

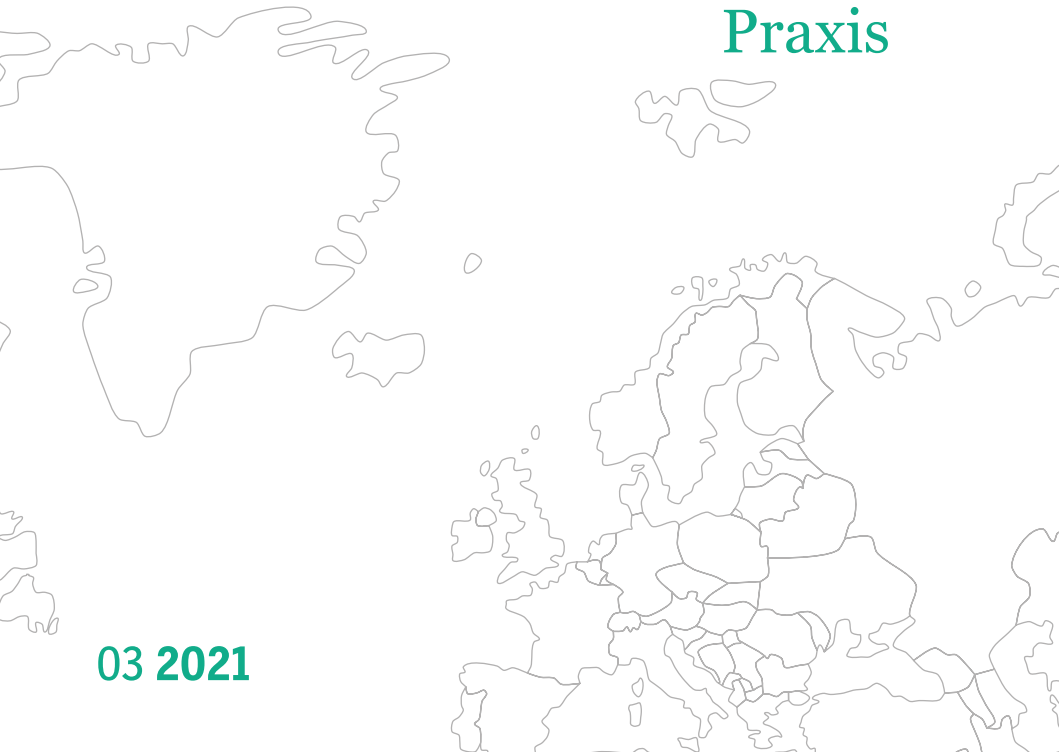
Globale Trends Analysen



Amandeep Singh Gill

KI-Regulierung in einen globalen Einklang bringen: Lehren aus der aktuellen Praxis

03 2021



„Dr. Frankensteins Verbrechen bestand nicht darin, dass er durch eine Kombination aus Hybris und Hochtechnologie eine Kreatur erfand, sondern dass er die Kreatur sich selbst überließ.“ (Latour 2012)

IMPRESSUM

Herausgegeben von
Stiftung Entwicklung und Frieden (sef.)
Dechenstr. 2, 53115 Bonn, Deutschland
Bonn 2021

Herausgeber-Team

Internationale Mitglieder: Dr. Adriana E. Abdenur (Plataforma CIPÓ, Rio de Janeiro), Prof. Dr. Manjiao Chi (University of International Business and Economics, Beijing), Dr. Tamirace Fakhoury (Aalborg University, Kopenhagen), Prof. Dr. Siddharth Mallavarapu (Shiv Nadar University, Dadri/Uttar Pradesh), Nanjala Nyabola (politische Analystin, Nairobi)

Mitglieder der herausgebenden Institutionen: Prof. Dr. Lothar Brock (Goethe-Universität Frankfurt, Mitglied im Beirat der sef.), Dr. Michèle Roth (Geschäftsführerin der sef.), Dr. Cornelia Ulbert (Universität Duisburg-Essen, Wissenschaftliche Geschäftsführerin des INEF und Mitglied im Vorstand der sef.)

Koordinierende Herausgeberinnen:

Michèle Roth, Cornelia Ulbert

Übersetzung: Rosa Großmann, Rolf Großmann

Lektorat: Ingo Haltermann

Design und Grafik: DITHO Design, Köln

Satz: Gerhard Süß-Jung (sef.)

Druck: DCM Druck Center Meckenheim GmbH

Papier: Umweltzeichen Blauer Engel

GEDRUCKT IN DEUTSCHLAND

ISSN: 2568-8790

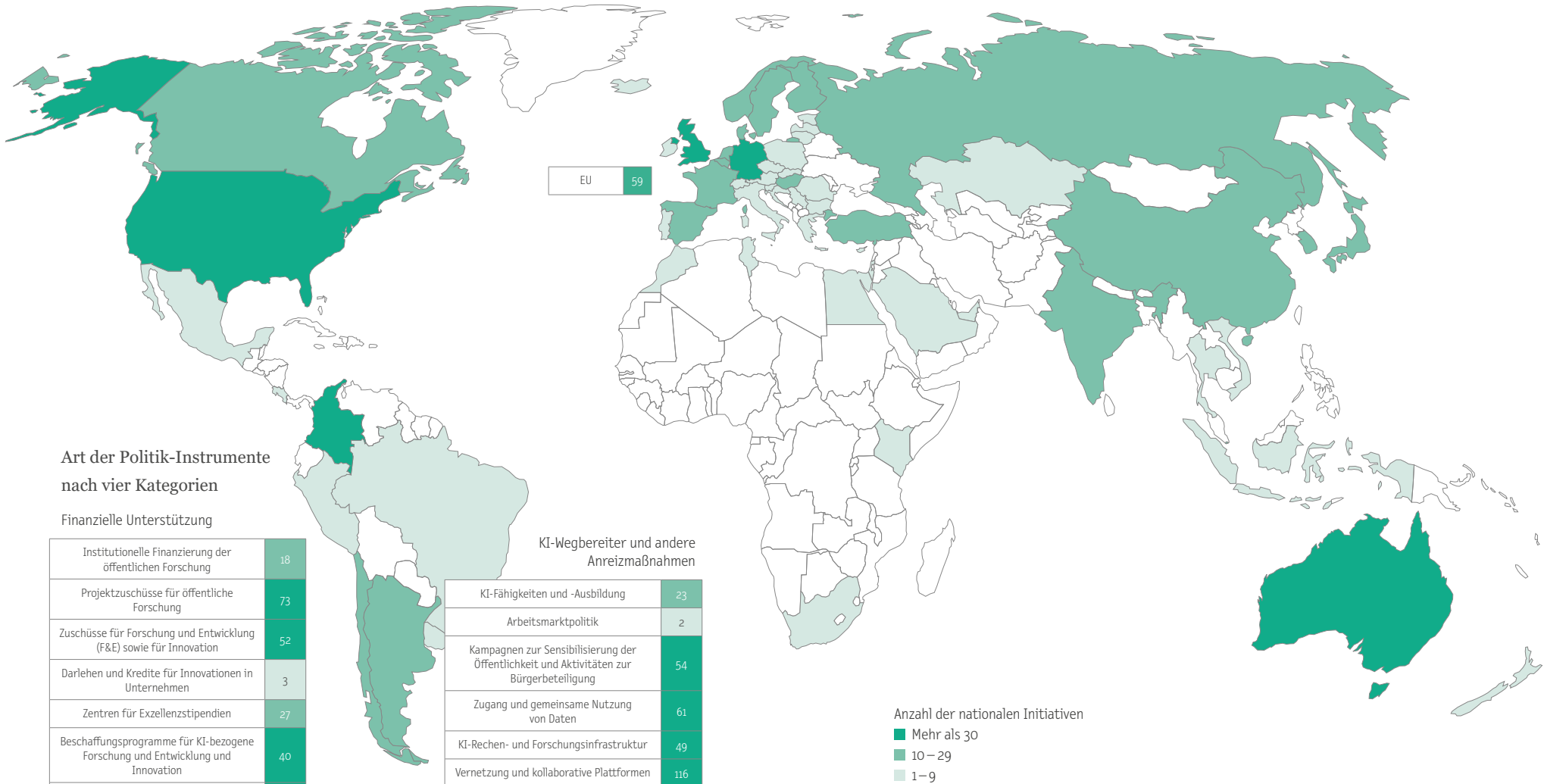
EINLEITUNG

Daten und künstliche Intelligenz (KI) als globale Gemeingüter anzusehen, könnte entscheidend dazu beitragen, dass diese Schlüsseltechnologien des 21. Jahrhunderts der gesamten Menschheit zugutekommen. Bemühungen, die Entwicklung und Regulierung von KI voranzutreiben, sind bisher jedoch stark fragmentiert und über den ganzen Globus verteilt. Dies setzt den Erfolg des Gemeingut-Ansatzes aufs Spiel. Die jüngsten Erfahrungen in den drei Bereichen bewaffnete Konflikte, Bildung und Gesundheit zeigen deutlich, dass das Problem der Governance von KI weit über unverfälschte Daten, Transparenz und Verständlichkeit von Algorithmen hinausgeht. Es bedarf eines systemischen Denkens, um die Grenzen und Zielkonflikte in jeder Phase der KI-Entwicklung auszuloten, die menschliche Verantwortung und Rechenschaftspflicht zu stärken und nach der Nutzung Rückmeldung an die Politik zu geben. In der öffentlichen Verwaltung gilt es, den gesamten Lebenszyklus künstlicher Intelligenz in den Blick nehmen. Es bedarf zudem neuer Methoden, um KI-Systeme auf nationaler Ebene kontinuierlich überprüfen zu können. Global gesehen gibt es mehrere Wege, die verschiedenen Ansätze der KI-Governance aufeinander abzustimmen und Anreize für die Einhaltung von Regeln zu schaffen. Hierzu gehören der Gemeingüter-Ansatz, ein gemeinsames Vokabular, geteilte Werte und Maßstäbe sowie digitale öffentliche Infrastrukturen.

ABBILDUNG 1

NATIONALE KI-STRATEGIEN UND -INITIATIVEN (STAND NOVEMBER 2021)

Anzahl der Initiativen pro Land



Governance

Nationale Strategien, Agenden und Pläne	238
KI-Koordinierungs- und/oder Überwachungsstellen	37
Öffentliche Konsultationen von Stakeholdern oder Experten	135
KI-Einsatz im öffentlichen Sektor	69

Vorgaben und Regulierung

Entwicklung KI-bezogener Regulierung	163
Regulierungsaufsicht und ethische Beratungsstellen	56
Regulierung der Arbeitsmobilität und Anreize	11
Normen und Zertifizierung für die Entwicklung und Einführung von Technologien	19

Art der Politik-Instrumente nach vier Kategorien

Finanzielle Unterstützung

Institutionelle Finanzierung der öffentlichen Forschung	18
Projektzuschüsse für öffentliche Forschung	73
Zuschüsse für Forschung und Entwicklung (F&E) sowie für Innovation	52
Darlehen und Kredite für Innovationen in Unternehmen	3
Zentren für Exzellenzstipendien	27
Beschaffungsprogramme für KI-bezogene Forschung und Entwicklung und Innovation	40
Stipendien sowie Postgraduierten-Darlehen und -Stipendien	31
Eigenkapitalfinanzierung	8
Indirekte finanzielle Unterstützung	9

KI-Wegbereiter und andere Anreizmaßnahmen

KI-Fähigkeiten und -Ausbildung	23
Arbeitsmarktpolitik	2
Kampagnen zur Sensibilisierung der Öffentlichkeit und Aktivitäten zur Bürgerbeteiligung	54
Zugang und gemeinsame Nutzung von Daten	61
KI-Rechen- und Forschungsinfrastruktur	49
Vernetzung und kollaborative Plattformen	116
Wissenstransfer und Unternehmensberatungsdienste	21
Wissenschafts- und Innovationswettbewerbe, Preise und Auszeichnungen	21

Anzahl der nationalen Initiativen

- Mehr als 30
- 10 – 29
- 1 – 9

Quelle: <https://oecd.ai/en/dashboards>, 29.11.2021

1. KÜNSTLICHE INTELLIGENZ VERSTEHEN

Herkömmliche Computersoftware verwendet quantitative Messwerte („Daten“) und ein Regelset („Code“), um ein Ergebnis („Output“) zu erzielen. Sie ist also regelbasiert und deterministisch. Die Daten passen oder sie passen nicht, auch wenn statistische Methoden Wahrscheinlichkeitsanalysen ermöglichen. Im Falle der KI wird der gewünschte Output wiederum zum Input – zusammen mit den als Trainingsdaten genutzten quantitativen Inputs. Das Vorgehen ähnelt dem eines Bäckers, der etwas Vorteig verwendet, um den Teig aufgehen zu lassen. Diese Umkehrung von In- und Output unterstreicht eine wichtige Dimension im Verständnis von künstlicher Intelligenz: KI-Modelle sind subjektiv bestimmt. Sie spiegeln menschliche Entscheidungen über die Art der von Menschen gewünschten Welt wider und sind keine Artefakte, die ausschließlich durch objektive Messwerte bestimmt werden. Sobald die gewünschten Ergebnisse mit den dafür relevanten Daten gepaart sind, wird ein Modell („Code“) ausgegeben – unser KI-Algorithmus. Auch bei diesem Schritt sind die subjektiven Entscheidungen des Menschen ausschlaggebend. Wir können dieses Modell mit neuen Daten füttern, Muster erkennen und mit einem gewissen Grad an Sicherheit vorhersagen, ob das gewünschte Ergebnis eintreten wird [siehe Abbildung 3].

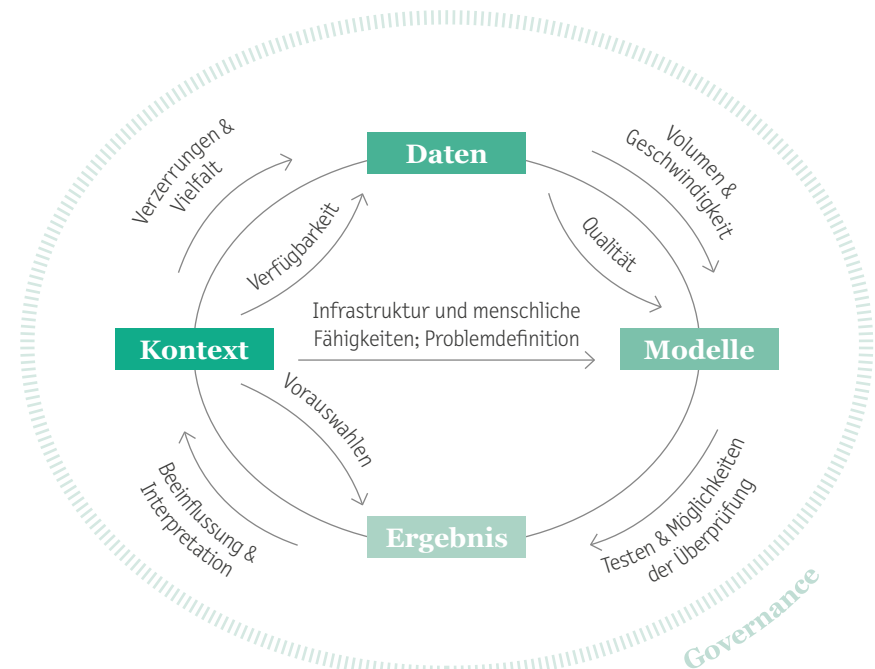
Im Gegensatz zu herkömmlicher Software ist KI also nicht-deterministisch und ihre Ergebnisse sind nicht vollständig vorhersehbar. Grob gesagt, ähneln KI-Modelle den kognitiven Modellen der Welt, wie sie alle Menschen erstellen und sie dann mit den eingehenden Daten abgleichen, um über ihre Handlungen zu entscheiden. In Wartofskys Worten sind alle kognitiven Artefakte, die wir schaffen, „Modelle: Darstellungen dessen, was wir tun, was wir wollen und was wir erhoffen. Das Modell ist also nicht einfach ein Abbild oder eine Kopie eines Sachverhalts, sondern darüber hinaus eine mutmaßliche Handlungsweise“ (Wartofsky 1979, S. xv).

Dies führt uns zum zweiten wichtigen Aspekt des Verständnisses von KI. Obwohl es sich bei KI-Modellen um lernende Systeme handelt, deren Lehrer/-innen Daten und gewünschte Ergebnisse sind, lernen sie (noch) nicht wie Menschen und sind auch nicht „intelligent“ im menschlichen Sinne. Tatsächlich kann KI einige kognitive Funktionen des Menschen, wie z.B. die Wahrnehmung, sehr gut imitieren, aber ihre „Intelligenz“ ist äußerst begrenzt. Viele Routineaspekte menschlicher Intelligenz, etwa kreative und kollaborative Intelligenz, liegen außerhalb ihrer Reichweite. Darüber hinaus wissen wir trotz

des seit mehr als hundert Jahren bekannten Zwei-Faktoren-Modells von Spearman und trotz der unglaublichen Fortschritte beim Verständnis kognitiver Funktionen sowie neuro-morphologischer und neuro-biologischer Mechanismen nur sehr wenig über das berühmte „g“, die allgemeine Intelligenz mit ihren unglaublichen Variationen bei Menschen (Barbey 2018). Es ist offensichtlich: Wenn wir nicht einmal wissen, was menschliche Intelligenz ist, wie können wir dann behaupten, dass irgendein menschliches Artefakt ihr nahekommt?

Der dritte Aspekt zum besseren Verständnis von KI ist das Zusammenspiel von Kontext, Daten, Modellen und Ergebnissen [siehe Abbildung 2]. Der Kontext bestimmt das Problem, das durch KI gelöst werden soll. Er ist zentral für das Verständnis, welche Arten von Daten für dieses Problem relevant sind und ob und wie wir auf sie zugreifen können. Damit werden der Tiefe und Vielfalt des Dateninputs Grenzen gesetzt. Die Art und Weise, wie das Problem kontextualisiert wird, hat wiederum Auswirkungen auf die Gültigkeit und

ABBILDUNG 2
KI und das Zusammenspiel von Kontext, Daten, Modellen und Ergebnissen



Quelle: Autor

DER GRUNDLEGENDE UNTERSCHIED ZWISCHEN HERKÖMMLICHER SOFTWARE UND KÜNSTLICHER INTELLIGENZ

Wie steuert man eine Reihe von Ampeln entlang einer stark befahrenen Straße?

Daten + Code = Output
(herkömmliche Software)

Durch den Einsatz herkömmlicher Software:

- 1) Ein Programm liest Daten von Sensoren an jeder Kreuzung, die ihm die Verkehrsdichte an bestimmten Punkten entlang der Strecke mitteilen (= **DATEN**).
- 2) Dementsprechend generiert es **OUTPUT**-Befehle, um die Wartezeit an jeder Kreuzung nach vorher festgelegten Regeln (= **CODE**) anzupassen, die in die Software geschrieben wurden, um einen reibungsloseren Verkehrsfluss zu gewährleisten.

ABER: Mit zunehmender Komplexität, d. h. mit mehr Kreuzungen und Einmündungen, wird die Anwendung vorgegebener Regeln schwieriger, und ständige Aktualisierungen und Anpassungen werden notwendig.

Durch den Einsatz künstlicher Intelligenz:

- 1) Anhand von Daten über eine vordefinierte Durchschnittsgeschwindigkeit auf verschiedenen Straßenabschnitten (= **OUTPUT**), der Anzahl der an einer bestimmten Ampel wartenden Fahrzeuge und der Wartezeiten oder Unfälle und Pannen entlang verschiedener Streckenabschnitte (= **DATEN**) wird ein KI-Modell erstellt.
- 2) Dieses Modell hilft uns, in Echtzeit den optimalen Schaltzeitpunkt für jede Ampel vorherzusagen (= **CODE**).

UND: Es ließen sich sogar noch komplexere Modelle erstellen. Die Notwendigkeit von Aktualisierungen sinkt durch das Hinzufügen historischer Daten über die jährliche Zulassung von Fahrzeugen in der Stadt und die Entwicklung des Transits von Auswärtigen.

Output + Daten = Code
(künstliche Intelligenz)

Zuverlässigkeit des Modells. Gleiches gilt für die Qualität und Repräsentativität der Daten. Die Modelle spiegeln auch den mathematischen und rechnerischen Sachverstand sowie die Auswahl und Beschränkungen der jeweiligen Personalausstattung und der IT-Infrastruktur wider. Sie müssen getestet oder mit anderen Worten „verifiziert“ (auf Übereinstimmung des implementierten Modells mit dem konzeptionellen Modell) und „validiert“ werden (ob die Ergebnisse des implementierten Modells das reale Problem adressieren, das wir zu lösen versuchen). Die Ergebnisse gehen wiederum in den Kontext ein. Ohne ein angemessenes statistisches Verständnis des Kontexts können selbst die besten Modellergebnisse falsch interpretiert oder falsch angewendet werden. Wichtig ist, dass jede dieser vier Dimensionen in unterschiedlichem Maße von Regulierung beeinflusst werden kann. Governance von KI muss daher als ein Eingriff ins algorithmische System betrachtet werden.

Der erste Schritt zur KI-Governance ist demnach ein umfassendes Verständnis. Der grundlegende Charakter von KI ist die Erweiterung menschlicher Fähigkeiten und nicht deren Nachbildung. Die menschliche Intelligenz, die wir bislang immer noch wenig verstehen, lässt sich nicht auf Datenströme in Bits und Bytes oder auf Strukturen in einem bestimmten Körperorgan reduzieren. Sie ist analytisch, aber sie ist auch emotional und ethisch, somatisch und haptisch und vor allem kollektiv. Vor diesem Hintergrund können wir uns nun mit den Versprechen und Grenzen der KI befassen.

2. DAS POTENZIAL DER KI UND IHRE RISIKEN

Künstliche Intelligenz bietet gegenüber herkömmlicher Datenverarbeitung und -analyse erhebliche Vorteile. Sie kann mehr Verknüpfungen und nichtlineare Zusammenhänge berücksichtigen als traditionelle statistische Methoden oder Rechentools. Anstatt Erkenntnisse nur aus historischen Daten zu gewinnen, kann KI – basierend auf einem Lernen aus historischen Daten – Echtzeit- oder Zukunftsdaten für Vorhersagen heranziehen. Darüber hinaus vermag sie mit unterschiedlichen Daten wie Zahlen, Bildern, Videos und unstrukturierten Texten leichter umzugehen als herkömmliche Datenverarbeitung. Dies ist ein enormer Vorteil in der heutigen Internetumgebung mit ihren unzähligen Wegen, Nutzer/-innen einzubinden und Geräte zu vernetzen. Zu guter Letzt bieten KI-Systeme mehr Flexibilität und Möglichkeiten zum Experimentieren als herkömmliche Software. Insgesamt verleiht KI dem Menschen mehr datengesteuerte Problemlösungskompetenz. In einer

Welt, die von Daten überschwemmt wird, hilft sie den Menschen, eine Informationsflut zu vermeiden und Routineentscheidungen und Vorhersagen Maschinen zu überlassen.

Ein weiterer wichtiger sozioökonomischer Aspekt künstlicher Intelligenz wird oft unterschätzt: KI senkt die Einstiegshürden in Expertenbereiche. So können Menschen mit wenig operativem Fachwissen über den Erdölsektor durch den einfachen Zugriff auf historische Daten von Ölquellen und deren Einbindung in intelligente KI-Modelle durchaus zutreffende Vorhersagen über die Optimierung der Ölförderung machen (Koroteev/Tekic 2021). Dies hat Auswirkungen auf traditionelle Geschäftspraktiken. Es führt bestehende Wirtschaftsbereiche zusammen, indem es sie horizontal durchdringt, und schafft neue Wachstumsmöglichkeiten. Das gilt ebenfalls für die Lösung sozialer Probleme im Kontext nationaler und internationaler Entwicklungen. Seit einigen Jahren beschäftigt sich eine Reihe neuer Akteure im Bereich der nachhaltigen Entwicklungsziele (Sustainable Development Goals, SDGs) mit KI-basierten Lösungen für die seit langem bestehenden Herausforderungen in den Sektoren Landwirtschaft, Gesundheit, Bildung und Umwelt. Allein die erwarteten wirtschaftlichen Auswirkungen sind beträchtlich: Schätzungen des zusätzlichen globalen Bruttoinlandsprodukts (BIP) bis 2030 schwanken von unter 1,5 Billionen bis zu mehr als 15 Billionen US-\$ (Szczepański 2019).

Dieses Potenzial ist mit erheblichen Risiken verbunden. Es ist diese Janusköpfigkeit der KI, die die Notwendigkeit ihrer Regulierung vorgibt. Betrachten wir die Fallstricke der KI anhand von Beispielen aus drei spezifischen Bereichen: bewaffneter Konflikt, Bildung und Gesundheit.

2.1 FALLBEISPIEL: BEWAFFNETER KONFLIKT

Die in Waffen und ihre Systeme eingebettete Technologie ist ein entscheidender Faktor für die nationale Sicherheit. Digitale Technologien sind ein Merkmal fast aller modernen Waffensysteme. Cyberwaffen sind als eigene Waffenklasse in der Grauzone zwischen Frieden und Krieg sowie zwischen staatlichen und nichtstaatlichen Akteuren in den Vordergrund gerückt. In den letzten Jahren haben Kriegsstrategen auch damit begonnen, den Einsatz von KI für eine Vielzahl von Aufgaben wie Aufklärung, Ausbildung, Verteidigung und Angriff zu planen (Horowitz et al. 2018). Dieses wachsende Interesse ist verständlich, bedenkt man die Vorteile künstlicher Intelligenz

bei der Verarbeitung riesiger Datenmengen, bei der Einbindung verschiedener Sensoren und Plattformen und bei der Verbesserung der Geschwindigkeit und Genauigkeit der menschlichen Reaktion in sich schnell bewegenden, multidimensionalen Kampfszenarien. Mit KI ausgestattete Waffensysteme stellen jedoch neue Herausforderungen an die menschliche Kontrolle und die Verantwortlichkeit für den Einsatz tödlicher Gewalt dar. Die Besorgnis über die sogenannten tödlichen autonomen Waffensysteme (*lethal autonomous weapons systems*, LAWS), die aus Hollywood bekannten „Killer-Roboter“ und „Terminatoren“, ist gewachsen. Diese Befürchtungen sind zum Teil rechtlicher Natur: LAWS könnten dem Anwendungsbereich des humanitären Völkerrechts entgehen, indem sie die menschliche Verantwortung für die Anwendung von Gewalt verschleiern. Beispiele sind die Differenzierung zwischen zivilen und militärischen Zielen oder die Verhältnismäßigkeit und Vorsichtsmaßnahmen zum Schutz der Zivilbevölkerung. Doch die Bedenken sind auch ethischer Natur: Die Ausstattung von Maschinen mit der Fähigkeit, Entscheidungen über Leben und Tod zu treffen, verstößt gegen seit langem geltende Positionen von menschlicher Würde und Handlungsfähigkeit. Und schließlich geht es um die internationale Sicherheit: KI schafft neue Machtasymmetrien zwischen technologisch fortgeschrittenen und anderen Ländern, senkt die Hemmschwelle für den Einsatz von Gewalt und kann zu weiterer Unsicherheit in instabilen und umkämpften Regionen führen (Gill 2019).

2.2 FALLBEISPIEL: BILDUNG

KI bietet spannende Möglichkeiten für personalisiertes Lernen und für die Verlagerung von Frontalunterricht auf stärker lernzentrierte Ansätze. KI kann auch die Verlagerung vom „Classroom“-Lernen zu mobilen Mikro-Lernmomenten und von Lehrbüchern zu offenen Bildungsressourcen (Open Educational Resources, OER) erleichtern, die neu angeordnet, wiederverwendet, überarbeitet, weiterverteilt und digital gespeichert werden können. Es ist zwar unwahrscheinlich, dass Roboter in absehbarer Zeit die Lehrer/-innen ersetzen werden, aber KI-basierte virtuelle Assistentinnen und Assistenten sind in den kommenden Jahrzehnten wahrscheinlich. Sie können bei der Korrektur von Hausaufgaben helfen und die Fortschritte der einzelnen Schüler/-innen verfolgen. Es birgt jedoch zahlreiche Risiken, die Kontrolle über das Lernen an Maschinen abzugeben. KI-Systeme, die mit den besten Absichten entwickelt wurden, könnten am Ende Lehrende und Lernende gleichermaßen in ihrer Handlungsfreiheit beeinträchtigen. KI könnte die digitale

Abhängigkeit weiter vervielfachen und oberflächliches Denken und Einsamkeit verfestigen. Personalisierte Nachhilfe könnte eine neue Kluft zwischen Besitzenden und Besitzlosen schaffen. Der Einsatz virtueller Lehrassistenten wäre ohne große Mengen an Trainingsdaten und intensives Tracking von Lehrenden und Lernenden nicht möglich. Es stellen sich unangenehme Fragen zum Datenschutz und zum Eigentum der Daten der Kinder. Ein massiver Einsatz von KI im Bildungswesen könnte uns weiter in eine überforderte und überwachte Gesellschaft treiben.

2.3 FALLBEISPIEL: GESUNDHEIT

Wissenschaft und Technologie waren zentral für den Fortschritt im Bereich der persönlichen und öffentlichen Gesundheit. KI bietet die Gelegenheit für einen Paradigmenwechsel. Sie ermöglicht es, die Gesundheitsversorgung neu zu erfinden, Kosten zu senken, Diagnose und Behandlung zu personalisieren und die Kommunikation zwischen Patient/-in und Arzt/Ärztin zu verändern (Topol 2019). So ist es dem Clalit Research Institute in Israel unter Verwendung von Längsschnittdaten aus 20 Jahren und von 4,5 Millionen Patientinnen und Patienten sowie vernetzten elektronischen Krankenakten und Leistungsansprüchen gelungen, das Risiko eines Nierenversagens bei Diabetiker/-innen fünf Jahre im Voraus vorherzusagen (Balicer 2018). BloodCounts!, ein an der Universität Cambridge angesiedeltes Netzwerk, wendet KI an, um alle Dateneinträge von routinemäßig durchgeführten Complete-Blood-Count-Tests (CBC-Tests) zu analysieren. Es verwandelt sie in ein umfassendes Überwachungsnetzwerk zur Entdeckung von Ausbrüchen von Infektionskrankheiten, ohne dass neue Instrumente oder Reagenzien benötigt werden. Der Wert einer solchen Vorwarnung für die Gesundheitsbehörden ist angesichts der COVID-19-Pandemie unschätzbar. In ressourcenarmen Umgebungen bietet KI die Chance für einen Durchbruch in der universellen Gesundheitsversorgung (Wahl et al. 2018).

Auf der anderen Seite gibt es Bedenken hinsichtlich der Sicherheit von Gesundheitsdaten, des Eigentums, des Datenschutzes und der sachkundigen Zustimmung. Datensätze, die zur Entwicklung von KI für das Gesundheitswesen verwendet werden, können verzerrt und nicht repräsentativ sein, was zu problematischen Ergebnissen führt. Eine Studie mit zwei großen Patientengruppen zeigte, dass Pulsoximeter bei farbigen Patientinnen und Patienten eine bestehende verdeckte Sauerstoff-Unterversorgung fast drei-

mal seltener feststellten als bei weißen Patientinnen und Patienten (Sjoding et al. 2020). Technischer „Lösungsglaube“ ohne ausreichende Berücksichtigung des Kontexts und ohne Einbeziehung des Gesundheitspersonals ist nach wie vor weit verbreitet. Es besteht ein großer Hype um Lösungen, die sich in der Praxis nicht bewähren und daher das Vertrauen der Ärzte und Ärztinnen nicht verdient haben. Während beispielsweise die Telemedizin während der COVID-19-Pandemie aufgrund der verhängten Kontaktbeschränkungen einen Aufschwung erlebte, konnten Apps zur Kontaktverfolgung wegen einiger Fehlentwicklungen das Vertrauen der Bürger/-innen nicht gewinnen (Lewis 2020). Zwar gehörte eine KI-gestützte Plattform zur Überwachung von Krankheitsausbrüchen zu den ersten, die die Welt über Berichte zu einer neuartigen grippeähnlichen Krankheit in China informierte (Niiler 2020), doch die nachfolgenden KI-basierten Vorhersagen über die Ausbreitung der Krankheit und die Belastung der Gesundheitssysteme waren von uneinheitlicher Qualität und Wirkung. Schließlich besteht die Sorge, dass digitale Gesundheit und KI nur bestimmten Teilen der globalen Gesellschaft zugutekommen und die bestehenden Ungleichgewichte im Gesundheitswesen weiter verfestigen könnten (Kickbusch et al. 2021).

3. ZIEL(E) DER REGULIERUNG VON KI

Die im letzten Abschnitt durchgeführte Risikoanalyse dreier Fallbeispiele für KI-Anwendungen unterstreicht die entscheidende Bedeutung der Regulierung von KI. Dies ist inzwischen allgemein anerkannt und führte zu einer Vielzahl von fundierten Initiativen zum Management entsprechender Risiken. Eine wachsende Zahl von Ländern setzt KI-bezogene Richtlinien ein, wie etwa die von der EU vorgeschlagene Verordnung zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) (Europäische Kommission 2021b). Vor der Untersuchung spezifischer Aspekte dieser Governance-Initiativen lohnt sich ein Blick auf die Ziele, die wir als Gesellschaften und als internationale Gemeinschaft mit der Regulierung von KI verfolgen sollten.

Nehmen wir ein konkretes Beispiel. Vor einiger Zeit gab es bei automatischen Wasserhähnen ein Problem. Sie funktionierten bei dunklerer Haut nicht so gut. Der Grund dafür ist, dass ihre Sensoren mit einer begrenzten Anzahl von Daten trainiert wurden, die primär auf hellhäutigen Personen basierten. Es ist nicht so, dass die Entwickler/-innen rassistisch oder vorein-

genommen waren, sie waren einfach blind für die fehlende Vielfalt in ihren Datensätzen. Das kann ärgerlich sein, in einem anderen Zusammenhang ist das gleiche Problem jedoch potenziell tödlich. Pulsoximeter wurden während der Covid-19-Pandemie eingesetzt, um den Sauerstoffgehalt infizierter Menschen zu überwachen. Sie halfen zu entscheiden, ob Patientinnen und Patienten zu Hause mit Sauerstoff versorgt oder auf die Intensivstation eines Krankenhauses verlegt werden sollten. Aus einer Reihe von Gründen ist die Wahrscheinlichkeit, dass sie den Sauerstoffgehalt des Blutes falsch diagnostizieren, bei Farbigen dreimal so hoch wie bei Weißen (Sjoding et al. 2020). Ein grundlegendes Ziel der Anstrengungen im Bereich der Governance muss also die Verhütung solcher Gefahren sein. Zur Schadensverhütung gehört auch ein gewisses Maß an Verbraucherschutz, um unbedarfte Nutzer/-innen vor irreführenden Behauptungen über gesundheitliche Vorteile von KI-Anwendungen zu schützen.

Ein weiteres Ziel der Regulierung von KI ist die Verhinderung von Datenmissbrauch und die Gewährleistung eines angemessenen Werts für die Dateneigentümer/-innen bei der Nutzung ihrer Daten. Dazu gehört nicht nur der Schutz persönlicher oder öffentlicher Daten vor Verletzungen der Privatsphäre und Sicherheit, sondern auch die Kontrolle der Nutzer/-innen darüber, wie und von wem ihre Daten verwendet werden. User-Daten werden von Unternehmen oftmals ohne deren Wissen und Zustimmung genutzt, um ihnen Waren oder Dienstleistungen zu verkaufen. Beiträge in sozialen Medien können dazu verwendet werden, Profile von Personen zu erstellen und sie für politische Zwecke zu nutzen bzw. zu manipulieren. KI-gestützte, auf persönlichen Daten basierende Vorhersagen über das Risiko eines Kreditausfalls können verwendet werden, um einer Person Optionen für einen Kredit zu unterbreiten. In einigen Szenarien könnte dies einzelnen Personen einen fairen Zugang zu Krediten verwehren. Auch Apotheken verfügen über viele sogenannte „Lieferkettendaten“, d.h. sie wissen, wer welche Arzneimittel von wem, zu welchem Preis und aus welchem Grund kauft. In diesen Daten sind wertvolle Erkenntnisse verborgen, zum Beispiel über Arzneimittelreaktionen oder Krankheitsausbrüche (Bacry/Gaïffas 2020). Die Unternehmen machen diese Daten zu Geld, ohne dass die Patientinnen und Patienten einen Gegenwert erhielten. Das ist nicht fair, zumal sie die Nachteile der weitergegebenen Daten zu tragen haben. Verhinderung von Missbrauch und die Förderung der fairen Datennutzung ist daher ein wichtiger Bestandteil des Leitbilds für KI-Governance.

Ein drittes Ziel betrifft die „entgangene Nutzung“ („missed use“) und „fehlende Daten“ („missed data“) (Gill/Germann 2021). Betrachten wir die entgangene Nutzung am Beispiel der CBC-Geräte, die jedes Jahr für die Erstellung von 3,6 Milliarden Blutbildern eingesetzt werden (University of Cambridge 2021). Es werden Daten generiert, aber nur ein Teil davon wird von den verschreibenden Ärztinnen und Ärzten genutzt. Der Rest, der von den Festplatten gelöscht wird, sollte mit Hilfe gesetzlicher Regulierung „gerettet“ werden. Er könnte wertvolle Erkenntnisse über Pandemieausbrüche oder antimikrobielle Resistenzen enthalten. Ebenso könnten Emissionsdaten über verschiedene Katalysatoren in Autos gemeinsam genutzt werden, um in der gesamten Autoindustrie optimale Entscheidungen zu treffen. Aber kommerzielle und Eigentums-Erwägungen verhindern die gemeinsame Nutzung dieser Daten. Insbesondere in einkommensstarken Ländern gibt es eine Fülle solcher Daten, die in geschlossenen Systemen oder unter kommerzieller/öffentlicher Kontrolle liegen. Sie können zumeist nicht genutzt werden, weil es an Richtlinien und Regelungen fehlt, die dies ermöglichen. Die Gesellschaft als Ganzes trägt die Opportunitätskosten für die entgangene Nutzung dieser Daten.

Schließlich gibt es schlichtweg fehlende Daten sowohl in einkommensstarken als auch in ressourcenarmen Umfeldern, z.B. über Behinderungen bei jungen Menschen im Kontext COVID-bedingter Erkrankungen (Briggs/Vassall 2021). Was wir nicht zählen oder nicht zählen können, hat Auswirkungen darauf, was wir mit KI tun können oder nicht. Wenn Leitlinien und Regulierung die Beseitigung der Datenarmut und die Analyse zum Wohle der Allgemeinheit nicht erleichtern, geht eine Chance zur Nutzung von Daten als öffentliches Gut verloren.

4. DIE GEGENWÄRTIGE PRAXIS DER REGULIERUNG VON KI

Wie die beiden vorangegangenen Abschnitte gezeigt haben, ist die Regulierung von KI aus verschiedenen Gründen wichtig und dringend notwendig. Mit der zunehmenden Verbreitung von KI sind mehrere grundlegende Governance-Initiativen entstanden, die auf spezifische Bereiche wie Sicherheit, Bildung und Gesundheit zugeschnitten sind. Wie tragen diese Governance-Initiativen allgemein dazu bei, einerseits Schaden und Missbrauch von KI zu verhindern und andererseits entgangene Nutzungschancen zu vermeiden?

Und wie kann die Regulierungspraxis den Anforderungen des Einsatzes in spezifischen Bereichen wie dem Gesundheitswesen gerecht werden?

Aus einer Vielzahl von Initiativen zu allgemeinen Grundsätzen und Verhaltenskodizes ragen einige wenige heraus. Auf internationaler Ebene formulieren die KI-Prinzipien der OECD, die anschließend von der G20 auf dem Gipfel in Osaka im Juni 2019 übernommen wurden, fünf sich gegenseitig verstärkende wertebasierte Leitlinien für einen verantwortungsvollen und vertrauenswürdigen Umgang mit KI [siehe Kasten] (OECD 2019).

DIE KI-GRUNDSÄTZE DER OECD

- KI sollte für die Menschen und den Planeten Nutzen bringen, indem sie ein inklusives Wachstum, eine nachhaltige Entwicklung und die Lebensqualität fördert.
- KI-Systeme sollten so konzipiert werden, dass sie das Prinzip der Rechtsstaatlichkeit, die Menschenrechte, demokratische Werte und die Vielfalt achten, und sie sollten angemessene Schutzmaßnahmen – z.B. bei Bedarf die Möglichkeit menschlichen Eingreifens – vorsehen. Das Ziel ist dabei eine faire und gerechte Gesellschaft.
- KI-Systeme sollten Transparenz und verantwortungsvolle Offenlegung gewährleisten um sicherzustellen, dass die Menschen KI-basierte Ergebnisse verstehen und hinterfragen können.
- KI-Systeme müssen über ihren gesamten Lebenszyklus robust und sicher funktionieren, und potenzielle Risiken sollten kontinuierlich beurteilt und kontrolliert werden.
- Die Organisationen und Personen, die KI-Systeme entwickeln, einführen oder betreiben, sollten für den einwandfreien Betrieb gemäß den oben aufgeführten Grundsätzen rechenschaftspflichtig sein.

Quelle: <https://www.oecd.org/berlin/presse/Empfehlung-des-Rats-zu-kuenstlicher-Intelligenz.pdf>

Der Bericht der Hochrangigen Kommission für digitale Zusammenarbeit des UN-Generalsekretärs vom Juni 2019 fordert in seiner Empfehlung 3C, dass autonome intelligente Systeme so konzipiert werden sollten, dass ihre Entscheidungen verständlich und Menschen für ihre Nutzung verantwortlich sind (UN Secretary-General's High-level Panel on Digital Cooperation 2019). Das Gremium verlangt, dass die Praxis diesen Grundsätzen durch Audit- und Zertifizierungssysteme nachkommt. Diese sollen die Einhaltung technischer und ethischer Standards durch KI-Systeme überwachen. Diese Standards und Grundsätze wie Transparenz und Unvoreingenommenheit, sollten in verschiedenen gesellschaftlichen Umfeldern anwendbar sein und im Rahmen von Multi-Stakeholder- und multilateralen Ansätzen entwickelt werden.

Kürzlich wurde von einer 24-köpfigen internationalen Ad-hoc-Expertengruppe (AHEG) der Organisation der Vereinten Nationen für Bildung, Soziales und Kultur (UNESCO) eine Empfehlung zur Ethik künstlicher Intelligenz erarbeitet. Der Text wurde von den Mitgliedstaaten geprüft und auf der Generalkonferenz im November 2021 angenommen (UNESCO 2021). Neben einer umfassenden Reihe von Grundsätzen enthält der Text Leitlinien zur Anwendbarkeit dieser Grundsätze auf elf Themenfelder sowie einige Empfehlungen zur Überwachung und Bewertung.

Auf regionaler Ebene hat die Hochrangige KI-Sachverständigengruppe der Europäischen Union Leitlinien für vertrauenswürdige KI erarbeitet. Sie identifiziert sieben Schlüsselanforderungen, die KI-Anwendungen erfüllen müssen. Diese sind menschliche Verantwortung und Kontrolle, technische Robustheit und Sicherheit, Datenschutz und Datenkontrolle, Transparenz, Vielfalt, Nichtdiskriminierung und Fairness, gesellschaftliches und ökologisches Wohlergehen sowie Rechenschaftspflicht (Europäische Kommission 2019). Es wurde auch eine Checkliste vorgeschlagen, um zu bewerten, ob diese Anforderungen erfüllt werden. Sowohl die EU-Leitlinien als auch die zuvor aufgeführten internationalen Leitlinien sind jedoch nicht verbindlich und müssen an den jeweiligen Kontext angepasst werden. Als nächsten Schritt hat die EU einen Rechtsrahmen für KI vorgelegt, der die Vorschriften mit einer vierstufigen Risikohierarchie verknüpft: unannehmbares Risiko, hohes Risiko, begrenztes Risiko und minimales Risiko (Europäische Kommission 2021a).

Nach diesen vorgeschlagenen EU-Rechtsvorschriften würde eine unabhängige Stelle prüfen, ob ein mit hohem Risiko behaftetes KI-Produkt Qualitätskriterien in Bezug auf die verwendeten Daten erfüllt. Dabei geht es

um die Verringerung von Risiken und diskriminierenden Ergebnissen. Des Weiteren sind Anforderungen an die technische Dokumentation und Rückverfolgbarkeit, an Transparenz und die Bereitstellung von Informationen für die Nutzer/-innen, ein angemessenes Maß an menschlicher Aufsicht und ein hohes Maß an Cybersicherheit, Genauigkeit und Robustheit zu erfüllen (Europäische Kommission 2021b).

Neben Regierungen und internationalen Organisationen ist auch der private Sektor als ein äußerst mächtiger Akteur im digitalen Bereich aktiv an Industriestandards und Audits zur Regulierung von KI beteiligt. Facebooks Community Standards oder Algorithmen zur Inhaltsmoderation sind Regulierungsmechanismen mit grenzüberschreitender Wirkung (Meta 2019). Microsoft hat mit seinen eigenen Richtlinien Druck auf die Regulierung von KI im Bereich der Gesichtserkennung auf der ganzen Welt ausgeübt (Smith 2018). Technologieverbände wie das Institute of Electrical and Electronics Engineers (IEEE) haben unabhängige Gruppen gegründet, um die Terminologie in verschiedenen Bereichen zu vereinheitlichen (IEEE Ethically Aligned Design, First Edition) oder verantwortungsvolle Entwicklungs- und Beschaffungspraktiken der Industrie im Hinblick auf risikobehaftete KI-Technologien zu fördern (Bloch et al. o.J.). Sie haben das Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) ins Leben gerufen, um Transparenz und Verantwortlichkeit zu verbessern und algorithmische Verzerrung (*algorithmic bias*) zu verringern. In ähnlicher Weise hat das Weltwirtschaftsforum im April 2021 die Responsible AI (RAI) Certification Beta ins Leben gerufen, um sich der Themen Verantwortlichkeit, algorithmische Verzerrung und Fairness, Datenqualität, Erklär- und Interpretierbarkeit sowie Robustheit anzunehmen (Responsible Artificial Intelligence Institute 2021).

Die oben beschriebenen Initiativen sind wichtig, um Missbrauch und entgangene Nutzungschancen von KI zu verhindern. Die Umsetzung allgemeiner Leitlinien in Vorschriften für einen spezifischen Geltungsbereich kann sich jedoch als schwierig erweisen. Dies gilt vor allem, wenn die entsprechenden Fachleute an Governance-Mechanismen gewöhnt sind, die der Vor-KI-Ära entstammen. Die Herausforderung einer adäquaten Umsetzung in einem konkreten Anwendungsbereich lassen sich exemplarisch am Beispiel des Gesundheitswesens untersuchen.

Im Juni 2021 veröffentlichte die Weltgesundheitsorganisation (WHO) einen Bericht über Ethik und Governance der künstlichen Intelligenz für

die Gesundheit sowie sechs Leitprinzipien für ihre Gestaltung und Nutzung (WHO 2021). Diese Grundsätze ähneln den OECD-, UNESCO- oder EU-Grundsätzen, da sie beispielsweise die Autonomie des Menschen, Transparenz, Erklärbarkeit, Verantwortung und Rechenschaftspflicht betonen. Darüber hinaus enthält der WHO-Bericht wichtige Überlegungen, die für den Gesundheitsbereich relevant sind. So wird etwa davor gewarnt, den Nutzen der KI für die Gesundheit zu überschätzen. Das gilt insbesondere, wenn dies auf Kosten von Kerninvestitionen und -strategien geschieht, die für das Erreichen einer allgemeinen Gesundheitsversorgung erforderlich sind. Außerdem wird davor gewarnt, die Interessen von Patientinnen und Patienten und Gemeinschaften kommerziellen oder staatlichen Interessen unterzuordnen. Besondere Beachtung findet das problematische Vertrauen auf Datensätze aus Ländern mit hohem Einkommen beim Training von KI-Lösungen, die in Ländern mit niedrigem und mittlerem Einkommen eingesetzt werden sollen.

Zentral bei den derzeitigen Regulierungsansätzen im Gesundheits- und anderen spezifischen Bereichen ist die Frage, was die regulierende Instanz genehmigt. Handelt es sich um einen „abgeschlossenen“ Algorithmus, der weder lernt noch sich im Laufe der Zeit verändert? Oder geht es um ein wirklich anpassungsfähiges System, das dann entweder in einem vorab genehmigten Bereich arbeitet oder nach einiger Zeit in der Praxis einer erneuten Prüfung durch die Regulierungsbehörde unterzogen werden muss? Dies ist einer der Gründe, warum die US-amerikanische Food and Drug Administration (FDA) in ihrem auf künstlicher Intelligenz und maschinellem Lernen basierenden neuen Aktionsplan für Medizinprodukte einen Produktlebenszyklus-Ansatz (TPLC) für KI-Systeme vorgeschlagen hat (U.S. Food & Drug Administration 2021).

Die wichtigsten Trends, die sich aus dieser Übersicht über die aktuellen Governance-Praktiken ergeben, sind folgende: Erstens haben nur wenige Länder hochrangige KI-Strategiedokumente und Governance-Initiativen verabschiedet [Abbildung 1].

Zweitens bleibt die Verhütung von Schäden und Missbrauch das vorrangige Ziel der Governance-Bemühungen. Die „entgangene Nutzung“ wird weniger betont. Mit anderen Worten: Regulierung wird weitgehend getrennt von Entwicklung gesehen, auch wenn die EU begonnen hat, ihre KI-Regulierungsbemühungen mit den Bestrebungen im Bereich digitaler Wirtschaft zu verknüpfen. Im Gesundheitssektor hat sie die Schaffung eines Europäischen

Gesundheitsdatenraums (European Health Data Space) vorangetrieben, um die gemeinsame Nutzung von Daten für die öffentliche Gesundheit, Behandlung, Forschung und Innovation zu erleichtern. Auch bei der WHO laufen die jüngsten Governance-Bemühungen parallel zur Einführung einer Strategie für digitale Gesundheit und zum Aufbau einer neuen Abteilung für digitale Gesundheit und Innovation. Ziel ist die Förderung der verantwortungsvollen Nutzung von Daten und KI für die Gesundheit.

Innerhalb des Paradigmas der Missbrauchsvermeidung liegt der Schwerpunkt weiterhin auf dem zustimmungspflichtigen Datenschutz. Weniger Beachtung findet die Befähigung zur Datennutzung, bei der die Bürger/-innen im Mittelpunkt der Datenströme stehen und fundierte Entscheidungen über die Freigabe von Daten treffen können (Nilekani 2018).

Der dritte wesentliche Trend ist das Aufkommen von nationalen Governance- und Politikaktivitäten, die auf entgangene Nutzung und fehlende Daten abzielen. Sie bleiben bisher jedoch auf einige wenige Länder beschränkt. Ein gutes Beispiel ist das finnische Gesetz über die Sekundärnutzung von Gesundheits- und Sozialdaten (552/2019), das am 1. Mai 2019 in Kraft getreten ist. Mit Findata wurde eine neue Agentur zur Datenverarbeitung eingerichtet. Sie sammelt Datensätze, kombiniert, pseudoanonymisiert oder anonymisiert diese und verringert die Komplexität bei der Einholung von Genehmigungen zur Datennutzung.

Im Gegensatz dazu gibt es in Ländern mit niedrigem und mittlerem Einkommen, die zwar reich an Potenzial, aber arm an Daten sind, relativ wenig Gestaltungsanstrengungen, die sich auf entgangene Nutzung oder fehlende Daten konzentrieren. Im Grunde geht es hier um Integration und Gerechtigkeit. Wenn eine große Zahl von Menschen „nicht gezählt“ wird und keine Möglichkeit hat, an KI-/Datenchancen teilzuhaben, entsteht eine neue digitale Kluft, die zu den bestehenden Gräben bei Konnektivität, Inhalten und Geräten hinzukommt. Ein Auszug aus einem aktuellen Bericht zeigt, wie sich die Forschung und Entwicklung im Bereich der digitalen Gesundheit und der KI auf einige wenige Länder in Asien, Nordamerika und Westeuropa konzentriert, in denen Daten und KI-Fachwissen im Überfluss vorhanden sind (I-DAIR 2021).

Schließlich werden im Gegensatz zu den allgemeinen Werten und Grundsätzen, die etwa auf internationaler und regionaler Ebene im Vordergrund stehen, die Themen Erfahrungsaustausch und Kapazitätsaufbau

für Regulierungsbehörden und Entscheidungsträger/-innen nach wie vor zu wenig beachtet. Einige nationale Regulierungsbehörden haben angefangen, bestehende Regulierungsmaßnahmen für Geräte und Dienstleistungen auf KI-basierte Produkte auszuweiten. Wenige, wie die britische Wettbewerbs- und Marktaufsichtsbehörde (CMA), haben begonnen, Begriffe wie „algorithmische Systeme“ zu verwenden. Damit wird impliziert, dass die Regulierung neben Daten, Modellen, Prozessen, Algorithmen und Zielen auch die Art und Weise umfasst, wie Menschen die Systeme nutzen. Dies unterstreicht die Herausforderung, vor der die Regulierungsbehörden in Bezug auf Ausbildung und Techniken zur kontinuierlichen Bewertung von KI-Systemen stehen werden (UK Competition & Markets Authority 2021).

5. DER WEG IN DIE ZUKUNFT: REGULIERUNG VON KI WELTWEIT IN EINKLANG BRINGEN

Nach dem „Warum“ und „Was“ der KI-Governance, geht es nachfolgend um das „Wie“. Dabei stehen speziell Methoden und Verfahren im Fokus, die in verschiedenen geografischen Umgebungen funktionieren.

Die erste wichtige Dimension für die Zukunft ist der Ansatz der „globalen Gemeingüter“ (*global commons*). In einer Vielzahl von Bereichen mit grenzüberschreitenden Auswirkungen, wie dem Weltraum, den Ozeanen und der Biosphäre, hat die internationale Gemeinschaft diesen Governance-Ansatz gewählt, um internationale Zusammenarbeit und Maßnahmen zu erleichtern (Lambert et al. 2021). Jüngste Überlegungen zu Leitlinien wie diejenigen der Hochrangigen Kommission für Digitale Zusammenarbeit des UN-Generalsekretärs unterstreichen die Nützlichkeit eines Gemeingüter-Ansatzes auch in der digitalen Sphäre. Die Leitlinien sollten gemeinsame Hauptpfade (*common rails*) und Wegbegrenzungen (*guard rails*) umfassen, um die Vorteile digitaler Technologien weiter zu verbreiten und eine neuerliche „Tragödie der Gemeingüter“ durch Missbrauch zu verhindern [Abbildung 4].

Wie im Bericht des Gremiums des UN-Generalsekretärs dargelegt, dienen die gemeinsamen Hauptpfade dazu, das globale digitale Ökosystem inklusiv zu machen sowie Innovation und Skalierung zu unterstützen, um die nachhaltigen Entwicklungsziele (SDGs) einzulösen. Die Wegbegrenzungen sorgen dafür, dass niemand zurückbleibt und sozialer Schaden vermieden wird (UN Secretary-General’s High-level Panel on Digital Cooperation 2019). Darü-

ber hinaus kann eine globale Architektur digitaler Gemeingüter den Dialog über neue Aufgaben fördern. Anwendungsfälle und zu lösende Probleme lassen sich besser an diverse Interessengruppen kommunizieren. Die zugrundeliegende Architektur aus Multi-Stakeholder-Tracks oder -Plattformen könnte auch neue Daten und Erfahrungen zu den Auswirkungen künstlicher Intelligenz und anderen aufkommenden Technologien verbreiten. Dies hilft, Diskussionen über Governance „faktenreicher“ zu gestalten.

Die zweite Dimension des „Wie“ ist ein werteorientiertes, vielfältiges Governance-Instrumentarium. Am weicheren Ende des Spektrums können Werte und Grundsätze, wie sie kürzlich von der OECD und der UNESCO entwickelt wurden, die Gestaltung und Umsetzung von Richtlinien leiten. Sie können internationale Normen länderübergreifend kanalisieren und las-

ABBILDUNG 4
Voraussetzungen und Potenziale eines Gemeingüter-Ansatzes zur Regulierung von KI



Quelle: UN Secretary-General’s High-level Panel on Digital Cooperation 2019, S. 10f.

sen sich flexibler über Kulturen und Grenzen hinweg einsetzen. Werte- und grundsatzbasierte Leitlinien ermöglichen es zudem, die Governance auf frühe/unklare Phasen des Technologieentwicklungszyklus auszuweiten. Allerdings müssen die Werte und Leitprinzipien in jedem Nutzungskontext „entdeckt“ und eindeutig mit den Vorgaben der Regulierung verknüpft werden. Sie müssen auch sichtbar gemacht werden, um eine potenzielle Manipulation durch kommerzielle oder politische Interessen zu vermeiden (*ethics wash*).

Die dritte Dimension ist die mehrstufige Governance. Eine Steuerung der KI durch eine zentrale Aufsicht über Daten und Algorithmen wäre sowohl auf nationaler als auch auf internationaler Ebene unklug. Dies könnte Innovationen hemmen und neue Risiken für persönlichen Freiheiten schaffen. Stattdessen sollten die Governance-Kriterien stufenweise angewendet werden [siehe Abbildung 5]. Wo es notwendig ist, z.B. im Hinblick auf den Einsatz von KI in Waffensystemen, können Regulierungen auf internationaler Ebene erfolgen. Wo verbindliche Normen schwer zu erreichen sind, z.B. in den Bereichen Bildung und Gesundheit, können weit gefasste normative Vorgaben von zwischenstaatlichen Foren hilfreich sein, um die nationalen Gesetze auf einen respektablen Standard auszurichten. Nationale oder regionale Gesetze und normative Rahmenwerke zur Verhinderung von Schaden und Missbrauch wie die EU-Datenschutzgrundverordnung von 2016 müssen ergänzt werden durch digitale Architekturen für die Datenermächtigung der Bürger/-innen sowie durch Industriestandards und -verfahren. Letztere bieten eine gesonderte Möglichkeit, globale Praktiken anzugleichen, da sie unpolitisch sind. Governance-Maßnahmen auf der Makro-, Meso- und Mikroebene können sich somit gegenseitig verstärken, um das KI-Ökosystem sicher, inklusiv und zum Nutzen der Allgemeinheit zu gestalten. Dieser abgestufte Ansatz erleichtert auch eine kluge Multi-Stakeholder-Regulierung, indem Regierung, Industrie und Zivilgesellschaft in einem flexiblen Rahmen zusammengebracht werden, der Lernen erleichtert (Eising 2002).

Die vierte Dimension des „Wie“ bei der Regulierung von KI ist ein Systemansatz. Das Problem geht weit über die Frage unbalancierter Daten oder der Transparenz und Erklärbarkeit von Algorithmen hinaus. Am besten ist es, KI als „KI-Systeme“ zu betrachten, die Problemdefinition, Datenmodelle, Datensammlung und -kuratierung, Algorithmen, Tests und Validierung ebenso umfassen wie Folgenabschätzungen und retrospektive Bewertungen. Öffentliche Funktionsträger/-innen müssen den Lebenszyklus-Ansatz in Bezug auf KI verinnerlichen. Es gilt sich nicht nur auf eine algorithmische oder

Daten-Governance, sondern auf eine durchgängige Regulierung solcher Systeme vorzubereiten [Abbildung 5]. Sie sollten zum einen in der Lage sein, die Grenzen und Kompromisse in jeder Phase der Systementwicklung zu erkennen. Zum anderen bedarf es ausreichender Kompetenzen zur Stärkung der menschlichen Verantwortung und Rechenschaftspflicht für die Nutzung und für Abhilfemaßnahmen und Anpassungen der Vorschriften nach der Einführung. Das Richtlinien-Instrumentarium des 20. Jahrhunderts wird für einen Systemansatz nicht ausreichen. Die Behörden bräuchten darüber hinaus neue Audit-Verfahren zur laufenden Überprüfung von KI-Systemen.

Zusammenfassend lässt sich sagen, dass die Praxis der Regulierung von KI durch ein globales gemeinsames Dach, durch die Verwendung international vereinbarter Werte und Grundsätze sowie durch einen abgestuften Systemansatz grenzüberschreitend ausgerichtet werden kann. Die globale Angleichung der KI-Governance kann durch die Nutzung einer gemeinsamen Infrastruktur und den Aufbau von Kapazitäten weiter gefördert werden. Eine dezentralisierte Infrastruktur, die z.B. von Wissenschaftler/-innen aus verschiedenen Ländern gemeinsam genutzt wird, wäre hilfreich. So wäre es möglich, verantwortungsvolles Regieren in die Technologieentwicklung einzubringen, ähnlich wie nationale digitale öffentliche Infrastrukturen Mechanismen für die Datenermächtigung der Bürger enthalten können.

6. SCHLUSSFOLGERUNG

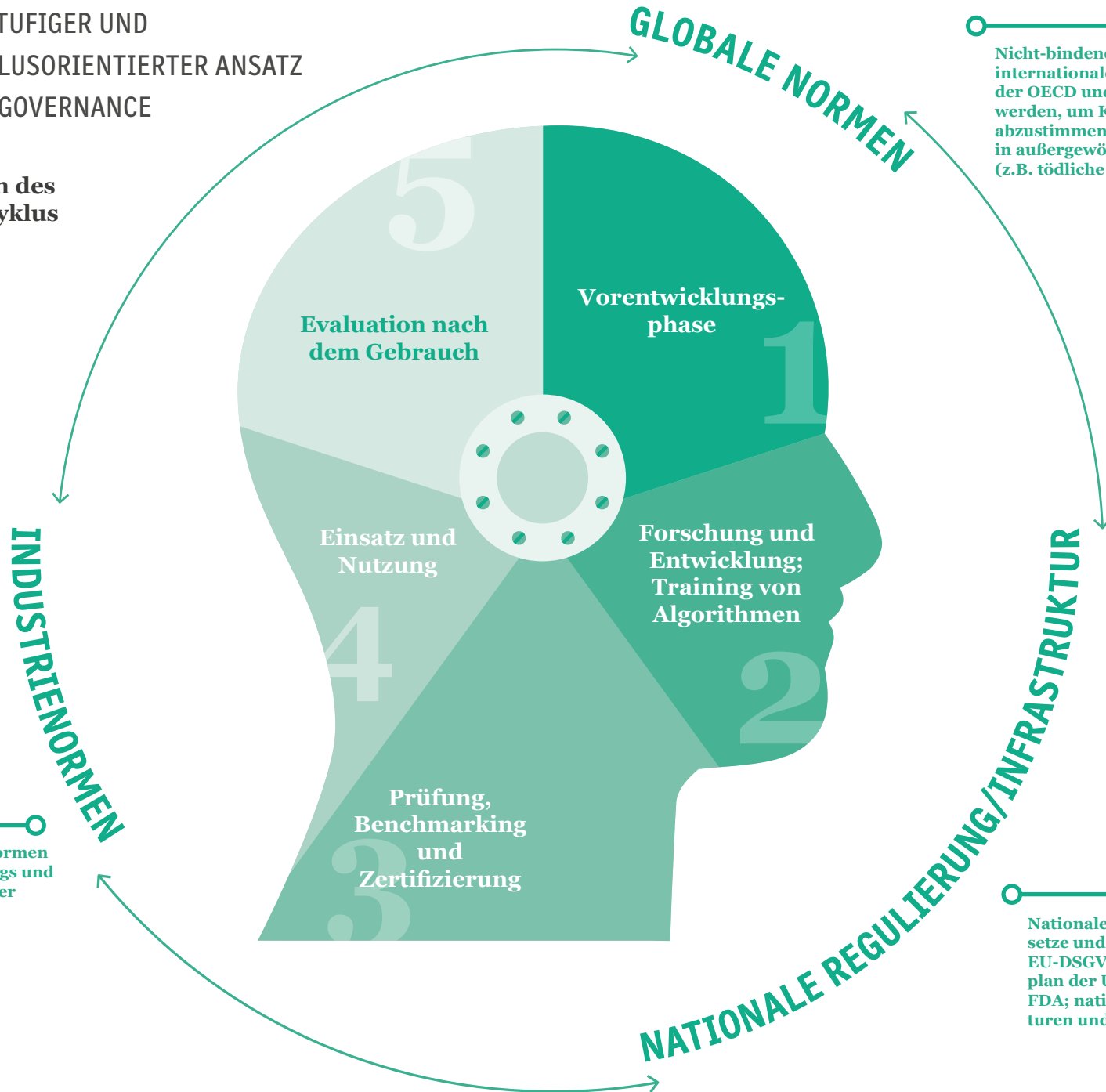
Die größte Herausforderung, vor der wir heute bei der Regulierung der digitalen Wirtschaft stehen, besteht darin, dass sie in verschiedenen Regionen schnell gewachsen ist, ohne dass wir ihre Auswirkungen überblicken oder vollständig verstehen. Die Tech-Giganten haben die politischen Entscheidungsträger/-innen vor vollendete Tatsachen gestellt und eine Machtfülle angesammelt, die regulatorische Entscheidungen einschränkt. Sie haben sich für ihre Belange günstige juristische Zuständigkeiten verschafft, um freie Hand zu haben. Regulierungsbehörden aus verschiedenen Ländern und Regionen wie der EU, Asien und den Vereinigten Staaten bemühen sich, ihren Rückstand aufzuholen.

Bevor die Entwicklung von KI-Technologien und -Anwendungen die Regulierungsbehörden erneut vor vollendete Tatsachen stellt, ist es wichtig, KI-Governance-Grundsätze und bewährte Verfahren (*best practice*) weltweit

ABBILDUNG 5

EIN MEHRSTUFIGER UND
LEBENSZYKLUSORIENTIERTER ANSATZ
FÜR DIE KI-GOVERNANCE

Fünf Phasen des
KI-Lebenszyklus



Nicht-bindende Grundsätze, die von internationalen Foren wie der WHO, der OECD und der UNESCO festgelegt werden, um Konzepte und Prioritäten abzustimmen; verbindliche Vorschriften in außergewöhnlichen Risikofällen (z.B. tödliche autonome Waffen)

Nicht-bindende Industrienormen und -praktiken sowie Settings und Präferenzen einzelner Nutzer

Nationale und regionale Gesetze und Vorschriften wie die EU-DSGVO oder der Aktionsplan der US-amerikanischen FDA; nationale Infrastrukturen und öffentliche Güter

anzugleichen. Angleichung bedeutet nicht, dass jedes Land genau dasselbe Regulierungssystem anwendet. Es bedeutet jedoch, dass wir grenzüberschreitend anerkennen, dass die KI-Domäne ein globales Gemeinwesen ist und dass Schaden und Missbrauch letztlich zum Nachteil aller sind. Wenn Daten fehlen oder in geschlossenen Systemen verbleiben, werden sie nicht zur kollaborativen Entwicklung von KI beitragen. Dies betrifft etwa auch Wissenschaftler/-innen, die versuchen, die nächste Infektionspandemie zu verstehen.

Ein wichtiger Schritt zur Angleichung der globalen Regulierung von KI ist die Entwicklung einer gemeinsamen Sprache. Es bedarf eines gemeinsamen Vokabulars für Technologie und ihre Auswirkungen sowie einer klar verständlichen Taxonomie bezüglich der Einordnung in die Vorgaben von Richtlinien und Governance. Hochschulen und Forschungseinrichtungen werden in dieser Hinsicht eine wichtige Rolle spielen.

Als Nächstes müssen Werte und Grundsätze angewandt werden, um Governance grenzüberschreitend auszurichten und um Entscheidungen über Technologien bereits in frühen und unklaren Phasen ihrer Entwicklung mitzuprägen. Werte haben eine kulturübergreifende Resonanz und können im jeweiligen Kontext wiederentdeckt werden. So dienen sie dazu, verschiedene Interessengruppen mit unterschiedlichen Perspektiven und Vorbehalten einzubinden. Eben deshalb haben sie sich in multilateralen Kontexten bewährt, in denen harte Normen schwierig zu erarbeiten sind und relativ viel Zeit in Anspruch nehmen.

Ein weiterer Weg zur globalen Angleichung ist die Institutionalisierung des Austauschs von Governance-Innovationen zwischen den verschiedenen Rechtsordnungen. Das Teilen von bewährten wie untauglichen Verfahren kann das Peer-to-Peer-Lernen fördern und dafür sorgen, dass die Bemühungen mit der Einführung von Technologien Schritt halten. UN-Foren in Bereichen wie Gesundheit (WHO) oder Bildung (UNESCO) können in dieser Hinsicht eine wichtige Rolle spielen. Eine Bestandsaufnahme für bewährte Verfahren in verschiedenen Bereichen kann auch als Grundlage für Programme zum Kapazitätsaufbau dienen.

Gemeinsame Risiko- und Wirkungskennzahlen sind weitere fruchtbare Bereiche für eine globale Zusammenarbeit, um Vertrauen aufzubauen und die Regulierung von KI grenzüberschreitend abzustimmen. Harmonisierte Maßstäbe helfen den Regulierungsbehörden zu erfahren, ob etwas durchgängig und über verschiedene Kontexte hinweg wie angegeben funktioniert.

Schließlich ist der Aufbau und die gemeinsame Nutzung von Technologie-Infrastruktur über Grenzen hinweg ein wirksames Mittel, um Anreize für sinnvolle Verfahren in der vergleichenden Bewertung und Regulierung von KI zu schaffen. Dies gilt insbesondere für Forscher/-innen und Entwickler/-innen im Globalen Süden, die nicht immer Zugang zu Hochleistungsrechnern für KI haben. Datensicherheit, Systemstabilität und Qualitätskontrolle können durch eine gemeinsam genutzte und dezentralisierte Infrastruktur gefördert werden. Dies hat den zusätzlichen Vorteil, dass die KI-Entwicklung integrativer erfolgt.

LITERATUR

- BACRY, EMMANUEL/GÄIFFAS, STÉPHANE** 2020: Machine Learning and Massive Health Data, in: Nordlinger, Bernard/Cédric, Villani/Rus, Daniela (Hg.) Healthcare and Artificial Intelligence, Cham: Springer, S. 23–31.
- BALICER, RAN** 2018: The Doctor Will See Your Future Now (Forbes, 16.04.2018), o.O. (<https://www.forbes.com/sites/startupnationcentral/2018/04/16/for-predictive-medicine-its-back-to-the-future/?sh=1d29014f3525>, 08.12.2021).
- BARBEY, ARON K.** 2018: Network Neuroscience Theory of Human Intelligence, in: Trends in Cognitive Science, Jg. 22/1, S. 8–20.
- BLOCH, EMMANUEL/CONN, ARIEL/GARCIA, DENISE/GILL, AMANDEEP/LLORENS, ASHLEY/NOORMA, MART/ROFF, HEATHER** o.J.: Ethical and technical challenges in the development, use, and governance of autonomous weapons systems (Report by an independent group of experts convened by the IEEE Standards Association), o.O., (<https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ethical-technical-challenges-autonomous-weapons-systems.pdf>, 08.12.2021).
- BIGGS, ANDREW/VASSALL, ANNA** 2021: Count the cost of disability caused by COVID-19, in: Nature, Bd. 593, S. 502–505.
- EISING, RAINER** 2002: Policy Learning in Embedded Negotiations: Explaining EU Electricity Liberalization, in: International Organization, Jg. 56/1, S. 85–120.
- EUROPÄISCHE KOMMISSION** 2019: Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen: Schaffung von Vertrauen in eine auf den Menschen ausgerichtete künstliche Intelligenz (COM(2019) 168 final), Brüssel (<https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52019DC0168&qid=1642060029783&from=EN>, 13.01.2022).
- EUROPÄISCHE KOMMISSION** 2021a: Ein Europa für das digitale Zeitalter: Kommission schlägt neue Vorschriften und Maßnahmen für Exzellenz und Vertrauen im Bereich der künstlichen Intelligenz vor (Pressemitteilung 21.04.2021), Brüssel (https://ec.europa.eu/commission/presscorner/detail/de/IP_21_1682, 13.01.2022).
- EUROPÄISCHE KOMMISSION** 2021b: Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union (COM/2021/206 final), Brüssel (<https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52021PC0206&qid=1642060384520&from=EN>, 13.01.2022).
- GILL, AMANDEEP S.** 2019: Artificial Intelligence and International Security: The Long View, in: Ethics and International Affairs, Jg. 33/2, S. 169–179.
- GILL, AMANDEEP S./GERMANN, STEFAN** 2021: Conceptual and Normative Approaches to AI governance for a global digital ecosystem supportive of the UN Sustainable Development Goals (SDGs), in: AI and Ethics 06.05.2021 (<https://doi.org/10.1007/s43681-021-00058-z>, 08.12.2021).
- HOROWITZ, MICHAEL/ALLEN, GREGORY C./SARAVALLE, EDOARDO/CHO, ANTHONY/FREDERICK, KARA/SCHARRE, PAUL** 2018: Artificial Intelligence and International Security, Washington DC: Centre for New American Security (https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS-AI-and-International-Security-July-2018_Final.pdf?mtime=20180709122303&focal=none, 08.12.2021).
- I-DAIR** 2021: The Global Research Map for digital health, o.O. (<https://grm.i-dair.org/#/report>, 08.12.2021).
- KICKBUSCH, ILONA ET AL.** 2021: The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world, in: The Lancet, Bd. 398/10312, S. 1727–1776.
- KOROTEEV, DMITRY/TEKIC, ZELJKO** 2021: Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future, in: Energy and AI, Jg. 3/März 2021 (<https://doi.org/10.1016/j.egyai.2020.100041>, 08.12.2021).
- LAMBERT, DOMINIQUE/REICHBERG, GREG/THELIS-SON, EVA** 2021: Human Fraternity in Cyberspace: Ethical Challenges and Opportunities (Caritas in Veritate Foundation), o.O. (im Erscheinen).
- LATOUR, BRUNO** 2012: Love Your Monsters: Why We Must Care for Our Technologies As We Do Our Children, The Breakthrough Institute (<https://thebreakthrough.org/journal/issue-2/love-your-monsters>, 08.12.2021).
- LEWIS, DYANI** 2020: Where Covid Contact-Tracing Went Wrong, in: Nature, Bd. 588, S. 384–388.
- META** 2019: Writing Facebook’s Rulebook, o.O. (<https://about.fb.com/news/2019/04/inside-feed-community-standards-development-process/>, 08.12.2021).
- NIILER, ERIC** 2020: An AI Epidemiologist Sent the First Warnings of the Wuhan Virus (Wired, 25.01.2020), o.O. (<https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/>, 08.12.2021).
- NILEKANI, NANDAN** 2018: Data to the People. India’s Inclusive Internet, in: Foreign Affairs, Jg. 97/5, S. 19–26.
- OECD** 2019: Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449.
- RESPONSIBLE ARTIFICIAL INTELLIGENCE INSTITUTE** 2021: Responsible Artificial Intelligence (RAI) Certification Beta, o.O., (<https://assets.ctfassets.net/rz1q59puyoa/w1myaH-22mA16YoeIXQND3qv/7974df6bd0973e65f-100d327b93129a2/Whitepaper.pdf>, 08.12.2021).
- SJODING, MICHAEL W./DICKSON, ROBERT P./IWASHYNA, THEODORE J./GAY, STEVEN E./VALLEY, THOMAS S.** 2020: Racial Bias in Pulse Oximetry Measurement, in: New England Journal of Medicine, Jg. 383/25, S. 2477–2478.
- SMITH, BRAD** 2018: Facial recognition technology: The need for public regulation and corporate responsibility (Microsoft Blog, 13.07.2018), o.O., (<https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/>, 08.12.2021).
- SZCZEPAŃSKI, MARCIN** 2019: Economic impacts of artificial intelligence (AI), Briefing to the European Parliament (European Parliamentary Research Service, PE 637.967), o.O. ([https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637967/EPRS_BRI\(2019\)637967_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637967/EPRS_BRI(2019)637967_EN.pdf), 08.12.2021).
- TOPOL, ERIC** 2019: Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again, New York: Basic Books.
- UK COMPETITION & MARKETS AUTHORITY** 2021: Algorithms: how they can reduce competition and harm consumers, o.O. (<https://www.gov.uk/government/publications/algorithms-how-they-can-reduce-competition-and-harm-consumers/algorithms-how-they-can-reduce-competition-and-harm-consumers>, 08.12.2021).
- UN SECRETARY-GENERAL’S HIGH-LEVEL PANEL ON DIGITAL COOPERATION** 2019: The Age of Digital Interdependence, o.O. (<https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf>, 08.12.2021).

UNESCO (UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION) 2021:

Draft Text of the Recommendation on the Ethics of Artificial Intelligence (SHS/IGM-AIETHICS/2021/APR/), o.O. (<https://unesdoc.unesco.org/ark:/48223/pf0000376713>, 08.12.2021).

UNIVERSITY OF CAMBRIDGE 2021: BloodCounts!

Breakthrough in disease detection, (University of Cambridge News, 25.06.2021), o.O. (<http://www.haem.cam.ac.uk/blog/bloodcounts-breakthrough-in-disease-detection/>, 08.12.2021).

U.S. FOOD & DRUG ADMINISTRATION 2021: Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan, o.O. (<https://www.fda.gov/media/145022/download>, 08.12.2021).

WAHL, BRIAN/COSSY-GANTNER, ALICE/GERMANN, STEFAN/SCHWALBE, NINA 2018: Artificial

Intelligence (AI) and Global Health: how can AI contribute to health in resource-poor settings?, in: BMJ Global Health (<http://dx.doi.org/10.1136/bmjgh-2018-000798>, 08.12.2021).

WARTOFSKY, MARX W. 1979: Models. Representation and the Scientific Understanding, Dordrecht: Springer.

WHO (WORLD HEALTH ORGANIZATION) 2021: Ethics and governance of artificial intelligence for health: WHO guidance, Genf.

DER AUTOR

AMANDEEP SINGH GILL

Projektleiter/CEO der International Digital Health & AI Research Collaborative (I-DAIR) und Professor für angewandte Forschung, The Graduate Institute of International and Development Studies, Genf



WEITERE AUSGABEN

Alle Ausgaben sind kostenlos abrufbar unter www.sef-bonn.org



GLOBALE TRENDS. ANALYSEN 02|2021

Sprechen wir durch das Recht:
Für einen rechtlich verankerten Multilateralismus
Heike Krieger
Dezember 2021, 27 Seiten

Seit einiger Zeit ist der im Völkerrecht verankerte Multilateralismus unter Druck geraten. Im Wechselspiel gegensätzlicher Kräfte eröffnen sich jedoch auch Handlungsspielräume für politische Akteure. In den GLOBALE TRENDS. ANALYSEN 2|2021 fordert Heike Krieger die EU-Mitgliedstaaten dazu auf, sich für Entwicklungen und Trends zu engagieren, die die internationale Ordnung stabilisieren. Zu diesem Zweck sollen sie einen rechtlich institutionalisierten Multilateralismus informellen Netzwerkstrukturen vorziehen. Grundlage ist eine glaubwürdige und konsistente Einhaltung des Völkerrechts sowie die Erarbeitung eines gemeinsamen Verständnisses dieser Rechtsordnung gerade mit Staaten des Global Südens.



GLOBALE TRENDS. ANALYSEN 01|2021

Finanzpolitischen Spielraum schaffen:
Ein Gebot der Menschenrechte in Zeiten von COVID-19
Ignacio Saiz
Mai 2021, 29 Seiten

Die Ungleichheit zwischen Staaten hat sich durch die COVID-19-Pandemie verstärkt. Die wirtschaftlichen Folgen der Pandemie sind in Ländern des Globalen Südens besonders verheerend. Die Ressourcen, die sie zur Bewältigung der Krise mobilisieren können, sind völlig unzureichend. Umso wichtiger ist es, dass die wohlhabenderen Länder und die internationalen Finanzinstitutionen die Hindernisse beseitigen, die sich aus ihrer Schulden- und Steuerpolitik für den finanzpolitischen Spielraum von Ländern mit niedrigem und mittlerem Einkommen ergeben. Eine solche Zusammenarbeit ist nicht nur ein Gebot der globalen öffentlichen Gesundheit. Sie ist auch eine bindende menschenrechtliche Verpflichtung, wie Ignacio Saiz erläutert.

GLOBALE TRENDS. ANALYSEN

untersuchen gegenwärtige und künftige Herausforderungen einer globalisierten Welt vor dem Hintergrund langfristiger politischer Trends. Die Reihe widmet sich Fragen von hoher politischer Relevanz für künftige Entwicklungen auf regionaler oder globaler Ebene. GLOBALE TRENDS. ANALYSEN deckt ein breites Themenfeld in den Bereichen Global Governance, Frieden und Sicherheit, nachhaltige Entwicklung, Weltwirtschaft und Weltfinanzsystem, Umwelt und natürliche Ressourcen ab. Die Reihe zeichnet sich durch Perspektiven aus verschiedenen Weltregionen aus.

