stärken ihre KI-Forschungskapazitäten durch verschiedene Maßnahmen, darunter: 1. Aufbau von Forschungseinrichtungen, 2. Einführung neuer KI-bezogener Master- und Doktorandenstudiengänge und Anpassung der bestehenden Studiengänge durch Aufnahme von KI-Kursen, z. B. in naturwissenschaftlichen Disziplinen, und 3. Anwerbung in- und ausländischer Talente, z. B. durch eine erleichterte Visumerteilung für KI-Experten.
- **Nachfragebedingungen**: Manche Länder erklären bestimmte Wirtschaftszweige zu strategischen Sektoren für die KI-Entwicklung, insbesondere Verkehr, Gesundheitsversorgung und öffentlicher Dienst. Sie ergreifen dann Maßnahmen zur Förderung der inländischen Verbrauchernachfrage nach KI-Dienstleistungen in diesen Branchen. In einigen Ländern stellen die staatlichen Stellen über das Beschaffungswesen sicher, dass die KI-Systeme im öffentlichen Dienst bestimmte Standards erfüllen, z. B. im Hinblick auf Genauigkeit und Robustheit.
- **Verwandte und unterstützende Branchen**: Die Wettbewerbsfähigkeit im KI-Bereich erfordert den Zugang zu digitalen Infrastrukturangeboten und Diensten, Daten, Rechenleistung und Breitbandverbindungen. Einige Länder planen KI-zentrierte Technologie-Cluster und Förderstrukturen für kleine und mittlere Unternehmen (KMU).
- **Firmenstrategie, Struktur und Wettbewerb**: Zur Förderung von Privatinvestitionen und Wettbewerb im KI-Bereich ergreifen die Länder u. a. folgende

Maßnahmen: 1. Ausarbeitung von Roadmaps für die Förderung von Privatinvestitionen, 2. Schaffung von Anreizen für internationale KI-Unternehmen, im Inland zu investieren, z. B. durch die Einrichtung von KI-Laboratorien, sowie 3. experimentelle Politikansätze, z. B. „regulatorische Sandkästen" für KI-Anwendungen, die den Unternehmen Anreize für Innovation geben sollen.

Um die nationalen KI-Initiativen effektiv umzusetzen, prüfen viele Länder außerdem geeignete Governancemechanismen für einen koordinierten ressortübergreifenden Ansatz. Frankreich hat beispielsweise eine KI-Koordinierungsstelle beim Premierminister angesiedelt, die mit der Umsetzung der französischen KI-Strategie betraut ist.

Quelle: Porter (1990[4]), "The competitive advantage of nations", https://hbr.org/1990/03/the-competitive-advantage-of-nations.

Argentinien

Die argentinische Regierung kündigte für Juli 2019 eine zehnjährige nationale KI-Strategie an. Dem war eine Beurteilung vorangegangen, die das argentinische Ministerium für Wissenschaft, Technologie und Innovation im Rahmen der argentinischen Digitalagenda 2030 und des argentinischen Innovationsplans 2030 (Plan Argentina Innovadora 2030) 2018 durchgeführt hatte. Die thematischen Prioritäten der nationalen KI-Strategie lauten: Talente und Bildung, Daten, FuE und Innovation, Supercomputer-Infrastruktur, Maßnahmen zur Erleichterung des Arbeitsplatzwechsels sowie Förderung der Zusammenarbeit zwischen öffentlichem und privatem Sektor bei der Datennutzung. Weitere Prioritäten sind die öffentlichen Dienstleistungen und das Verarbeitende Gewerbe (als Zielbereiche der KI-Entwicklung). Die Strategie umfasst folgende Querschnittsthemen: 1. Investitionstätigkeit, Ethik und Regulierung, 2. Kommunikation und Sensibilisierung sowie 3. internationale Zusammenarbeit.

Sieben Ministerien sollen an der Strategie mitwirken. Zudem ist die Entwicklung eines nationalen KI-Innovationshubs vorgesehen, in dem Projekte aus den einzelnen thematischen Bereichen durchgeführt werden sollen. Für jede thematische Priorität soll es einen Lenkungsausschuss aus Sachverständigen geben, der dafür zuständig ist, die Ziele und die Methoden zur Fortschrittsmessung zu definieren.

Australien

Die australische Regierung hat im Haushalt 2018/2019 mehr als 28 Mio. AUD (21 Mio. USD) veranschlagt, um KI-Kapazitäten aufzubauen und eine verantwortungsvolle KI-Entwicklung in Australien zu fördern. Damit sollen die folgenden Projekte finanziert werden: 1. Projekte der Cooperative Research Centres mit Schwerpunkt KI (18 Mio. USD), 2. KI-zentrierte Promotionsstipendien (1 Mio. USD), 3. Entwicklung von Online-Ressourcen für den KI-Unterricht in Schulen (1,1 Mio. USD), 4. Untersuchung der Auswirkungen von KI-Technologien auf die einzelnen Branchen, die Chancen und Herausforderungen für die Erwerbsbevölkerung und den Bildungssektor (250 000 USD), 5. Ausarbeitung eines Ethikrahmens für KI auf der Basis von Fallstudien (367 000 USD) und 6. Aufstellung einer Roadmap für KI-Standards (72 000 USD mit Beteiligung der Wirtschaft).

Das Ministerium für Industrie, Innovation und Wissenschaft führt ebenfalls KI-bezogene Projekte durch. Der Australian Council of Learned Academies wurde beauftragt, die Chancen, Risiken und Folgen zu untersuchen, die mit einer weitreichenden Einführung von

KI in Australien in den nächsten zehn Jahren verbunden sind. Außerdem hat die Australian Human Rights Commission im Juli 2018 ein wichtiges Projekt zum Themenkomplex Menschenrechte und Technologie eingeleitet. Dazu wurde ein Diskussionspapier erstellt und eine internationale Konferenz veranstaltet; der Abschlussbericht war für 2019/2020 geplant.

Brasilien

Brasiliens Strategie für den digitalen Wandel (E-Digital) vom März 2018 sieht vor, verschiedene staatliche Initiativen zu digitalen Fragen zu harmonisieren und zu koordinieren, um die Ziele für nachhaltige Entwicklung in Brasilien voranzubringen. Im Hinblick auf KI zielt die E-Digital-Strategie darauf ab, „die potenziellen wirtschaftlichen und sozialen Auswirkungen [...] der künstlichen Intelligenz und von Big Data zu evaluieren und Maßnahmen vorzuschlagen, die ihre negativen Effekte abschwächen und die positiven Ergebnisse maximieren“ (Brasilien, 2018[5]). Zu den weiteren Prioritäten der E-Digital-Strategie gehören Mittelzuweisungen für Forschung, Entwicklung und Innovation (FEuI) sowie für den Kapazitätsaufbau im KI-Bereich. Brasilien hat die Absicht, 2019 eine KI-Strategie zu starten. Das Land nimmt aktiv an internationalen Diskussionen über die Entwicklung technischer Standards und politischer Rahmenkonzepte für KI teil.

Zwischen 2014 und Anfang 2019 hat das Ministerium für Wissenschaft, Technologie, Innovation und Kommunikation 16 verschiedene KI-Projekte und 59 KI-Start-ups durch gezielte Anreize und finanzielle Förderung unterstützt. Darüber hinaus wird KI auf Bundesebene in 39 E-Government-Initiativen eingesetzt. Diese Initiativen zielen insbesondere darauf ab, die Verwaltungs- und Beurteilungsverfahren zu verbessern, z. B. in Bereichen wie soziale Dienste, Bürgerdienste oder Stellenausschreibungen. 2019 wurde ein neues Institut für KI-Forschung – das Advanced Institute for Artificial Intelligence – gegründet. Es fördert Partnerschaften zwischen Universitäten und Unternehmen zu gemeinsamen FEuI-Projekten im KI-Bereich. Schwerpunkte sind u. a. Landwirtschaft, Smart Cities, digitale Governance, Infrastruktur, Umwelt, natürliche Ressourcen sowie Sicherheit und Verteidigung.

China

Im Mai 2016 veröffentlichte die chinesische Regierung einen dreijährigen nationalen KI-Plan, der gemeinsam von der Nationalen Entwicklungs- und Reformkommission, dem Ministerium für Wissenschaft und Technologie, dem Ministerium für Industrie und Informationstechnik sowie der chinesischen Cyberspace-Verwaltung erstellt wurde. Die Internet-Plus-Initiative, die 2015 als nationale Strategie eingerichtet wurde, um im Zeitraum 2016-2018 das Wirtschaftswachstum durch innovative internetgestützte Technologien zu fördern, wurde anschließend auf KI ausgeweitet (Jing, M. und S. Dai, 2018[6]). Im Mittelpunkt stehen folgende Elemente: 1. Ausbau der KI-Hardwarekapazitäten, 2. starke Plattform-Ökosysteme, 3. KI-Anwendungen in wichtigen sozioökonomischen Bereichen sowie 4. Auswirkungen von KI auf die Gesellschaft. Bis 2018 sollte mit dieser Initiative durch FuE ein Markt mit einem Volumen von 15 Mrd. USD für die chinesische KI-Branche geschaffen werden (China, 2016[7]).

Mitte 2017 veröffentlichte der chinesische Staatsrat den New Generation of Artificial Intelligence Development Plan, mit dem ein Zeitrahmen für die langfristige KI-Entwicklung aufgestellt wurde: 1. KI-gestütztes Wirtschaftswachstum in China bis 2020, 2. bahnbrechende Fortschritte bei den theoretischen Grundlagen und beim Aufbau einer

intelligenten Gesellschaft bis 2025 sowie 3. Positionierung Chinas als globales KI-Innovationszentrum bis 2030 und Aufbau einer KI-Industrie mit einem Gesamtvolumen von 1 Bill. RMB (150 Mrd. USD) (China, 2017[8]). Der Plan scheint ressort- und ebenenübergreifend umgesetzt zu werden. Dank staatlicher Unterstützung und dynamischen privaten Unternehmen übernimmt China allmählich eine Führungsrolle im KI-Bereich. Der chinesische Staatsrat hat Ziele für „eine neue Generation der Informationstechnologie" festgelegt und plant, das Volumen dieser strategischen Branche bis 2020 auf 15 % des Bruttoinlandsprodukts auszuweiten.

In ihrem 13. Fünfjahresplan (2016-2020) bekundete die chinesische Regierung die Absicht, das Land in einen führenden Wissenschafts- und Technologiestandort zu verwandeln. Dazu sollen 16 „Science and Technology Innovation 2030 Megaprojects", z. B. „AI 2.0", umgesetzt werden. Der Plan sieht auch Maßnahmen im öffentlichen Sektor vor (Kania, 2018[9]). Die Unternehmen werden aufgefordert, die FuE im Bereich der KI-Hardware und -Software zu beschleunigen, u. a. im Hinblick auf KI-gestütztes Sehen, Spracherkennung und biometrische Erkennung, Mensch-Maschine-Schnittstellen und intelligente Steuerungen.

Am 18. Januar 2018 setzte China eine nationale Standardisierungsgruppe für KI und eine nationale KI-Sachverständigengruppe ein. Zugleich veröffentlichte das National Standardisation Management Committee des Zweiten Industrieministeriums ein Weißbuch zu KI-Standardisierung, an dem auch das China Electronic Standardisation Institute (eine Abteilung des Ministeriums für Industrie und Informationstechnologie) mitgewirkt hatte (China, 2018[10]).

Chinesische Privatunternehmen zeigten bereits großes Interesse an KI, bevor dieser Bereich Ziel intensiver staatlicher Förderung wurde. Unternehmen wie Baidu, Alibaba und Tencent haben beträchtliche Anstrengungen und Investitionen in KI getätigt. Während die chinesische Industrie ihr Augenmerk vor allem auf Anwendungen und Datenintegration richtet, konzentriert sich die Zentralregierung auf grundlegende Algorithmen, Open Data und konzeptionelle Arbeiten. Die Stadtverwaltungen interessieren sich in erster Linie für Anwendungen und offene Daten auf kommunaler Ebene.

Dänemark

Dänemark stellte im Januar 2018 seine digitale Wachstumsstrategie vor. Diese Strategie soll sicherstellen, dass das Land im Digitalbereich eine Vorreiterrolle einnehmen kann und die gesamte Bevölkerung von der digitalen Transformation profitiert. Die Strategie enthält Initiativen, um die mit KI, Big Data und dem Internet der Dinge (Internet of Things – IoT) verbundenen Wachstumschancen zu nutzen. Sie umfasst folgende Elemente: 1. Einrichtung eines digitalen Hubs für öffentlich-private Partnerschaften, 2. Unterstützung von KMU bei der datengestützten Geschäftsentwicklung und der Digitalisierung, 3. Gründung von Bildungseinrichtungen im Rahmen eines Technologiepakts zur Förderung der technischen und digitalen Kompetenzen, 4. Stärkung der Cybersicherheit in Unternehmen und 5. Entwicklung einer agilen Regulierung, die neue Geschäftsmodelle und Experimente erleichtert. Die dänische Regierung hat bis 2025 1 Mrd. DKK (160 Mio. USD) für die Durchführung der Strategie zugesagt. Davon entfallen 75 Mio. DKK (12 Mio. USD) auf 2018 und 125 Mio. DKK (20 Mio. USD) auf den Zeitraum 2019-2025. Der größte Teil dieses Budgets ist für die Kompetenzentwicklung vorgesehen, gefolgt von der Einrichtung des digitalen Hubs und der Unterstützung für KMU (Dänemark, 2018[11]).

Deutschland

Die Bundesregierung hat im Dezember 2018 Deutschlands nationale Strategie für künstliche Intelligenz ins Leben gerufen (Deutschland, o.J.[12]). Deutschland will zu einem führenden KI-Standort werden. Forschungsergebnisse sollen schnell und umfassend in konkreten Anwendungen umgesetzt werden. „Artificial Intelligence Made in Germany" soll zu einem Exportschlager und weltweit anerkannten Gütesiegel werden. Erreicht werden soll dies durch neue Forschungszentren, eine verstärkte deutsch-französische Forschungszusammenarbeit, Finanzhilfen für den Aufbau von Clustern sowie KMU-Förderung. Die Strategie befasst sich zudem mit Themen wie Infrastrukturanforderungen, verbesserter Datenzugang, Kompetenzentwicklung und Sicherheitsvorkehrungen gegen Missbrauch sowie mit ethischen Fragen.

Im Juni 2017 legte die Ethik-Kommission Automatisiertes und Vernetztes Fahren, die vom Bundesminister für Verkehr und digitale Infrastruktur eingesetzt wurde, ihren Bericht vor. Dieser Bericht, der einen Katalog von ethischen Regeln enthält, befasst sich u. a. eingehend mit Fragen der Entscheidungsprogrammierung. Dabei geht es insbesondere darum, ob es zulässig sein kann, Menschenleben gegeneinander abzuwägen (sogenanntes „Weichenstellerproblem"), was der Bericht klar vereint. Bei unausweichlichen Unfallsituationen sollte die Handlungsvariante gewählt werden, die möglichst wenig Menschenleben kostet. Eine Qualifizierung nach persönlichen Merkmalen (Alter, Geschlecht, körperliche oder geistige Konstitution) sollte dabei strikt untersagt sein. Die Kommission stellt außerdem klar, dass dem Einzelnen keine Solidarpflichten auferlegt werden dürfen, sich für andere aufzuopfern (Deutschland, 2017[13]).

Estland

Estland plant derzeit den nächsten Schritt seines KI-gestützten E-Governancesystems, um Kosten zu sparen und die Effizienz zu steigern. Außerdem erprobt das Land neue Einsatzmöglichkeiten in den Bereichen E-Health und Situationsbewusstsein. Ziel ist es, die Lebensqualität, vor allem in den Städten, zu verbessern und menschliche Werte zu stärken. Im Bereich der Rechtsdurchsetzung konzentriert sich Estland auf die wesentlichen Aspekte Ethik, Haftung, Integrität und Verantwortlichkeit. Dabei stützt sich das Land auf die neuesten Technologien und baut ein auf Blockchain basierendes Durchsetzungssystem auf, das die Risiken im Hinblick auf Integrität und rechtliche Verantwortlichkeit mindern soll. Für 2018 ist ein Pilotprojekt geplant.

Seit März 2017 werden im Rahmen von StreetLEGAL selbstfahrende Autos auf den Straßen des Landes getestet. Estland ist außerdem das erste Land, das darüber diskutiert, KI-Systemen eine eigene Rechtspersönlichkeit zu verleihen. Das würde bedeuten, Algorithmen Vertretungsrechte zu übertragen und sie in die Lage zu versetzen, im Namen ihrer Eigentümer Dienstleistungen zu kaufen und zu verkaufen. 2016 setzte die Regierung Estlands eine Taskforce ein, um das Problem der rechtlichen Verantwortlichkeit beim Einsatz von ML-Algorithmen und die Notwendigkeit gesetzlicher Regelungen zu prüfen. Das Ministerium für wirtschaftliche Angelegenheiten und Kommunikation und die Stabsstelle der Regierung sind dabei ebenfalls einbezogen (Kaevats, 2017[14]; Kaevats, 2017[15]).

Finnland

Finnland will KI nutzen, um eine sichere und demokratische Gesellschaft zu schaffen, die weltweit besten öffentlichen Dienstleistungen zu bieten und Wohlstand, Wachstum und Produktivität zu steigern. Die im Oktober 2017 veröffentlichte KI-Strategie Finland's Age of Artificial Intelligence zielt darauf ab, das hohe Bildungsniveau der Bevölkerung, den

fortgeschrittenen Grad der Digitalisierung und den großen Datenbestand des öffentlichen Sektors bestmöglich zu nutzen. Zugleich sieht sie vor, internationale Verbindungen bei Forschung und Investitionen aufzubauen und Anreize für Privatinvestitionen zu setzen. Finnland hofft, sein Wirtschaftswachstum dank KI bis 2035 zu verdoppeln. Dabei sind acht Handlungsachsen vorgesehen, um Wachstum, Produktivität und Lebensqualität durch KI zu steigern: 1. Wettbewerbsfähigkeit der Unternehmen stärken, 2. Daten in allen Sektoren nutzen, 3. Einführung von KI beschleunigen und vereinfachen, 4. erstklassige fachliche Kompetenz sichern, 5. Anreize für mutige Entscheidungen und Investitionen setzen, 6. besten öffentlichen Dienst der Welt schaffen, 7. neue Kooperationsmodelle einführen und 8. Finnland als Vorreiter im KI-Zeitalter positionieren. Der Strategiebericht geht besonders darauf ein, wie KI genutzt werden kann, um die öffentlichen Dienstleistungen zu verbessern. So soll die finnische Einwanderungsbehörde beispielsweise das intelligente nationale Kundenservice-Netzwerk Aurora nutzen, um Anfragen mehrsprachig zu beantworten (Finnland, 2017[16]).

Außerdem wurde im Februar 2018 eine Förderagentur für Forschung und kommerzielle Projekte im KI-Bereich eingerichtet. Die Agentur wird 200 Mio. EUR an Zuschüssen und Anreizen für den Privatsektor und insbesondere KMU bereitstellen. Amtlichen Angaben zufolge sind rd. 250 Unternehmen in der KI-Entwicklung tätig. Die KI-Anstrengungen, die mit tief greifenden Reformen einhergehen, sollen auch den Fachkräften und Patienten im finnischen Gesundheitssektor zugutekommen (Sivonen, 2017[17]). Außerdem soll die Rolle des staatlich finanzierten Technical Research Centre of Finland und der Finnish Funding Agency for Technology and Innovation erweitert werden.

Frankreich

Der französische Staatspräsident Emmanuel Macron hat am 29. März 2018 Frankreichs KI-Strategie vorgestellt. Sie sieht vor, bis 2022 öffentliche Mittel in Höhe von 1,5 Mrd. EUR bereitzustellen, damit Frankreich zu einem führenden Akteur in der KI-Forschung und -Innovation werden kann. Die Maßnahmen basieren zum großen Teil auf den Empfehlungen im Bericht des Parlamentsabgeordneten Cédric Villani (Villani, 2018[18]). Die Strategie fordert Investitionen in öffentliche Forschung und Bildung, den Aufbau von Forschungshubs von Weltformat, die durch öffentlich-private Partnerschaften mit der Wirtschaft verbunden sind, sowie die Anwerbung führender ausländischer oder im Ausland tätiger französischer KI-Forscher. Um in Frankreich ein KI-Ökosystem aufzubauen, sollen bestehende Branchen technologisch aufgerüstet werden. Die Impulse sollen dabei zunächst von Anwendungen im Gesundheits-, Umwelt-, Verkehrs- und Verteidigungsbereich ausgehen. Der Datenzugang soll Priorität haben. Dazu ist geplant, „Datenallmenden“ für eine wohlfahrtssteigernde gemeinsame Datennutzung von öffentlichen und privaten Akteuren einzurichten, das Urheberrecht anzupassen, um Data Mining zu erleichtern, und Daten des öffentlichen Sektors, wie beispielsweise des Gesundheitssektors, Partnern aus der Wirtschaft zugänglich zu machen.

Außerdem skizziert die Strategie erste Lösungsansätze für KI-induzierte Disruptionen. Für Datentransfers in außereuropäische Länder sieht sie strenge Regeln vor (Thompson, 2018[19]). Vorgesehen ist auch die Einrichtung einer zentralen Datenagentur mit einem Team von rd. 30 Sachverständigen für KI-Anwendungen in der staatlichen Verwaltung. Im Zentrum der ethischen und philosophischen Grundsätze der Strategie steht das Prinzip der Algorithmentransparenz. So sollen z. B. von französischen Behörden oder mit öffentlichen Mitteln entwickelte Algorithmen offengelegt werden. Der Schutz der Privatsphäre und anderer Menschenrechte soll *by design*, d. h. durch Technikgestaltung gewährleistet

werden. Für von KI bedrohte Berufe sind verstärkte Weiterbildungsmaßnahmen vorgesehen. Die Strategie fordert zudem Experimentalprogramme für beschäftigungspolitische Innovationen und einen Dialog über die Frage, wie der durch KI geschaffene Mehrwert entlang der Wertschöpfungskette verteilt werden kann. Ende März 2018 wurde in Frankreich außerdem ein Bericht zum Thema „KI und Arbeit“ veröffentlicht (Benhamou, S. und L. Janin, 2018[20]).

Indien

Indien hat seine KI-Strategie im Juni 2018 veröffentlicht. Mit ihr will Indien eine führende Nation im KI-Bereich werden. Unter dem Motto #AIFORALL liegt der Schwerpunkt dabei auf Kompetenzentwicklung sowie sozialem und inklusivem Wachstum. Besonders im Fokus stehen KI-Anwendungen für Gesundheitsversorgung, Landwirtschaft, Bildung, Smart Cities und Verkehr. Aufgrund schwacher Forschungskapazitäten und fehlender Daten-Ökosysteme ist es in Indien schwierig, das Potenzial von KI voll auszuschöpfen. Die Strategie enthält mehrere Empfehlungen, um das zu ändern: Indien sollte zweigliedrige Forschungseinrichtungen schaffen (für Grundlagenforschung und angewandte Forschung) und Lernplattformen einrichten, um die Kompetenzen der Erwerbsbevölkerung zu entwickeln. Außerdem sollten gezielt Datensätze und Inkubationshubs für Start-ups aufgebaut werden. Des Weiteren sollte ein Rechtsrahmen für Datenschutz und Cybersicherheit geschaffen werden (Indien, 2018[21]).

Italien

Italien veröffentlichte im März 2018 ein Weißbuch zum Thema „Künstliche Intelligenz im Dienst des Bürgers”, das von einer Taskforce der Agenzia per l'Italia Digitale erstellt wurde. Im Mittelpunkt dieses Berichts steht die Frage, wie die öffentliche Verwaltung KI-Technologien nutzen kann, um die Servicequalität für Bürger und Unternehmen, die Effizienz des öffentlichen Dienstes sowie die Nutzerzufriedenheit zu steigern. Der Bericht befasst sich mit den Herausforderungen, die sich beim Einsatz von KI für öffentliche Dienstleistungen stellen, insbesondere in Bezug auf Ethik, Technologie, Datenverfügbarkeit und Wirkungsmessung. Außerdem enthält er Empfehlungen zu folgenden Themen: Förderung einer nationalen Plattform für gelabelte Daten, Algorithmen und Lernmodelle, Kompetenzentwicklung sowie Gründung eines nationalen Kompetenzzentrums und eines transdisziplinären Zentrums für KI. Darüber hinaus fordert der Bericht Leitlinien und Verfahren zur Verbesserung der Kontrolle und zur Erleichterung des Datenaustauschs zwischen allen europäischen Ländern im Hinblick auf von bzw. auf KI ausgeübte Cyberangriffe (Italien, 2018[22]).

Japan

Das japanische Kabinettsamt hat im April 2016 einen Strategierat für KI-Technologie eingerichtet, um FuE und gewerbliche Anwendungen im KI-Bereich zu fördern. Im März 2017 veröffentlichte der Rat die Artificial Intelligence Technology Strategy, in der zentrale Fragen behandelt werden. Dazu gehört die Notwendigkeit, die Investitionen zu steigern, Datenzugang und Datennutzung zu erleichtern und die Zahl der KI-Forscher und -Ingenieure zu erhöhen. Außerdem wurden strategische Bereiche aufgezeigt, in denen KI einen erheblichen Nutzen bringen könnte: Produktivität, Gesundheit, medizinische Versorgung und Lebensqualität, Mobilität und Informationssicherheit (Japan, 2017[23]).

Die vom Kabinettsamt im Juni 2018 veröffentlichte *Integrated Innovation Strategy* sieht mehrere KI-Maßnahmen vor (Japan, 2018[24]), darunter Diskussionen mit Beteiligung

mehrerer Interessengruppen über ethische, rechtliche und gesellschaftliche Fragen der KI. Im Anschluss an diese Diskussionen veröffentlichte das Kabinettsamt im April 2019 die *Social Principles for Human-centric AI* (Japan, 2019[25]).

Auf der Tagung der zuständigen Minister für IKT der G7, die im April 2016 in Takamatsu stattfand, schlug Japan vor, gemeinsame Grundsätze für KI-Forschung und -Entwicklung zu erarbeiten. Die Expertengruppe Conference toward AI Network Society erstellte daraufhin die *Draft AI R&D Guidelines for International Discussions*, die das japanische Ministerium für Innere Angelegenheiten und Kommunikation im Juli 2017 veröffentlichte. Diese Leitlinien zielen in erster Linie darauf ab, die Vorteile und Risiken von KI-Netzen abzuwägen, Technologieneutralität zu sichern und übermäßige Belastungen für die Entwickler zu vermeiden. Die Leitlinien bestehen aus neun Grundsätzen, die Forscher und Entwickler von KI-Systemen berücksichtigen sollten (Japan, 2017[26]). Sie sind in Tabelle 5.2 zusammengefasst. Im Anschluss an die Diskussionen veröffentlichte die Expertengruppe im Juli 2018 die *Draft AI Utilization Principles* (Japan, 2018[27]).

Tabelle 5.2. Grundsätze für FuE im KI-Bereich

Grundsatz	Empfehlungen für Entwickler
I. Kooperation	Interkonnektivität und Interoperabilität der KI-Systeme im Blick haben
II. Transparenz	Auf die Überprüfbarkeit der Inputs und Outputs von KI-Systemen und die Nachvollziehbarkeit ihrer Entscheidungen achten
III. Kontrollierbarkeit	Auf die Kontrollierbarkeit von KI-Systemen achten
IV. Schutz vor unbeabsichtigten Bedrohungen	Sicherstellen, dass von Aktoren oder sonstigen Geräten von KI-Systemen keine Gefahren für Leib und Leben sowie das Eigentum von Nutzern oder Dritten ausgehen
V. Schutz vor böswilligen Angriffen	Sicherheit von KI-Systemen im Blick haben
VI. Privatsphäre	Berücksichtigen, dass KI-Systeme die Privatsphäre von Nutzern oder Dritten nicht beeinträchtigen dürfen
VII. Ethik	Die Menschenwürde und die Autonomie des Einzelnen bei der FuE von KI-Systemen respektieren
VIII. Nutzerunterstützung	Berücksichtigen, dass KI-Systeme die Nutzer unterstützen und ihnen angemessene Wahlmöglichkeiten geben sollen
IX. Rechenschaftspflicht	Rechenschaftspflicht gegenüber den betroffenen Akteuren und insbesondere den Nutzern von KI-Systemen erfüllen

Quelle: Japan (2017[26]), *Draft AI R&D Guidelines for International Discussions*, www.soumu.go.jp/main_content/000507517.pdf.

Kanada

Kanada will sich als führender Akteur im KI-Bereich positionieren, was sich insbesondere an der im März 2017 eingeleiteten Pan-Canadian AI Strategy zeigt (CIFAR, o.J.[28]). Diese Strategie wird von dem gemeinnützigen Canadian Institute for Advanced Research geleitet und mit staatlichen Mitteln in Höhe von 125 Mio. CAD (100 Mio. USD) gefördert. Die für einen Fünfjahreszeitraum bereitgestellten Mittel dienen dazu, das Kompetenzniveau der kanadischen Bevölkerung anzuheben, die KI-Forschung in Kanada zu unterstützen und KI-Forschungsergebnisse im öffentlichen und privaten Sektor anzuwenden. Die Pan-Canadian AI Strategy hat folgende Ziele:

1. die Zahl der Forscher und Hochschulabsolventen im KI-Bereich erhöhen
2. vernetzte wissenschaftliche Exzellenzzentren in Kanadas drei großen KI-Instituten einrichten: Edmonton (Alberta Machine Intelligence Institute), Montreal (Montreal Institute for Learning Algorithms) und Toronto (Vector Institute for Artificial Intelligence)

3. ein globales Programm für „KI in der Gesellschaft“ entwickeln und weltweit eine Vordenkerrolle in Bezug auf die ökonomischen, sozialen, ethischen, politischen und rechtlichen Auswirkungen der Fortschritte im KI-Bereich einnehmen
4. eine nationale Forschungsgemeinschaft im KI-Bereich aufbauen

Die kanadische Bundesregierung plant, dem National Research Council Canada (NRC) in einem Siebenjahreszeitraum Mittel für Forschungsinvestitionen von insgesamt 50 Mio. CAD (40 Mio. USD) bereitzustellen. Dabei geht es um die Anwendung von KI in folgenden Bereichen: Datenanalyse, AI for Design, Cybersicherheit, kanadische indigene Sprachen, Förderung von Superclustern und Kooperationszentren auf Bundesebene unter Mitwirkung kanadischer Universitäten sowie strategische Partnerschaften mit internationalen Partnern.

Zusätzlich zu den Zuschüssen auf Bundesebene werden auch Mittel auf regionaler Ebene vergeben: Die Provinzregierung von Quebec stellt 100 Mio. CAD (80 Mio. USD) für KI-Akteure in Montreal bereit und Ontario fördert das Vector Institute for Artificial Intelligence mit 50 Mio. CAD (40 Mio. USD). 2016 vergab der Canada First Research Excellence Fund Mittel in Höhe von 93.6 Mio. CAD (75 Mio. USD) an drei Universitäten, die Spitzenforschung im Bereich Deep Learning betreiben: die Université de Montréal, die Polytechnique Montréal und die HEC Montréal. Facebook und andere dynamische Privatunternehmen wie Element AI sind ebenfalls in Kanada aktiv.

Die Regierung von Quebec plant, ein Weltobservatorium für die sozialen Auswirkungen von KI und digitalen Technologien einzurichten (Fonds de recherche du Québec, 2018[29]). Im März 2018 wurde ein Workshop organisiert, bei dem über den Auftrag und den Aufbau dieses Observatoriums, seine Governance und Finanzierung, Möglichkeiten der internationalen Zusammenarbeit sowie prioritäre Bereiche und Themen diskutiert wurde. Die Regierung von Quebec hat 5 Mio. CAD (3,7 Mio. USD) für die Einrichtung des Observatoriums bereitgestellt.

Außerdem arbeitet Kanada mit internationalen Partnern zusammen, um KI-Initiativen voranzubringen. Die kanadische und die französische Regierung kündigten beispielsweise im Juli 2018 an, sich gemeinsam für die Einrichtung eines Internationalen Panels für KI einzusetzen. Die Aufgabe des Panels soll darin bestehen, eine verantwortungsvolle Anwendung von KI zu unterstützen, die menschenzentriert ist und den Menschenrechten, den Prinzipien der Teilhabe aller und der Vielfalt, der Innovation und dem Wirtschaftswachstum verpflichtet ist.

Korea

Die koreanische Regierung veröffentlichte im März 2016 die Intelligent Information Industry Development Strategy. Sie kündigte an, bis 2020 öffentliche Investitionen in Höhe von 1 Bill. KRW (940 Mio. USD) im Bereich der KI und damit verbundener Informationstechnologien wie IoT und Cloud-Computing zu tätigen. Die Strategie zielt darauf ab, ein neues intelligentes Ökosystem für die IT-Branche zu schaffen und bis 2020 Privatinvestitionen in Höhe von 2,5 Bill. KRW (2,3 Mrd. USD) zu mobilisieren. Damit verfolgt die Regierung drei Ziele: Erstens plant sie die Einrichtung von Flagship-Projekten für die KI-Entwicklung, beispielsweise in den Bereichen sprachliche, visuelle, räumliche und emotionale Intelligenz. Zweitens hat sie die Absicht, die KI-bezogenen Kompetenzen der Arbeitskräfte zu stärken. Und drittens wird sie den Datenzugang und die Datennutzung durch staatliche Stellen, Unternehmen und Forschungseinrichtungen fördern (Korea, 2016[30]).

Im Dezember 2016 veröffentlichte die koreanische Regierung den Mid- to Long-Term Master Plan in Preparation for the Intelligence Information Society. Der Plan enthält nationale Maßnahmen, um auf die Veränderungen und Herausforderungen der „Vierten industriellen Revolution" zu reagieren. Um ihre Vision einer „menschenzentrierten intelligenten Gesellschaft" zu verwirklichen, will die koreanische Regierung die Voraussetzungen für eine intelligente IT von Weltformat schaffen. Diese IT könnte branchenübergreifend angewendet werden und auch die Sozialpolitik modernisieren. Um den Plan umzusetzen, richtet die Regierung derzeit großangelegte Testbeds für die Entwicklung neuer Dienstleistungen und Produkte ein, wobei es auch um bessere öffentliche Dienstleistungen geht (Korea, 2016[31]).

Im Mai 2018 veröffentlichte die koreanische Regierung einen nationalen Plan, um die FuE-Kapazitäten im KI-Bereich zu stärken. Der Plan, in dessen Umsetzung bis 2022 2,2 Bill. KRW (2 Mrd. USD) investiert werden sollen, sieht vor, sechs KI-Forschungsinstitute einzurichten, KI-Talente durch 4 500 KI-Stipendien und kurze Intensivschulungen zu fördern und die Entwicklung von KI-Chips zu beschleunigen (Peng, 2018[32]).

Mexiko

In Mexiko hat der Nationale Rat für Wissenschaft und Technologie (Consejo Nacional de Ciencia y Tecnología) 2004 ein Forschungszentrum für künstliche Intelligenz gegründet, das die Entwicklung intelligenter Systeme leitet.

Im Juni 2018 wurde ein Weißbuch mit dem Titel „Towards an AI Strategy in Mexico: Harnessing the AI Revolution" veröffentlicht.[1] In diesem Bericht wird festgestellt, dass Mexiko im „AI Readiness Index" von Oxford Insights unter 35 OECD-Ländern auf Platz 22 rangiert. Dieser Gesamtwert ergibt sich aus dem Durchschnitt von 9 Messgrößen, die von digitalen Kompetenzen bis zu Innovation in der staatlichen Verwaltung reichen. In den Bereichen Open Data und digitale Infrastruktur schneidet Mexiko gut ab, in Bereichen wie fachliche Kompetenzen, Digitalisierung und Innovation im öffentlichen Sektor dagegen schlecht. Der Bericht empfiehlt politische Maßnahmen, um KI in Mexiko weiterzuentwickeln und zu verbreiten. Diese Empfehlungen erstrecken sich über fünf Bereiche der staatlichen Politik: staatliche Verwaltung und öffentliche Dienstleistungen; FuE; Kapazitäten, Kompetenzen und Bildung; Daten und digitale Infrastruktur; Ethik und Regulierung (Martinho-Truswell, 2018[33]).

Niederlande

Die Regierung der Niederlande verpflichtete sich 2018 durch die Verabschiedung ihrer nationalen Digitalisierungsstrategie auf zwei Ziele: Erstens will sie soziale und wirtschaftliche Chancen nutzen. Zweitens will sie für günstige Rahmenbedingungen sorgen. Dazu will sie insbesondere in den Bereichen Kompetenzen, Datenpolitik, Vertrauen und Resilienz, Grundrechte und Ethik (z. B. Einfluss von Algorithmen auf Autonomie und Gleichbehandlung) sowie KI-zentrierte Forschung und Innovation ansetzen. Im Oktober 2018 veröffentlichte AINED, ein Zusammenschluss von Vertretern aus Wirtschaft und Wissenschaft, Ziele und Maßnahmen für einen nationalen KI-Plan. Dabei geht es um einen besseren Zugang zu KI-Talenten und -Kompetenzen sowie zu hochwertigen öffentlichen Daten. Ein weiteres Ziel besteht darin, die KI-gestützte Geschäftsentwicklung zu erleichtern und eine umfassende Nutzung von KI in der staatlichen Verwaltung zu fördern. Außerdem ist vorgesehen, sozioökonomische und ethische Regelwerke für KI zu schaffen, die Zusammenarbeit zwischen dem öffentlichen und dem privaten Sektor in wichtigen

Bereichen und Wertschöpfungsketten zu fördern und die Niederlande als KI-Forschungszentrum von Weltrang zu etablieren. Die niederländische Regierung plant, bis Mitte 2019 einen ressortübergreifenden strategischen Aktionsplan vorzulegen. Dabei will sie den AINED-Bericht, den koordinierten Plan der EU und die Diskussionen der von der Europäischen Kommission ernannten hochrangigen Expertengruppe für KI (HEG-KI) berücksichtigen.

Norwegen

Norwegen führt KI-Initiativen im Rahmen der Digital Agenda for Norway und eines langfristigen Plans für Forschung und Hochschulbildung durch. Dabei geht es um Folgendes:

- Gründung mehrerer KI-Laboratorien, darunter das Norwegian Open AI-Lab an der Norwegian University for Science and Technology. Das Open AI-Lab wird von mehreren Unternehmen unterstützt. Es befasst sich mit Bereichen wie Energie, Meerestechnologie, Aquakultur, Telekommunikation, digitales Banking sowie Gesundheit und Biomedizin, in denen Norwegen eine starke internationale Position innehat.
- Reform des Bildungssystems zur Förderung des lebenslangen Lernens. Das Reformprogramm, für das ein Budget von 130 Mio. NOK (16 Mio. USD) vorgesehen ist, soll den Arbeitskräften helfen, Kompetenzen in KI, Gesundheitsversorgung und anderen Bereichen zu entwickeln oder zu aktualisieren.
- Open-Data-Strategie. Diese Strategie sieht vor, dass staatliche Stellen ihre Daten über Anwendungsprogrammschnittstellen in maschinenlesbaren Formaten zugänglich machen und die verfügbaren Datensätze in einem gemeinsamen Katalog registrieren.
- Plattform zur Entwicklung von Leitlinien und ethischen Grundsätzen für die Nutzung von KI.
- Regulierungsreform, um Tests für selbstfahrende Fahrzeuge auf der Straße zu erlauben – darunter Testfahrten ohne Fahrer im Fahrzeug.

Russische Föderation

Die russische Regierung legte im Juli 2017 ihre Strategie für die digitale Wirtschaft vor. Zu den Prioritäten dieser Strategie gehört die Nutzung der Chancen der KI-Entwicklung, wozu u. a. günstige rechtliche Bedingungen zur Erleichterung von FuE-Tätigkeiten geschaffen werden sollen. Außerdem sieht die Strategie Anreize für staatliche Unternehmen vor, sich an den nationalen Forschungsgemeinschaften (Kompetenzzentren) zu beteiligen. Darüber hinaus wird die Entwicklung nationaler Standards für KI-Technologien gefördert (Russische Föderation, 2017[34]). Vor der Digitalstrategie hatte die Regierung bereits in verschiedene KI-Projekte investiert und Instrumente für den Aufbau öffentlich-privater Partnerschaften geschaffen. Der russische KI-Verband fördert die Zusammenarbeit zwischen Wissenschaft und Wirtschaft, um den Technologietransfer an die Unternehmen zu erleichtern.

Saudi-Arabien

Saudi-Arabien hat 2016 seine „Vision 2030“ vorgestellt. Sie enthält einen wirtschaftlichen Reformplan zur Stimulierung neuer Branchen, zur Diversifizierung der Wirtschaft und zur

Erleichterung öffentlich-privater Geschäftsmodelle. Ziel ist es, die Abhängigkeit des Landes von Öleinnahmen zu reduzieren. In der „Vision 2030“ wird der digitale Wandel als Schlüsselinstrument betrachtet, um die Wirtschaft durch die Nutzung von Daten, KI und industrieller Automatisierung weiterzuentwickeln. Prioritäre Sektoren, u. a. für die Einrichtung von Innovationszentren, sind Gesundheitsversorgung, staatliche Dienstleistungen, nachhaltige Energie- und Wasserwirtschaft, das Verarbeitende Gewerbe sowie Mobilität und Verkehr. Die Regierung entwirft derzeit ihre nationale KI-Strategie, die darauf abzielt, bis 2030 ein innovatives und ethisches KI-Ökosystem in Saudi-Arabien aufzubauen.

Eine wichtige Rolle spielen dabei der Breitband- und 5G-Netzausbau sowie der Datenzugang und die Datensicherheit. Außerdem führt das Land mehrere Smart-City-Projekte durch, um Anreize für die rasche Einführung von KI-Konzepten und -Anwendungen zu setzen und neue Lösungen zu fördern. Ausgangspunkt ist das Smart-City-Megaprojekt NEOM, das 2017 eingeleitet wurde und für das Investitionen in Höhe von 1,8 Bill. SAR (500 Mio. USD) vorgesehen sind. Darüber hinaus beteiligt sich Saudi-Arabien aktiv an den globalen Diskussionen über mögliche Governance-Rahmen für KI.

Schweden

Im Mai 2018 veröffentlichte die schwedische Regierung einen Bericht mit dem Titel *Artificial Intelligence in Swedish Business and Society*. Der Bericht, der die KI-Forschung und -Innovation in Schweden stärken soll, beschreibt sechs strategische Schwerpunkte: 1. Industrie, 2. Tourismus und Verkehr, 3. nachhaltige und intelligente Städte, 4. Gesundheitsversorgung, 5. Finanzdienstleistungen und 6. Sicherheit, einschließlich Polizei und Zoll. Er betont die Notwendigkeit, im Bereich der Forschung, Bildung und Innovation eine kritische Masse zu erreichen. Außerdem wird ein kooperativer Ansatz in Bezug auf Investitionen in Forschung und Entwicklung, Bildung, Regulierung und Arbeitskräftemobilität gefordert (Vinnova, 2018[35]).

Singapur

Die Infocomm Media Development Authority veröffentlichte im Mai 2018 den *Digital Economy Framework for Action*. Dieser Handlungsrahmen zielt darauf ab, Singapur zu einer führenden digitalen Wirtschaft zu machen. KI gehört dabei zu den Spitzentechnologien, die den digitalen Wandel in Singapur vorantreiben sollen (Singapur, 2018[36]).

Die Personal Data Protection Commission veröffentlichte im Januar 2019 ein Musterregelwerk für KI-Governance, um eine verantwortungsvolle Einführung von KI in Singapur zu fördern. Das Musterregelwerk enthält konkrete Leitlinien für die praktische Umsetzung ethischer Grundsätze. Es basiert auf einem Diskussionspapier und den Beiträgen des Regulators’ Roundtable, einer Expertengruppe aus Vertretern von Regulierungsbehörden und öffentlichen Stellen. Das Musterregelwerk kann auf freiwilliger Basis umgesetzt werden. Es dient außerdem als Basis für die Entwicklung sektorspezifischer Governance-Rahmen für KI.

Im Juni 2018 wurde der Advisory Council on the Ethical Use of AI and Data gegründet, in dem verschiedene Akteure vertreten sind. Er berät die Regierung von Singapur zu ethischen, rechtlichen, regulatorischen und politischen Fragen, die sich aus der kommerziellen Einführung von KI ergeben. Das Ziel besteht darin, eine rechenschaftspflichtige und verantwortungsvolle Einführung von KI-Produkten und -Dienstleistungen in den einzelnen Branchen zu fördern.

Im September 2018 hat Singapur ein fünfjähriges Forschungsprogramm zur Governance der künstlichen Intelligenz und der Datennutzung eingeleitet, um das Land zu einem führenden Wissenszentrum mit internationaler Expertise im Bereich der KI-Politik und -Regulierung zu entwickeln. Zuständig für die Durchführung des Programms ist das Centre for AI and Data Governance der Singapore Management University School of Law. Das Zentrum befasst sich mit praxisbezogener KI-Forschung, die für Industrie, Gesellschaft und Unternehmen relevant ist.

Tschechische Republik

Die tschechische Regierung gab 2018 eine Studie über die Umsetzung von KI in Auftrag, um strategische Ziele festzulegen und die Verhandlungen auf europäischer und internationaler Ebene zu unterstützen. Ein Team, bestehend aus Wissenschaftlern des Technologiezentrums der tschechischen Akademie der Wissenschaften, der Technischen Hochschule in Prag und des Instituts für Staat und Recht der tschechischen Akademie der Wissenschaften, legte einen Bericht mit dem Titel *Analysis of the Development Potential of Artificial Intelligence in the Czech Republic* vor (OGCR, 2018[37]). Dieser Bericht untersucht: 1. den derzeitigen Stand der KI-Einführung in der Tschechischen Republik, 2. die möglichen Auswirkungen von KI auf den tschechischen Arbeitsmarkt und 3. die ethischen, rechtlichen und regulatorischen Aspekte der KI-Entwicklung im Land.

Türkei

Die TÜBITAK, die führende staatliche Einrichtung für das Management und die Finanzierung der Forschung in der Türkei, hat zahlreiche FuE-Projekte zu künstlicher Intelligenz finanziert. Sie plant eine multilaterale Ausschreibung für KI-Projekte im Rahmen des zwischenstaatlichen Innovationsnetzwerks EUREKA. Das türkische Ministerium für Wissenschaft und Technologie hat im Kontext der Industrial Digital Transformation Platform eine nationale digitale Roadmap entwickelt. Darin geht es auch um technische Fortschritte in neuen digitalen Technologien wie KI.

Ungarn

In Ungarn wurde im Oktober 2018 eine KI-Koalition gegründet. Es handelt sich um eine Partnerschaft zwischen staatlichen Stellen, führenden IT-Unternehmen und Universitäten. Die Koalition erarbeitet derzeit eine KI-Strategie, um Ungarn als KI-Innovator zu positionieren. Außerdem untersucht sie die sozialen und wirtschaftlichen Auswirkungen von KI. Zu ihren Mitgliedern zählen etwa 70 universitäre Forschungszentren, Unternehmen und Behörden. Sie gilt als Forum für sektorübergreifende Zusammenarbeit im Bereich der KI-Forschung und -Entwicklung. Die Budapest University of Technology and Economics und die Eötvös Loránd Universität sind Teil des Konsortiums, das das mit 20 Mio. EUR ausgestattete Projekt AI4EU leitet, mit dem eine europäische Plattform für KI auf Abruf aufgebaut werden soll.

Vereinigtes Königreich

Die im März 2017 veröffentlichte *UK Digital Strategy* würdigt KI als entscheidenden Faktor für das Wachstum der Digitalwirtschaft des Vereinigten Königreichs (Vereinigtes Königreich, 2017[38]). Sie stellt britischen Universitäten Fördermittel in Höhe von 17,3 Mio. GBP (22,3 Mio. USD) für die Entwicklung von KI- und Robotertechnologien zur Verfügung. Die Regierung hat die Investitionen in KI-Forschung und -Entwicklung für die

nächsten vier Jahre um 4,7 Mrd. GBP (6,6 Mrd. USD) erhöht. Diese Mittel sollen teilweise über ihren Industrial Strategy Challenge Fund vergeben werden.

Im Oktober 2017 veröffentlichte die Regierung einen unter Federführung der Wirtschaft erstellten Bericht über die KI-Branche des Vereinigten Königreichs. Diesem Bericht zufolge ist das Vereinigte Königreich ein internationales Zentrum für KI-Expertise, was u. a. der Leistung von Pionieren der Informatik wie Alan Turing zu verdanken sei. Laut Schätzungen der britischen Regierung könnte KI 579,7 Mrd. GBP (814 Mrd. USD) zur inländischen Wirtschaft beitragen. Zu den im Vereinigten Königreich genutzten KI-Tools gehören ein persönlicher Gesundheitsführer (Your.MD), ein für Bankkunden entwickelter Chatbot und eine Plattform, die Kindern beim Lernen hilft und Lehrer unterstützt, personalisierte Bildungsprogramme zu entwickeln. Der Bericht enthält 18 Empfehlungen, darunter beispielsweise die, den Datenzugang und Datenaustausch durch die Entwicklung von Datentreuhändern zu verbessern. Außerdem wird empfohlen, die KI-Kompetenzen durch von der Wirtschaft geförderte KI-Masterabschlüsse zu verbessern. Darüber hinaus werden folgende Prioritäten genannt: die KI-Forschung durch eine Koordinierung der Nachfrage nach Rechnerkapazitäten zwischen den relevanten Institutionen maximieren, die Einführung von KI durch einen britischen KI-Rat unterstützen und ein Rahmenkonzept entwickeln, um die Transparenz und die Verantwortlichkeit KI-gestützter Entscheidungen zu verbessern (Hall, W. und J. Pesenti, 2017[39]).

Im November 2017 hat die britische Regierung auch eine Industriestrategie veröffentlicht. Darin wird KI als eine von vier „Grand Challenges“ bezeichnet, die das Vereinigte Königreich bewältigen muss, um in Zukunftsbranchen eine Spitzenposition einzunehmen und von den weitreichenden Veränderungen zu profitieren, zu denen es weltweit kommt (Vereinigtes Königreich, 2017[40]). Im April 2018 veröffentlichte das Vereinigte Königreich den *AI Sector Deal*, der ein Investitionspaket von 950 Mio. GBP (1,2 Mrd. USD) vorsieht. Dabei soll auf den Stärken des Vereinigten Königreichs aufgebaut werden, um ein KI-Ökosystem von Weltformat zu errichten. Das Programm hat drei Schwerpunkte: Kompetenzen und Talente, Anreize für die Einführung von KI sowie Daten und Infrastruktur (Vereinigtes Königreich, 2018[41]).

Um den AI Sector Deal umzusetzen und die Einführung von KI allgemein zu fördern, hat die Regierung das Office for Artificial Intelligence (OAI) eingerichtet. Außerdem hat sie das Centre for Data Ethics and Innovation gegründet. Das Zentrum hat die Aufgabe, die Governance zu verbessern, um Innovationen zu ermöglichen und zugleich das Vertrauen der Öffentlichkeit zu sichern. Es soll die Regierung unabhängig und sachkundig zu den Maßnahmen beraten, die für sichere, ethische und bahnbrechende Innovationen in datengestützten und KI-basierten Technologien erforderlich sind. Das Zentrum plante, zu diesem Zweck bis Ende 2019 ein Datentreuhandmodell zu erproben. Ein KI-Rat, der sich auf das Fachwissen der Wirtschaft stützt, arbeitet zudem eng mit dem OAI zusammen (Vereinigtes Königreich, 2018[42]).

Vereinigte Staaten

Am 11. Februar 2019 unterzeichnete Präsident Trump die Executive Order 13859 über die „Aufrechterhaltung der amerikanischen Führungsrolle im Bereich der künstlichen Intelligenz“, mit der die amerikanische KI-Initiative eingeleitet wurde. Diese Initiative ist auf fünf Handlungsachsen aufgebaut: 1. in KI-Forschung und -Entwicklung investieren, 2. KI-Ressourcen mobilisieren, 3. Leitlinien für KI-Regulierung und technische Standards festlegen, 4. Arbeitskräfte für KI ausbilden sowie 5. ein internationales Umfeld schaffen,

das die amerikanische Forschungs- und Innovationstätigkeit im KI-Bereich fördert und Märkte für amerikanische KI-Unternehmen öffnet.

Dieser Initiative ging eine Reihe anderer Regierungsmaßnahmen zur Stärkung der amerikanischen Führungsrolle im KI-Bereich voran. Im Mai 2018 fand im Weißen Haus der erste KI-Gipfel (Summit on AI for American Industry) statt, an dem Akteure aus der Wirtschaft, Wissenschaftler und führende Regierungsvertreter teilnahmen. Die Teilnehmer betonten, wie wichtig es sei, Hindernisse für KI-Innovationen in den Vereinigten Staaten zu beseitigen und die Zusammenarbeit im FuE-Bereich mit den amerikanischen Verbündeten zu fördern. Außerdem verwiesen sie auf die Notwendigkeit, das Bewusstsein für KI zu schärfen, damit die Öffentlichkeit besser verstehe, wie diese Technologien funktionieren und wie die Menschen in ihrem täglichen Leben von ihnen profitieren können. Im gleichen Monat veröffentlichte das Weiße Haus ein Informationsblatt mit dem Titel *Artificial Intelligence for the American People* (Vereinigte Staaten, 2018[43]), in dem die KI-bezogenen Regelungen und Maßnahmen der Regierung aufgelistet wurden. Diese Maßnahmen umfassen höhere öffentliche Mittel für KI-Forschung und -Entwicklung und eine Regulierungsreform, um die Entwicklung und Nutzung von Drohnen und selbstfahrenden Autos zu erleichtern. Zu den weiteren Prioritäten gehören die Ausbildung von Fachleuten in den Bereichen Natur- und Ingenieurwissenschaften, Technik und Mathematik (wobei der Schwerpunkt besonders auf Informatik liegt) sowie bessere Möglichkeiten zur Nutzung von Daten der Bundesbehörden für KI-Forschung und -Anwendungen.

In den FuE-Budgets des Präsidenten für die Finanzjahre 2019 und 2020 werden KI und maschinelles Lernen als Prioritäten genannt. Konkret geht es dabei u. a. um KI-Grundlagenforschung in der National Science Foundation und angewandte FuE im Verkehrsministerium. Weitere Forschungsprioritäten sind die Analyse von Gesundheitsdaten in den National Institutes of Health und die KI-Rechnerinfrastruktur im Energieministerium. Insgesamt sind die Investitionen der US-Regierung in FuE für künstliche Intelligenz und verwandte Technologien seit 2015 um mehr als 40 % gestiegen.

Im September 2018 begann das Select Committee on AI des National Science and Technology Council, den National Artificial Intelligence Research and Development Strategic Plan zu aktualisieren. Seit der Veröffentlichung des Plans im Jahr 2016 haben sich die Technologien, Anwendungsfälle und kommerziellen Nutzungsformen von KI rasch weiterentwickelt. Das Select Committee konsultiert derzeit die Öffentlichkeit zur Verbesserung des Plans, wobei insbesondere Akteure einbezogen werden, die im Bereich der KI-Forschung und -Entwicklung tätig sind oder von ihr betroffen sind.

Die Aus- und Weiterbildung der amerikanischen Erwerbsbevölkerung gehört ebenfalls zu den Prioritäten der US-Regierung. Präsident Trump hat eine Executive Order zur Einführung einer von der Wirtschaft anerkannten betrieblichen Berufsausbildung unterzeichnet und eine auf Kabinettsebene angesiedelte Taskforce zur Förderung der betrieblichen Berufsausbildung (Task Force on Apprenticeship Expansion) eingerichtet. Ebenso wie in dem bereits erwähnten Informationsblatt wurde auch in einem Presidential Memorandum auf die große Bedeutung hochwertiger Bildungsangebote in Mathematik, Natur- und Ingenieurwissenschaften und besonders Informatik hingewiesen. Mit diesem Memorandum wurden 200 Mio. USD an Zuschüssen bereitgestellt, wobei die Privatwirtschaft 300 Mio. USD beisteuert.

Der US-Kongress hat im Mai 2017 die parteiübergreifende KI-Arbeitsgruppe Artificial Intelligence Caucus ins Leben gerufen, deren Vorsitz die Abgeordneten John K. Delaney und Pete Olson führen (Vereinigte Staaten, 2017[44]). Die Arbeitsgruppe, in der Experten

aus der Wissenschaft, staatlichen Behörden und dem Privatsektor vertreten sind, beschäftigt sich mit den Auswirkungen von KI-Technologien. Der Kongress prüft derzeit Rechtsvorschriften, um auf Bundesebene einen KI-Beratungsausschuss einzurichten und Sicherheitsstandards für selbstfahrende Autos einzuführen.

Zwischenstaatliche Initiativen

G7 und G20

Auf der Tagung der für IKT zuständigen G7-Minister in Takamatsu (Japan) im April 2016 legte der japanische Minister für Innere Angelegenheiten und Kommunikation eine Reihe von Grundsätzen zu KI-Forschung und -Entwicklung zur Diskussion vor (G7, 2016[45]).

Die G7-Tagung der IKT- und Industrieminister, die im September 2017 unter italienischem Vorsitz in Turin stattfand, schloss mit einer Erklärung, in der die G7-Länder die enormen potenziellen Vorteile der KI für die Gesellschaft und die Wirtschaft anerkannten und einen menschenzentrierten KI-Ansatz forderten (G7, 2017[46]).

Im März 2018 kamen die Innovationsminister der G7 unter dem Vorsitz Kanadas in Montreal zusammen. Sie sprachen sich für die Vision einer menschenzentrierten KI aus und betonten den Zusammenhang zwischen Wirtschaftswachstum und Innovation im KI-Bereich. Außerdem verwiesen sie auf die Notwendigkeit, das Vertrauen in KI-Technologien zu stärken, ihre Verbreitung zu unterstützen und eine inklusive KI-Entwicklung und -Umsetzung zu fördern. Die G7-Mitglieder beschlossen Maßnahmen mit folgenden Zielen:

- in Grundlagenforschung und angewandte FuE in der Frühphase investieren, um KI-Innovationen herbeizuführen, unternehmerische Initiative im KI-Bereich fördern und die Arbeitskräfte auf die Automatisierung vorbereiten
- Forschungsförderung fortsetzen, insbesondere im Hinblick auf die Bewältigung gesellschaftlicher Herausforderungen, die Steigerung des Wirtschaftswachstums und die Prüfung ethischer Aspekte von KI sowie allgemeiner Fragen, beispielsweise in Bezug auf automatisierte Entscheidungssysteme
- die Öffentlichkeit für die tatsächlichen und potenziellen Vorteile und die weitreichenden Auswirkungen von KI sensibilisieren
- geeignete technische, ethische und technologieneutrale Ansätze voranbringen
- den freien Informationsfluss durch den Austausch von Best Practices und Anwendungsfällen in Bezug auf einen offenen, interoperablen und sicheren Zugang zu Verwaltungsdaten für die KI-Programmierung unterstützen
- diese G7-Erklärung weltweit verbreiten, um die Entwicklung und Zusammenarbeit im KI-Bereich auf internationaler Ebene zu fördern (G7, 2018[47])

Im Juni 2018 veröffentlichte die G7 in Charlevoix (Kanada) ein Kommuniqué, um eine menschenzentrierte KI und die kommerzielle Nutzung von KI zu fördern. Die G7-Mitglieder kamen außerdem überein, sich weiter für geeignete technische, ethische und technologieneutrale Ansätze einzusetzen.

Die Innovationsminister der G7 beschlossen, im Dezember 2018 in Kanada eine Multi-Stakeholder-Konferenz über KI zu veranstalten. Auf dieser Konferenz sollte erörtert werden, wie die positiven Transformationskräfte von KI genutzt werden können, um ein inklusives und nachhaltiges Wachstum zu fördern. Frankreich sollte im Rahmen seiner G7-Präsidentschaft im Jahr 2019 weitere KI-bezogene Initiativen vorschlagen.

Die G20 beschäftigt sich ebenfalls zunehmend mit KI, was sich auch an den Diskussionsvorschlägen Japans im Rahmen des G20-Vorsitzes im Jahr 2019 zeigt (G20, 2018[1]). Auf der G20-Tagung der Digitalminister in Salta 2018 wurden die Länder aufgefordert, Menschen und Unternehmen dazu zu befähigen, von der Digitalisierung und den neuen Technologien, beispielsweise 5G, IoT und KI, zu profitieren. Japan wurde ermutigt, seine Präsidentschaft im Jahr 2019 zu nutzen, um die 2018 von der G20 geleistete Arbeit, insbesondere im Hinblick auf KI, fortzusetzen.

OECD

OECD-Grundsätze zur Förderung des Vertrauens in KI und ihrer Verbreitung

Im Mai 2018 gründete der OECD-Ausschuss für digitale Wirtschaft eine Sachverständigengruppe für künstliche Intelligenz in der Gesellschaft (AIGO). Sie wurde eingerichtet, um Grundsätze für die Politik und die internationale Zusammenarbeit zu erarbeiten, die das Vertrauen in KI sowie deren Verbreitung fördern. Diese Grundsätze bildeten anschließend die Basis für die Empfehlung des Rats der OECD zu künstlicher Intelligenz (OECD, 2019[3]), der am 22. Mai 2019 vierzig Länder beitraten. Bereits 2018 hatte der Vorsitz der Tagung des Rats auf Ministerebene die OECD aufgefordert, „Multi-Stakeholder-Gespräche über die mögliche Ausarbeitung von Grundsätzen zu führen, die die Entwicklung und ethische Anwendung künstlicher Intelligenz im Dienste der Menschen unterstützen sollten“.

Der AIGO gehörten mehr als fünfzig Sachverständige aus verschiedenen Sektoren und Bereichen an, darunter Vertreter staatlicher Stellen, der Wirtschaft, der Fachwelt, der Arbeitnehmerorganisationen und der Zivilgesellschaft sowie der Europäischen Kommission und der UNESCO. Sie hielt vier Sitzungen ab: zwei im OECD-Hauptsitz in Paris, vom 24.-25. September und am 12. November 2018, eine beim Massachusetts Institute of Technology (MIT) in Cambridge, vom 16.-17. Januar 2019, und eine Abschlusstagung in Dubai, vom 8.-9. Februar 2019, am Rande des World Government Summit. Die von der AIGO festgelegten Grundsätze für eine verantwortungsvolle Steuerung vertrauenswürdiger KI sind für alle betroffenen Akteure relevant. Zu diesen Grundsätzen gehören: Achtung der Menschenrechte, Fairness, Transparenz und Erklärbarkeit, Robustheit und Sicherheit sowie Rechenschaftspflicht. Außerdem formulierte die Gruppe konkrete Empfehlungen, wie die Grundsätze auf nationaler Ebene umgesetzt werden können. Diese Arbeiten bildeten die Grundlage für die Ausarbeitung der *Empfehlung des Rats zu künstlicher Intelligenz*, die die OECD in der ersten Jahreshälfte 2019 vorlegte (OECD, 2019[3]).

Das KI-Observatorium der OECD

Die OECD unternahm auch Anstrengungen zur Einrichtung eines KI-Observatoriums, das die aktuellen und voraussichtlichen Entwicklungen im KI-Bereich und deren Politikimplikationen untersuchen soll. Es sollte 2019 seine Arbeit aufnehmen. Ziel ist es, die Umsetzung der weiter oben aufgeführten KI-Grundsätze durch die Einbindung eines breiten Spektrums externer Akteure zu unterstützen, z. B. von Vertretern staatlicher Stellen, der Wirtschaft, der Wissenschaft und der Fachwelt wie auch der breiten Öffentlichkeit. Das Observatorium ist als multidisziplinäres und evidenzbasiertes Zentrum konzipiert, das Regierungen politisch relevante Daten, Denkanstöße und Orientierungshilfen liefern soll. Zugleich soll es eine Anlaufstelle bieten, bei der sich externe Partner über politikrelevante KI-Aktivitäten und -Erkenntnisse im OECD-Raum informieren können.

Europäische Kommission und andere europäische Institutionen

Im April 2018 veröffentlichte die Europäische Kommission eine Mitteilung über künstliche Intelligenz für Europa, in der drei wichtige Prioritäten dargelegt wurden: 1. Förderung der technologischen und industriellen Leistungsfähigkeit der EU sowie der Verbreitung von KI in der gesamten Wirtschaft, 2. Vorbereitung auf die mit KI verbundenen sozioökonomischen Veränderungen und 3. Gewährleistung eines geeigneten ethischen und rechtlichen Rahmens. Im Dezember 2018 legte die Kommission einen koordinierten Plan über die Entwicklung von KI in Europa vor. Dieser Plan zielt in erster Linie darauf ab, die Wirkung von Investitionen zu maximieren und gemeinsam den Weg nach vorn zu definieren. Er soll bis 2027 laufen und enthält rd. 70 Einzelmaßnahmen in den folgenden Bereichen:

- **Strategische Maßnahmen und Koordinierung**: Erstellung nationaler KI-Strategien in den Mitgliedstaaten unter Angabe der Investitionssummen und Umsetzungsmaßnahmen
- **Maximierung der Investitionen durch Partnerschaften**: Förderung der Investitionen in strategische KI-Forschung und Innovation durch öffentlich-private Partnerschaften, eine Führungsgruppe sowie einen speziellen Investitionsfonds für Start-ups und innovative KMU
- **Vom Labor bis zum Markt**: Stärkung der Spitzenforschungszentren und digitalen Innovationszentren (Digital Innovation Hubs), Aufbau von Testeinrichtungen und möglicherweise Einrichtung von „regulatorischen Sandkästen“
- **Kompetenzen und lebenslanges Lernen**: Förderung von Talenten, Kompetenzen und lebenslangem Lernen
- **Daten**: Aufbau eines gemeinsamen europäischen Datenraums, um den Zugang zu Daten von öffentlichem Interesse und digitalen industriellen Plattformen zu erleichtern, was auch Gesundheitsdaten umfasst
- **Integrierte Ethik und regulatorischer Rahmen**: Einführung eines angemessenen ethischen und regulatorischen Rahmens für KI (der Sicherheits- und Haftungsfragen umfasst). Der ethische Rahmen für KI sollte auf den Ethikleitlinien der unabhängigen HEG-KI basieren. Die Europäische Kommission verpflichtet sich außerdem, das Grundprinzip einer „integrierten Ethik“ in ihren Aufforderungen zur Einreichung von Vorschlägen zu verankern.
- **KI für den öffentlichen Sektor**: Maßnahmen für den öffentlichen Sektor, wie beispielsweise die gemeinsame Beschaffung von KI-Lösungen oder die Bereitstellung automatisierter Übersetzungen
- **Internationale Zusammenarbeit**: Internationale Ausrichtung, Verankerung der KI in den Rahmen der Entwicklungspolitik und Ankündigung eines internationalen Treffens auf Ministerebene im Januar 2019

Im Rahmen ihrer KI-Strategie hat die Kommission außerdem im Juni 2018 die HEG-KI eingesetzt. Die HEG-KI, die sich aus Vertretern der Wissenschaft, der Zivilgesellschaft und der Wirtschaft zusammensetzt, hat zwei Aufgaben: Erstens soll sie Ethikleitlinien für die Akteure erarbeiten, die sich mit der Entwicklung, Einführung und Nutzung von KI befassen, um eine „vertrauenswürdige KI“ sicherzustellen. Zweitens soll sie für die Europäische Kommission und die Mitgliedstaaten Politik- und Investitionsempfehlungen („Empfehlungen“) im Hinblick auf die mittel- bis langfristigen KI-bezogenen Entwicklungen erarbeiten, um die Wettbewerbsfähigkeit Europas zu stärken. Parallel dazu hat die Kommission ein Multi-Stakeholder-Forum, die Europäische KI-Allianz, eingerichtet, um

eine breit gefasste Diskussion über die KI-Politik in Europa anzuregen. Jeder kann über die Plattform zur Arbeit der HEG-KI beitragen und so die Politik der EU mitgestalten.

Die HEG-KI veröffentlichte im Dezember 2018 einen ersten Entwurf ihrer Ethikleitlinien, um Kommentare einzuholen. Der Leitlinienentwurf skizziert einen Rahmen für eine vertrauenswürdige KI, die die EU-Grundrechte respektiert. Um vertrauenswürdig zu sein, muss KI geltendes Recht achten und ethisch, technisch robust und zuverlässig sein. Die Leitlinien legen eine Reihe ethischer Grundsätze für KI fest. Außerdem werden wichtige Anforderungen für eine vertrauenswürdige KI und Methoden zur Umsetzung dieser Anforderungen aufgeführt. Darüber hinaus enthalten die Leitlinien eine nichterschöpfende Bewertungsliste, die die einzelnen Anforderungen auf konkrete Anwendungsfälle bezieht, um den verschiedenen Akteuren zu helfen, die Grundsätze umzusetzen. Zum Zeitpunkt der Abfassung dieses Berichts prüfte die HEG-KI die Kommentare zu den Leitlinien, um den Text der Europäischen Kommission am 9. April 2019 offiziell vorzulegen. Die Europäische Kommission wird anschließend die nächsten Schritte der Leitlinien und die Entwicklung eines globalen ethischen Rahmens für KI näher erläutern. Das zweite Arbeitsergebnis der HEG-KI, die Empfehlungen, sollte im Sommer 2019 vorgelegt werden.

2017 veröffentlichte die Parlamentarische Versammlung des Europarats eine Empfehlung mit dem Titel *Technologische Konvergenz, künstliche Intelligenz und Menschenrechte*. In dieser Empfehlung wird das Ministerkomitee aufgefordert, die Gremien des Europarats anzuweisen zu untersuchen, welche Herausforderungen neue Technologien wie KI für die Menschenrechte bedeuten. Außerdem wurden darin Leitlinien zu Themen wie Transparenz, Rechenschaftspflicht und Profiling gefordert. Im Februar 2019 verabschiedete das Ministerkomitee des Europarats eine Erklärung zur Manipulationskraft von Algorithmen. Die Erklärung verweist auf die Gefahren für demokratische Gesellschaften, die sich daraus ergeben, dass ML-Systeme Gefühle und Gedanken beeinflussen können, und sie fordert die Mitgliedstaaten auf, diesen Gefahren entgegenzuwirken. Im Februar 2019 veranstaltete der Europarat eine hochrangige Konferenz mit dem Titel „Die neuen Spielregeln beherrschen: Wie sich die Entwicklung der künstlichen Intelligenz auf Menschenrechte, Demokratie und Rechtsstaatlichkeit auswirkt“.

Darüber hinaus verabschiedete die Europäische Kommission des Europarats für die Wirksamkeit der Justiz im Dezember 2018 die erste europäische Ethik-Charta zur Verwendung künstlicher Intelligenz in Justizsystemen. Darin werden fünf Grundsätze für die Entwicklung von KI-Tools in den europäischen Justizsystemen vorgelegt. 2019 beschloss der Ausschuss für Recht und Menschenrechte, einen Unterausschuss für KI und Menschenrechte einzurichten.

Im Mai 2017 verabschiedete der Europäische Wirtschafts- und Sozialausschuss (EWSA) eine Stellungnahme zu den gesellschaftlichen Auswirkungen der KI. In der Stellungnahme wurden die Interessenträger der EU aufgefordert sicherzustellen, dass die Entwicklung, der Einsatz und die Nutzung von KI einen gesellschaftlichen Nutzen hat und das Leben der Menschen verbessert. Der EWSA betont, dass Menschen die Kontrolle darüber behalten sollten, wann und wie KI im täglichen Leben eingesetzt wird, und nennt zwölf Bereiche, in denen KI gesellschaftliche Herausforderungen entstehen lässt. Dazu gehören Ethik, Sicherheit, Transparenz, Schutz der Privatsphäre, Standards, Arbeit, Bildung, (Un-)Gleichheit und Inklusion, Rechts- und Verwaltungsvorschriften, Regierungsführung und Demokratie, aber auch Kriegsführung und Superintelligenz. Die Stellungnahme fordert gesamteuropäische Standards für KI-Ethik, angepasste Arbeitsmarktstrategien und eine europäische KI-Infrastruktur mit quelloffenen Lernumgebungen (Muller, 2017[48]). Der EWSA hat eine befristete Studiengruppe zu KI eingerichtet, um diese Fragen zu prüfen.

Nordische und baltische Staaten

Im Mai 2018 unterzeichneten die Minister der nordischen und baltischen Staaten eine gemeinsame Erklärung zum Thema „KI in der nordisch-baltischen Region“. Zu den Ländern dieser Gruppe gehören Dänemark, Estland, Finnland, die Faröer Inseln, Island, Lettland, Litauen, Norwegen, Schweden und die autonome Region Åland. Sie kamen überein, ihre Zusammenarbeit im KI-Bereich zu stärken und zugleich ihre Position als Europas führende Region im Bereich der digitalen Entwicklung aufrechtzuerhalten (Nordic Council of Ministers, 2018[49]). In der Erklärung werden sieben Handlungsachsen genannt, um eine stärker auf den Menschen ausgerichtete KI zu entwickeln und zu fördern. Erstens sollen die Möglichkeiten der Kompetenzentwicklung verbessert werden, damit mehr Behörden, Unternehmen und Organisationen KI nutzen. Zweitens soll der Datenzugang verbessert werden, damit KI zur Erhöhung der Qualität der Dienstleistungen für die Bürger und Unternehmen der Region eingesetzt werden kann. Drittens sollen ethische und transparente Leitlinien, Standards, Grundsätze und Werte zu der Frage erarbeitet werden, wann und wie KI-Anwendungen genutzt werden sollten. Viertens soll dafür gesorgt werden, dass Infrastruktur, Hardware, Software und Daten – alles Faktoren, die für die Nutzung von KI von zentraler Bedeutung sind – auf Standards basieren, die Interoperabilität, den Schutz der Privatsphäre, Sicherheit, Vertrauen, Nutzerfreundlichkeit und Übertragbarkeit ermöglichen. Fünftens wollen die Länder sicherstellen, dass KI bei den Diskussionen über den digitalen Binnenmarkt der EU und der Umsetzung entsprechender Initiativen ein hoher Stellenwert eingeräumt wird. Sechstens werden sie unnötige Regulierung in diesem sich rasch entwickelnden Bereich vermeiden. Siebtens werden sie den Nordischen Ministerrat nutzen, um die Zusammenarbeit in relevanten Politikbereichen zu erleichtern.

Vereinte Nationen

Im September 2017 unterzeichnete das Interregionale Forschungsinstitut der Vereinten Nationen für Kriminalität und Rechtspflege das Gastlandabkommen zur Eröffnung eines Zentrums für künstliche Intelligenz und Robotik im Rahmen des VN-Systems in Den Haag.[2]

Die Internationale Fernmeldeunion hat mit mehr als 25 anderen Unterorganisationen der VN zusammengearbeitet, um den Weltgipfel von „AI for Good“ auszurichten. Außerdem arbeitet sie mit Organisationen wie der XPRIZE Foundation und der Association for Computing Machinery zusammen. Im Anschluss an den ersten Gipfel im Juni 2017 veranstaltete die Internationale Fernmeldeunion im Mai 2018 in Genf ein zweites Gipfeltreffen.[3]

Die UNESCO hat einen globalen Dialog über die ethischen Aspekte von KI eingeleitet, um der Komplexität und den Auswirkungen von KI auf Gesellschaft und Menschheit Rechnung zu tragen. Im September 2018 veranstaltete sie eine öffentliche Gesprächsrunde mit Experten, auf die im März 2019 eine globale Konferenz mit dem Titel „Principles for AI: Towards a Humanistic Approach? – AI with Human Values for Sustainable Development” folgte. Diese Veranstaltungen zielten darauf ab, das Bewusstsein für die Chancen und Herausforderungen von KI und ähnlichen Technologien zu schärfen und eine Reflektion über sie anzustoßen. Vorbehaltlich der Zustimmung des Exekutivrats der UNESCO im April 2019 sollte im November 2019 auf der 40. Generalkonferenz der UNESCO die Möglichkeit erörtert werden, im Zeitraum 2020-2021 eine Empfehlung über KI zu erarbeiten.

Internationale Organisation für Normung

Die Internationale Organisation für Normung (ISO) und die Internationale Elektrotechnische Kommission (IEC) haben 1987 den gemeinsamen Fachausschuss ISO/IEC JTC 1 gegründet. Der Ausschuss hat die Aufgabe, IT-Standards für Unternehmens- und Verbraucheranwendungen zu entwickeln. Im Oktober 2017 wurde im Rahmen des JTC 1 der Unterausschuss 42 (SC 42) eingerichtet, um KI-Standards zu entwickeln. Der Unterausschuss SC 42 berät die ISO- und IEC-Ausschüsse im Hinblick auf die Entwicklung von KI-Anwendungen. Zu seinen Aufgaben gehören die Ausarbeitung eines Referenzrahmens und eines gemeinsamen Vokabulars, die Ermittlung der Berechnungsansätze und -architekturen von KI-Systemen sowie die Evaluierung entsprechender Gefahren und Risiken (Price, 2018[50]).

Initiativen privater Akteure

Nichtstaatliche Akteure haben zahlreiche Partnerschaften gegründet und Initiativen eingeleitet, um KI-Fragen zu erörtern. In vielen dieser Initiativen ist ein breites Spektrum von Akteuren vertreten. Dieser Abschnitt beschreibt jedoch in erster Linie die Initiativen, an denen die Fachwelt, der Privatsektor, Arbeitnehmerorganisationen und die Wissenschaft beteiligt sind. Die Liste ist nicht erschöpfend.

Fachkreise und Wissenschaft

Das Institute for Electrical and Electronics Engineers (IEEE) leitete im April 2016 die Global Initiative on Ethics of Autonomous and Intelligent Systems ein. Sie zielt darauf ab, die öffentliche Diskussion über den Einsatz von KI-Technologien zu fördern und wichtige Werte und ethische Maßstäbe festzulegen. Im Dezember 2017 veröffentlichte das IEEE die zweite Fassung seiner Design-Leitlinien (*Ethically Aligned Design*) und bat die Öffentlichkeit um Stellungnahmen. Die Veröffentlichung der endgültigen Fassung dieser Leitlinien war für 2019 geplant (Tabelle 5.3) (IEEE, 2017[51]). Im Juni 2018 gründete das IEEE zusammen mit dem MIT Media Lab den Council for Extended Intelligence. Diese Initiative hat das Ziel, eine verantwortungsvolle Entwicklung intelligenter Systeme zu fördern, die Kontrolle über personenbezogene Daten zurückzugewinnen und Messgrößen für den wirtschaftlichen Wohlstand zu entwickeln, die über das Bruttoinlandsprodukt hinausgehen (Pretz, 2018[52]).

Tabelle 5.3. Grundprinzipien der IEEE-Leitlinien *Ethically Aligned Design (Version 2)*

Grundprinzipien	Ziele
Menschenrechte	Sicherstellen, dass autonome intelligente Systeme (AIS) die international anerkannten Menschenrechte nicht verletzen
Wohlergehen in den Vordergrund rücken	Bei Design und Nutzung von AIS größeres Augenmerk auf Messgrößen der Lebensqualität richten, weil die traditionellen Messgrößen des (materiellen) Wohlstands die Auswirkungen von KI-Technologien auf das menschliche Wohlergehen nicht ausreichend erfassen
Verantwortlichkeit	Sicherstellen, dass AIS-Designer und -Operatoren verantwortlich handeln und rechenschaftspflichtig sind
Transparenz	Sicherstellen, dass AIS transparent arbeiten
Sensibilisierung für Missbrauchsrisiken	Das Risiko des Missbrauchs von AIS-Technologie minimieren

Quelle: IEEE (2017[51]), *Ethically Aligned Design (Version 2)*, http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf.

Bei den KI-Leitsätzen von Asilomar, die auf der Konferenz des Future of Life Institute im Januar 2017 entwickelt wurden, handelt es sich um 23 Grundsätze, die kurz- und längerfristig eine sichere und sozialverträgliche Entwicklung von KI sicherstellen sollen. Die auf der Asilomar-Konferenz entwickelten Leitsätze basieren auf Diskussionen, Beiträgen und Dokumenten des IEEE, der Wissenschaft und gemeinnütziger Organisationen.

Die Themen sind in drei Bereiche unterteilt. Unter „Forschungsthemen" werden mehrere wichtige Aspekte genannt, die es zu gewährleisten gilt: Forschungsgelder für eine gesellschaftlich positiv wirkende KI unter Berücksichtigung schwieriger Fragen der Computerwissenschaft, Wirtschaft, Rechtswissenschaft, Ethik und Sozialwissenschaft, ein konstruktiver Austausch zwischen Wissenschaft und Politik und eine Forschungskultur, die auf Kooperation, Vertrauen und Transparenz basiert. Im Abschnitt „Ethik und Werte" wird gefordert, dass KI-Systeme so entwickelt und betrieben werden sollten, dass Sicherheit, Transparenz, Verantwortung, Freiheit und Privatheit, Menschenwürde, Menschenrechte und kulturelle Vielfalt, Teilhabe und Vorteile für alle gewährleistet sind. Im Abschnitt „Längerfristige Probleme" wird darauf hingewiesen, dass starke Annahmen im Hinblick auf die künftige Leistungsfähigkeit von KI vermieden werden sollten und dass die Entwicklung einer allgemeinen künstlichen Intelligenz sorgfältig vorausgeplant werden sollte (FLI, 2017[53]). Tabelle 5.4 enthält eine Liste der KI-Grundsätze von Asilomar.

Tabelle 5.4. Die KI-Leitsätze von Asilomar (Auszug)

	Forschungsthemen	Ethik und Werte	Längerfristige Probleme
Leitsätze	– Forschungsziel – Forschungsgelder – Verbindung von Wissenschaft und Politik – Forschungskultur – Vermeidung eines Wettlaufs	– Sicherheit – Transparenz bei Fehlfunktionen – Juristische Transparenz – Verantwortung – Wertorientierung – Menschliche Werte – Privatsphäre – Freiheit und Privatheit – Geteilter Nutzen – Geteilter Wohlstand – Menschliche Kontrolle	– Vorsichtige Annahmen zur Leistungsfähigkeit – Tragweite – Risiken – Rekursive Selbstverbesserung – Allgemeinwohl

Quelle: FLI (2017[53]), *Asilomar AI Principles*, https://futureoflife.org/ai-principles/.

Das gemeinnützige KI-Forschungsunternehmen OpenAI wurde Ende 2015 gegründet. Seine sechzig Vollzeitforscher haben den Auftrag, die Voraussetzungen für eine sichere allgemeine künstliche Intelligenz zu schaffen und sicherzustellen, dass deren Nutzen so breit und ausgewogen wie möglich verteilt wird.[4]

Die Future Society hat 2015 eine KI-Initiative eingeleitet, um an der Gestaltung des globalen KI-Regelwerks mitzuwirken. Sie hat eine Online-Plattform für einen multidisziplinären Bürgerdialog eingerichtet. Diese Plattform soll dazu beitragen, dass die Dynamik, die Vorteile und die Risiken von KI-Technologien besser verstanden werden, und so die Ausarbeitung von Politikempfehlungen unterstützen.[5]

Darüber hinaus gibt es in allen OECD-Ländern ebenso wie vielen Partnervolkswirtschaften zahlreiche Initiativen der Wissenschaft. Die Internet Policy Research Initiative des MIT soll beispielsweise helfen, die Kluft zwischen Fachwelt und Politik zu schließen. Das

Berkman Klein Center an der Harvard-Universität startete 2017 die Initiative „Ethics and Governance of Artificial Intelligence“. Das MIT Media Lab befasst sich mit Themen wie Algorithmen und Justiz, autonome Fahrzeuge sowie Transparenz und Erklärbarkeit von KI.

Initiativen der Privatwirtschaft

Im September 2016 starteten Amazon, DeepMind, Google, Facebook, IBM und Microsoft die Partnership on Artificial Intelligence to Benefit People and Society (PAI). Diese Partnerschaft zielt darauf ab, Best Practices für KI-Technologien zu untersuchen und zu formulieren und der Öffentlichkeit zu einem besseren Verständnis von KI zu verhelfen. Außerdem bietet sie eine offene Plattform für Diskussionen über KI und ihren Einfluss auf Menschen und Gesellschaft. Seit ihrer Gründung hat sich die PAI zu einer multidisziplinären Stakeholder-Gemeinschaft mit mehr als achtzig Mitgliedern entwickelt. Dabei handelt es sich um kommerzielle Technologieunternehmen ebenso wie um Vertreter der Zivilgesellschaft, von Hochschulen und Forschungseinrichtungen sowie um Start-ups.

Der Information Technology Industry Council (ITI) ist ein Verband aus Technologieunternehmen mit Sitz in Washington, DC. Er hat mehr als sechzig Mitglieder. Im Oktober 2017 veröffentlichte der ITI die *AI Policy Principles* (Tabelle 5.5). Diese Grundsätze unterstreichen die Verantwortung der Wirtschaft in bestimmten Bereichen und fordern staatliche Unterstützung für KI-Forschung und öffentlich-private Partnerschaften (ITI, 2017[54]). Die Unternehmen ergreifen auch einzeln Maßnahmen.

Tabelle 5.5. Die KI-Politikgrundsätze des ITI

Verantwortung: Verantwortungsvolle Entwicklung und Nutzung fördern	Chancen für den Staat: Investieren und das KI-Ökosystem stärken	Chancen für öffentlich-private Partnerschaften: Lebenslange Bildung und Vielfalt fördern
– Verantwortung bei Design und Einsatz – Sicherheit und Kontrollierbarkeit – Robuste und repräsentative Daten – Interpretierbarkeit – Haftung von KI-Systemen aufgrund ihrer Autonomie	– In KI-Forschung und -Entwicklung investieren – Flexible Regulierung – Förderung von Innovation und Internetsicherheit – Cybersicherheit und Schutz der Privatsphäre – Globale Standards und Best Practices	– Demokratisierung des Zugangs und Chancengleichheit – Natur-/ingenieurwissenschaftliche, technische und mathematische Bildung – Arbeitskräfte – Öffentlich-private Partnerschaften

Quelle: ITI (2017[54]), *AI Policy Principles*, https://www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf.

Zivilgesellschaft

Die vom Electronic Privacy Information Center gegründete Public Voice Coalition hat im Oktober 2018 ihre *Allgemeinen Richtlinien für die künstliche Intelligenz* (Universal Guidelines on Artificial Intelligence – UGAI) vorgelegt (The Public Voice, 2018[55]). Die UGAI lenken die Aufmerksamkeit auf die zunehmenden Herausforderungen, die mit intelligenten Computersystemen verbunden sind, und enthalten konkrete Empfehlungen zur Verbesserung ihres Designs. Sie fördern Transparenz und Rechenschaftspflicht von KI-Systemen und sollen sicherstellen, dass die Menschen die Kontrolle über die von ihnen geschaffenen Systeme behalten.[6] Die zwölf Grundsätze beziehen sich auf verschiedene Rechte und Pflichten. Sie umfassen das Recht auf Transparenz und menschliche Bestimmung sowie Verpflichtungen in Bezug auf Identifikation, Fairness, Folgenabschätzung und Rechenschaft, Richtigkeit, Zuverlässigkeit und Gültigkeit, Datenqualität, öffentliche

Sicherheit, Cybersicherheit und Abschaltung. Zudem verbieten sie geheime Profilerstellung und umfassendes Scoring.

Arbeitnehmerorganisationen

Die UNI Global Union vertritt mehr als 20 Millionen Beschäftigte des Dienstleistungssektors aus über 150 Ländern. Zu ihren Kernanliegen gehört die Sicherung einer Zukunft, die die Arbeitnehmer zu mehr Teilhabe befähigt und menschenwürdige Arbeit bietet. Sie hat zehn wichtige Grundsätze für ethische KI festgelegt. Ziel ist es sicherzustellen, dass Tarifverträge, globale Rahmenabkommen und multinationale Allianzen unter Einbeziehung von Gewerkschaften, Vertrauensleuten und globalen Allianzen die Arbeitnehmerrechte respektieren (Tabelle 5.6) (Colclough, 2018[56]).

Tabelle 5.6. Die zehn wichtigsten Grundsätze für ethische künstliche Intelligenz (UNI Global Union)

1. Transparenz der KI-Systeme:	Die Arbeitnehmer müssen das Recht haben, Transparenz bei den Entscheidungen/Ergebnissen von KI-Systemen und ihren Algorithmen einzufordern. Sie müssen außerdem bezüglich der Implementierung, Entwicklung und Verbreitung von KI-Systemen konsultiert werden.
2. Ausrüstung von KI-Systemen mit einer „Ethik-Black-Box“:	Diese Box sollte nicht nur maßgebliche Daten zur Gewährleistung der Transparenz und Erklärbarkeit des Systems, sondern auch klare Daten und Informationen über die darin eingebauten ethischen Überlegungen enthalten.
3. KI im Dienst des Menschen und des Planeten:	Dies umfasst Ethikcodes für die Entwicklung, Anwendung und Nutzung von KI, sodass KI-Systeme über ihren gesamten operativen Prozess hinweg kompatibel bleiben und die Menschenwürde, die Integrität und die Freiheit, den Schutz der Privatsphäre und der kulturellen und geschlechtlichen Vielfalt sowie die grundlegenden Menschenrechte fördern.
4. Gesamtsteuerung durch den Mensch (*human in command*):	Die Entwicklung von KI muss verantwortungsbewusst, sicher und nützlich sein. Maschinen müssen dabei den rechtlichen Status von Werkzeugen behalten und Menschen müssen stets die Kontrolle über sie haben. Die Verantwortung für die Maschinen liegt beim Menschen.
5. Sicherung einer geschlechterneutralen, unvoreingenommenen KI:	Bei Gestaltung und Betrieb von KI-Systemen ist es wichtig, dass diese auf negative oder schädliche menschliche Voreingenommenheiten hin kontrolliert werden und dass alle Voreingenommenheiten, sei es im Hinblick auf Geschlecht, Hautfarbe, sexuelle Orientierung oder Alter, erkannt und nicht von den Systemen weiterverbreitet werden.
6. Teilhabe aller an den Vorteilen von KI-Systemen:	Der von KI geschaffene wirtschaftliche Wohlstand sollte breit und gleichmäßig verteilt werden, damit die gesamte Menschheit davon profitiert. Deshalb bedarf es globaler wie auch nationaler politischer Strategien zur Schließung wirtschaftlicher, technologischer und gesellschaftlicher Gräben.
7. Gerechter technologischer Wandel, grundlegende Freiheiten und Rechte:	Mit der Weiterentwicklung der KI-Systeme und der Augmented Reality verschwinden Arbeitsplätze und Tätigkeitsfehler. Entscheidend sind daher politische Strategien für einen gerechten Übergang in die digitale Realität, einschließlich spezifischer Maßnahmen, um freigesetzten Arbeitskräften bei der Beschäftigungssuche zu helfen.
8. Einrichtung globaler Governancemechanismen:	Nötig sind Governancegremien für menschenwürdige Arbeit und ethische KI auf globaler und regionaler Ebene, in denen verschiedene Akteure vertreten sind, darunter KI-Programmierer, -Hersteller, -Besitzer, -Entwickler und -Forscher, Arbeitgeber, Juristen, zivilgesellschaftliche Organisationen und Gewerkschaften.
9. Verbot der Übertragung von Verantwortung an Roboter:	Roboter sollten so entwickelt und betrieben werden, dass sie geltende Gesetze einhalten und die Grundrechte und -freiheiten, einschließlich der Persönlichkeitsrechte, achten.
10. Verbot eines Wettrüstens mit KI-Waffen:	Tödliche autonome Waffen, einschließlich Cyber Warfare, sollten verboten werden. UNI Global Union fordert ein globales Abkommen über ethische KI, das die unbeabsichtigten negativen Folgen von KI bewältigen hilft bzw. verhindert und gleichzeitig ihre Vorteile für Arbeitnehmer und Gesellschaft vergrößert. Wir betonen, dass Menschen und Unternehmen die verantwortlichen Akteure sind.

Quelle: Colclough (2018[56]), “Ethical artificial intelligence – 10 essential ingredients”, https://www.oecd-forum.org/channels/722-digitalisation/posts/29527-10-principles-for-ethical-artificial-intelligence.

Literaturverzeichnis

Benhamou, S. und L. Janin (2018), *Intelligence artificielle et travail*, France Stratégie, http://www.strategie.gouv.fr/publications/intelligence-artificielle-travail. [20]

Brasilien (2018), *Brazilian digital transformation strategy "E-digital"*, Ministry of Science, Technology, Innovation and Communications, http://www.mctic.gov.br/mctic/export/sites/institucional/sessaoPublica/arquivos/digitalstrategy.pdf. [5]

China (2018), *AI Standardisation White Paper*, Government of China, übersetzt ins Englische von Jeffrey Ding, Forscher im Future of Humanity's Governance of AI Program, https://baijia.baidu.com/s?id=1589996219403096393. [10]

China (2017), *Guideline on Next Generation AI Development Plan*, Government of China, State Council, http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm. [8]

China (2016), *Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry (2018-2020)*, Chinese Ministry of Industry and Information Technology, http://www.miit.gov.cn/n1146290/n1146392/c4808445/content.html. [7]

CIFAR (o.J.), "CIFAR Pan-Canadian Artificial Intelligence Strategy", https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy. [28]

Colclough, C. (2018), "Ethical artificial intelligence – 10 essential ingredients", *A.Ideas Series*, The Forum Network, OECD, Paris, https://www.oecd-forum.org/channels/722-digitalisation/posts/29527-10-principles-for-ethical-artificial-intelligence. [56]

Dänemark (2018), *Strategy for Denmark's Digital Growth*, Ministry of Industry, Business and Financial Affairs, https://em.dk/english/publications/2018/strategy-for-denmarks-digital-growth. [11]

Deutschland (2017), *Automatisiertes und Vernetztes Fahren*, Bundesministerium für Verkehr und digitale Infrastruktur, Ethik-Kommission, https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile. [13]

Deutschland (o.J.), *KI Made in Germany – Die nationale KI-Strategie der Bundesregierung*, Bundesregierung, https://www.ki-strategie-deutschland.de/home.html. [12]

Finnland (2017), "Finland's Age of Artificial Intelligence – Turning Finland into a Leader in the Application of AI", *MEAE reports*, No. 47/2017, Ministry of Economic Affairs and Employment, http://julkaisut.valtioneuvosto.fi/handle/10024/160391. [16]

FLI (2017), *Die KI-Leitsätze von Asilomar*, Future of Life Institute, https://futureoflife.org/ai-principles-german/. [53]

Fonds de recherche du Québec (2018), "Québec lays the groundwork for a world observatory on the social impacts of artificial intelligence and digital technologies", Pressemitteilung, 29. März, https://www.newswire.ca/news-releases/quebec-lays-the-groundwork-for-a-world-observatory-on-the-social-impacts-of-artificial-intelligence-and-digital-technologies-678316673.html. [29]

G20 (2018), *G20 Digital Economy Ministerial Declaration*, G20-Digitalministertreffen, 24. August, Salta, Argentinien, http://www.g20.utoronto.ca/2018/2018-08-24-digital_ministerial_declaration_salta.pdf. [1]

G7 (2018), *Chairs' Summary: G7 Ministerial Meeting on Preparing for Jobs of the Future*, https://g7.gc.ca/en/g7-presidency/themes/preparing-jobs-future/. [47]

G7 (2017), "G7 Multistakeholder Exchange on Human Centric AI for our Societies" Anhang 2 der Abschlusserklärung des G7-IKT- und Industrieministertreffens, 26.-27. September, Turin, Italien, http://www.g7italy.it/sites/default/files/documents/ANNEX2-Artificial_Intelligence_0/index.pdf. [46]

G7 (2016), "Proposal of Discussion toward Formulation of AI R&D Guideline" G7-IKT-Ministertreffen, 29.-30. April, Takamatsu, Kagawa, Japan, http://www.soumu.go.jp/joho_kokusai/g7ict/english/index.html. [45]

Hall, W. und J. Pesenti (2017), *Growing the Artificial Intelligence Industry in the UK*, Department for Business, Energy & Industrial Strategy und Department for Digital, Culture, Media & Sport, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf. [39]

IEEE (2017), *Ethically Aligned Design – Version 2 – For Public Discussion*, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf. [51]

Indien (2018), "National Strategy for Artificial Intelligence #AI for All", Discussion Paper, NITI Aayog, http://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf. [21]

Italien (2018), "Artificial Intelligence at the service of citizens", White Paper, Agenzia per l'Italia Digitale (AGID), https://libro-bianco-ia.readthedocs.io/en/latest/. [22]

ITI (2017), *AI Policy Principles*, Information Technology Industry Council, https://www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf. [54]

Japan (2019), *Social Principles for Human-Centric AI*, Japan Cabinet Office, https://www8.cao.go.jp/cstp/stmain/aisocialprinciples.pdf. [25]

Japan (2018), *Draft AI Utilization Principles*, Ministry of Internal Affairs and Communications, http://www.soumu.go.jp/main_content/000581310.pdf. [27]

Japan (2018), *Integrated Innovation Strategy*, Japan Cabinet Office, Juni, https://www8.cao.go.jp/cstp/english/doc/integrated_main.pdf. [24]

Japan (2017), *Artificial Intelligence Technology Strategy*, Strategic Council for AI Technology, New Energy and Industrial Technology Development Organization, http://www.nedo.go.jp/content/100865202.pdf. [23]

Japan (2017), *Draft AI R&D Guidelines for International Discussions*, Ministry of Internal Affairs and Communications, http://www.soumu.go.jp/main_content/000507517.pdf. [26]

Jing, M. und S. Dai (2018), "Here's what China is doing to boost its artificial intelligence capabilities", *South China Morning Post*, https://www.scmp.com/tech/science-research/article/2145568/can-trumps-ai-summit-match-chinas-ambitious-strategic-plan. [6]

Kaevats, M. (2017), "Estonia considers a 'kratt law' to legalise artifical intelligence (AI)", E-residency blog, 25. September,, https://medium.com/e-residency-blog/estonia-starts-public-discussion-legalising-ai-166cb8e34596. [15]

Kaevats, M. (2017), "Estonia's Ideas on Legalising AI", Präsentation bei der OECD-Konferenz "AI: Intelligent Machines, Smart Policies", Paris, 26.-27. Oktober, https://prezi.com/yabrlekhmcj4/oecd-6-7min-paris/. [14]

Kania, E. (2018), "China's AI agenda advances", *The Diplomat,* 14. Februar, https://thediplomat.com/2018/02/chinas-ai-agenda-advances/. [9]

Korea (2016), *Mid- to Long-term Master Plan in Preparation for the Intelligent Information Society*, Ministry of Science and ICT, Government of Korea, Interdepartmental Exercise, http://english.msip.go.kr/cms/english/pl/policies2/__icsFiles/afieldfile/2017/07/20/Master%20Plan%20for%20the%20intelligent%20information%20society.pdf. [31]

Korea (2016), "MSIP announces development strategy for the intelligence information industry", *Science, Technology & ICT Newsletter,* No. 16, Ministry of Science and ICT, https://english.msit.go.kr/english/msipContents/contentsView.do?cateId=msse44&artId=1296203. [30]

Martinho-Truswell, E. (2018), *Towards an AI strategy in Mexico: Harnessing the AI Revolution*, British Embassy in Mexico, Oxford Insights und C Minds, http://go.wizeline.com/rs/571-SRN-279/images/Towards-an-AI-strategy-in-Mexico.pdf. [33]

Muller, C. (2017), "Künstliche Intelligenz – die Auswirkungen der künstlichen Intelligenz auf den (digitalen) Binnenmarkt sowie Produktion, Verbrauch, Beschäftigung und Gesellschaft" Stellungnahme, Europäischer Wirtschafts- und Sozialausschuss, Brüssel, https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence. [48]

Nordic Council of Ministers (2018), *AI in the Nordic-Baltic region*, Erklärung des Nordic Council of Ministers, https://www.regeringen.se/49a602/globalassets/regeringen/dokument/naringsdepartementet/20180514_nmr_deklaration-slutlig-webb.pdf. [49]

OECD (2019), *Empfehlung des Rats zu künstlicher Intelligenz*, OECD, Paris, http://www.oecd.org/berlin/presse/Empfehlung-des-Rats-zu-kuenstlicher-Intelligenz.pdf. [3]

OECD (2019), *Scoping Principles to Foster Trust in and Adoption of AI – Proposal by the Expert Group on Artificial Intelligence at the OECD (AIGO)*, OECD, Paris, http://oe.cd/ai. [2]

OGCR (2018), *Analysis of the Development Potential of Artificial Intelligence in the Czcech Republic*, Office of the Government of the Czcech Republic, https://www.vlada.cz/assets/evropske-zalezitosti/aktualne/AI-Summary-Report.pdf. [37]

Peng, T. (2018), "South Korea aims high on AI, pumps $2 billion into R&D", Medium, 16. Mai, https://medium.com/syncedreview/south-korea-aims-high-on-ai-pumps-2-billion-into-r-d-de8e5c0c8ac5. [32]

Porter, M. (1990), "The competitive advantage of nations", *Harvard Business Review,* März-April, https://hbr.org/1990/03/the-competitive-advantage-of-nations. [4]

Pretz, K. (2018), "IEEE Standards Association and MIT Media Lab form council on extended intelligence", *IEEE Spectrum*, http://theinstitute.ieee.org/resources/ieee-news/ieee-standards-association-and-mit-media-lab-form-council-on-extended-intelligence. [52]

Price, A. (2018), "First international standards committee for entire AI ecosystem", *IE e-tech,* Issue 03, https://iecetech.org/Technical-Committees/2018-03/First-International-Standards-committee-for-entire-AI-ecosystem. [50]

Russische Föderation (2017), *Digital Economy of the Russian Federation*, Government of the Russian Federation, http://pravo.gov.ru. [34]

Singapur (2018), *Digital Economy Framework for Action*, Infocomm Media Development Authority, https://www.imda.gov.sg/-/media/imda/files/sg-digital/sgd-framework-for-action.pdf?la=en. [36]

Sivonen, P. (2017), "Ambitious Development Program Enabling Rapid Growth of AI and Platform Economy in Finland" Präsentation bei der OECD-Konferenz, AI Intelligent Machines, Smart Policies", Paris, 26.-27. Oktober, http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-sivonen.pdf. [17]

The Public Voice (2018), *Universal Guidelines for Artificial Intelligence*, The Public Voice Coalition, October, https://thepublicvoice.org/ai-universal-guidelines/memo/. [55]

Thompson, N. (2018), "Emmanuel Macron talks to WIRED about France's AI strategy", *WIRED,* 31. März, https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy. [19]

Vereinigtes Königreich (2018), "Artificial Intelligence Sector Deal", Department for Business, Energy & Industrial Strategy und Department for Digital, Culture, Media & Sport, Government of the United Kingdom, https://www.gov.uk/government/publications/artificial-intelligence-sector-deal. [41]

Vereinigtes Königreich (2018), "Centre for Data Ethics and Innovation Consultation", Department for Digital, Culture, Media & Sport, Government of the United Kingdom, https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation/centre-for-data-ethics-and-innovation-consultation. [42]

Vereinigtes Königreich (2017), "Industrial Strategy: Building a Britain fit for the future", Government of the United Kingdom, https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future. [40]

Vereinigtes Königreich (2017), "UK Digital Strategy" Government of the United Kingdom, https://www.gov.uk/government/publications/uk-digital-strategy/uk-digital-strategy. [38]

Vereinigte Staaten (2018), "Artificial Intelligence for the American People", Executive Office of the President, Government of the United States, https://www.whitehouse.gov/briefings-statements/artificial-intelligence-american-people/. [43]

Vereinigte Staaten (2017), "Delaney launches bipartisan artificial intelligence (AI) caucus for 115th Congress", Pressemitteilung, Congressional Artificial Intelligence Caucus, 24. Mai, https://artificialintelligencecaucus-olson.house.gov/media-center/press-releases/delaney-launches-ai-caucus. [44]

Villani, C. (2018), *For a Meaningful Artificial Intelligence – Towards a French and European Strategy*, AI for Humanity, https://www.aiforhumanity.fr/ (Abruf: 15. Dezember 2018). [18]

Vinnova (2018), "Artificial Intelligence in Swedish Business and Society: Analysis of development and potential – Summary", *Vinnova Report*, VR 2018:09, https://www.vinnova.se/contentassets/29cd313d690e4be3a8d861ad05a4ee48/vr_18_09.pdf. [35]

Anmerkungen

[1] Der Bericht wurde von der britischen Botschaft in Mexiko in Auftrag gegeben, vom Prosperity Fund des Vereinigten Königreichs finanziert und von Oxford Insights und C Minds in Zusammenarbeit mit der mexikanischen Regierung unter Mitwirkung von Experten aus ganz Mexiko erstellt.

[2] Vgl. www.unicri.it/news/article/2017-09-07_Establishment_of_the_UNICRI.

[3] Vgl. https://www.itu.int/en/ITU-T/AI/.

[4] Vgl. https://openai.com/about/#mission.

[5] Vgl. http://ai-initiative.org/ai-consultation/.

[6] Wegen weiterer Informationen vgl. https://thepublicvoice.org/ai-universal-guidelines/memo/.

Künstliche Intelligenz in der Gesellschaft

1950 stellte Alan Turing erstmals die Frage, ob Maschinen denken können. Seitdem wurden im Bereich der künstlichen Intelligenz (KI) gewaltige Fortschritte erzielt. Heute verändert KI Gesellschaft und Wirtschaft. KI ermöglicht Produktivitätssteigerungen, kann die Lebensqualität erhöhen und sogar bei der Bewältigung globaler Herausforderungen wie Klimawandel, Ressourcenknappheit und Gesundheitskrisen helfen. Mit der weltweiten Einführung von KI-Anwendungen wachsen aber auch die Bedenken, die ihre Nutzung aufwirft, u. a. in Bezug auf menschliche Werte und Fragen wie Fairness, menschliche Entscheidungsfreiheit, Datenschutz, Sicherheit und Haftung. Dieser Bericht soll helfen, ein gemeinsames Verständnis von künstlicher Intelligenz aktuell und für die nahe Zukunft zu entwickeln. Dazu beschreibt er technische, wirtschaftliche und politische Entwicklungen, geht auf konkrete Anwendungsfälle ein und befasst sich mit grundlegenden Fragen der Politikgestaltung. Zudem soll er für mehr Kohärenz in den Diskussionen sorgen, die in verschiedenen nationalen und internationalen Foren zum Thema künstliche Intelligenz geführt werden.

Diese Publikation ist Teil des OECD-Projekts „Going Digital“. Ziel dieses Projekts ist es, den Politikverantwortlichen Instrumente an die Hand zu geben, um das wirtschaftliche und gesellschaftliche Wohlergehen in einer zunehmend digitalen und datengesteuerten Welt zu erhöhen.

Weitere Informationen unter ***www.oecd.org/going-digital***

#GoingDigital

PRINT ISBN 978-92-64-18146-5
PDF ISBN 978-92-64-43961-0





