Friederike Rohde, Josephin Wagner, Philipp Reinhard, Ulrich Petschow
Andreas Meyer, Marcus Voß, Anne Mollen

# Nachhaltigkeitskriterien für künstliche Intelligenz

## Entwicklung eines Kriterien- und Indikatorensets für die Nachhaltigkeitsbewertung von KI-Systemen entlang des Lebenszyklus

Schriftenreihe des IÖW 220/21


i | ö | w INSTITUT FÜR ÖKOLOGISCHE WIRTSCHAFTSFORSCHUNG

sustain
Der Nachhaltigkeits-Index für Künstliche Intelligenz

Friederike Rohde, Josephin Wagner, Philipp Reinhard, Ulrich Petschow
Andreas Meyer, Marcus Voß, Anne Mollen

# Nachhaltigkeitskriterien für künstliche Intelligenz

## Entwicklung eines Kriterien- und Indikatorensets für die Nachhaltigkeitsbewertung von KI-Systemen entlang des Lebenszyklus

Gefördert durch das Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit (BMU), im Rahmen der Förderinitiative KI-Leuchttürme für Umwelt, Klima, Natur und Ressourcen unter dem Förderkennzeichen 67KI2060.

Schriftenreihe des IÖW 220/21
Berlin, Dezember 2021

ISBN 978-3-940920-24-9

# Impressum

**Herausgeber:**
Institut für ökologische
Wirtschaftsforschung GmbH, gemeinnützig
Potsdamer Straße 105
D-10785 Berlin
Tel. +49 – 30 – 884 594-0
Fax +49 – 30 – 882 54 39
E-Mail: mailbox@ioew.de
www.ioew.de

Gefördert durch:

Bundesministerium
für Umwelt, Naturschutz
und nukleare Sicherheit

In Kooperation mit:

DAI-Labor
Distributed Artificial Intelligence Laboratory

Technische Universität Berlin / DAI-Labor
Fakultät IV für Elektrotechnik und Informatik
Sekretariat TEL 14
Ernst-Reuter-Platz 7
D-10587 Berlin
E-Mail: sekretariat@dai-labor.de

und

ALGORITHM
WATCH

AW AlgorithmWatch gGmbH
Linienstr. 13
D-10178 Berlin
E-Mail: info@algorithmwatch.org

# Zusammenfassung

Technische Dynamiken, die zunehmende Durchdringung vieler gesellschaftlicher und wirtschaftlicher Bereiche mit Informations- und Kommunikationstechnologien sowie die Verfügbarkeit großer Rechenkapazitäten zur Datenverarbeitung haben zur Entwicklung immer leistungsfähigerer Modelle des maschinellen Lernens (ML) geführt. Diese werden häufig mit dem Überbegriff „Künstliche Intelligenz“ bezeichnet, obwohl sie korrekterweise nur als schwache künstliche Intelligenz einzustufenen sind. Technologien basierend auf ML werden mittlerweile in vielen gesellschaftlichen Bereichen und Industriesektoren eingesetzt und sind mit großen Erwartungen in Bezug auf die Optimierung von Prozessen oder der Entscheidungsfindung verbunden. Weil die zugrundeliegenden Modelle für sehr unterschiedliche Optimierungsaufgaben eingesetzt werden können, werden sie auch als *general purpose technology* bezeichnet.

Die zunehmende Nutzung dieser immer komplexer werdenden Systeme hat weltweit Debatten über diskriminierende Effekte, intransparente Entscheidungs- und Optimierungsprozesse oder die Reproduktion gesellschaftlicher Ungleichheiten aufgeworfen. Zunehmend werden auch die Energieverbräuche und Treibhausgasemissionen in der KI-Modellentwicklung und -anwendung diskutiert sowie weitreichendere Folgen auf Arbeitsmärkte, Konsummuster oder die Marktmacht großer Unternehmen. Dementsprechend sind Systeme künstlicher Intelligenz mit vielfältigen gesellschaftlichen, ökologischen und ökonomischen Herausforderungen verbunden. Dieses Diskussionspapier entwickelt eine übergreifende Nachhaltigkeitsperspektive auf künstliche Intelligenz und basiert auf den Arbeiten im Forschungsprojekt „SustAIn – Nachhaltigkeitsindex für Künstliche Intelligenz“. Das Ziel ist, aktuelle Diskussionen zu verantwortungsvoller KI aufzugreifen und zu einer übergreifenden Perspektive auf nachhaltige KI zu erweitern. Die Gestaltung dieser sozio-technischen Systeme ist dabei genauso relevant wie die Auswirkungen ihrer Anwendung. Betrachtet werden die sozialen, ökologischen und ökonomischen Auswirkungen entlang des gesamten KI-Lebenszyklus sowie die organisationale Einbettung dieser Systeme.

Dafür werden zunächst die begrifflichen Grundlagen erläutert und eine Nachhaltigkeitsperspektive auf KI entwickelt. Basierend auf einer Analyse bestehender wissenschaftlicher und gesellschaftlicher Diskurse über die sozialen, ökologischen und ökonomischen Auswirkungen von KI werden dreizehn Nachhaltigkeitskriterien mit entsprechenden Nachhaltigkeitsindikatoren vorgeschlagen sowie Querschnittsindikatoren, die sich neben weiteren Indikatoren auf die organisationale Einbettung dieser Systeme beziehen. Abschließend wird skizziert, wie dieses Kriterien- und Indikatorenset im weiteren Forschungsprozess in Bewertungsinstrumente übersetzt wird. Ziel der hier vorgestellten Forschung ist es, den gesellschaftlichen Diskurs zur Nachhaltigkeit von KI-Systemen zu stärken sowie eine systematische Nachhaltigkeitsbewertung von KI-basierten Systemen zu ermöglichen, um deren Entwicklung und Nutzung im Sinne der Nachhaltigkeit zu stärken.

# Abstract

Technical dynamics and advances in the field of so-called artificial intelligence (AI) and especially machine learning (ML) have led to its ubiquitous application in many societal domains and industrial sectors. Due to its pervasiveness it is sometimes even referred to as a *general purpose technology*. The increasing use of AI systems raises debates about their societal, environmental and economic impacts, such as non-transparent decision-making processes, discrimination, increasing inequalities, rising energy consumption and greenhouse gas emissions in AI model development and application. Other discussions also relate to broader consequences on labour markets, consumption patterns or the market power of large corporations.

From a sustainability perspective AI systems are thus associated with multiple challenges. We argue that these socio-technical systems, with their design, their organizational embedding and implementation processes, can unfold significant impacts on society, the environment and economic actions. The goal of this paper is to expand current discussions on responsible AI into an overarching perspective on sustainable AI. This includes a systematic overview of the social, environmental, and economic impacts along the entire AI lifecycle as well as along the organizational embedding of these systems.

This publication is based on work conducted in the research project “SustAIn – Sustainability Index for Artificial Intelligence”. It presents conceptional ideas for a comprehensive sustainability assessment along the AI lifecycle. For this purpose, first the conceptual foundations and a sustainability perspective on AI are being explained. We then suggest thirteen sustainability criteria with corresponding indicators based on an analysis of existing scientific and societal discourses on the social, ecological and economic impacts of AI. Eventually, we will shortly sketch how in the ongoing research process these indicators will be translated into assessment tools. This paper aims at strengthening discourses on the sustainability of AI as well as at enabling a systematic assessment of the sustainability of AI systems in order to support their development and application in a sustainable manner.

# Die Autorinnen und Autoren

**Friederike Rohde** wissenschaftliche Mitarbeiterin, Institut für ökologische Wirtschaftsforschung, Digitaler Wandel, Technikzukünfte, sozio-technische Transformationsprozesse
**Kontakt: Friederike.Rohde@ioew.de**
**Tel. +49 – 30 – 884 594-57**

**Josephin Wagner** wissenschaftliche Mitarbeiterin, Institut für ökologische Wirtschaftsforschung, Digitaler Wandel, Ökonomie und Governance, sozial-ökologische Transformation
**Kontakt: Josephin.Wagner@ioew.de**
**Tel. +49 – 30 – 884 594-45**

**Philipp Reinhard,** studentischer Mitarbeiter, Institut für ökologische Wirtschaftsforschung, Software & Digital Business, Ethik in KI und natürlicher Sprachverarbeitung
**Kontakt: Philipp.Reinhard@ioew.de**

**Ulrich Petschow,** wissenschaftlicher Mitarbeiter, Institution, Innovations- und Technikanalysen, Institut für ökologische Wirtschaftsforschung, Ökonomische Instrumente und neue Steuerungsformen, Transformationsstrategien.
**Kontakt: Vorname.Name@ioew.de**
**Tel. +49 – 30 – 884 594-39**

**Andreas Meyer,** wissenschaftlicher Mitarbeiter, Distributed Artificial Intelligence Laboratory, TU Berlin, Nachhaltigkeit von KI, Machine Learning
**Kontakt: Andreas.Meyer@dai-labor.de**

**Marcus Voß** leitete das Anwendungszentrum Smart Energy Systems des DAI-Labors an der TU Berlin. Er ist als KI-Experte bei der Birds on Mars GmbH tätig.
**Kontakt: Marcus.Voss@tu-berlin.de**

**Dr. Anne Mollen,** Policy and Advocacy Managerin sowie Projektmanagerin, AlgorithmWatch, Arbeitsschwerpunkte Nachhaltigkeit von KI, automated decision making in der Arbeitswelt und im öffentlichen Sektor, Plattformregulierung.
**Kontakt: Mollen@algorithmwatch.org**

# Inhaltsverzeichnis

**1 Hintergrund und Ziele ..... 11**

**2 Betrachtungsgegenstand ..... 11**

2.1 Charakterisierung von „Künstlicher Intelligenz“ ..... 11
Lernverfahren des maschinellen Lernens ..... 13
Künstliche neuronale Netze ..... 13
2.2 Anwendungsfelder ..... 15

**3 Künstliche Intelligenz und Nachhaltigkeit ..... 19**

3.1 KI für Nachhaltigkeit oder nachhaltige KI? ..... 19
3.2 Verantwortung entlang des KI-Lebenszyklus ..... 22
3.3 Wirkungsebenen und relevante Akteure für nachhaltige KI ..... 23
3.4 Nachhaltigkeitsverständnis ..... 25
3.5 Soziale Nachhaltigkeit ..... 28
3.6 Ökologische Nachhaltigkeit ..... 29
3.7 Ökonomische Nachhaltigkeit ..... 30
3.8 Zusammenfassende Definition für nachhaltige KI ..... 30

**4 Nachhaltigkeitskriterien und -indikatoren für künstliche Intelligenz ..... 31**

4.1 Organisatorische Verankerung: Querschnittskriterien ..... 33
4.1.1 Festgelegte Verantwortlichkeiten ..... 33
4.1.2 Code of Conduct ..... 33
4.1.3 Stakeholder-Analyse & -Beteiligung ..... 34
4.1.4 Dokumentation der KI-Systeme ..... 34
4.1.5 Risikomanagement ..... 35
4.2 Soziale Kriterien und Indikatoren ..... 36
4.2.1 Transparenz und Verantwortungsübernahme ..... 36
4.2.2 Nicht-Diskriminierung und Fairness ..... 37
4.2.3 Technische Verlässlichkeit und menschliche Aufsicht ..... 38
4.2.4 Selbstbestimmung und Datenschutz ..... 39
4.2.5 Inklusives und partizipatives Design ..... 40
4.2.6 Kulturelle Sensibilität ..... 41
4.3 Ökologische Kriterien und Indikatoren ..... 43
4.3.1 Energieverbrauch ..... 43
4.3.2 $CO_2$- und Treibhausgasemissionen ..... 45
4.3.3 Nachhaltigkeitspotenziale in der Anwendung ..... 47
4.3.4 Indirekter Ressourcenverbrauch ..... 49
4.4 Ökonomische Kriterien und Indikatoren ..... 52
4.4.1 Marktvielfalt & Ausschöpfung des Innovationspotenzials ..... 52
4.4.2 Verteilungswirkung in Zielmärkten ..... 54
4.4.3 Arbeitsbedingungen und Arbeitsplätze ..... 55
4.5 Übersicht Kriterien- und Indikatorenset für nachhaltige KI ..... 57
4.6 Mapping der Nachhaltigkeitsindikatoren entlang des KI-Lebenszyklus ..... 64

**5 Herausforderungen und Grenzen der KI-Nachhaltigkeitsbewertung ..... 67**

5.1 Nutzen und Grenzen eines Indikatorenbasierten Ansatzes ..... 67
5.2 Bewertungsinstrumente zur Anwendung in der Praxis ..... 68

**6 Fazit ..... 70**

**7 Literaturverzeichnis ..... 72**

# Abbildungsverzeichnis

Abbildung 1: Abgrenzung der Begriffe „Künstliche Intelligenz“, „Maschinelles Lernen“ und „Deep Learning“ ..... 12
Abbildung 2: Grundaufbau eines künstlichen neuronalen Netzes (KNN) ..... 14
Abbildung 3: Nachhaltige KI und KI für nachhaltige Entwicklung ..... 21
Abbildung 4: Phasen des KI Lebenszyklus ..... 22
Abbildung 5: Wirkungsebenen und relevante Akteure für nachhaltige KI ..... 23
Abbildung 6: Nachhaltigkeitsverständnis im Hinblick auf KI ..... 28
Abbildung 7: Übersicht über die Nachhaltigkeitskriterien ..... 32
Abbildung 8: Anzahl der Parameter populärer ML-Modelle zwischen 2012 und 2021 ..... 44
Abbildung 9: $CO_2$-Emissionen des NLP-Modelltrainings ..... 46
Abbildung 10: Kriterien und Indikatoren entlang des KI-Lebenszyklus ..... 64

# Tabellenverzeichnis

Tabelle 1: Übersicht über Anwendungsfelder von künstlicher Intelligenz ..... 16
Tabelle 2: Übersicht der Nachhaltigkeitskriterien und Indikatoren ..... 57

# 1 Hintergrund und Ziele

Das Thema „Künstliche Intelligenz“ (KI) steht im Fokus zahlreicher Debatten und die Technologieentwicklung ist mit einer großen Innovationsdynamik verbunden. Auf der einen Seite werden Chancen gesehen, durch den Einsatz von KI-basierten Verfahren beispielsweise die Gesundheitsversorgung zu verbessern oder zu Klimaschutz und Klimaanpassung beizutragen (bspw. UBA 2019; Reset.org 2020). Auf der anderen Seite bestehen beim Einsatz dieser Systeme auch Risiken wie intransparente Entscheidungsprozesse, Diskriminierung (Europäische Kommission 2019), steigende Energieverbräuche (Strubell et al. 2019) oder die Erhöhung von Konsum und Produktion, z. B. durch personalisierte Werbung (Kingaby 2021). Während die Debatte um ethisch-soziale Aspekte bereits etabliert ist – und in Regulierungsprojekte wie die geplante KI-Verordnung[1] auf EU-Ebene einfließt –, steht die Diskussion um die ökologischen Aspekte und sozio-ökonomischen Technikfolgen noch ganz am Anfang (vgl. UBA 2019). Ziel unseres Ansatzes ist es, Systeme künstlicher Intelligenz (im Folgenden als „KI-Systeme“ bezeichnet) einer umfassenden Nachhaltigkeitsbewertung zugänglich zu machen. Das bedeutet erstens, bestehende KI-Systeme entlang ihres Lebenszyklus[2] anhand ihrer sozialen, ökologischen und ökonomischen Auswirkungen zu bewerten. Zweitens, gilt es, mithilfe von Indikatoren Ansatzpunkte für Verbesserung vorzuschlagen. Hierzu machen wir zunächst auf Basis der bestehenden Literatur einen Vorschlag, welche ökologischen, sozialen und ökonomischen Aspekte im Kontext von KI-Systemen überhaupt relevant sind. Schließlich stellen wir Kriterien und mögliche Indikatoren vor, die es ermöglichen sollen, diese Auswirkungen sinnvoll zu bewerten. Die Nachhaltigkeitsbewertung soll Akteure aus dem Bereich der Entwicklung und Anwendung von KI-Systemen in die Lage versetzen, die mit diesen Systemen verbundenen sozialen, ökologischen und ökonomischen Risiken einzuschätzen und Maßnahmen zu deren Minimierung zu ergreifen. Auf diese Weise leistet der Ansatz einen Beitrag, um Entwicklung und Nutzung dieser Technologien aus einer Nachhaltigkeitsperspektive zu betrachten und die Entwicklung nachhaltiger KI voranzutreiben.

# 2 Betrachtungsgegenstand

## 2.1 Charakterisierung von „Künstlicher Intelligenz“

Unter dem Begriff „Künstliche Intelligenz“ wird im Allgemeinen die maschinelle Nachbildung von kognitiven menschlichen Fähigkeiten verstanden, indem beispielsweise Entscheidungsstrukturen des Menschen durch Programmierung nachempfunden werden. Auch die Fähigkeit zum Lernen oder zur Improvisation wird dazu gezählt. Während die umfassende Nachbildung menschlicher Intelligenz, die in der Regel als *starke KI* bezeichnet wird (z. B. CogPrime, vgl. hierzu Goertzel et al. 2014) bisher nur in theoretischen Überlegungen besteht, werden *schwache KI-Systeme* in Form von Verfahren des maschinellen Lernens und „Deep Learnings“ in einer Vielzahl von Anwendungsgebieten eingesetzt.

---

1 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

2 Wir nutzen den Begriff im Sinne des gesamten Lebensweges eines KI-Systems von der Forschung und Entwicklung bis zur Anwendung unter Beachtung der Anwendungsebene, aber auch der technischen Ebene und Hardware.

Die Begriffe „Künstliche Intelligenz“, „Maschinelles Lernen“ und „Deep Learning“ werden dabei gelegentlich fälschlicherweise synonym verwendet. Letztere stellen dabei nur Teilgebiete der künstlichen Intelligenz dar (Abbildung 1), die sich insbesondere in den letzten Jahren stark weiterentwickelt haben. Das maschinelle Lernen umfasst Algorithmen, die mithilfe statistischer Methoden Computersysteme befähigen, aus Daten zu lernen und Entscheidungen zu treffen. Dabei unterscheiden sie sich von traditionellen Computer-Algorithmen dadurch, dass Regeln nicht explizit in der Programmierung formuliert werden. Stattdessen werden aus den Daten Muster und Gesetzmäßigkeiten gelernt, um so Vorhersagen auch für neue Daten treffen zu können. Hierbei kommen, abhängig von der Beschaffenheit der Daten und der Zielstellung der Anwendung, verschiedene Verfahren zum Einsatz. Dazu zählen zum Beispiel Algorithmen wie *Random Forest*, *Support Vector Machine*, *k-Nächste Nachbarn* und *künstliche neuronale Netze* (KNN). Mit dem Begriff „Deep Learning“ wird wiederum eine spezielle Klasse von künstlichen neuronalen Netzen mit besonders tiefen, komplexen Architekturen und einer großen Anzahl von zu lernenden Parametern bezeichnet. Diese Unterscheidung ist bedeutsam, weil die Komplexität und die Möglichkeiten, die vom System generierten Entscheidungspfade nachzuvollziehen, sehr unterschiedlich sind.

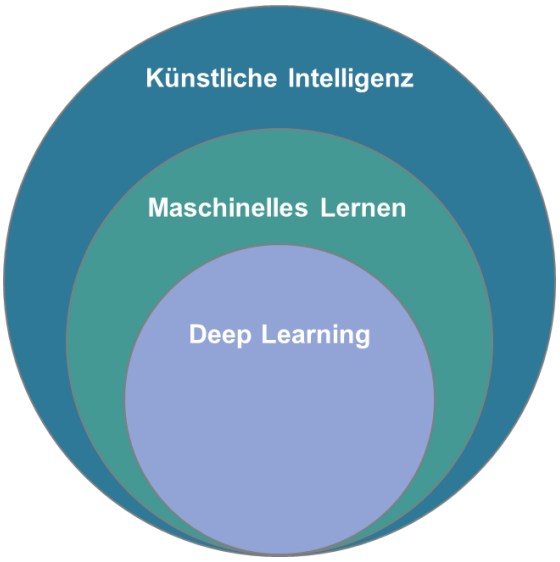


**Abbildung 1: Abgrenzung der Begriffe „Künstliche Intelligenz“, „Maschinelles Lernen“ und „Deep Learning“**
Quelle: Eigene Darstellung

Methoden aus dem Bereich der schwachen KI zeichnen sich gegenüber der starken KI dadurch aus, dass sie jeweils nur für ein sehr konkretes eng umrissenes Problem eingesetzt werden können – z. B. der Erkennung von Objekten in Bildern, der Annotation von Wortkategorien in Texten oder der Vorhersage von Zeitreihen. Als ein aktueller Trend zeichnet sich dabei ab, dass zunehmend Modelle zum Einsatz kommen, die für unterschiedliche Aufgaben trainiert werden können und dabei auch mehrere Modalitäten – wie Text (auch in verschiedenen Sprachen), Bild und Video – unterstützen können. Diese sogenannten *foundation models* (bspw. BERT, GPT3 oder DALL-E) sind mit einer Vielzahl von Möglichkeiten, aber auch Risiken verbunden – angefangen bei ihren Fähigkeiten (z. B. Spracherkennung, Bilderkennung, Robotik, logisches Denken, menschliche Interaktion) über ihre technische Ausgestaltung (z. B. Modellarchitekturen, Trainingsverfahren, Daten, Systeme, Sicherheit, Konzeption und Evaluation) bis hin zu ihren Anwendungen (z. B. Recht, Gesundheitswesen, Bildung) (Bommasani et al. 2021).

## Lernverfahren des maschinellen Lernens

Die Lernverfahren des maschinellen Lernens können in *überwachtes Lernen* (*supervised learning*), *unüberwachtes Lernen* (*unsupervised learning*) und *bestärkendes Lernen* (*reinforcement learning*) unterteilt werden. Beim überwachten Lernen werden dem Algorithmus zum Lernen Daten zur Verfügung gestellt, sodass für viele Beispiele die Eingabeparameter mit dem gewünschten Ergebnis (z. B. Zuordnung zu einer Klasse oder ein Zielwert) vorliegen (sogenannte Label). Während der Lernphase werden fortlaufend die Vorhersagen des Modells mit den erwünschten Ausgaben verglichen, um die Entscheidungsfindung des Modells gegebenenfalls durch Anpassung seiner Parameter zu korrigieren. Beim unüberwachten Lernen werden dem Algorithmus Datensätze ohne zugeordnete Ausgabewerte zur Verfügung gestellt. Ziel des Lernverfahrens ist es, Strukturen in den Daten zu identifizieren, aus denen neue Informationen abgeleitet werden können. Die gängigsten Anwendungen hierfür sind das *Clustering*, bei dem ähnliche Datenpunkte anhand ihrer Eigenschaften gruppiert werden (z. B. zur Kundensegmentierung), und die Assoziationsanalyse, bei der Regeln gelernt werden mithilfe derer Assoziationen gezeichnet werden können (z. B. für Produktempfehlungen). Beim bestärkenden Lernen versucht ein Agent, ein geeignetes Verhalten bzw. eine Strategie in seiner Umgebung zu lernen. Dabei kann der Agent verschiedene vordefinierte Aktionen durchführen, um ein bestimmtes Ziel zu erreichen. Durch Belohnung erwünschter Verhaltensweisen und Bestrafung unerwünschter Verhaltensweisen wird die bisherige Vorgehensweise angepasst. Beispielhafte Anwendungen finden sich unter anderem in der Robotik, indem z. B. ein Bewegungsablauf gelernt wird, in der Optimierung von Kontrollproblemen, wie z. B. Heizungssteuerung oder in Spielen, wie z. B. Go oder Schach.

Eine weitere zuletzt populär gewordene Kategorie an Algorithmen aus dem Bereich des maschinellen Lernens sind *generative Modelle*, die es ermöglichen sehr komplexe Verteilungen, wie z. B. von Bildern und Text, zu modellieren, sodass von dieser gelernten Verteilung neue Instanzen generiert werden können. So können z. B. real aussehende Bilder[3] oder korrekt formulierter Text (siehe z. B. GPT3) generiert werden, sodass diese den Verteilungen aus den Trainingsdaten folgen.

## Künstliche neuronale Netze

Während für viele einfachere Probleme mit strukturierten Datensätzen auch einfache Algorithmen im Einsatz sind, sind künstliche neuronale Netze die zurzeit populärste Gruppe von Algorithmen für eine Vielzahl von Problemen. Die Spannweite der Anwendungen, und damit auch der konkreten Architekturen, ist dabei sehr groß. Sie reicht von kleinen Modellen, die beispielsweise auf strukturierten Daten in Sekunden auf einem handelsüblichen Laptop trainiert werden, bis zu sehr großen Modellen wie GPT3, die z. T. auf spezialisierter Hardware mit tausenden GPUs (Graphics Processing Unit) oder TPUs (Tensor Processing Units) für Tage und Wochen trainiert werden.

Gemein ist diesen künstlichen neuronalen Netzen, dass sie lose vom Aufbau des Gehirns inspiriert sind und dessen Lernprozesse nachbilden sollen. Sie bestehen aus miteinander verbundenen *Neuronen*, die basierend auf einem Eingangssignal durch interne Berechnungen ein Ausgabesignal erzeugen. Innerhalb eines Neurons werden die Eingabesignale summiert. Mithilfe einer Aktivierungsfunktion wird dann ermittelt, welches Ausgabesignal weitergegeben wird. Künstliche neuronale Netze bestehen aus einer Vielzahl an einzelnen Neuronen, die in Schichten angeordnet sind: Eingabeschicht, verborgene Schichten und Ausgabeschicht. Die Informationen fließen von der Eingabeschicht über die verborgenen Schichten zur Ausgabeschicht. Diese Verbindungen werden im

[3] siehe z. B. https://thispersondoesnotexist.com/ unter Nutzung von StyleGAN2

Berechnungsprozess als Gewichte (Verbindungsintensität) dargestellt. Die Gewichte spielen eine wichtige Rolle bei der Ausbreitung des Signals im Netz. Sie enthalten das Wissen des neuronalen Netzes über die Eingabe-Ausgabe-Beziehung. Sowohl die Anzahl der versteckten Schichten als auch die Anzahl der Neuronen bestimmen die Komplexität des Modells und sind wichtige zu wählende Parameter bei der Entwicklung eines KNNs. Ziel des KNN-Trainings ist es, die Fehler zwischen den gewünschten Zielwerten und den vom Modell berechneten Werten zu minimieren.

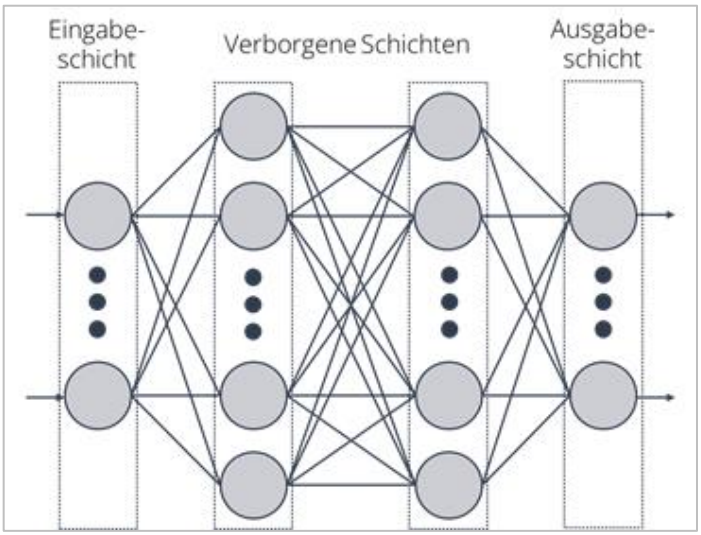


**Abbildung 2: Grundaufbau eines künstlichen neuronalen Netzes (KNN)**
Quelle: Eigene Darstellung

Bei einfachen neuralen Netzen sind alle Neuronen einer Schicht vollständig mit den Neuronen der darauffolgenden Schicht verbunden. Die Informationen fließen dabei nur in eine Richtung. Moderne Architekturen, mit sehr vielen verborgenen Schichten (tiefe neuronale Netze), weichen hiervon ab, da sonst die Anzahl der zu trainierenden Parameter exponentiell ansteigen würde. Ein populärer Ansatz sind hier z. B. *Convolutional Neural Nets* (CNNs – faltende neuronale Netze), die spezielle *Filter*- und *Pooling*-Schichten verwenden. Sie eignen sich besonders für strukturierte Daten wie Bilder oder Zeitreihen. Sie erfassen deren räumliche oder zeitliche Abhängigkeiten und reduzieren dabei die Anzahl der Parameter des Netzes. *Recurrent Neural Nets* (RNNs – rekurrente neuronale Netze) nutzen zirkuläre Verbindungen, und sind so z. B. für die Verarbeitung von Zeitreihendaten oder sequenziellen Daten geeignet. Noch modernere Architekturen verwenden Mechanismen wie *Attention* (Aufmerksamkeit), um Netze mit vielen Parametern trainieren zu können. Oder sie versuchen geometrische Zusammenhänge, wie z. B. in Grafen, direkt in der Struktur abzubilden (*Graph Neural Networks*, Graf-Neuronale Netze). Diese verschiedenen Formen von maschinellen Lernprozessen zeigen die Komplexität des Betrachtungsgegenstands auf.

Wir sprechen im Folgenden von KI-Systemen und definieren diese als Systeme, bei denen die Regeln nicht in der Programmierung des Algorithmus von Menschen festgelegt werden, sondern durch einen selbständigen Lernprozess (aus Daten) entstehen. KI-Systeme umfassen sowohl die zugrundeliegenden Machine Learning-Modelle als auch die zum Lernen genutzten Daten.

Mit unseren Nachhaltigkeitskriterien adressieren wir sowohl KI-Systeme mit einfacheren ML-Modellen als auch mit komplexeren Modellen, die auf Deep Learning basieren. Unsere Nachhaltigkeitskriterien zielen darauf ab, möglichst viele Modelle in ihren Wirkungen zu erfassen und Ansatzpunkte für eine nachhaltige Gestaltung entlang ihres Lebenszyklus aufzuzeigen.

## 2.2 Anwendungsfelder

Die Anwendungsfelder für die Methoden und Technologien, die unter den Begriff „Künstliche Intelligenz“ fallen, sind entsprechend ihrer Rolle als *general purpose*-Technologie im Grunde unbegrenzt (Cockburn et al. 2019). Prinzipiell können sie für alle Zwecke eingesetzt werden, bei denen Schlüsse oder Erkenntnisse aus Daten gezogen werden sollen. Dabei können unterschiedliche Ziele und Erwartungen mit dem Einsatz von KI-Systemen verbunden sein. Aus ökonomischer Perspektive ist die Erwartung, dass sich Geschäftsmodelle unter dem Einsatz von KI-Systemen signifikant verändern und gleichzeitig neue Geschäftsmodelle entwickeln. Dabei wird der Einsatz dieser neuen Technologien mit der Notwendigkeit begründet, Big Data und damit verbundene Marktveränderungen zu managen (Simon 2019). Entsprechend versprechen sich Führungskräfte, laut einer Umfrage von Deloitte (2017), auf der einen Seite marketingrelevante Vorteile, wie die Verbesserung von Produkten, die Entwicklung neuer Produkte und die Erschließung neuer Märkte. Auf der anderen Seite spielen vor allem die Optimierung von Entscheidungsprozessen sowie von internen Geschäftsabläufen und die Entlastung von Mitarbeitenden mit Blick auf operative Aufgaben eine entscheidende Rolle.

Der Verbreitungsgrad von KI-Systemen unterscheidet sich dabei in den unterschiedlichen Branchen – mit der stärksten Verbreitung in Telekommunikations- und Technologieunternehmen, Finanzinstitutionen und der Automobilbranche (Simon 2019). Laut der Umfrage von Deloitte (2017) spielt in den befragten Unternehmen die robotergestützte Prozessautomatisierung beim Einsatz von KI-Systemen die bisher größte Rolle (dicht gefolgt von ML-basierten statistischen Analysen und natürlicher Sprachverarbeitung oder -generierung). Entsprechend weisen insbesondere die Sektoren, deren Wertschöpfungsprozesse überwiegend vorhersehbare körperliche Tätigkeiten beinhalten, ein hohes technisches Potenzial für den Einsatz von KI-Systemen im Kontext von Automatisierung auf (Manyika et al. 2017).

Prinzipiell können KI-Systeme innerhalb der verschiedenen Anwendungsfelder auch mit der Zielstellung einer nachhaltigen Entwicklung zum Einsatz kommen (Vinuesa et al. 2020). Hierbei liegt aktuell ein Fokus auf Umweltschutz- und Klimaschutzzielen. Tabelle 1 gibt einen Überblick über typische Anwendungsfelder – inklusive, soweit vorhanden, Angaben über den Verbreitungsgrad von KI-Systemen in diesen Bereichen und beispielhaften Technologien und Anwendungsfällen sowie Akteuren in diesem Feld.

**Tabelle 1: Übersicht über Anwendungsfelder von künstlicher Intelligenz**

Quelle: Eigene Darstellung

| Bereich /Branche | Verbreitung | Technologien und Anwendungsfälle | Beispielhafte Akteure |
|---|---|---|---|
| Finanzwirtschaft | Laut BMWi (2019) verwenden rund 12,2% aller Unternehmen in der Finanzwirtschaft in Deutschland KI. | Anwendungen für den Kundenverkehr (Bonitätsbewertung, Policen, Chatbots) | Deutsche Bank, Commerzbank (Auswertung von Kundendaten), DZ Bank (Chatbot) |
| | | Anwendungen für interne Prozesse (Riskmanagement, Betrugserkennung) | Alpha-Dig (Deutsche Bank), Commerzbank, DekaBank (Prozessoptimierungen), Sparkasse |
| | | Handel und Portfoliomanagement | PWC, KPMG |
| | | Einhaltung rechtlicher Vorschriften (RegTech, SupTech) | PWC, EY, KPMG |
| Gesundheitswesen | In einer Umfrage der Beratungsgesellschaft PWC bestätigten 30% der CEOs im Gesundheitswesen den Einsatz von KI. | Medizinische Bildgebung | IBM, Merantix, Siemens, Fuji, GeneralElectrics, Philips |
| | | Entscheidungsunterstützung/Patientenversorgung | IBM, PeakProfiling, Lenovo, Philips |
| | | Selbstüberwachung | Ada Health, Maya, Woom, FitBit |
| | | Laborautomatisierung | Siemens Healthineers, Faulhaber, Festo |
| | | KI-gestützte Robotik | |
| | | Telemedizin | Kry, TeleClinic |
| Logistik | In der Logistik setzen nur 1,5% der Unternehmen KI ein. (BMWi 2019). | Prozessoptimierungen | IBM, msg, inconso/Körber, Amazon |
| | | Routen-/Tourenplanung | Synaos, SSI Schäfer, Flaschenpost |
| | | Platooning | Continental/Knorr-Bremse, MAN/DB Schenker, Volvo/Scania |
| Handel | Im Großhandel kommt KI nur bei 1% der Unternehmen zum Einsatz (BMWi 2019). | Prognosen, Platzierungs- und Bestandsmanagement, Finanzierung, Buchhaltung | Intel, alexanderthamm, Ailio, adesso |
| | | Dynamische Preisoptimierung | SAP, 7learnings, Prudsys, Minderest |
| Industrie 4.0 | Laut Bitkom setzen 12% aller deutschen Industrieunternehmen KI ein. | Internet of Things (IoT) | Bosch, IBM, Ericsson, Siemens, SAP, Software AG |
| | | Wartung / Predictive Maintenance | IBM, SAP, Siemens, Microsoft, General Electric, Intel, Bosch, |
| | | Robotik in der Produktion | ABB, Kuka, Kawasaki, Fanuc |
| Bildung | Gemäß der Zukunftsstudie Münchner Kreis Band VIII sehen 60% aller befragten Experten & Expertinnen einen großen Einfluss von KI. | Persönliche Lernassistenten | |
| | | E-Learning | |
| | | Plagiats-Software | |

| | | | |
|---|---|---|---|
| Smart Home | Etwa 20% der Haushalte in Deutschland nutzen Anwendungen, die auf KI basieren (z. B. Sprachassistenten) (Statista 2021). | Energiemanagement Pflege, z. B. Vorhersage von Unfällen | mitpflegeleben |
| | | Personal Assistant | Apple Siri, Amazon Alexa, Microsoft Cortana, Google, Bixby |
| Marketing & Soziale Medien | Im Marketing geben 7% der Führungskräfte an, dass ihre Unternehmen laut einer Studie der SRH Berlin University of Applied Sciences KI intensiv nutzen. | Empfehlungssysteme | Amazon, Spotify, Netflix, Youtube, Tinder, Facebook, |
| | | Personalisierte Werbung | Richrelevance, Kibo-Personalization |
| | | Soziale Medien | Facebook, Twitter, Instagram, SnapChat, TikTok |
| | | ChatBots | IBM Bluemix, Google Cloud AI, Microsoft Cognitive Services, Amazon AWS AI, Rasa |
| Recht | 9% aller KI einsetzenden Unternehmen nutzen KI im Bereich *Legal & Compliance* (Deloitte 2020). | RegTech | Deloitte, EY, PWC, KPMG, Oracle |
| | | Vertragsüberprüfungen | Deloitte, EY, PWC, KPMG, Oracle |
| | | Compliance | Deloitte, EY, PWC, KPMG, Oracle |
| Sicherheit und Verteidigung | Im Bereich IT-Sicherheit geben 16% aller befragten IT-Verantwortlichen an, dass in ihren Unternehmen KI für Cybersecurity eingesetzt wird (Deloitte 2020). | Cybersecurity | Capgemini, IBM, Deloitte axians, |
| | | Physische Wach- und Sicherheitsdienste | Ciborius, Security Robotics |
| | | Autonome Waffensysteme | |
| Energie | Eine Umfrage der Deutschen Energie-Agentur zeigt, dass 82% der Befragten davon überzeugt sind, das KI im Energiesektor eine große Bedeutung zukommen wird. (Dena 2019) | Smart Buildings | Leanheat, Bosch, IBM, Telekom, Siemens, Cisco, Arup Neuron, infineon, dormakaba, Recogizer |
| | | Smart Grid | PowerTAC, PSIngo, Siemens - Spectrum Power Solutions, DAI-Labor – AC Smart Energy Systems, Huawai, gridX |
| | | Forecast | Carbon Intensity Forecast (API), ElectricityMap, myst AI, enercast, peltarion, Dexter |
| Mobilität | Laut einer Bitkom Umfrage (2020) nutzen 19% der Befragten KI zur Planung von Transportrouten. | Autonomes Fahren | AAI (Berlin), SIGRA Deep Einstein (München), Phantasma Labs (Berlin), MotorAI (Berlin), AImotive, Artisense, Waymo, VW, Daimler/Bosch, Continental, EDAG CityBot, Tesla |
| | | Mobility as a Service | Hacon, PTV Group, Jelbi, Google, Uber, Free2move, Vulog AiMA, Axon Vibe, Moovit, EasyMile, Zoox, ReideCell |
| | | Treibstoffplanung Carsharing | AI-Rescue, arvata systems, Brodtmann Consulting, YUNEX Traffic, pickwings |

| | | | |
|---|---|---|---|
| | | Fahrgastanalysen<br>Smart Parking | |
| Anwend-ungen im Umwelt bereich | 1,3 % der Start Ups einer Crunchbase Analyse entwickeln KI mit direktem oder indirektem Nachhaltigkeits bezug (UBA 2019) | Monitoring & Schutz von Ökosystemen<br>Klimaschutz<br>Energiewende<br>Kreislaufwirtschaft & Res-sourceneffizienz<br>Landwirtschaft 4.0 | KI Leuchttürme des BMUs: I4C, KI am Zug, AuSeSol …<br>Reset-Green Book: FishFace, Project Zamba, The Guardian, Tesselo, Hawa Dawa, Smarter Sorting, OpenSurface, Hyper-ganic, REIF, Plantix, Eco-robotix, AeroFarms<br>Climate Change AI |

# 3 Künstliche Intelligenz und Nachhaltigkeit

Bezüglich des Zusammenhangs von KI-Systemen und Nachhaltigkeit existieren in aktuellen Debatten unterschiedliche Perspektiven und Zugänge, die untereinander jedoch viele Schnittstellen aufweisen. Grundsätzlich kann unterschieden werden zwischen den Bereichen KI für nachhaltige Entwicklung (*AI for sustainability*) und nachhaltiger KI (*sustainable AI* oder *sustainability of AI*) (vgl. dazu auch Rohde et al. 2021; Wynsberghe 2021). Auf der einen Seite werden KI-Systeme entwickelt, die explizit positive Beiträge für soziale und ökologische Zielsetzungen leisten wollen und damit auch die Erreichung der Sustainable Development Goals (SDGs) vorantreiben sollen (z. B. Vinuesa et al. 2020) (KI für nachhaltige Entwicklung). Auf der anderen Seite werden die Auswirkungen von KI-Systemen auf Menschen und Umwelt in Diskursen über ethische und verantwortungsvolle KI diskutiert (z. B. Cockelberg 2020) sowie die Frage, welche Umweltauswirkungen im Hinblick auf Energieverbrauch und Treibhausgasemissionen von den KI-Systemen selbst (z. B. durch Training und Inferenz) verursacht werden (nachhaltige KI). Dennoch stehen die beiden Perspektiven – KI für Nachhaltigkeit und nachhaltige KI – grundsätzlich in enger Verbindung zueinander. Denn die Auswirkungen von KI-Systemen auf Menschen und Umwelt können freilich positive oder negative Konsequenzen für die nachhaltige gesellschaftliche Entwicklung beinhalten.

## 3.1 KI für Nachhaltigkeit oder nachhaltige KI?

Im wachsenden Diskurs um die Rolle von KI für die Erreichung der Ziele der nachhaltigen Entwicklung (SDGs) werden große Erwartungen in die Nutzung von KI-Systemen gesetzt. Zunehmend werden unter Begriffen wie *AI for Earth* oder *AI for social Good* die Potenziale von KI-Systemen zur Lösung sozialer und ökologischer Problemlagen hervorgehoben. Besondere Potenziale für die Erreichung von Nachhaltigkeitszielsetzungen werden in Anwendungsfeldern wie dem Monitoring von Ökosystemen, Klimaschutz, Kreislaufwirtschaft- und Ressourceneffizienz, Agrar- und Lebensmitteltechnologien oder der Energiewende (Netzsteuerung, Prognosen, Gebäude- und Energieeffizienz) gesehen (Reset.org 2020; UBA 2019; Rolnick et al. 2019; Clutton-Brock et al. 2021). Eine umfassende Übersicht über die Anwendungsmöglichkeiten im Bereich Klimaschutz hat die Initiative *Climate Change AI* in einem Überblickspaper (Rolnick et al. 2019) sowie einem Wiki[4] zusammengetragen. Zudem haben erste Studien begonnen sich mit den Wirkungen von KI im Hinblick auf die Erreichung der SDGs zu beschäftigen und sowohl Risiken als auch Potenziale identifiziert (Vinuesa et al. 2020; Sætra 2021; Kaack et al. 2021).

Eine wesentliche Herausforderung, die sich im Hinblick auf die Analyse der Beiträge von KI-Systemen für die Erreichung der SDGs ergibt, ist, dass ganz grundsätzlich Wechselwirkungen zwischen den unterschiedlichen Zielen zum Teil nicht klar sind und Zielkonflikte häufig nicht ausreichend adressiert werden (Nash et al. 2020). Auch die Bewertung der Auswirkungen von KI auf die SDGs steht vor der Herausforderung, Interdependenzen zwischen einzelnen Zielen ausreichend berücksichtigen zu müssen. Sætra (2021) unterscheidet dafür zwischen direkten und indirekten sowie Mikro-, Meso- und Makrolevel-Auswirkungen von KI auf verschiedene SDGs. Direkte Auswirkungen bedeuten in diesem Fall, dass der Einsatz von KI sich direkt auf das Erreichen eines SDGs auswirkt (z. B. KI zur Integration von erneuerbaren Energien). Indirekte Auswirkungen beziehen

[4] https://wiki.climatechange.ai

sich darauf, inwieweit der Einsatz von KI zur Erreichung eines bestimmten Ziels (z. B. Wirtschaftswachstum (SDG 8)) Auswirkungen auf ein anderes hat (z. B. Verschärfung von Ungleichheiten (SDG 10). Außerdem gilt es zu betrachten, auf welcher Ebene diese Auswirkungen stattfinden. So könnte Wirtschaftswachstum innerhalb einer Region oder eines Landes (hier Meso-Perspektive) zwar positiv sein, gleichzeitig könnten sich die Ungleichheiten innerhalb des Landes (Mikro) und zwischen Ländern (Makro) weiter verstärken (Sætra 2021). Die Frage, inwieweit KI-Systeme positive oder negative Auswirkungen auf die Ziele der nachhaltigen Entwicklung haben, lässt sich daher nicht unbedingt ohne Weiteres beantworten.

Wir wollen im Folgenden unsere Perspektive auf nachhaltige KI, die wir als *eingebettete Perspektive* bezeichnen, näher erläutern (Abbildung 3). Im Zusammenhang mit den gesellschaftlichen Auswirkungen von KI existieren verschiedene Diskurse. Wir knüpfen mit unserem Ansatz sowohl an die Diskussion zu KI und Ethik (z. B. Jobin et al. 2019; Cockelbergh 2020) als auch zu *Green AI* (Schwarz 2020) und *sustainability of AI* (Whynsberghe 2021) an. Die Diskussion über ethische Aspekte wird häufig nicht mit dem Verweis auf Nachhaltigkeit geführt. Mit einer umfassenden Definition sozialer Nachhaltigkeit (vgl. Kapitel 3.5.) sind jedoch auch diese Aspekte als Gegenstand einer Nachhaltigkeitsbewertung zu betrachten.

Im Hinblick auf die Nachhaltigkeitsbewertung von KI-Systemen selbst werden in den aktuellen Diskussionen eher soziale Nachhaltigkeitsaspekte (z. B. Kohl 2020) und verantwortungsvolle KI-Entwicklung fokussiert (z. B. Europäische Kommission 2019; Arrieta et al. 2020; Spiekermann 2021). Ökologische Aspekte halten zudem vermehrt Einzug in die Debatte – etwa in Form von Fragen nach dem ökologischen Fußabdruck der eingesetzten Modelle sowie der Größe und Rechenintensität der Modelle mit Blick auf das sinnvolle und erfolgreiche Erfüllen ihrer Aufgabe (vgl. van Wynsberghe 2021; Schwartz et al. 2020). Auch über die ökonomischen Dynamiken, die mit der zunehmenden Verbreitung von KI verbunden sind, gibt es einen wachsenden Diskurs (z. B. Bughin et al. 2018; Simon 2019; Lu & Zhou 2019; Altenried 2020; Klinova & Korinek 2021; Prause et al. 2021; Tratjenberg 2018), der jedoch nicht mit explizitem Bezug zu Nachhaltigkeit geführt wird. Wir argumentieren, dass auch die Verstärkung problematischer ökonomischer Dynamiken aus einer ökonomischen Nachhaltigkeitssicht nicht wünschenswert und zielführend ist.

Bei der wachsenden Diskussion um die Nutzung von KI-Anwendungen für soziale und ökologische Herausforderungen (*AI for Earth*, *AI for social good*, *Climate Change AI* oder A*I for Sustainable Development*) steht eine instrumentelle Perspektive im Vordergrund, die die Potenziale der Technologie für die nachhaltige Entwicklung stark in den Vordergrund stellt. Allerdings stellen diese Anwendungen bislang, wie unsere Übersicht in Tabelle 1 zeigt, nur eine kleine Nische im Bereich möglicher KI-Anwendungen dar. Es kann davon ausgegangen werden, dass der viel größere Teil der Entwicklung und Anwendung von KI-Systemen ohne direkten Nachhaltigkeitsbezug stattfindet, beziehungsweise nicht mit dem expliziten Ziel einer sozial-ökologischen Transformation antritt – etwa in den Bereichen Finanzwirtschaft, Industrie 4.0, Handel und Logistik oder im Marketing und in den sozialen Medien. Wir entwickeln daher eine *eingebettete Perspektive* auf nachhaltige KI, die Technologie als gesellschaftlich gestaltet und gestaltbar begreift und nicht als sogenanntes neutrales Werkzeug. KI-Systeme sind mit inhärenten sozialen, ökologischen und ökonomischen Wirkungen verbunden, die in der Entwicklung und Anwendung berücksichtigt werden sollten. Die Ebene der KI-Entwicklung lässt sich dabei nicht immer von der Ebene der Anwendung trennen, denn allein die Entscheidung, was das Ziel einer Optimierung sein soll – beispielsweise der schnellste oder der treibstoffsparendste Weg zu einem Ziel – kann schon Implikationen für die Nachhaltigkeitspotenziale in der Anwendung haben. Mit unserer eingebetteten Perspektive auf Nachhaltigkeitskriterien und -indikatoren möchten wir die verschiedenen Diskurse verknüpfen und zu einer übergreifenden Nachhaltigkeitsperspektive zusammenfassen.

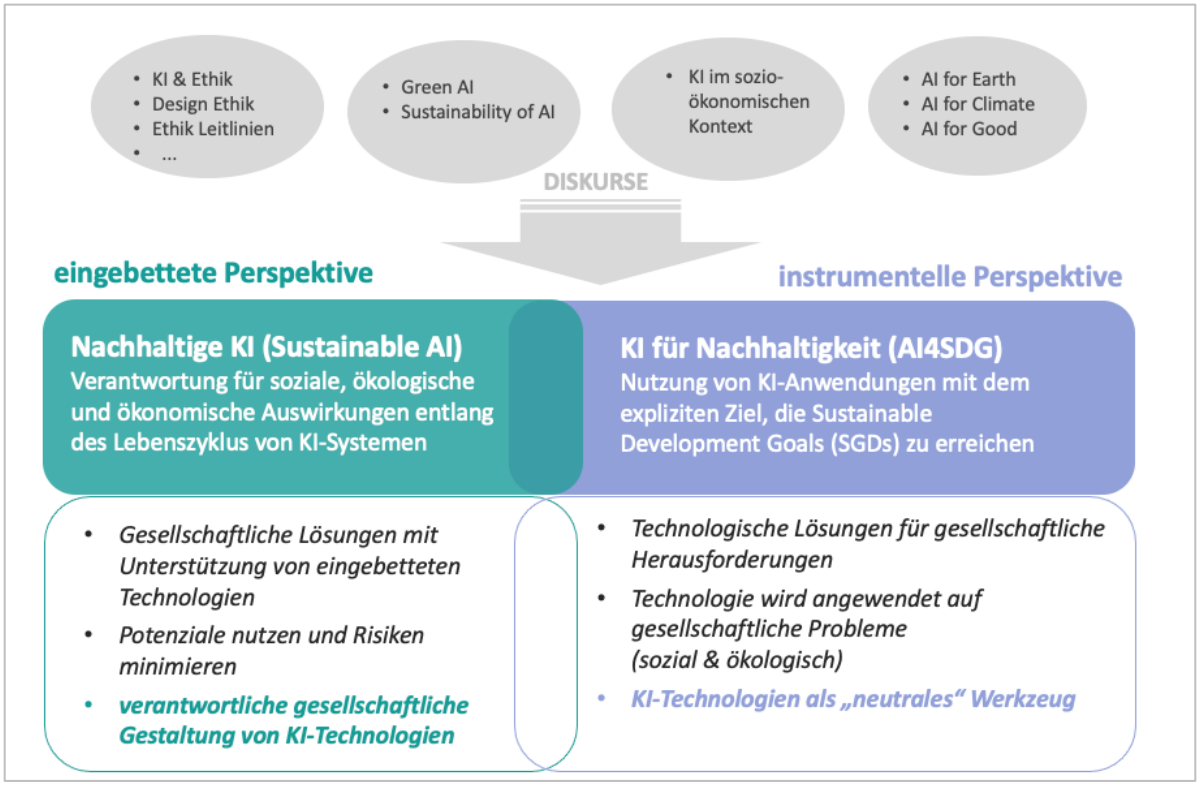


**Abbildung 3: Nachhaltige KI und KI für nachhaltige Entwicklung**
Quelle: Eigene Darstellung

Dass diese Perspektiven stärker integriert werden sollten, zeigen aktuelle Entwicklungen von Empfehlungs- und Standardisierungsprozessen. So adressieren die UNESCO-Empfehlungen für ethische KI neben gesellschaftlichen und menschenrechtlichen Aspekten auch weitreichende ökologische Auswirkungen. Darunter sind beispielsweise direkte und indirekte Umweltauswirkungen während des gesamten Lebenszyklus von KI-Systemen, z. B. der $CO_2$-Fußabdruck, der Energieverbrauch und die Umweltauswirkungen der Rohstoffgewinnung für die Herstellung von KI-Technologien (UNESCO 2021) sowie die Verringerung der Umweltauswirkungen von KI-Systemen und Dateninfrastrukturen. Auch der kürzlich veröffentlichte AI Prozessstandard IEEE 7000 adressiert die Lebenszyklusperspektive mit Blick auf verantwortungsvolles Engineering (Spikermann 2021). Aus der Community der Entwickelnden selbst kommen darüber hinaus kritische Reflexionen zum Einsatz von KI-Systemen, wie das prominent gewordene Paper zu diskriminierenden Effekten großer Sprachmodelle (Bender et al. 2021), das Working Paper zu Klimawirkungen von KI (Kaack et al. 2021) oder ein kürzlich erschienener Beitrag über die Risiken und Potenziale von *foundation models* (Bommasani et al. 2021).

Denn für die Frage, ob KI-Systeme positive oder negative Wirkungen im Hinblick auf die Ziele für nachhaltige Entwicklung entfalten, kann nicht allein darauf geschaut werden, in welchem Sektor KI-Systeme eingesetzt werden und ob sich daraus möglicherweise positive Beiträge für einzelne Aspekte nachhaltiger Entwicklung (z. B. Klimaschutz oder Armutsbekämpfung) ableiten lassen. Diese instrumentelle Perspektive greift zu kurz. Vielmehr müssen die Nachhaltigkeitswirkungen von KI-Systeme entlang des Lebenszyklus analysiert werden. Bei den Nachhaltigkeitskriterien und -indikatoren entwickeln wir daher eine eingebettete Perspektive auf nachhaltige KI, die in erster Linie die sozialen, ökologischen und ökonomischen Auswirkungen entlang des Lebenszyklus aller KI-Sys-

teme adressiert. Die Wirkungen von KI-Anwendungen, die mit dem Ziel antreten, soziale und ökologische Probleme zu lösen, haben wir teilweise im Indikatorenset abgebildet (Nachhaltigkeitspotenziale der Anwendung). Doch bislang zählt der Großteil der eingesetzten KI-Anwendungen nicht dazu. Darum sollte der Fokus zunächst auf einer grundsätzlichen Nachhaltigkeitsbewertung liegen.

## 3.2 Verantwortung entlang des KI-Lebenszyklus

Wir möchten mit diesem Diskussionsbeitrag dafür plädieren, Nachhaltigkeit als eine übergreifende Zielperspektive zu betrachten. Es geht uns darum, die verschiedenen Aspekte, die aktuell im Hinblick auf die verantwortungsvolle Entwicklung und Nutzung von KI sowie auf ihre sozialen, ökologischen und ökonomischen Wirkungen diskutiert werden, in einer übergreifende Nachhaltigkeitsperspektive zusammenzuführen. Dabei kann Nachhaltigkeit als globales normatives Leitbild verstanden werden, auf das sich die Weltgemeinschaft geeinigt hat und das in Form der SDGs konkretisiert wurde. Bei unserem Ansatz geht es jedoch nicht in erster Linie um den Beitrag einzelner Anwendungen für einzelne Aspekte nachhaltiger Entwicklung. Es geht auch nicht darum, dass Erreichen der SDGs lediglich an neue technische Möglichkeiten der Datenverarbeitung zu knüpfen. Sondern es geht im Sinne von *responsbile research and innovation* (RRI) auch darum sicherzustellen, dass die Innovationsrichtung in Bezug auf alle entwickelten und zu entwickelnden KI-Systeme nachhaltig ist.

Nachhaltige KI bezieht sich also in erster Linie darauf, dass KI-Systeme ganz grundsätzlich in einer Art und Weise entwickelt und genutzt werden, die es erlaubt, die vielfältigen Potenziale dieser Technologie (auch für die Ziele der nachhaltigen Entwicklung) zu nutzen und dabei die sozialen, ökologischen und ökonomischen Dimensionen dieser Technologie angemessen zu berücksichtigen. Die nachhaltige Gestaltung von KI-Systemen hat dementsprechend auf einer übergeordneten Perspektive auch Auswirkungen auf die SDGs und diese Gesamtsicht erübrigt sich dadurch nicht. Was wir mit unserem Kriterien- und Indikatorenset anstreben, ist jedoch auf der ganz konkreten Ebene der KI-Systeme mit Blick auf ihren Lebenszyklus geeignete Indikatoren zu identifizieren, um ihre Nachhaltigkeitswirkungen abschätzen zu können. Denn nur so können KI-Systeme entworfen, entwickelt und eingesetzt werden, die nicht auf der einen Seite Probleme lösen und auf der anderen Seite wieder neue Probleme schaffen. Die Wirkungsrichtungen und -dimensionen von KI-Systeme sind sehr komplex und die beteiligten Akteure vielfältig. Doch nur durch eine holistische Betrachtung dieses sozio-technischen Systems kann eine umfassende Nachhaltigkeitsperspektive erreicht werden. Abbildung 4 zeigt schematisch Phasen eines KI-Lebenszyklus, auf die wir bei der Entwicklung der Indikatoren bezog genommen haben.


**Abbildung 4: Phasen des KI Lebenszyklus**

Quelle: Eigene Darstellung in Anlehnung an Cheatham et al. (2019)

## 3.3 Wirkungsebenen und relevante Akteure für nachhaltige KI

Um diese Wirkungsdimensionen und Betrachtungsebenen genauer abzubilden, liegt unseren Überlegungen eine Systematisierung der Wirkungsebenen von KI-Systemen zugrunde (Abbildung 5). Diese Systematisierung soll es ermöglichen, die Komplexität und gesellschaftliche Einbettung von KI als sozio-technischem System ein Stück weit abzubilden. Wir haben auf eine sehr differenzierte Darstellung von Wirkungsrichtungen zwischen verschiedenen Ebenen zunächst verzichtet, weil wir die grundsätzlichen Wirkungsebenen sowie die jeweils relevanten Akteure aufzeigen wollen. Denn bloße Leitlinien, dass wird im Diskurs um KI und Ethik zunehmend deutlich (z. B. Resseguier & Rodrigues 2020), sind ein zahnloser Tiger, wenn sie nicht mit konkreten Verantwortungszuschreibungen und Handlungsmöglichkeiten verbunden sind. Hintergrund dieser Systematisierung ist deshalb vor allem die Frage, wer eigentlich Adressat von Verantwortungsübernahme sein kann. Denn soziale, ökologische und ökonomische Wirkungen, die mit der Entwicklung und Nutzung von KI verbunden sind, liegen zum Teil auf verschiedenen Ebenen. Ziel ist es im weiteren Projektverlauf aufzuzeigen, wo Akteure, die diese Systeme entwickeln, nutzen oder ihre Entwicklung innovationspolitisch vorantreiben und regulativ flankieren, steuernd ansetzen können.

Wir unterscheiden zwischen drei verschiedenen Wirkungsebenen, die von Bedeutung für die Nachhaltigkeitsbewertung von KI-Systemen sind: die Ebene der KI-Systeme selbst, die Anwendungsebene und die gesamtgesellschaftliche Ebene.

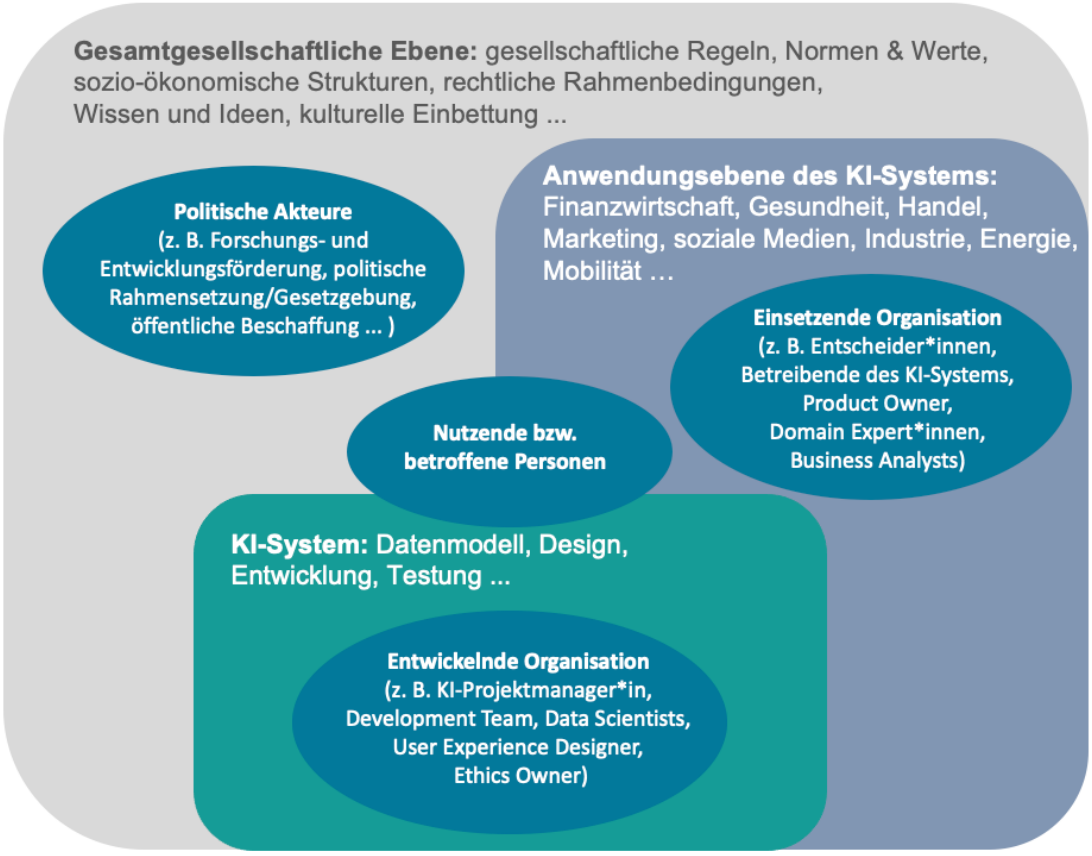


**Abbildung 5: Wirkungsebenen und relevante Akteure für nachhaltige KI**
Quelle: Eigene Abbildung

Die **gesellschaftliche Ebene** umfasst vor allem die übergeordneten Strukturen – im Sinne von Regeln und Vorgaben sowie die kulturelle Einbettung und gesellschaftliche Normen und Werte. All diese Aspekte haben einerseits einen Einfluss darauf, mit welchem Wissen, welchen Zielen und unter welchen Rahmenbedingungen KI-Systeme entwickelt werden. Andererseits wirken diese Systeme auch in viele gesellschaftliche Bereiche hinein, wenn sie zur Vergabe von Sozialleistungen oder von Krediten, für Einstellungsverfahren oder zur Unterstützung juristischer Leistungen genutzt werden. Die gesellschaftlichen Folgen dieser Systeme des automatisierten Entscheidens (automated decision making systems; ADM-Systeme), die immer häufiger auf KI-Systemen basieren, sind sehr vielfältig (Chiuisi et al. 2020). Im den von uns entwickelten Indikatorenset sind vor allem kulturelle Aspekte sowie die ökonomischen Dimensionen dieser Betrachtungsebene abgebildet.

In enger Verbindung zur gesamtgesellschaftlichen Ebene steht die **Anwendungsebene** von KI-Systemen. Denn nur in der konkreten Anwendung zeigt sich, ob die Wirkungen eines KI-Systems als kritisch einzustufen ist oder nicht. In ihrer Charakterisierung als *general purpose*-Technologie, können derartige Systeme und Verfahren für eine Vielzahl von Aufgaben herangezogen werden, die auch im Hinblick auf Nachhaltigkeit, ganz unterschiedliche Wirkungen entfalten können. So können KI-Systeme beispielsweise durch Fernerkundungsalgorithmen für die Analyse von Satellitenbildern genutzt werden, um Informationen über landwirtschaftliche Produktivität zu sammeln oder den Energieverbrauch von Gebäuden vorherzusagen. Gleichzeitig werden KI-Verfahren eingesetzt, um beispielsweise Öl- und Gasexploration zu beschleunigen (Kaack et al. 2020). KI-basierte Verfahren können genutzt werden, um Menschen im öffentlichen Raum zu überwachen (Jansen 2021), aber auch um moderne Sklaverei und Menschenhandel nachzuvollziehen (vgl. AI4Good[5]). Die Ziele, die mit dem Einsatz verbunden sind, sowie die Frage, was genau mit diesem System optimiert wird, spielen daher eine wichtige Rolle, wenn es um die Frage geht, in welchem Zusammenhang KI und Nachhaltigkeit zueinanderstehen. In unserem Kriterien- und Indikatorenset fokussieren wir stärker auf den KI-Lebenszyklus. Doch weil die Anwendungsebene bei einer Nachhaltigkeitsbetrachtung nicht ausgeklammert werden kann, beziehen sich einige unserer Kriterien – wie ökologische Nachhaltigkeitspotenziale in der Anwendung oder Arbeitsmarkteffekte – auch auf den Anwendungskontext.

Die **Ebene des KI-Systems** selbst adressiert vor allem den Entwicklungsprozess von Modellen, der über Datenbeschaffung und Datenmanagement, Konzeptualisierung und Training auch Testung und möglicherweise Re-Training sowie vor allem die Anwendung (auch Inferenz bezeichnet) beinhaltet. Um die Einbettung des KI-Systems als sozio-technisches Systems zu verdeutlichen, liegt dieses Feld innerhalb der gesamtgesellschaftlichen Ebene, denn die Entwicklung von KI-Systemen wird in großem Maße von übergeordneten gesellschaftlichen Normen, Denkmustern und natürlich auch den regulativen Rahmenbedingungen beeinflusst. Innerhalb des KI-Lebenszyklus spielt diese Wirkungsebene somit eine besondere Rolle, insbesondere wenn es darum geht, welche Akteure für eine Veränderung auf dieser Ebene relevant sind. Denn innerhalb dieser Ebene werden sehr viele Entscheidungen getroffen, die für die Ausgestaltung einer nachhaltigen KI von großer Bedeutung sind.

Im Hinblick auf die **relevanten Akteure** ist es wichtig zu betonen, dass die Verantwortung auf viele verschiedene Akteure verteilt ist, denn es handelt sich um ein komplexes sozio-technisches System. Den **entwickelnden Organisationen** kommt insofern eine bedeutende Rolle zu, da sie über

[5] https://aiforgood.itu.int/

sehr viele Eigenschaften des technischen Systems entscheiden, die im späteren Verlauf mit verschiedenen Risiken (Diskriminierung, Privatsphäre, Energieverbrauch usw.) verbunden sind. Dies kann die einsetzende Organisation selbst sein, ist aber aktuell häufig ein externer Dienstleister. Alternativ kann es sich um eine innerhalb desselben Unternehmens dezidiert zuständige Abteilung handeln.

**Einsetzende Organisationen:** Bei den einsetzenden Organisationen kann es sich um Unternehmen, die öffentliche Verwaltung und sonstige Organisationen, wie NGOs oder zivilgesellschaftliche Akteure handeln. Wie die Diskussionen um den geplanten AI Act der Europäischen Kommission verdeutlichen, sollen viele Fragen im Hinblick auf die Risiken, die mit dem Einsatz dieser Systeme verbunden sind, auf der Ebene der entwickelnden Organisationen geregelt werden. Gleichzeitig können die entwickelnde Organisationen möglicherweise nicht über alle Aspekte des KI-Systems Auskunft geben und den einsetzenden Organisationen zur Verfügung stellen, weil kein angemessener Dokumentationsprozess bei der Entwicklung stattgefunden hat. Für die Risiken, die sich im Hinblick auf den Einsatz der Systeme ergeben, sollten geeignete Vorkehrungen und organisationsspezifische Maßnahmen getroffen werden, die wir u. a. in den Querschnittsindikatoren (Kapitel 4.1) adressieren.

**Nutzende und betroffene Personen:** Insbesondere, wenn Menschen mit KI-Systemen interagieren oder von deren automatisiert getroffenen Entscheidungen abhängig sind, müssen sie im Sinne der Nachhaltigkeit als relevante Stakeholder berücksichtigt werden. Sie können sowohl als Wissensberechtige sowie als Betroffene gesehen werden, mit einem unter Umständen formulierten Recht auf Partizipation, Transparenz und Erklärbarkeit der eingesetzten KI-Systeme. Die Perspektive von Nutzenden und Betroffenen kann in der Planung und Entwicklung berücksichtigt werden. Für eine informierte und selbstbestimmte Nutzung von KI-Systemen sind ausreichend Informationen über die eingesetzten Systeme Voraussetzung. Zudem können Redress-Mechanismen eine aktive Rolle von Betroffenen sichern, indem diese automatisierte Entscheidungen anfechten können. Instrumente zur Umsetzung sind beispielsweise Risiko-Analysen, Technikfolgenabschätzungen[6], Auditierungsverfahren sowie Beschwerdeverfahren[7].

**Politische Akteure:** Einige Nachhaltigkeitsindikatoren überschreiten den individuellen und organisationsbezogenen Verantwortungsspielraum, in dem Sinne, dass nicht einzelne entwickelnde oder anwendende Organisationen hier handlungsfähig sind. An dieser Stelle braucht es umfassende regulatorische Ansätze und politische Lösungen. Dies betrifft beispielsweise die Datenpolitik, wie z. B. der Datenzugang für kleine und mittelständische Unternehmen (KMU), das klimaneutrale Betreiben von Datenzentren oder auch Transparenzverpflichtungen für KI-einsetzende Organisationen, um eine Bewertung der Nachhaltigkeit von KI überhaupt möglich zu machen.

## 3.4 Nachhaltigkeitsverständnis

Die Debatte um Nachhaltigkeit und die Interpretation des Nachhaltigkeitsbegriffs hat mit dem Bericht der Brundtlandkommission 1987 einen zentralen Impuls erfahren. Nachhaltigkeit wurde wie folgt interpretiert: „Sustainable development meets the needs of the present without compromising the ability of future generations to meet their own needs.“ (Brundtland 1987) Damit wurden bereits

[6] https://algorithmwatch.org/de/adms-impact-assessment-public-sector-algorithmwatch

[7] https://unding.de

die Grundfragen einer nachhaltigen Entwicklung, nämlich die Frage nach der intra- und intergenerationalen Gerechtigkeit gestellt. Mit der Rio-Konferenz 1992 wurden die Dimensionen Ökologie, Soziales und Ökonomie betont und insbesondere auch ihre Interdependenzen. In den neunziger Jahren des letzten Jahrhunderts wurden unterschiedliche Nachhaltigkeitskonzepte diskutiert, wobei die Diskussionen um starke und schwache Nachhaltigkeit eine wichtige Rolle spielten. Wesentliches Unterscheidungskriterium dabei war unter anderem die Frage nach der Substituierbarkeit des sogenannten Naturkapitals. Die Vertreter*innen der starken Nachhaltigkeit gingen im Wesentlichen davon aus, dass eine Substituierbarkeit nicht gegeben sei, während die Vertreter*innen der schwachen Nachhaltigkeit von einer weitgehenden Substituierbarkeit von Naturkapital durch Menschen gemachtes Kapital (z. B. in Form von technologischen Entwicklungen) ausgingen. Aus diesen unterschiedlichen Interpretationen ergeben sich weitgehende Unterschiede in den Handlungsorientierungen.

Die deutsche Diskussion fokussierte sich auf ein Nachhaltigkeitsverständnis, dass die drei miteinander eng verbunden Dimensionen der Nachhaltigkeit (Ökonomie, Soziales, und Ökologie) als Drei-Säulen-Modell konzeptualisierte und letztlich darauf fokussierte, dass alle drei Säulen in sich nachhaltig sein sollten und mithin Zielkonflikte zu beachten sind: Demnach bestünde die Gefahr, dass eine zu starke Betonung der ökologischen Säule zu einer Gefährdung der ökonomischen Säule führen würde. Nachhaltige Entwicklung wurde dementsprechend als „regulative Idee“ (Homann 1996) verstanden und die Vorstellung „harter“ Grenzen verneint. In Forschungen und Diskussionen zum Klimawandel wurde diese Interpretation zunehmend infrage gestellt und mit dem Budgetansatz des WBGU (2009) verdeutlicht, dass eine Übernutzung natürlicher Ressourcen – in diesem Fall der Aufnahmekapazität der Atmosphäre für Treibhausgase – zu massiven Schäden führen werde. Fast zeitgleich wurde das Konzept der „planetary boundaries“ (Rockström et al. 2009 und später weiterentwickelt Steffen et al. 2015) entwickelt. Es benennt die ökologischen Grenzen der Erde, deren Überschreitung die Stabilität der Ökosysteme und die Lebensgrundlagen der Menschheit gefährden würde. Das Konzept zeigt gleichzeitig den sogenannten sicheren Handlungsspielraum (*safe operating space for humanity*) auf, der die Einhaltung der definierten Grenzen achtet (Rockström et al. 2009). Im Hinblick auf die Rolle der Ökonomie wurde diese Perspektive mit dem Modell der sogenannten Donut-Ökonomie weiterentwickelt (Raworth 2017), indem es um ein soziales Fundament ergänzt wurde, das zwölf Dimensionen sozialer Gerechtigkeit umfasst – beispielsweise den Zugang zu ausreichend Nahrung, zu sauberem Wasser und zu Energie sowie zu einem Gesundheitssystem, die Verfügbarkeit von angemessenem Wohnraum, ordentlicher Arbeit oder auch Gleichberechtigung und politische Teilhabe an der Gesellschaft. Demnach müssen sich menschliches Handeln und insbesondere ökonomische Aktivitäten nicht nur in einem sicheren, sondern auch in einem gerechten Handlungsspielraum bewegen („*safe* and *just* operating space“ (Raworth 2017)). Die ökonomischen Aktivitäten müssen dabei so ausgestaltet sein, dass sie gleichzeitig menschliches Wohlergehen sichern und planetare Grenzen respektieren – somit kommt der ökonomischen Dimension eine sogenannte dienende Funktion zu.

Die im Kontext der Rio+20-Konferenz entwickelten und 2015 auf höchster politischer Ebene beschlossenen SDGs der Vereinten Nationen greifen innerhalb von 17 Zielen mit 169 Indikatoren die drei Dimensionen der Nachhaltigkeit auf. Sie bieten prinzipiell einen Orientierungsrahmen für gesellschaftliche Entwicklungsprozesse und in diesem Zuge auch wirtschaftliche Aktivitäten. Die Problematik der Zielkonflikte zwischen Dimensionen und einzelnen SDGs bleibt auch hier größtenteils ungelöst (vgl. z. B. Kroll et al. 2019). Eine explizite Priorisierung der ökologischen und sozialen Dimensionen gegenüber der ökonomischen findet nicht statt. Vielmehr wird mit Blick auf das Ziel 9 „Menschenwürdige Arbeit und Wirtschaftswachstum“ davon ausgegangen, dass eine Entkopplung von Wirtschaftswachstum und Ressourcenverbrauch erreicht werden kann. Hiermit wird suggeriert,

dass in diesem Fall kein Zielkonflikt zwischen ökonomischer und ökologischer Dimension existiert.[8] Ein indikatorenbasierter Ansatz kann dieses Problem nur unzureichend adressieren. Bei unserer Analyse sind wir daher von den Auswirkungen der Technologie, die bereits diskutiert werden, ausgegangen, anstatt einzelne Anwendungen anhand der SDGs bewerten zu wollen. Denn wir verfolgen einen generischen Ansatz, der Aussagen darüber treffen soll, wie die Technologie grundsätzlich nachhaltiger gestaltet werden kann. Das bedeutet nicht, dass wir keinen Zusammenhang zwischen den SDGs und KI sehen. Die SDGs sind als übergreifender Orientierungsrahmen auf für wirtschaftliche Aktivitäten der KI-Entwicklung und deren Anwendung bedeutsam. Wir sind vielmehr der Meinung, dass man Nachhaltigkeit entlang des gesamten Lebenszyklus betrachten, muss um die übergeordneten Ziele einer nachhaltigen Entwicklung zu erreichen.

Nachhaltigkeit wird dabei als Prozess begriffen, der sich um die Frage nach der gerechten Verteilung zwischen den heute lebenden Menschen und zukünftigen Generationen dreht und um den gerechten Umgang der Menschen miteinander sowie mit der Natur. Es ist ein normatives Konzept, dessen Gegenstück als *Kollaps* bezeichnet werden kann. Damit ist Nachhaltigkeit zur Grundlage „für die Bestimmung von Fortschritt und Verantwortung, Freiheit und Kultur geworden“ (Bachmann 2010, S. 2). Unser Nachhaltigkeitsverständnis im Kontext von KI (siehe auch Abbildung 6) orientiert sich an den bereits vorgestellten drei zentralen Dimensionen: soziale, ökonomische und ökologische Nachhaltigkeit. Die Unterscheidung der Dimensionen dient im Folgenden der Abgrenzung und Strukturierung der Nachhaltigkeitskriterien. Uns ist klar, dass dies nur eine idealtypische Unterscheidung sein kann und in den realen Prozessen vielfältige Wechselwirkungen zwischen diesen Dimensionen zu beobachten sind. Gleichzeitig ist das Verständnis dieser Wechselwirkungen grundlegend, um die Frage nach der konkreten nachhaltigen Ausgestaltung von KI-Systemen, überhaupt zu beantworten.

Es wird deutlich, dass ökologische und soziale Nachhaltigkeit im Grunde zwei Seiten einer Medaille sind. Denn ohne gesellschaftlichen Zusammenhalt und menschenwürdige Lebensbedingungen kann auch der Schutz von Umwelt und Natur nicht gelingen und ohne eine intakte Umwelt können menschenwürdige Lebensbedingungen langfristig nicht aufrechterhalten werden. Deshalb wird heutzutage auch von sozial-ökologischer Transformation gesprochen (Brand 2014), um zu verdeutlichen, dass nur mit tiefgreifenden Veränderungsprozessen eine sozial und ökologisch gerechte Gesellschaft ermöglicht werden kann. Die Ökonomie, die wir ins Zentrum unserer Abbildung gesetzt haben, ist insbesondere im Hinblick auf KI *der* entscheidende Einflussfaktor. Er nimmt wesentlich auf die Erreichung dieser Zielsetzungen Einfluss. Denn es sind die wirtschaftlichen Aktivitäten, also der Produktions- und Konsumzyklus, der einen wesentlichen Anteil daran hat, ob planetare Grenzen und menschenwürdiges und selbstbestimmtes Leben in Einklang gebracht werden können. Die Verteilungs- und Gerechtigkeitsfragen, die sich im Zuge einer sozial-ökologischen Transformation stellen, werden häufig im ökonomischen Kontext wirksam. Deshalb lautet die Frage der kommenden Jahrzehnte vor allem, wie wir wirtschaftliche Aktivitäten so ausgestalten können, dass wir soziale Ziele erreichen und ökologische Grenzen nicht sprengen. Die politische Steuerung, die in den Diensten dieser normativen Zielsetzungen stehen muss, ist dabei von außerordentlicher Bedeutung.

[8] Diese Annahme ist umstritten. Sichtweisen, die dieser Annahme entgegenstehen und welche Schussfolgerungen gezogen werden können, wenn eine rechtzeitige Entkopplung von Wirtschaftswachstum und ressourcenverbrauch nicht möglich ist, werden in Petschow et al. 2018 ausführlich dargelegt.

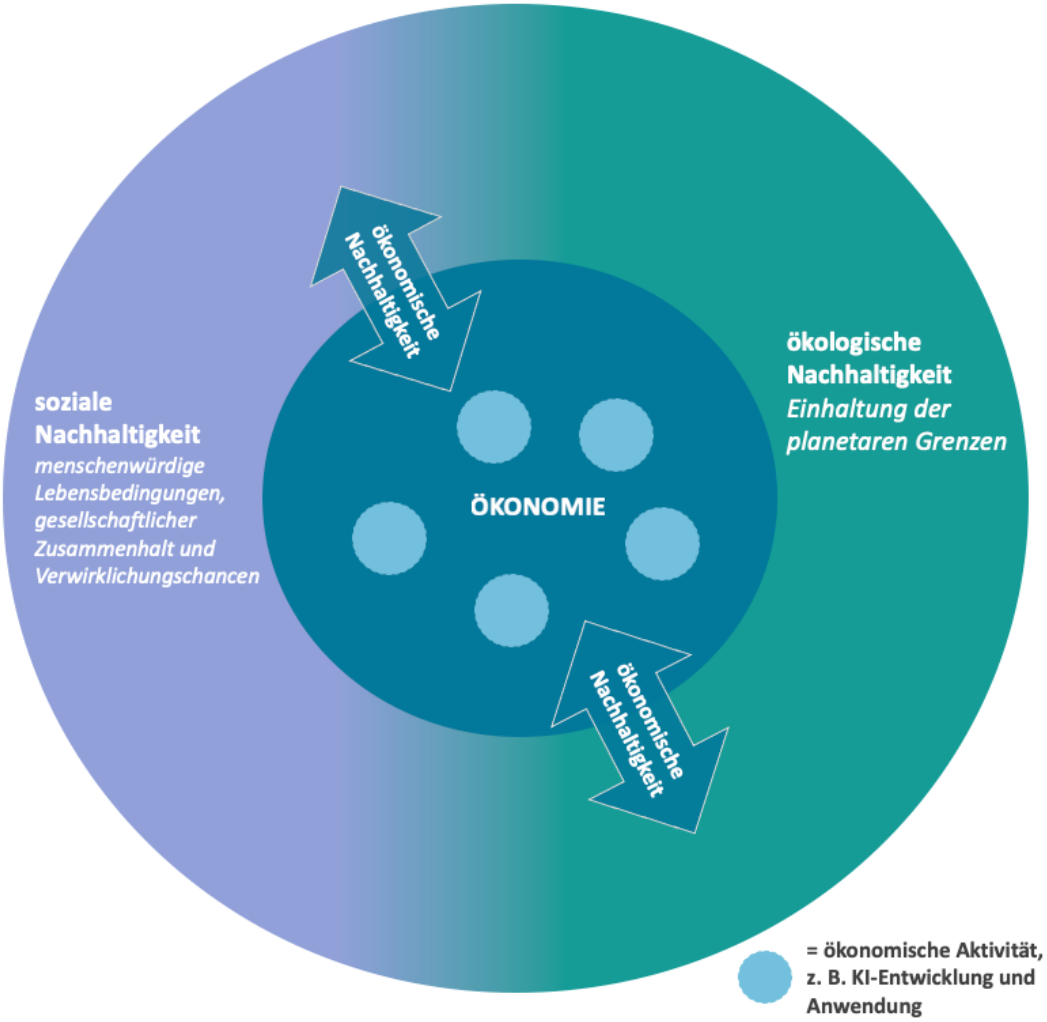


**Abbildung 6: Nachhaltigkeitsverständnis im Hinblick auf KI**

Unser Nachhaltigkeitsverständnis mit Blick auf KI baut auf diesen Konzeptionen von Nachhaltigkeit auf. Wir verstehen dabei die Entwicklung und Anwendung von KI als einen Teil der ökonomischen Aktivitäten, die sowohl zum Erhalt eines sozialen Fundaments beitragen als auch die Einhaltung planetarer Grenzen gewährleisten müssen. Dabei gilt es zu beachten, dass KI-Entwicklung und KI-Anwendung sowohl unmittelbare soziale und ökologische Auswirkungen haben als auch Dynamiken und Strukturen innerhalb der Ökonomie verändern. Somit wirken sie über diese Veränderungen auf die Bedingungen für soziale und ökologische Nachhaltigkeit. Gleichzeitig bestehen auch zwischen der sozialen und ökologischen Dimension Wechselwirkungen und mitunter Zielkonflikte, die im weiteren Verlauf der Debatte um nachhaltige KI angemessen berücksichtigt werden müssen (z. B. Datensparsamkeit für Datenschutz und Ressourceneffizienz vs. Diversität der Datenbasis). Um die Auswahl unserer Kriterien und deren Herleitung zu verdeutlichen, nehmen wir im Folgenden eine Definition der jeweiligen Nachhaltigkeitsdimensionen vor und beschreiben die Implikationen für die Betrachtung von KI-Systemen.

## 3.5 Soziale Nachhaltigkeit

Soziale Nachhaltigkeit zielt auf die Erfüllung der grundlegenden Bedürfnisse von Menschen (Nahrung, Wohnen, Versorgung), auf die Lebensbedingungen (Einkommen, Bildung) sowie auf den Zu-

gang zu sozialen Infrastrukturen (z. B. Littig and Grießler 2005) und auf die Sicherstellung der gesellschaftlichen Integrität und des gesellschaftlichen Zusammenhalts ab. Das bedeutet, dass besonders vulnerable Gruppen vor Benachteiligung und Exklusion geschützt werden müssen (Vavik & Keitsch 2010) und dass eine inter- und intragenerationale Gerechtigkeit hergestellt wird (Vallance et al. 2011). Darüber hinaus wird Diversität als ein fundamentaler Wert von sozioökonomischer und menschlicher Entwicklung gesehen (Matutinović 2001). In einem umfassenden Verständnis von sozialer Nachhaltigkeit wird auch der Befähigungsansatz (*Capability Approach*) (Sen 2000; Nussbaum 2006) als mögliche Grundlage verwendet. Er besagt, dass nicht nur die Erfüllung der grundlegenden menschlichen Bedürfnisse, sondern auch die Verwirklichungschancen der Menschen von enormer Bedeutung für eine nachhaltige gesellschaftliche Entwicklung sind. Menschen müssen demnach auf ein Fundament von Freiheiten und materiellen und kulturellen Handlungsressourcen zurückgreifen können, um ihre Rechte wahrzunehmen. Diese Perspektive zielt darauf ab, soziale Nachhaltigkeit positiv zu formulieren und mit gesellschaftlichen Verwirklichungschancen zu verknüpfen.

Im Hinblick auf die Entwicklung, Nutzung und den Einsatz von KI-Systemen bedeutet Nachhaltigkeit vor allem, dass die Würde des Menschen respektiert wird, keine Menschen ausgeschlossen, benachteiligt oder diskriminiert werden und die menschliche Autonomie und Handlungsfreiheit durch KI-Systeme nicht eingeschränkt werden darf. In einer erweiterten Perspektive auf Nachhaltigkeit bedeutet soziale Nachhaltigkeit auch, dass neben körperlicher Unversehrtheit und menschenwürdigen Lebensbedingungen, auch die Fähigkeit auf menschliche Art und Weise zu denken, zu argumentieren und zu handeln, nicht eingeschränkt werden sollte. Hier zeigt sich schon, dass ein umfassendes Verständnis von sozialer Nachhaltigkeit sehr weitreichende Konsequenzen für die Gestaltung von KI-Systemen hat. Dies wird zwar in Ansätzen von unseren Kriterien und Indikatoren abgebildet, kann jedoch nie vollumfänglich mit einem Indikatoren-Ansatz adressiert werden. Denn bei sozialen Nachhaltigkeitsaspekten handelt es sich um sehr komplexe soziale Zusammenhänge, die mit Indikatoren teils nur unzureichend erfasst werden können.

## 3.6 Ökologische Nachhaltigkeit

Ziel der ökologischen Nachhaltigkeit ist es, den sicheren Handlungsraum der Menschheit – der durch die planetaren Grenzen definiert wird – nicht zu verlassen. Durch die planetaren Belastbarkeitsgrenzen werden ökologische Grenzen verschiedener Kontrollvariablen festgelegt, deren Über- oder Unterschreitung irreversible Umweltveränderungen und -schäden zur Folge hätte. Durch menschliche Aktivitäten, auch im Rahmen der Digitalisierung und Nutzung von energieintensiven Rechenzentren, kann die Stabilität des Ökosystems und die Lebensgrundlage der Menschheit gefährdet werden. Rockström et al. (2009) diskutieren hierzu neun planetare Belastungsgrenzen: Klimawandel, neue Substanzen und modifizierte Lebensformen, Ozonverlust in der Stratosphäre, Aerosolgehalt der Atmosphäre, Versauerung der Meere, biogeochemische Flüsse, Süßwassernutzung, Landnutzungswandel, Intaktheit der Biosphäre (unterteilt in funktionale Vielfalt und genetische Vielfalt) (Rockström et al 2009; Steffen et al 2015). Auf viele dieser Belastungsgrenzen wirken die Entwicklung und Anwendung von KI-Systemen direkt (Energieverbrauch, Ressourcen für Hardware, Entsorgung der Hardware) oder indirekt (Energieeffizienz, nachhaltige Produktion und Agrarwirtschaft) ein. Darüber hinaus geht ökologische Nachhaltigkeit sowie Nachhaltigkeit als übergreifende Perspektive intergenerationaler Gerechtigkeit damit einher, dass ein System dann nachhaltig ist, wenn es selber überlebt und langfristig Bestand hat (Carnau 2011). Aus ökologischer Perspektive ist Nachhaltigkeit also immer damit verbunden, nicht mehr Ressourcen zu nutzen, als ein System in der Lage ist zu reproduzieren. Vor allem gilt es dabei sicherzustellen, dass künftige Generationen die gleichen Chancen auf ein gutes Leben haben wie heutige.

## 3.7 Ökonomische Nachhaltigkeit

Ökonomische Nachhaltigkeit richtet sich vor dem Hintergrund der *Grand Challenges* an der Einhaltung der planetaren Grenzen (vgl. Rockström et al. 2009; Steffen et al. 2015) aus. Sie setzt ökonomischen Aktivitäten das Ziel, die Befriedigung der Bedürfnisse aller heute und zukünftig lebenden Menschen zu ermöglichen. Damit wird die Ökonomie zwischen sozialen und ökologischen Leitplanken eingebettet. Daraus ergibt sich seitens der Wirtschaftsakteure die Notwendigkeit, beim Treffen ökonomischer Entscheidungen über Produktion, Verteilung und Verwendung von Gütern, Verantwortung für das Gemeinwohl und für den Erhalt sowie die Regeneration von Ökosystemen zu übernehmen. Vor diesem Hintergrund sind auch die Wirkungen, die die Entwicklung und Anwendung von KI-Systemen auf die ökonomischen Strukturen und Dynamiken haben, zu berücksichtigen. Entwicklung und Anwendung von KI-Systemen sind dabei als in der Ökonomie eingebettet zu verstehen.

## 3.8 Zusammenfassende Definition für nachhaltige KI

Unser Konzept für nachhaltige künstliche Intelligenz zielt darauf ab, die soziale, ökologische und ökonomische Dimension in eine Lebenszyklusbetrachtung zu integrieren und mittels konkreter Kriterien und Indikatoren Ansatzpunkte aufzuzeigen, wie die Entwicklung nachhaltiger KI – und zwar in allen Einsatzbereichen – gefördert werden kann. Mit dem Ansatz der nachhaltigen KI sollen also Mindestanforderungen formuliert werden, die bei der Entwicklung und Implementierung aller KI-Systeme vor dem Hintergrund gesellschaftlich gesteckter Nachhaltigkeitsziele zu beachten sind. Damit soll auch der Tatsache Rechnung getragen werden, dass KI-Systeme zwar für die Verfolgung von Nachhaltigkeitszielen eingesetzt werden können, solche Zielsetzungen aber bei weitem nicht bei der Entwicklung von allen KI-Systemen den Ausgangspunkt bilden. Die Auswirkungen von KI und deren Verknüpfung mit der digitalen Infrastruktur sollten daher in einer übergreifenden Nachhaltigkeitsperspektive entlang des KI-Lebenszyklus adressiert werden.

Eine nachhaltige KI ist aus unserer Perspektive vorhanden, wenn Entwicklung und Einsatz dieser Systeme die planetaren Grenzen respektiert, keine problematischen ökonomischen Dynamiken verstärkt und den gesellschaftlichen Zusammenhalt nicht gefährdet. Mit den Nachhaltigkeitskriterien für künstliche Intelligenz wollen wir erste Ansatzpunkte aufzeigen und eine Diskussion anstoßen, wie eine umfassende Perspektive auf die nachhaltige Gestaltung dieses sozio-technischen Systems aussehen kann.

# 4 Nachhaltigkeitskriterien und -indikatoren für künstliche Intelligenz

Die Erarbeitung der Nachhaltigkeitskriterien und -indikatoren basiert auf einer umfassenden Bestandsaufnahme und Literaturanalyse bestehender Konzepte, Bewertungsverfahren und Analysen im Bereich der gesellschaftlichen, ökologischen und ökonomischen Auswirkungen von KI-Systemen. Diese verschiedenen Diskurse wurden innerhalb der drei Nachhaltigkeitsdimensionen soziale, ökologische und ökonomische Nachhaltigkeit verortet und mit Querschnittsindikatoren zur organisationalen Einbettung ergänzt. Es ergeben sich dreizehn Nachhaltigkeitskriterien und fünf Querschnittsindikatoren (Abbildung 7). Die Herleitung dieser Kriterien und Indikatoren wird im Folgenden beschrieben und es wird erläutert, auf welchen Ansätzen und Konzepten diese Indikatoren basieren. Grundlage für die Bewertungssystematik bilden die drei Nachhaltigkeitsdimensionen, die eine Strukturierung des breiten Diskurses um mögliche Auswirkungen erlauben sollen und gleichzeitig die Zieldimensionen aufzeigen. Die Entwicklung, Überarbeitung und Clusterung der Kriterien und Indikatoren erfolgte in einem interdisziplinären Forschungsprozess in Zusammenarbeit mit Akteuren aus der Informatik, den Wirtschaftswissenschaften, der Medien- und Kommunikationswissenschaft, Soziologie sowie Nachhaltigkeitswissenschaften. Die von uns entwickelten Nachhaltigkeitsindikatoren sollen nicht direkt sichtbare gesellschaftliche, ökologische und ökonomische Auswirkungen von KI-Systemen erfassen und in vereinfachter Form darstellen. Die Operationalisierung der Kriterien in Indikatoren und Subindikatoren dient dazu, eine Orientierung zu geben, welche Ansatzpunkte, Kennzahlen oder Bewertungsmaßstäbe es für nachhaltige KI in den jeweiligen Bereichen gibt.

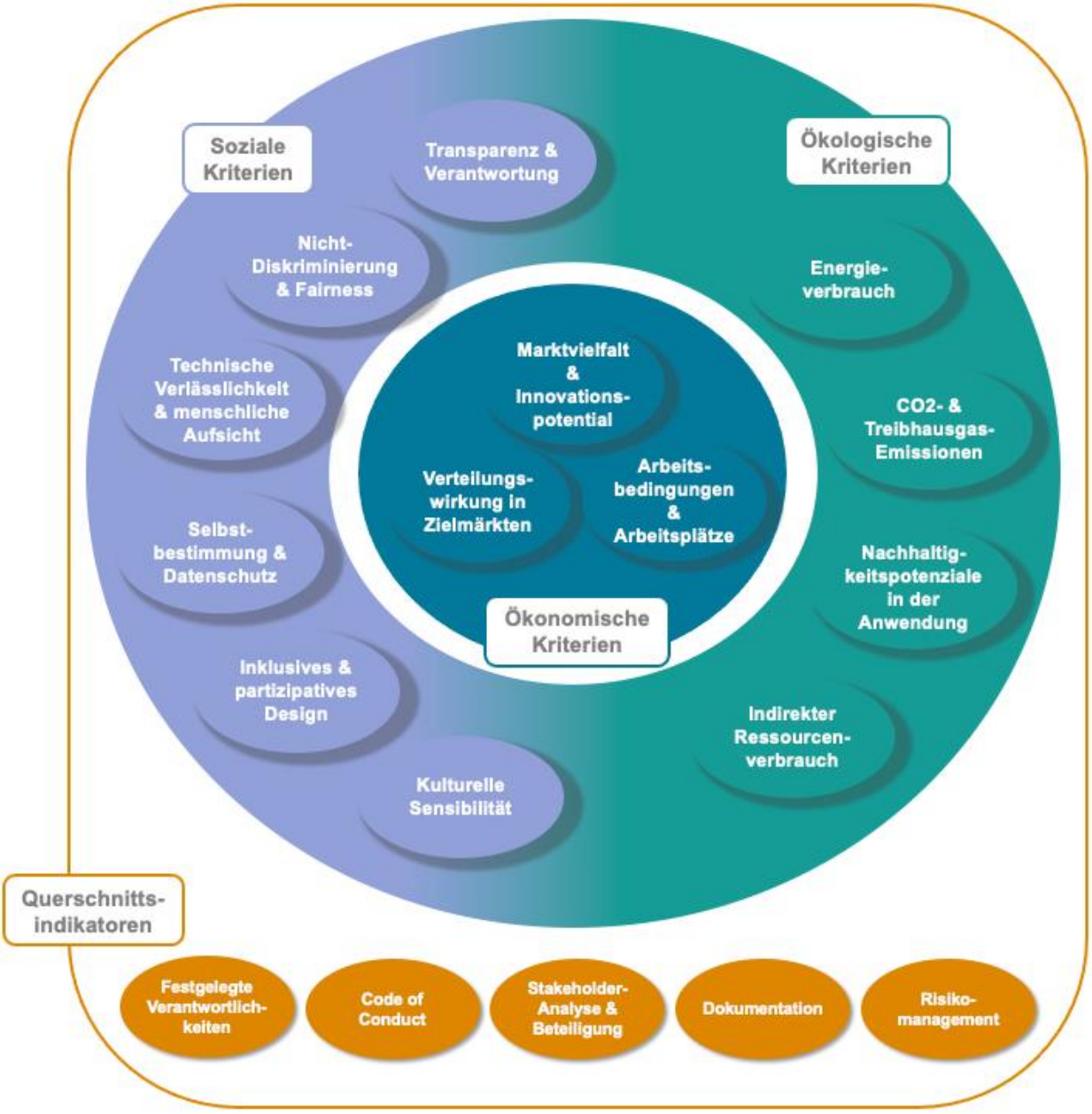


**Abbildung 7: Übersicht über die Nachhaltigkeitskriterien**
Quelle: Eigene Darstellung

## 4.1 Organisatorische Verankerung: Querschnittskriterien

Zwar bilden die drei Nachhaltigkeitsdimension eine gute Abgrenzung für die Systematisierung der Indikatoren. Jedoch ergab unsere Analyse, dass eine Vielzahl an Kriterien und Indikatoren nicht eindeutig einer Dimension zugeordnet werden kann. Somit ergibt sich eine Gruppe von organisatorischen Querschnittsindikatoren, die auf alle drei Bereiche der Nachhaltigkeit Einfluss nehmen.

### 4.1.1 Festgelegte Verantwortlichkeiten

Eine zentrale organisatorische Voraussetzung für die Sicherstellung von Verantwortung im Sinne der sozialen, ökologischen und ökonomischen Nachhaltigkeit ist das Festlegen von Verantwortlichkeiten. Im Sinne von Verantwortlichkeit sind Ansprechpersonen, Zuständigkeiten und Haftungsregelungen in Bezug auf die Entwicklung und die Anwendung von KI gemeint (vgl. AI Ethics Impact Group 2019).

**Indikatoren für Verantwortungsübernahme**

- **Ansprechpartner*innen für ethische Belange** setzen die Nachhaltigkeitsziele und konkret die Nachhaltigkeitskriterien, insbesondere in Bezug auf Transparenz, Fairness und Gerechtigkeit, im Unternehmen um. Die Ansprechpartner*innen werden als eine oder mehrere vermittelnde Instanzen installiert, um Konflikte im Rahmen von Transparenz, Fairness, Verantwortung, Datenschutz, Autonomie etc. zu regeln und zu regulieren (AI Ethics Impact Group 2019). Hierfür wird gegebenenfalls ein KI-Ethikausschuss oder Ähnliches gegründet. Von großer Bedeutung ist die Machtposition dieser Ansprechpartner*innen. Die Berechtigungen und Einflussmöglichkeiten dieser Funktionsrolle erlauben daher eine qualitative Bewertung hinsichtlich der Verantwortlichkeit (*Responsibility*).
- Die **Definition und Dokumentation von Zuständigkeiten** stellt sicher, dass Aufgaben, Rollen und Verantwortungsbereiche klar zueinander abgegrenzt werden. Die Zuteilung von Verantwortung zwischen einsetzender und entwickelnder Organisation für produzierte Ergebnisse ist klar und transparent geregelt, dokumentiert und kommuniziert (Felländer-Tsai 2020). Positiven Einfluss haben regelmäßige Updates der Dokumentation.
- Neben den Zuständigkeiten existieren auch **Regelungen zu Haftungsaspekten** im Falle von möglichen Schäden. Hierbei spielt die Absicherung durch finanzielle Rücklagen, Versicherungen und andere Kompensationsformen eine entscheidende Rolle.

### 4.1.2 Code of Conduct

Ein Code of Conduct soll die Werte und Normen für die Implementierung und Nutzung der KI-Systeme einer Organisation festlegen. Dieser enthält beispielsweise Werte, Prinzipien und bestimmte Standards, die von Entwickler*innen verinnerlicht werden sollten (Dignum 2018; AI Ethics Impact Group 2019). In dem Code of Conduct werden die grundlegenden Werte definiert, an denen sich die Organisation bei der Entwicklung und/oder Implementation von KI-Systemen orientiert. In der Praxis wird auch von Code of Ethics gesprochen. Im Sinne unseres mehrdimensionalen Nachhaltigkeitsansatzes sollte der Code of Conduct über ethische Prinzipien hinaus auch weitere Aspekte adressieren, wie Energie- und ressourcenschonende Entwicklung und die Berücksichtigung der sozialen Wirkungen von KI-Systemen.

### 4.1.3 Stakeholder-Analyse & -Beteiligung

Die Bedeutung von Multistakeholdergruppen, aus z. B. Zivilgesellschaft, Politik, dem privaten Sektor, der Wissenschaft und insbesondere Endnutzer*innen, nimmt mit der zunehmenden Vernetzung von Menschen und der Digitalisierung von Prozessen stetig zu. Aus diesem Grund müssen auch Betroffene in die Governance von KI-Systemen integriert werden (The Internet Society 2017). Das bedeutet, dass Stakeholder (je nach KI-System in unterschiedlichen Ausprägungen) und Zielgruppen (Endnutzer*innen und Betroffene) definiert, analysiert und in den Entwicklungs- und Testprozess zur Sicherstellung von Transparenz involviert werden sollten. Auch im Hinblick auf Erklärbarkeit, Autonomie und Fairness sollten Stakeholder (inkl. Zielgruppen) einbezogen werden, damit die Einbindung verschiedener Perspektiven sichergestellt werden kann (AI Ethics Impact Group 2019).

#### Indikatoren für Stakeholder-Analyse & -Beteiligung

- Stakeholder werden für jedes KI-System dezidiert **identifiziert** und sinnvollerweise **klassifiziert**. Eine Systematisierung und Analyse von Stakeholdern erleichtert ihre Einbindung in die Entwicklung des KI-Systems und stellt die Berücksichtigung von Stakeholdernarrativen sicher.
- Die Betroffenen werden nicht nur in den vorgelagerten Designprozessen integriert, sondern während des gesamten KI-Lebenszyklus. Das beinhaltet vor allem die Testphase und spätere Releases.
- Als weiterer Indikator lässt sich auch **der Umfang der Einbindung** operationalisieren. Hierzu können möglichst viele verschiedene Nachhaltigkeitskriterien als Input für die Stakeholderbeteiligung dienen.

### 4.1.4 Dokumentation der KI-Systeme

Die Dokumentation des KI-Systems bildet eine der Grundvoraussetzungen auf deren Basis die Berichterstattung über weitere Indikatoren überhaupt möglich ist. Eine vollständige und detaillierte Dokumentation ermöglicht Transparenz über die Funktionsweise des Systems, die Prüfung von Fairness oder auch die Sicherstellung der technischen Verlässlichkeit (Felzmann et al. 2019; AI Ethics Impact Group 2019). Speziell der Bereich der offenen KI-Entwicklung (Open Source, offene Standards, offene KI-Ökosysteme) erfordert eine transparente Aufzeichnung der technischen Eigenschaften.

#### Indikatoren für die Dokumentation des KI-Systems

- Um die Maßnahmen zur Dokumentation der KI-entwickelnden Organisation zu bewerten, sind **Vollständigkeit und Umfang** der Dokumentation ausschlaggebend. Mindestanforderungen sind die Aufzeichnung von Informationen hinsichtlich der Zielsetzung, Domain, Nutzer*innen, Daten, Model, Feature-Selektion-Prozesse, Eingaben, Tests, Metriken etc. Eine Forschungsgruppe von Google hat hierfür eine übersichtliche Model Card entwickelt, die als maximale Anforderung an die Dokumentation dienen kann. Entsprechend der Model Card sollte eine KI-Dokumentation folgende Inhalte beinhalten (Mitchell et al. 2019):
    - Modell-Details (Typ, Datum, Version, Lizenz, Training etc.)
    - Anwendungsfall (Nutzer*innen, Out-of-Scope)
    - Faktoren (Soziales, Umwelt)
    - Metriken

- Testdaten
- Trainingsdaten
- Ethische Aspekte
- Bedenken und Empfehlungen

- Neben der Prüfung anhand der Model Cards kann eine Prüfung der Datendokumentation (z. B. mit Datasheets) Aufschluss über die Herkunft, Aktualität, Repräsentativität und Vollständigkeit der Daten geben (Gebru et al. 2018).

### 4.1.5 Risikomanagement

Der Einsatz von KI ist in verschiedenen Anwendungsbereichen und im Vergleich zu konventionellen Software-Produkten mit einem erhöhten Risikopotenzial verbunden. Die Risiken können je nach Kontext sehr unterschiedlich sein: sowohl in der Art der Risiken (sozial, ethisch, ökologisch, ökonomisch etc.) als auch in deren Ausprägung (gering, mittel, hoch etc.) (Cheatham et al. 2019). Ziel eines Risikomanagements ist es, die potenziellen Risiken zu identifizieren und angemessene Maßnahmen zu deren Minimierung oder Vermeidung zu ergreifen. Ein Risikomanagementsystem, dass eine umfassende Risikoanalyse und -bewertung vornimmt, kann die Vorteile von KI-Systemen maximieren und gleichzeitig die Risiken minimieren (Europäische Kommission 2019). Darüber hinaus kann ein solches System die Einhaltung rechtlicher Vorgaben verbessern und dabei helfen, Schwachstellen zu erkennen. Viele Unternehmen werden ab 2023 im Zuge der Überarbeitung der Non-Financial Reporting Directive (NFRD) hin zur Corporate Sustainability Reporting Directive (CSRD) verpflichtet sein, Nachhaltigkeitsinformationen (sogenannte ESG-Indikatoren) zu veröffentlichen. Unternehmen, die in großem Maße KI-Systeme nutzen, könnten diese Aspekte beispielsweise in das Risikomanagementsystem und die Berichterstattung integrieren. Mit der Implementierung eines Risikomanagementsystems können Risiken (Diskriminierung, Privatsphäre etc.), Schwachstellen und Angriffsmöglichkeiten auf das System je nach Anwendungskontext identifiziert, kategorisiert und vorab bewertet werden.

#### Ausgestaltung des Risikomanagements

Die im *Risk Assessment* identifizierten Risiken und die Genauigkeit sollten regelmäßig im System überprüft werden. Nach Identifizierung und Klassifizierung der Risiken werden Maßnahmen entwickelt, um die Risiken zu minimieren. Die Maßnahmen werden anschließend erfolgreich umgesetzt. Für den Fall des Risikoeintritts wurden ein oder mehrere Fallback-/Notfallpläne basierend auf dem Risk Assessment entwickelt. Im Fall von Problemen oder Angriffen wird dieser Fallback-Plan ausgelöst. Beispiele für Fallback-Optionen sind Abbruch, Regel-basierte Vorgänge oder auch menschliche Kontrolle (Jain et al. 2020). Ein optimales Risikomanagement findet in regelmäßigen Zyklen statt und basiert auf einem definierten und implementierten Prozess sowie auf spezifischen Trainings und Schulungen.

#### Indikatoren für das Risikomanagement

- Die Organisation hat ein Risikomanagementsystem für die Entwicklung und Nutzung von KI-Systemen implementiert bzw. KI-bezogene Risiken werden über das unternehmensweite Risikomanagementsystem abgedeckt

## 4.2 Soziale Kriterien und Indikatoren

Im Bereich der ethisch-sozialen Auswirkungen von KI-Systemen existieren zahlreiche Diskussionen und eine sehr große, lebhafte Community. Ethikforscher*innen, Unternehmer*innen und Entwickler*innen-Communities sind an der Erarbeitung von Frameworks für eine ethische und verantwortungsvolle Programmierung und Anwendung von KI-Systemen beteiligt. Gleichzeitig sehen immer mehr Instanzen – allen voran die Europäische Union – die immensen Einflüsse von KI auf das Leben und die Würde der Menschen. Dieses zunehmende Bewusstsein resultiert in einer Vielfalt an Leitfäden, Frameworks, Guidelines und Tools im Bereich „AI Ethics“. Im „AI Ethics Guidelines Global Inventory“[9] von AlgorithmWatch sind bspw. bereits 171 Richtlinien für ethische künstliche Intelligenz enthalten. Unser Kriterien- und Indikatorenset integriert die diversen Diskussionen und zielt darauf ab, die wichtigsten Aspekte zu adressieren und gleichzeitig praktikable Indikatoren daraus abzuleiten. Darüber hinausgehend haben wir Indikatoren wie inklusives und partizipatives Design sowie kulturelle Sensibilität als wichtige Indikatoren für die soziale Nachhaltigkeit von KI-Systemen definiert. Unser Kriterienset stellt einen ersten Vorschlag dar und soll als Diskussionsgrundlage dienen.

### 4.2.1 Transparenz und Verantwortungsübernahme

Systeme, die auf Algorithmen basieren, können durch Undurchsichtigkeit und mangelnde Nachvollziehbarkeit gekennzeichnet sein (Europäische Kommission 2019). Der Anspruch, Transparenz herzustellen, zielt darauf ab, die Erklärbarkeit, Interpretierbarkeit oder andere Formen der Kommunikation und Offenlegung zu erhöhen (Jobin et al. 2019). Transparenz bezieht sich auf die Notwendigkeit, die KI-Algorithmen und -Ergebnisse beschreiben, überprüfen und reproduzieren zu können, sowie verwendete Daten fair zu verwalten (Schneider & Ziyal 2019). Transparenz bedeutet außerdem, dass Akteure, die KI nutzen bzw. mit ihr interagieren, Kenntnis davon haben, dass KI eingesetzt wird, wie die Entscheidungsfindung erfolgt und welche Konsequenzen dies für sie haben kann. Die „Association for Computing Machinery“, eine wissenschaftliche Gesellschaft für Informatik, führt in ihren Grundsätzen für algorithmische Transparenz und Rechenschaftspflicht an, dass es notwendig sei, Bewusstsein für den Umgang mit analytischen Systemen zu schaffen im Hinblick auf mögliche Voreingenommenheit, die mit dem Design, der Implementierung und der Verwendung einhergeht (Association for Computing Machinery US Public Policy Coucil 2017).

#### Indikatoren für Transparenz und Verantwortungsübernahme

- Viele Richtlinien schlagen eine verstärkte Offenlegung von Informationen durch diejenigen vor, die KI-Systeme entwickeln oder einsetzen. Allerdings variieren die Spezifikationen darüber, was offengelegt werden sollte stark. Dies kann beispielsweise den Einsatz von KI, den Quellcode, die Datennutzung, die Beweisgrundlage für den KI-Einsatz, Einschränkungen und Gesetze, Investitionen in KI und deren Auswirkungen umfassen. Audits und von Menschen durchgeführte Überprüfungen werden vor allem von Datenschutzbehörden und Non-Profit-Organisationen vorgeschlagen, während die Privatwirtschaft technische Lösungen präferiert (Jobin et al. 2019). Darüber hinaus geben die Modell-inhärenten Eigenschaften, wie die Art des Modells, der Reifegrad des Modells, die Anzahl der Parameter und so weiter, Aufschluss über die Transparenz des Systems (Deloitte 2019; AI Ethics Impact Group 2019; Felzmann et al. 2019). Hier sollten entwickelnde und einsetzende Organisationen ansetzen.

[9] https://inventory.algorithmwatch.org/

- In den Algo.Rules wird angeführt, dass eine Kennzeichnung eingeführt werden sollte, wenn Menschen mit algorithmischen Systemen interagieren, um dies für sie erkennbar zu machen. Außerdem sollte ein algorithmisches System „mit seinen direkten oder mittelbaren Wirkungen und seiner Funktionsweise für Menschen leicht verständlich gemacht werden, damit diese es hinterfragen und überprüfen können“ (iRights.Lab & Bertelsmann Stiftung 2019). Dies beinhaltet „Informationen über die dem System zugrundeliegenden Daten und Modelle, seine Architektur sowie die möglichen Auswirkungen“ (iRights.Lab & Bertelsmann Stiftung 2019). Die kanadische Regierung empfiehlt ebenfalls, dass Hinweise in einfacher Sprache darüber bereitgestellt werden sollten, wenn eine Entscheidung von einem automatisierten System getroffen wird und wie sie zustande kommt (Government of Canada 2019).
- In den Algo.Rules wird zudem empfohlen, dass „externe Prüfstellen […] unter Wahrung legitimer Geschäftsgeheimnisse durch entsprechende technische Vorkehrungen in die Lage versetzt werden [sollten], ein algorithmisches System tatsächlich und umfassend unabhängig prüfen zu können“, um die beabsichtigte Wirkung zu gewährleisten (iRights.Lab & Bertelsmann Stiftung 2019).

### 4.2.2 Nicht-Diskriminierung und Fairness

KI-Systeme können Menschen aufgrund von Alter, Geschlecht oder Hautfarbe diskriminieren, z. B. weil die Daten, mit denen die Modelle trainiert wurden, einen *Bias* (also eine Verzerrung) enthalten und somit gesellschaftliche Vorurteile reproduzieren. Oder die Daten bilden nicht das ab, was sie abbilden sollen (z. B. die gesellschaftliche Vielfalt). Dies kann passieren, wenn sie beispielsweise nur in einem spezifischen Kulturraum erhoben wurden. Die Gefahr besteht, dass der gesellschaftliche Zusammenhalt und die Integrität dadurch gefährdet werden. KI-spezifisch ist die Art und Weise wie die Daten für die Entscheidungsfindung genutzt werden. Denn bei Algorithmen des maschinellen Lernens werden (im Vergleich zu einem „einfachen“ Algorithmus), Entscheidungen auf Basis vorhergehender Entscheidungen getroffen, ohne dass diese Entscheidungsfindung vom Entwickler*innenteam direkt beeinflusst werden kann. Das System optimiert also selbständig seine Entscheidungsfindung. Dies macht es schwieriger, Ursachen für diskriminierende Entscheidungen zu ermitteln. Gerechtigkeit wird dabei hauptsächlich über Fairness definiert und zielt auf die Verhinderung, Überwachung oder Abschwächung unerwünschter Voreingenommenheit und Diskriminierung ab. In der Debatte wird auch der faire Zugang zu KI, Daten und den Vorteilen von KI betrachtet. Beauftragte aus dem öffentlichen Sektor betonen vor allem die Auswirkungen von KI auf den Arbeitsmarkt und die Notwendigkeit in diesem Zuge demokratische und gesellschaftliche Themen anzusprechen. Eine hervorgehobene Bedeutung hat dabei die Beschaffung und Verarbeitung von genauen, vollständigen und vielfältigen/diversen Daten, insbesondere Trainingsdaten (Jobin et al. 2019). Denn auch Daten stellen im Grunde nur eine Interpretation der Realität dar bzw. den Versuch, reale Zusammenhänge auf bestimmte Art und Weise zu strukturieren, was durch kulturspezifische Denkmuster stark beeinflusst sein kann.

Die Grundlage zur Umsetzung von Nicht-Diskriminierung und Fairness könnte durch technische Lösungen wie Standards oder explizite normative Kodierung, Transparenz – insbesondere durch die Bereitstellung von Informationen –, der Sensibilisierung der Öffentlichkeit für bestehende Rechte und Regelungen, Prüfverfahren sowie die Entwicklung oder Stärkung der Rechtsstaatlichkeit gewährleistet werden. Auch das Recht auf Einspruch und Wiedergutmachung, eine interdisziplinäre oder anderweitig vielfältigere Belegschaft sowie eine bessere Einbindung der Zivilgesellschaft oder anderer relevanter Stakeholder spielen eine entscheidende Rolle (Jobin et al. 2019). Campolo et al. (2017) empfehlen, dass vor der Freigabe eines KI-Systems Organisationen strenge Vorabtests durchführen sollten, um sicherzustellen, dass die Vorurteile und Fehler aufgrund von

Problemen mit den Trainingsdaten, Algorithmen oder anderen Elementen des Systemdesigns nicht verstärkt werden. Die zuvor skizzierten Querschnittsindikatoren haben somit eine wichtige Funktion für die Erreichung von Nicht-Diskriminierung und Fairness.

#### Indikatoren für Nicht-Diskriminierung und Fairness

Die Europäische Kommission (2018) empfiehlt auf institutioneller Ebene, dass für eine angemessene Definition von „Fairness“ gesorgt werden solle, die bei der Gestaltung des KI-Systems angewendet werden soll. Dazu soll betrachtet werden, ob die gewählte Definition allgemein gebräuchlich ist und ob zuvor andere Definitionen in Betracht gezogen wurden. Aufgabe eines Entwicklungsteams und des entsprechenden Managements ist es, ein Bewusstsein für unfaire Entscheidungen und Bias im Allgemeinen zu schaffen.

- Hierzu zählt erstens eine adäquate, meist auf den Anwendungsfall angepasste und kommunizierte **Definition von Fairness** (Robert et al. 2020).
- Zweitens sollte das Team vulnerable und möglicherweise **marginalisierte Gruppen** anhand von geschützten Attributen (*protected attributes*) wie Ethnizität, Hautfarbe, Herkunft, Religion, Geschlecht etc. definieren und in Test- und Evaluationsprozessen berücksichtigen (Mehrabi et al. 2021).
- In einem dritten Schritt müssen **Hilfsmittel und Tools** (Fairlearn, AI 360 usw.) sowie **Evaluationskriterien** (*Equalized Odds*, *Equal Opportunities* usw.) festgelegt und angewandt werden, um Bias in Trainingsdaten, Inputdaten, Modellen, Methoden und Design zu identifizieren und Fairness zu messen (vgl. Mehrabi et al. 2021; Robert et al. 2020; Ferrer et al. 2021).
- Abschließend müssen **Maßnahmen zur Beseitigung von Unfairness** umgesetzt werden. Dazu zählen beispielsweise das Entfernen von geschützten Attributen, das Hinzufügen von wichtigen, fehlenden Faktoren sowie das Entfernen von sogenannten Proxy-Attributen, die mit den geschützten Attributen korrelieren. Diese vier grundsätzlichen Indikatoren lassen sich nur schlecht quantifizieren. Umso wichtiger ist es, dass Unternehmen Einblick in ihre Fairness-Etikette geben und diese reflektieren.

### 4.2.3 Technische Verlässlichkeit und menschliche Aufsicht

Um vertrauenswürdig zu sein, müssen KI-Systeme und -Anwendungen, insbesondere jene mit hohem Risiko, technisch robust und genau sein. Das bedeutet, dass solche Systeme verantwortungsvoll und mit einer im Voraus angemessenen Berücksichtigung der Risiken, die sie erzeugen können, entwickelt werden. Ihre Entwicklung und Funktionsweisen müssen so beschaffen sein, dass KI-Systeme sich zuverlässig wie beabsichtigt verhalten. So soll das Risiko von Schäden minimiert werden (Europäische Kommission 2019). Technische Verlässlichkeit wird auch maßgeblich durch menschliche Aufsicht (*Human Oversight*) umgesetzt. Kontrollmechanismen und menschliche Aufsicht sollen helfen sicherzustellen, dass ein KI-System nicht die menschliche Autonomie untergräbt oder andere negative Auswirkungen verursacht. Das Ziel einer vertrauenswürdigen, ethischen und menschenzentrierten KI kann nur erreicht werden, wenn eine angemessene Beteiligung von Menschen in Bezug auf risikoreiche KI-Anwendungen sichergestellt werden (Europäische Kommission 2019).

#### Indikatoren für technische Verlässlichkeit und menschliche Aufsicht

- Die Europäische Kommission (2018) schlägt vor, dass Risiken und Angriffsmöglichkeiten auf das System vorab bewertet werden sollten, für die das System anfällig sein könnte. Es sollten außerdem mögliche Schwachstellen identifiziert werden und Maßnahmen und Systeme

eingerichtet werden, um die Integrität und Widerstandsfähigkeit des KI-Systems gegen potenzielle Angriffe sicherzustellen. Es wird außerdem empfohlen, eine Einschätzung darüber abzugeben, wie sich das System in unerwarteten Situationen und Umgebungen verhält. Zuletzt sollte bedacht werden, ob und in welchem Ausmaß das KI-System doppelverwendungsfähig (Stichwort *Dual Use*) sein könnte, um vorsorglich Präventivmaßnahmen zu treffen (Europäische Kommission 2018).

- Die Grundlage der technischen Verlässlichkeit stellt die Sicherstellung einer hohen **Datenqualität** dar. Entwickelnde Organisationen sollten Maßnahmen ergreifen, um aktuelle, vollständige, repräsentative und verlässliche Daten zu verwenden und zur Verfügung zu stellen (Wang & Strong 1996). Die Verlässlichkeit kann bspw. durch die Überprüfung der Herkunft und Supply Chain der Daten gewährleistet werden. Auch verschiedene Industriestandards wie ISO 25012 oder ISO 8000 geben Aufschluss über die Datenqualität und die intrinsische Verlässlichkeit des KI-Systems.
- The Internet Society (2017) führt an, dass zur verantwortlichen Umsetzung von KI signifikante Sicherheitschecks vor der Einführung durchgeführt sowie eine weitergehende Überwachung der Anwendungen durch Menschen während der Nutzung sichergestellt werden müssten. Außerdem wird empfohlen, dass Menschen zu jedem Zeitpunkt eingreifen können und die Kontrolle über die Anwendung haben sollten (iRights.Lab& Bertelsmann Stiftung 2019; The Internet Society 2017; Government of Canada 2019). Zu diesen **Eingriffsmöglichkeiten** zählen das Abbrechen, Pausieren aber auch das Überschreiben von Ergebnissen.

### 4.2.4 Selbstbestimmung und Datenschutz

KI-Systeme können eingesetzt werden, um personalisierte und qualitativ hochwertigere Dienste anzubieten. Doch sie können auch stark die menschliche Entscheidungsautonomie beeinträchtigen (Smuha 2019). Durch die zunehmende Digitalisierung großer Teile der Wirtschaft und der damit einhergehenden Erhebung, Analyse und Zusammenführung von Daten, besteht die Gefahr, Grundrechte wie das Recht auf informationelle Selbstbestimmung weiter auszuhöhlen (Initiative „Konzernmacht beschränken“ 2018). Zum Beispiel hat der Fall von Cambridge Analytica aufgezeigt, wie das systematische Erwerben und Auswerten von Daten genutzt werden kann, um Entscheidungsprozesse bei Individuen zu beeinflussen (Siggiqui, 2018).

Der Aspekt des Datenschutzes bezieht sich vor allem auf deie Wahrung der Privatsphäre. Sie wird im Rahmen von KI sowohl als Wert, den es zu bewahren gilt, als auch als ein Recht, das es zu schützen gilt, gesehen (Jobin et. al 2019). Durch den Einsatz von KI werden die Möglichkeiten, die Gewohnheiten von Menschen nachzuverfolgen und zu analysieren, erhöht. Ein potenzielles Risiko durch den Einsatz von KI besteht durch staatliche Behörden oder anderen Einrichtungen in Form von Massenüberwachung oder Beobachtung von Mitarbeitenden durch ihre Arbeitgeber*innen, auch wenn dies gegen den EU-Datenschutz und andere Vorschriften verstößt. Durch die Analyse großer Datenmengen und die Identifizierung von Verbindungen zwischen ihnen, kann KI auch dazu verwendet werden, Daten über Personen zurückzuverfolgen und zu de-anonymisieren. Hierdurch können neue Risiken für den Schutz personenbezogener Daten entstehen, sogar, wenn Datensätze keine personenbezogenen Daten enthalten (Europäische Kommission 2019). Die Aspekte Selbstbestimmung und Datenschutz stehen somit in einem sehr engen Zusammenhang. Denn durch die Verknüpfung von Datensätzen können neue Risiken im Hinblick auf die Verletzung von Persönlichkeitsrechten oder die informationelle Selbstbestimmung ausgehen, deren Umfang und Konsequenzen sich derzeit nur in Ansätzen erahnen lassen.

#### Indikatoren für Selbstbestimmung und Privatsphäre

- Zum Schutz der Privatsphäre und vor Eingriffen in die menschliche Autonomie durch Dritte müssen KI-entwickelnde und KI-einsetzende Organisationen **Privacy und Cyber Security Maßnahmen** vorbereiten und umsetzen. Ziel ist die Minimierung von Privatsphäre- und Internetsicherheits-Risiken (Europäische Kommission 2019). Zu den möglichen Maßnahmen zählen unter anderem Verwendung von *Privacy-by-Design*-Prinzipien (Verschlüsselung, Aggregation und Anonymisierung), die Umsetzung von Datenminimierung und die Positionierung eines *Data Protection Officers*. Des Weiteren kann die Ausdrucksstärke der Maßnahmen als Indikator für die Sicherheit des Systems verwendet werden. Hier lässt sich überprüfen, ob rechtliche Regeln eingehalten werden, ob ein Umgang mit Ungewissheit stattfindet und Konfliktlösungen berücksichtigt werden (Jafari et al. 2011).
- Ein professioneller Umgang mit persönlichen Daten zeigt sich zudem an den Benachrichtigungs- und Zustimmungs- bzw. Widerrufmechanismen. Nutzer*innen sollten informiert werden, sobald die KI-Systeme persönliche Daten verwenden oder durch Aufzeichnung von Nutzerverhalten erstellen. Darüber hinaus sollten Nutzer*innen die Möglichkeit haben, der Datennutzung zuzustimmen oder diese zu widerrufen. Zur Quantifizierung des Indikators können die sogenannten *Alert Ratio*, *Choice Ratio* und *Consent Ratio*, also die prozentualen Anteile der Operationen, die mit Benachrichtigung, Auswahl bzw. Zustimmung implementiert werden, herangezogen werden (Jafari et al. 2011).
- Die Nutzungsautonomie von KI-Systemen hängt maßgeblich von konzeptionellen Aspekten und der organisationalen Einbettung ab. Die Akzeptanz und die Adaption neuer KI-basierter Technologie sollte selbstbestimmt sein. Nutzer*innen dürfen nicht unter Druck gesetzt werden. Gleichzeitig müssen ihnen verfügbare Informationen zu Vorteilen und Risiken offengelegt werden, damit sie eine selbstbestimmte Entscheidung treffen. Während der Nutzung sollte das Verhalten der Nutzer*innen nicht durch Mechanismen wie *Nudging* oder *Dark Patterns* beeinflusst werden. Letztlich erzwingt nachhaltige KI keine Abhängigkeiten und ungesunde Nutzungsmuster (Rafael et al. 2020).

### 4.2.5 Inklusives und partizipatives Design

Vor dem Hintergrund, dass KI-Systeme zunehmend für die Bereitstellung wichtiger gesellschaftlicher Infrastrukturen (Sozialleistungen, Gesundheitsleistungen, Rechts- und Finanzdienstleistungen usw.) genutzt werden, ist die partizipative Gestaltung dieser Systeme von wachsender Bedeutung. Die Diversität der Gesellschaft und insbesondere die Perspektive vulnerabler Gruppen muss bereits bei der Gestaltung dieser Infrastrukturen berücksichtigt werden. Inklusive Teilhabe an der Gestaltung wichtiger Infrastrukturen kann als direkte Implikation intergenerationeller Gerechtigkeit (vgl. Vallance et al. 2011) gesehen werden. Gleichzeitig schützt partizipatives Design vor Benachteiligung und Exklusion, stärkt womöglich den gesellschaftlichen Zusammenhalt und die Diversität – und steht somit in einer instrumentellen Verbindung zu wichtigen Aspekten sozialer Nachhaltigkeit. Teil dieser inklusiven Teilhabe ist auch ein Beitrag zur Demystifizierung von KI inklusive der Reduktion von Informationsasymmetrien sowie einem Aufbau von gesellschaftlicher Evaluationskompetenz und Feedbackmechanismen.

KI basiert auf der probabilistischen Analyse von Daten und der Identifikation von Regelmäßigkeiten. Jeder Bias in der Auswahl der Daten, im Design der Analysemethode, aber auch in den sozial konstruierten und in den Daten abgebildeten Regelmäßigkeiten wird sich auch in den Ergebnissen der Analyse und den folgenden Entscheidungen widerspiegeln. Aus Perspektive der sozialen Nachhaltigkeit sollten solche Verzerrungen in der Auswahl der Daten, aber auch in der Gestaltung

der Analyse, die immer Ausdruck von normativen Vorannahmen und Werten ist, reduziert werden beziehungsweise auf gerechte Weise die Werte der Gesellschaft widerspiegeln (Ogolla & Gupta 2018). In diesem Zusammenhang spielen die Zusammensetzung des Entwicklerteams und die Integration von Stakeholdern, Nutzer*innen und Betroffenen eine zentrale Rolle. Eine diverse Zusammensetzung von Entwicklungsteams sowie Wissenstransfer zwischen Disziplinen, insbesondere Engineering und Sozialwissenschaften/Ethik, sowie die Berücksichtigung der Bedürfnisse sozial benachteiligter Gruppen sollten deshalb in den Designprozess einfließen.

**Indikatoren für inklusives und partizipatives Design**

- Entsprechen der „Assessment List for Trustworthy Artificial Intelligence“ werden bei der Gestaltung des *User Interfaces* verschiedene Präferenzen, Fähigkeiten und Bedürfnisse in der Gesellschaft und deren Diversität berücksichtigt. Auch von Ausgrenzung bedrohte Menschen können das KI-System problemlos und ohne technische Hürden nutzen (Europäische Kommission 2020). Dabei werden inklusive Designprinzipien, wie beispielsweise das *Universal Design*, angewandt. Zentrale Bestandteile der Barrierefreiheit und Zugänglichkeit sind der Zugang an sich (Hardware, Internet), die Verständlichkeit und Erklärung, der technische Support und Nutzungsanreize für benachteiligte Gruppen.
- Zur Umsetzung eines partizipativen Designs müssen Endnutzer*innen, Betroffene und weitere Stakeholder am Designprozess der KI-Systeme teilhaben. Dabei wird die Ausgestaltung des KI-Systems mit Anwender*innen und Betroffenen besprochen und reflektiert. In diesem Rahmen werden zudem Vorurteile gegenüber Gruppen und marginalisierten Kulturen diskutiert und berücksichtigt (Birhane 2021). Dieses Prinzip wird im Allgemeinen als Co-Design bezeichnet.

### 4.2.6 Kulturelle Sensibilität

KI-Systeme finden zunehmend weltweit Anwendung. Insbesondere entscheidungsunterstützende Prognosemodelle, die in der sozialen Sphäre zum Einsatz kommen, prägen mit den ihnen eingeschriebenen Werten und Normen gesellschaftliche Strukturen für eine Vielzahl von Menschen aus unterschiedlichsten kulturellen sowie sozioökonomischen Kontexten (Perdomo et al. 2020; Birhane et al. 2021). Nur wenn die in den KI-Systemen eingeschriebenen Werte und Normen sowie Interessen im Einklang sind mit den im Anwendungskontext gelebten Werten, Normen und Bedürfnissen, ist ein Beitrag zum gesellschaftlichen Zusammenhalt möglich. Zu berücksichtigen ist dabei, dass die zu lösenden Probleme historisch und kulturell eingebettet sind. Daher sollte bereits der Problemorientierung lokales Wissen als Ausgangspunkt für die ML-Modellentwicklung dienen (Birhane 2020). Die Deutungshoheit über soziale Fragen und Problemlösungsansätze verbleibt damit bei den Gemeinschaften von Individuen, die KI-Systeme anwenden. Aufgrund der derzeit vorherrschenden Produktionsweise von KI-Systemen, bei der die Technologien US-amerikanischer und chinesischer Technologieunternehmen den globalen Markt beherrschen, verbreiten sich entsprechende Werte, Normen und dominante Weltauffassungen, die der Diversität sozialer und kultureller Realitäten nicht gerecht werden können (Birhane 2020; Kwet 2019). Es besteht die Gefahr, dass bestehende Ungleichheiten, Abhängigkeitsbeziehungen und Dominanzstrukturen zwischen dem globalen Norden und dem globalen Süden reproduziert werden. Allen Gemeinschaften von Individuen sollte im Sinne sozialer Nachhaltigkeit somit die Deutungshoheit über eigene soziale Fragen und Problemlösungsansätze möglich sein. Dies kann jedoch durch die zunehmende Verbreitung von Systemen, die in westlich geprägten kulturellen Kontexten entwickelt wurden, unterminiert werden.

Ein dominantes Gütekriterium zur Bewertung von KI-Modellen ist ihre Generalisierbarkeit (Birhane et al. 2021). Generalisiert ein ML oder KI-Modell, dann liefert es nicht nur zuverlässige Ergebnisse für die gegebenen Trainingsdaten, sondern auch für weitere Datensätze. Üblicherweise streben Entwickler*innen die Übertragbarkeit eines ML-Modells auf möglichst viele Kontexte im Sinne von Stichproben, Datensätzen, Domänen oder Anwendungen an. Derzeit unterreflektiert bleiben dabei die von den Entwickler*innen eingeschriebene Werte und Normen in den ML-Modellen sowie die mit der Entwicklung verbundenen (ökonomischen) Interessen (ebd.). Eine Verallgemeinerung dieser Werte, Normen und Interessen für verschiedene kulturelle und sozioökonomische Kontexte ist nicht immer möglich oder sinnvoll, und wird der Diversität sozialer und kultureller Realitäten nicht gerecht. Die kulturelle Sensibilität ist somit ein wichtiges Kriterium für sozial nachhaltige KI-Systeme.

### Indikatoren für kulturelle Sensibilität

- Um den sozialen und kulturellen Kontext der Anwendung zu berücksichtigen, müssen **lokale Wissensbestände** in den Entwicklungsprozess integriert werden. Diese lokalen Wissensbestände können beispielsweise über Recherchen oder Umfragen aufgenommen werden. Im Idealfall werden lokale Expert*innen oder Stakeholder eingebunden, um den späteren Anwendungskontext bereits in die Entwicklung aufzunehmen (vgl. Hagendorff & Wezel 2020).
- Oftmals werden KI-Systeme für einen bestimmten Anwendungskontext oder in einem bestimmten Entwicklungskontext konzeptioniert und anschließend in diversen neuen und lokalen Anwendungskontexten angewandt. Aus diesem Grund sollten KI-Systeme eine gewisse **Anpassungsfähigkeit** besitzen: Die KI-Systeme lasse sich „umschulen“ (*retrain*) bzw. auf neue Parameter und Daten anpassen. Entweder findet diese Anpassung manuell statt, oder die KI-Systeme können ihr Lernverhalten automatisch an die neue Anwendungsform anpassen.
- Ausgangspunkt für eine hohe kulturelle Sensibilität sollte eine hohe **Team-Diversität** sein. Treffen Entwickler*innen aus unterschiedlichen Kulturen und Religionen, mit unterschiedlichem Geschlecht und diversen Alter aufeinander, können nicht nur besonders kreative und innovative Teams entstehen, sondern eine inhärente Sensibilität für andere Kulturen, Normen und Werte vorhanden sein oder gefördert werden. Diversität unter Mitarbeiter*innen in der KI-Forschung und -Entwicklung sollte daher angestrebt werden. Anhand der prozentuellen Anteile verschiedener Diversitätskategorien könnte dieses Kriterium beispielsweise analog zum Nachhaltigkeitsberichterstattungs-Standard (GRI G 4) bewertet werden.

## 4.3 Ökologische Kriterien und Indikatoren

In Folge des bedeutenden Fortschritts im Bereich der künstlichen Intelligenz – insbesondere beim maschinellen Lernen – und ihren vielfältigen Anwendungsmöglichkeiten, hat der Einsatz von KI stark zugenommen. Während KI-Anwendungen die Eindämmung des Klimawandels und die Anpassung an die daraus resultierenden Konsequenzen unterstützen können (Rolnick et al. 2019; Clutton-Brock et al. 2021), gerät auch der Ressourcenverbrauch von KI-Systemen und damit einhergehende ökologische Folgen vermehrt in den Fokus (Kaack et al. 2020; Kaack et al. 2021; Strubell et al. 2019; Schwartz et al. 2020). Durch den zunehmenden Einsatz von KI-Systemen und den Trend zur Verwendung immer größerer Modelle (Abbildung 8) haben auch die ökologischen Auswirkungen von KI zugenommen. Zum einen wird während der Entwicklung, dem Training und der Anwendung von KI-Systemen direkt Energie verbraucht, was abhängig von der verwendeten Energiequelle $CO_2$-Emissionen nach sich zieht. Zusätzliche Energie- und Ressourcenverbräuche entstehen durch die benötigte Hardwareinfrastruktur, während der Hardwareherstellung, dem Betrieb (v. a. durch die Kühlung der Hardware) und an ihrem Lebensende. Letztlich kann die Anwendung, in der ein KI-System eingesetzt wird, abhängig von Einsatzgebiet und -zweck zur Steigerung von Ressourcenverbräuchen und Treibhausgasemissionen führen.

### 4.3.1 Energieverbrauch

Besonders in den letzten Jahren ist ein rasanter Anstieg der Rechenleistung für das Training von KI-Modellen zu beobachten (Amodei & Hernandez 2018; Strubell et al. 2019). Dieser führt zum einen dazu, dass Forschung in diesem Bereich zunehmend teurer wird und so Eintrittsbarrieren für Akteure geschaffen werden (siehe Kapitel 3.7). Zum anderen führt es dazu, dass die Umwelt durch den gestiegenen Energiebedarf stärkeren Belastungen ausgesetzt ist.

Dabei führt *Moore's Law* grundsätzlich dazu, dass Computer Hardware (CPUs, GPUs, aber auch dedizierte Chips für ML, sogenannte *Tensor Processing Units*, TPUs) immer effizienter wird und so Berechnungen auch energieeffizienter ausgeführt werden können. Der exponentiell sinkende Energieverbrauch je Recheneinheit wird jedoch durch einen deutlich schneller wachsenden Trend zu größer werdenden ML-Modellen nichtig gemacht. In den Jahren 2012 bis 2018 hat sich der Bedarf an Rechenleistung für das Training moderner KI-Systeme um das 300.000-fache gesteigert, was insbesondere auf die Verwendung paralleler Algorithmen zurückzuführen ist (Amodei & Hernandez 2018). Seit 2018 hat sich diese Entwicklung besonders im Bereich der Sprachverarbeitungsmodelle weiter verstärkt (siehe Abbildung 8, zu beachten ist die logarithmische Skala), wie sich beispielsweise an der über hundertfachen Steigerung der Größe innerhalb eines Jahres zwischen GPT-2 (Radford et al. 2019) und GPT-3 (Brown et al. 2020), zwei auf Transformer-Netzwerken basierenden Modellen, zeigt. Dabei ging zwar anfänglich mit der steigenden Anzahl der Parameter auch ein starker Anstieg in der Vorhersagequalität in verschiedenen Problemklassen einher. Allerdings zeichnen sich nun vermehrt abnehmende Erträge ab, sodass Verbesserungen nun mit einem überproportionalen Energieverbrauch einhergehen (Thompson et al. 2021).

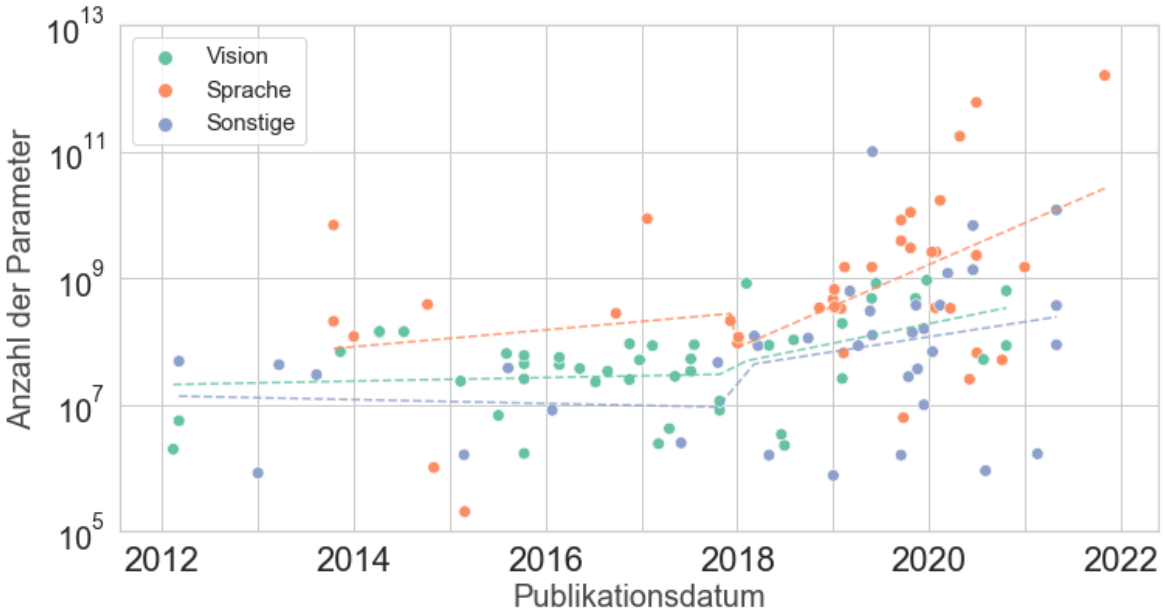


**Abbildung 8: Anzahl der Parameter populärer ML-Modelle zwischen 2012 und 2021**
Quelle: angelehnt an Sevilla et al. (2021)

Der direkte Energieverbrauch der Modelle aus dem Bereich des maschinellen Lernens verteilt sich auf die Lebenszyklusphasen Entwicklung, Training und Anwendung (Inferenz des ML-Modells und Datenmanagement). Die Entwicklung neuer Modelle, bei der in aufwendigen Experimenten neue Architekturen entworfen werden, ist dabei besonders energieintensiv. Hier werden häufig viele Trainingsdurchläufe mit unterschiedlichen Konfigurationen (sogenannten Hyperparametern) durchgeführt. Der Energieverbrauch der Trainings- und vor allem der Inferenzphase ist zwar deutlich geringer. Jedoch werden diese Phasen im Gegensatz zur Entwicklung, die lediglich einen einmaligen Vorgang darstellt, häufiger wiederholt.

Um steigende Umweltbelastungen durch die KI-Entwicklung zu verhindern, ist es nötig Effizienz als neues (zusätzliches) Leitziel der KI-Forschung zu etablieren. Die Energieeffizienz bezieht sich hier auf das Verhältnis der erreichten Modellleistung zum dafür verursachten Energieverbrauch. Aus diesem Grund ist es wiederum wichtig, dass Entwickler*innen in Forschung und Entwicklung in Zukunft die nötigen Informationen über den Entwicklungsprozess und das KI-Modell angeben. Denn nur so können Ergebnisse untereinander verglichen werden und Anreize, effizientere und klimafreundlichere Ergebnisse zu produzieren, geschaffen werden (Strubell et al. 2019). Nachdem der Effizienz der Modelle in der Modellauswahl bisher wenig Beachtung geschenkt wurde, zeigt sich dies nun als ein erster Trend in veröffentlichten Arbeiten (Schwartz et al. 2020).

### Indikatoren für Energieverbrauch

- Um Umweltauswirkungen zu reduzieren, ist es nötig, dass die **Energieeffizienz von KI-Modellen** während der Entwicklung und Auswahl berücksichtigt wird. Hierzu sollte die Entwicklung neuer Architekturen neben einer hohen Vorhersageleistung auf die Optimierung der Energieeffizienz abzielen (Schwartz et al. 2020). Einsetzende Organisationen sollten bei der Auswahl von KI-Systemen den Energieverbrauch als Kriterium berücksichtigen (Henderson et al. 2020). Dafür ist es relevant die Anforderungen an das System vorab zu definieren und Modelle mit niedriger Komplexität zu bevorzugen. Falls möglich sollten vortrainierte Modelle eingesetzt und lediglich angepasst werden (*Fine-Tuning*), um den Energieverbrauch während der Trainingsphase zu reduzieren (Strubell et al. 2019).

- Es stehen verschiedene Methoden zur Verfügung, um **Energieeffizienz und -verbrauch von KI-Entwicklung und KI-Einsatz aktiv zu optimieren**. Hierzu können unter anderem Maßnahmen zur Reduktion der Modellkomplexität (z. B. *pruning* und *quantization*), die Verwendung effizienter Lernverfahren (z. B. *distillation*) und eine effiziente Software- und Hardwareinfrastruktur beitragen (Menghani 2021). Des Weiteren kann durch Methoden zur Minimierung der zur Entwicklung verwendeten Daten (z. B. *data augementation*, *data minimalism*) der Energiebedarf weiter reduziert werden (Menghani 2021, Regneri et al. 2020). Um scheiternde Experimente schnellstmöglich zu identifizieren, sollten sorgfältige Unit-Tests, Integrationstests und ausführliches und frühzeitiges Debugging durchgeführt werden (Lacoste et al. 2019).
- Voraussetzung für die Berücksichtigung der Energieeffizienz sowie deren Optimierung ist die systematische und konsequente **Erfassung der Effizienz** während der Modellentwicklung, des Trainings sowie während des Einsatzes. Idealerweise wird der Energieverbrauch eines KI-Systems über alle Lebenszyklusphasen direkt gemessen und der Leistung des Systems zur Erfassung der Effizienz gegenübergestellt. Zur Annäherung an den Energiebedarf und die Energieeffizienz eines Modells können außerdem die Anzahl der Parameter und deren Ausnutzung, die Modelllaufzeit und die Anzahl der Fließkommaoperationen zur Berechnung einer Ausgabe hinzugezogen werden (Schwartz et al. 2020; Canziani et al. 2017). Jede der Metriken hat unterschiedliche Schwächen und prinzipiell ist ein Vergleich über unterschiedliche Systeme schwierig. Daher ist ein standardisiertes Reporting noch nicht etabliert (siehe z. B. García-Martín et al. 2019). Ein Ansatz zur Auswahl von Modellen sind *Budget/Accuracy*-Kurven (Dodge et al. 2019).

### 4.3.2 $CO_2$- und Treibhausgasemissionen

Der Einsatz von Energie und Ressourcen geht mit Treibhausgasen, d. h. mit Gasen, die zum Treibhausgaseffekt und so zur Erwärmung des Klimas beitragen, einher. Dabei ist vor allem Kohlenstoffdioxid ($CO_2$) als Konsequenz des Energieverbrauchs und seine langanhaltende Klimawirkung in der Atmosphäre relevant. Im Folgenden liegt der Fokus daher vor allem auf dem $CO_2$-Ausstoß, der entlang des Lebenszyklus des KI-Systems und der genutzten Hardware entsteht. Aber auch andere Treibhausgase entstehen indirekt. Gerade im Abbau und dem Transport fossiler Energieträger entsteht auch Methan, das um ein vielfaches klimawirksamer ist als $CO_2$. Daher wird häufig statt einem reinen $CO_2$-Ausstoß der Treibhausgasausstoß in $CO_2$-Äquivalenten angegeben ($CO_2e$).

Der $CO_2$-Ausstoß entsteht analog zum Energieverbrauch in allen Lebenszyklen des KI-Systems (Entwicklung, Training und Inferenz). Der $CO_2$-Ausstoß hängt neben dem Energieverbrauch selbst vom konkreten Zeitpunkt und Standort des Verbrauchs ab. Diese beiden Faktoren spiele eine Rolle für die $CO_2$-Intensität des Energiemixes des Rechenzentrums. So variiert die $CO_2$-Intensität der großen Betreiber aktuell stark. Jedoch definieren Amazon, Google und Microsoft alle Ziele von $CO_2$-Neutralität, $CO_2$-Freiheit oder gar $CO_2$-Negativität bis 2030[10,11]. Dabei beinhaltet der Begriff der $CO_2$-Neutralität, dass noch Emissionen ausgestoßen werden, aber diese vollständig kompensiert werden. Beim Ziel der $CO_2$-Freiheit fallen in einem Prozess gar keine Emissionen an. Im Falle von negativen $CO_2$-Emissionen werden bilanziell durch Maßnahmen wie Kompensation und z. B.

[10] https://sustainability.aboutamazon.com/environment/the-cloud?energy

https://sustainability.google/commitments/

[11] https://blogs.microsoft.com/blog/2020/01/16/microsoft-will-be-carbon-negative-by-2030/

Einspeisung erneuerbarer Energien (z. B. wenn diese nicht vom Rechenzentrum genutzt werden) positive Beiträge bezeichnet. Grundsätzlich muss auch berücksichtigt werden, dass auch bei bilanzieller Klimaneutralität trotzdem die entsprechenden zusätzlichen Erzeugungskapazitäten für erneuerbare Energien geschaffen werden müssen. Nur dann ist im Endeffekt auch ein positiver Beitrag für den Klimaschutz geleistet. In diesem Kontext hängt die Nachhaltigkeit sehr stark mit politischen Zielsetzungen wie der Energiewende zusammen. Es ergeben sich große Wechselwirkungen mit der Transformation des Energiesystems. Eine absolute Reduktion des Energieverbrauches – auch und gerade im Hinblick auf die Nutzung der digitalen Infrastruktur – ist deshalb mit Blick auf den Klimaschutz von besonderer Bedeutung.

Vor allem während der Modellentwicklung können hier analog zum Stromverbrauch, insbesondere für sehr große Modelle, auch große Mengen an $CO_2$-Emissionen entstehen. Strubell et al. haben etwa aufgezeigt, dass bei der Entwicklung großer Sprachmodelle über 280.000 kg $CO_2$e entstehen können, was dem $CO_2$ Ausstoß von 5 Autos über ihre gesamte Lebenszeit entspricht. Solche rechenintensiven Entwicklungsprozesse finden allerdings selten statt und werden nur von wenigen Organisationen durchgeführt (Kaack et al. 2021). Ebenso wie der Energieverbrauch, sind die während der Trainings- und Inferenzphase verursachten Emissionen deutlich geringer. Aber diese Prozesse werden weitaus häufiger durchgeführt. Hugging Face[12], eine Plattform die vortrainierte Modelle für Entwickler*innen bereitstellt, weist seit kurzem für Modelle im Rahmen der Model Cards (Mitchell et al. 2019) teilweise Emissionen aus dem Trainingsprozess mit aus. So lassen sich diese in der Auswahl der Modelle berücksichtigen. Abbildung 9 stellt die Emissionen durch das Training von Modellen, bei denen diese Information bereits ausgewiesen sind, dar. Dabei handelt es sich um einen Querschnitt aus Modellen, wie sie in verschiedenen NLP-Aufgaben (*natural language processing*, z. B. Text-Zusammenfassung und Entitäten-Erkennung) genutzt werden können. Hierbei wird deutlich, dass die $CO_2$-Emissionen insgesamt eher gering ausfallen – beispielsweise niedriger als der Ausstoß einer Stunde Video-Streaming in 4k-Auflösung auf einem Fernsehgerät.

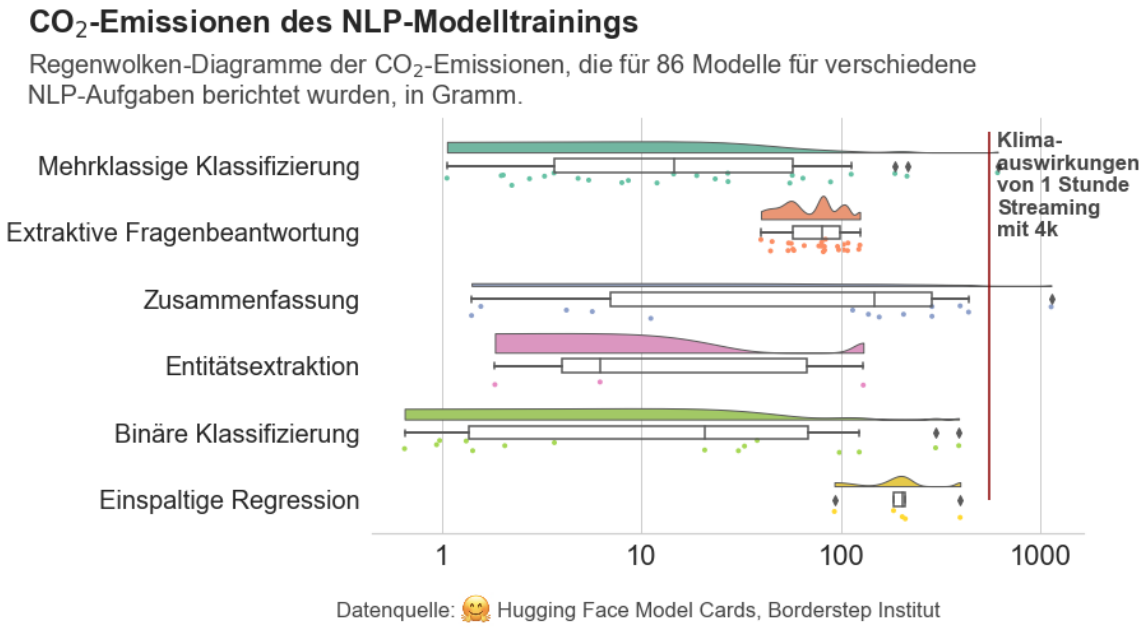


**Abbildung 9: $CO_2$-Emissionen des NLP-Modelltrainings**

Quelle: Eigene Darstellung auf Basis von Hugging Face Models Cards[11] und Borderstep (2020)

[12] https://huggingface.co/

#### Indikatoren für $CO_2$- und Treibhausgasemissionen

- Als wichtiger Indikator zur Evaluation der $CO_2$- und Treibhausgaus-Emissionen dient der **$CO_2$-Fußabdruck**. Die entwickelnde sowie die einsetzende Organisation sind für die Quantifizierung der direkten Emissionen, die durch das KI-System während der Entwicklung und während des Einsatzes verursacht werden, verantwortlich. Die Einheiten des Indikators für $CO_2$-Emissionen sind $CO_2$-Äquivalente (Lacoste et al. 2019, Lottick et al. 2019, Anthony et al. 2020). Je kleiner der Fußabdruck, desto geringer sind die negativen, direkten Auswirkungen auf die Umwelt. Zur Messung der $CO_2$-Emissionen können Werkzeuge wie beispielsweise *CodeCarbon.io*[13], der *ML CO2 Impact*- Rechner[14] (Lacoste et al. 2019) oder das *Microsofts Emissions Impact Dashboard*[15] herangezogen werden. Weitere Werkzeuge, wie das *Carbon Aware SDK*[16] der Green Software Foundation, befinden sich in der Entwicklung. Dieses Kriterium lässt sich im Vergleich zu anderen Kriterien sehr gut quantifizieren.
- Die **$CO_2$-Effizienz**, also das Verhältnis des Energiebedarfs zu den daraus erzeugten $CO_2$-Emissionen, sollte möglichst hoch sein. Dies wird durch einen hohen Anteil erneuerbarer Energien am Energieverbrauch erreicht. Durch eine geeignete Wahl von Trainingszeitpunkt und -standort kann die $CO_2$-Effizienz erhöht werden (Henderson 2020).
- Auch in naher Zukunft werden $CO_2$-Emissionen durch KI-Entwicklung entlang des Lebenszyklus nicht vermeidbar sein. Die Reduktion der Treibhausgasemissionen ist der erste Schritt. Die **verbleibenden Emissionen sollten zu einem möglichst hohen Anteil kompensiert werden**.

### 4.3.3 Nachhaltigkeitspotenziale in der Anwendung

Abhängig von Einsatzgebiet und -zweck können KI-Anwendungen einerseits in vielen Sektoren (z. B. Energie, Produktion, Land- und Forstwirtschaft und Katastrophenmanagement, vgl. UBA 2019) zur Bekämpfung der Klimaerwärmung beitragen. Anderseits können sie jedoch durch Anwendung in emissionsintensiven Sektoren oder Steigerung der Konsumnachfrage auch zum Anstieg der Treibhausgasemissionen beitragen (Kaack et al. 2020; Kaack et al. 2021; Rolnick et al. 2019). Es gilt daher die möglichen Auswirkungen von KI-Anwendungen auf Produktions- und Konsummuster zu betrachten.

Im Marketing wird KI unter anderem dafür eingesetzt, um auf Basis individueller Kundenprofilen inhaltlich und zeitlich personalisierte Werbeanzeigen zu erstellen. Hierdurch wird potenziell eine erhöhte Konsumnachfrage erzeugt, was einen Anstieg der damit einhergehenden Emissionen und Ressourcenverbräuche zur Folge hat. Andererseits kann KI im Bereich des Marketings auch ein wirkungsvolles Werkzeug bei der Förderung von Nachhaltigkeitsbemühungen auf der Angebots- und Nachfrageseite sein und so ein nachhaltiges Konsumverhalten stärken (Hermann 2021). Des Weiteren kann mithilfe von KI ein klimafreundlicheres Konsumverhalten angeregt werden und ein größeres Bewusstsein für die Auswirkungen von Konsumentscheidungen auf Umwelt und Klima

[13] https://codecarbon.io

[14] https://mlco2.github.io/impact

[15] https://www.microsoft.com/en-us/sustainability/emissions-impact-dashboard

[16] https://github.com/Green-Software-Foundation/carbon-aware-sdk

geschaffen werden (siehe z. B. KI-Leuchtturm Projekte GCA[17] & KI4NK[18]) (Coeckelbergh 2020; Rolnick et al 2019).

Neben dem Konsum erweist sich die Produktion als vielversprechendes Anwendungsfeld der KI. KI-Systeme werden in der Produktion dazu eingesetzt, um Prozesse entlang der Wertschöpfungskette zu optimieren. KI bietet also das Potenzial, die während der Produktion entstehenden Emissionen zu verringern – etwa durch das Verschlanken von Lieferketten, Verbesserung der Produktionsqualität, das Vorhersagen von Maschinenausfällen, optimierte Heiz- und Kühlsysteme und durch die Priorisierung von sauberer Energie gegenüber Energie aus fossilen Brennstoffen (Rolnick et al. 2019; Waltersmann et al. 2021). Auf diese Weise können KI-Systeme helfen, die Ressourceneffizienz zu steigern. Es ist jedoch anzumerken, dass eine höhere Effizienz durch Rebound-Effekte zu einer erhöhten Produktion von Gütern und damit zu einer Erhöhung der Treibhausgasemissionen führen kann (Rolnick et al. 2019).

Während der Konsum angetrieben durch die Digitalisierung bereits angestiegen ist (Wiedmann et al. 2020), wird durch den Einsatz von KI bis 2030 ein Anstieg des weltweiten BIP um 14 Prozent erwartet[19]. Dieser Anstieg soll einerseits durch die mittels KI ermöglichte Produktivitätssteigerung und andererseits durch eine erhöhte Konsumentennachfrage aufgrund von personalisierten und/oder qualitativ hochwertigeren KI-gestützten Produkten und Dienstleistungen ermöglicht werden.

### Indikatoren für Nachhaltigkeitspotenziale von KI-Anwendung auf den Konsum

- Damit KI-Systeme – insbesondere solche, die Empfehlungen für Produkte oder Dienstleistungen optimieren – die drei Nachhaltigkeitsdimensionen in der Anwendung maximieren, ist die **Berücksichtigung von Nachhaltigkeitskriterien** in der Entscheidungsfindung nötig. Dabei kann man die Art und Weise der Integration der Kriterien sowie die Anzahl und Art der Kriterien unterscheiden. Nachhaltigkeitskriterien wie $CO_2$-Emissionen, Arbeitsbedingungen und Gerechtigkeit könnten hierbei bspw. stärker gewichtet werden.
- In Analogie zur Berücksichtigung von Nachhaltigkeitskriterien sollten KI-Systeme im Handel, Marketing und Vertrieb zur **Förderung nachhaltiger Produkte** beitragen. Das System sollte die Position von nachhaltigen Produkten stärken (zeigt bspw. nachhaltige Alternativen) und zusätzliche Produktinformationen wie zum Beispiel $CO_2$-Emissionen ausweisen. Konsumenten können dadurch fundierte Entscheidungen treffen (Hermann 2021, Frick et al. 2019).
- Ein weiterer Indikator für nachhaltige KI-Systeme ist die **Förderung nachhaltiger Konsummuster**. Das KI-System fördert einen nachhaltigen Konsum der Nutzer. Bei Verbrauchsgütern, wie Strom oder Wasser, wird eine Verschwendung von Gütern und Ressourcen verhindert. Die Nutzer*innen werden darüber hinaus dazu angeregt, weniger aber qualitativ höherwertige Produkte zu erwerben (Hermann et al. 2021; Rolnick et al. 2019). Gleichzeitig werden nicht nachhaltige Nutzungsmuster (z. B. Binge-Watching, Lebensmittelverschwendung) vermieden oder reduziert. Das System regt zur Sparsamkeit und Suffizienz an.

### Indikatoren für *Nachhaltigkeitspotenziale von KI-Anwendung auf die Produktion*

- Für KI-Technologien, die in der Optimierung von Produktions-, Geschäfts- und Verwaltungsprozessen oder im Energiesektor eingesetzt werden, spiegelt eine **Reduzierung des Res-**

17 https://www.z-u-g.org/aufgaben/ki-leuchttuerme/projektuebersicht-fl2/gca

18 https://www.z-u-g.org/aufgaben/ki-leuchttuerme/projektuebersicht/ki4nk

19 https://www.pwc.co.uk/economic-services/assets/macroeconomic-impact-of-ai-technical-report-feb-18.pdf

**sourcenverbrauchs** den Grad an Nachhaltigkeit wider. Hierbei kann anhand der Art der Optimierung differenziert werden: Präzisere Nachfrage-Kalkulation, Predictive Maintenance, Optimierung der Produktionsprozesse (Qualitätskontrolle, Fehlerkontrolle etc.), automatisierte Materialbeschaffung, Optimierungen in der Logistik, Monitoring von Umwelteinflüssen, usw. (Rolnick et al. 2019). Die Ressourceneinsparung kann als Einsparung pro Produkteinheit bzw. Inputeinheit kalkuliert werden. Dabei kann zusätzlich auch zwischen den unterschiedlichen Ressourcenarten unterschieden werden (Waltersmann et al. 2021).

- Letztlich können KI-System einen positiven aber auch negativen Einfluss auf die **Produktqualität und -lebensdauer** nehmen. Ein bekanntes Problem ist in diesem Zusammenhang die softwarebedingte Hardware-Obsoleszenz, also das durch neue, performantere Software bedingte Austauschen von Hardwarekomponenten (Sætra 2021; Khakurel et al. 2018). Der Anspruch sollte es sein, dass das KI-System keine softwarebedingte Hardware-Obsoleszenz erzeugt. Stattdessen zielt die Anwendung des KI-Systems darauf ab, die Produktqualität (u. a. recyclebar, umweltfreundlich) und damit die Lebensdauer zu erhöhen.

### 4.3.4 Indirekter Ressourcenverbrauch

Durch die für die Entwicklung und den Einsatz von KI-Systemen benötigte Hardwareinfrastruktur werden zusätzlich zu den direkt verursachten Energieverbräuchen und $CO_2$-Emissionen weitere Umweltauswirkungen verursacht – vor allem durch Rechenzentren. Diese entstehen entlang der Wertschöpfungskette der eingesetzten Hardware (Herstellung, Betrieb und am Lebensende).

Zusätzlich zum Energiebedarf der Hardware selbst wird während ihres Betriebs zusätzliche Energie etwa durch Leistungsverluste, Kühlung, Beleuchtung und sonstige Verbräuche im Rechenzentrum aufgewendet (Whitehead et al. 2014). Derzeit wird durchschnittlich über 50 Prozent der Energie, die für den Betrieb der Hardware benötigt wird, zusätzlich für deren Bereitstellung benötigt (Uptime Institute 2021). Dieser Energieverbrauch hat abhängig vom vorliegenden Energiemix zusätzliche $CO_2$-Emissionen zur Folge. Um die Energieeffizienz von Rechenzentren zu erhöhen, wird vermehrt Wasser zur Kühlung verwendet. Vor allem in Gegenden in denen Wasserknappheit vorherrscht, stellt der erhebliche Wasserbedarf hierfür ein Problem dar. Des Weiteren ist das aus Rechenzentren abgeleitete Wasser mit diversen Verunreinigungen belastet, die eine negative Auswirkung auf die Umwelt haben können (Andrews et al. 2021). Während die Energieeffizienz des IT-Einsatzes in Rechenzentren zumeist erfasst wird, sind die $CO_2$-Emissionen und der Wasserverbrauch oft unbekannt (Uptime Institute 2021).

Angesichts von Fortschritten bei der Energieeffizienz der Systeme und der zunehmenden Nutzung erneuerbarer Energien stammt ein immer größerer Anteil der durch Rechenzentren entstehenden Emissionen nun aus dem Aufbau der benötigten Infrastruktur und der Herstellung der verwendeten Hardware. Neben den hierdurch verursachten Emissionen hat der angestiegene Einsatz von und die größeren Anforderungen an KI außerdem zu einer Vergrößerung der hierfür benötigten Infrastruktur geführt. Die für KI-Training und Inferenz eingesetzte Hardware von Facebook etwa ist innerhalb von weniger als zwei Jahren anteilig um einen Faktor von 4 beziehungsweise 3,5 angestiegen (Gupta et al. 2020; Naumov et al. 2020; Park et al. 2018). Um diese neuen Anwendungen zu unterstützen, werden in mobilen Endgeräten wie z. B. Smartphones außerdem mehr Transistoren und spezialisierte Schaltkreise eingebaut als in Vorgängermodellen (Gupta et al. 2020). Des Weiteren trägt KI zu einem beschleunigten und erhöhten Konsum von Technologieprodukten bei, etwa durch technische Obsoleszenz (Sætra 2021; Khakurel et al. 2018). Die Herstellung von elektronischen und nichtelektronischen Komponenten verbraucht Strom, Rohstoffe (u. a. Edelmetalle und

kritische Rohstoffe) Chemikalien, Wasser und erzeugt gefährliche Abfälle. All diese Faktoren können zu Umweltproblemen beitragen (Uddin & Rahman 2012). Die Auswirkungen werden weiterhin verstärkt, da die Geräte regelmäßig erneuert werden (Beispiel Rechenzentrum: Server alle 1-5 Jahre, Batterien alle 10 Jahre) (Andrews & Whitehead 2019).

Durch den gesteigerten Einsatz von KI und der hierfür benötigten Hardware, die aufgrund der sich rapide entwickelnden Technologien regelmäßig erneuert werden muss, ist davon auszugehen, dass die Menge des durch KI entstehenden Elektronikabfalls in Zukunft weiter zunehmen wird. Während etwa 70 bis 90 Prozent des Gewichts jedes elektronischen Produkts heute recycelt werden können, werden nur sehr wenige dieser vielen verschiedenen Elemente tatsächlich recycelt. Beispielsweise landen die meisten, wenn nicht sogar alle Seltenen Erden (REEs) aufgrund der geringen Menge, die in jedem Produkt verwendet wird, auf der Mülldeponie. Gründe für die niedrigen Recyclingquoten sind unter anderem die komplexe Materialzusammensetzung, die fehlende Infrastruktur in Form von hochtechnologischen Recyclinganlagen, fehlende Anreize für produzierende Unternehmen, die Langlebigkeit und Recyclingfähigkeit ihrer Produkte zu optimieren und ein Mangel an Systemen für die Rückgewinnung und das Recycling von Elektronikabfall sowie fehlende Anreiz für Konsumenten vorhandene Systeme zu Nutzen (Kreps & Fors 2020). Während Studien zur Entstehung von globalen Elektronikabfällen durchgeführt wurden (Forti et al. 2020), ist der Einfluss von KI darauf weitestgehend unerforscht.

### Indikatoren für den indirekten Ressourcenverbrauch

- Der **Anteil der zertifizierten Hardware** repräsentiert die Umweltfreundlichkeit der verwendeten Hardware-Komponenten. Renommierte Umweltzeichen und Zertifikate im Hardware-Bereich mit Fokus auf Green IT sind u. a. der Blaue Engel, das Europäische Umweltzeichen, der Energy Star oder das TCO-Gütesiegel.
- Analog zur verwendeten Hardware ermöglichen **Zertifizierungen der Rechenzentren**, auf denen das Training und die Inferenz stattfinden, eine Auskunft über die Nachhaltigkeit der Rechenleistung. Auch hier gibt es weitverbreitete Standards wie den Blauen Engel oder CEEDA (*Certification of Energy Efficiency for Data Centers*).
- Durch die Erfassung und Berechnung von **Effizienzmetriken** kann der ressourceneffiziente Betrieb von Rechenzentren und Hardware beurteilt werden. Hierzu kann eine Vielzahl an Metriken herangezogen werden, um die Effizienz auf unterschiedlichen Ebenen zu beurteilen (z. B. Effizienz von Energie-, $CO_2$- und Wassereinsatz) (Reddy et al. 2017). Die in der ISO Norm 30134 genannten Leistungskennzahlen für einen effektiven und effizienten Rechenzentrumsbetrieb umfassen dabei Kennzahlen zur Minimierung des Ressourcenbedarfs, Maximierung der IT-Leistung bei minimalem Energiebedarf, Nutzung von Abwärme und Einsatz von erneuerbaren Energien. Zusätzlich gibt die ISO Norm 30133 Hinweise und Richtlinien für einen ressourceneffizienten Rechenzentrumsbetrieb (Schödwell 2018).
- Als Indikatoren zur Bewertung des Hardwarelebensendes können die **Verwertungskennzahlen des Entsorgungsbetriebs** dienen. Grundsätzlich muss zwischen Recycling und Wiederverwendung unterschieden werden. Beim Recycling wird gebrauchte Hardware durch zertifizierte Fachbetrieb gesammelt. Komponenten (z. B. Batterie, Kabel, und Leiterplatte) werden separat entnommen und zum Teil für das Recycling weiter behandelt. Ein adäquater Subindikator ist hier die Recyclingquote – der gewichtsmäßige Anteil von recycelten Materialien an der entsorgten Hardware. Bei der Wiederverwendung – dem sogenannten *Reuse* – wird gebrauchte Hardware von *Original Equipment Manufacturers* (OEMs) oder Aufarbeitungsunternehmen gesammelt, wo aufbereitete Komponenten der Hardware wieder in neuen

oder gebrauchten Produkten eingesetzt werden. Dabei quantifiziert die Wiederverwendungsquote – also der gewichtsmäßige Anteil der Hardware, der aufbereitet oder wiederverwendet wird – die Nachhaltigkeit der Wiederverwendung. Letztlich lässt sich der Entsorgungsbetrieb an dem Anteil energetisch verwerteter Materialen an der entsorgten Hardware evaluieren (Peiró und Ardente 2015; Schödwell 2018).

## 4.4 Ökonomische Kriterien und Indikatoren

Legt man vor dem Hintergrund der gesellschaftlich vereinbarten Nachhaltigkeitsziele (insb. SDGs) die Notwendigkeit zugrunde, dass ökonomische Aktivitäten an der Einhaltung planetarer Grenzen sowie der Bereitstellung wesentlicher Lebensgrundlagen für alle Menschen zu orientieren sind, stellt sich die Frage, welche Entwicklungsrichtung von KI-Systemen mit Blick auf die Ökonomie zu beobachten sind. Das heißt welche Wirkungen gehen mit der Entwicklung und Anwendung von KI-Systemen einher und inwieweit befähigen oder verhindern diese Wirkungen eine nachhaltige Gestaltung der Ökonomie? Welche Stellschrauben lassen sich identifizieren, um die Entwicklung und Anwendung von KI-Systeme als dienlich für eine nachhaltige Ökonomie zu gestalten? Die hier zur Diskussion gestellten Kriterien für eine in diesem Sinne ökonomisch nachhaltige Gestaltung der Entwicklung und Anwendung von KI-Systemen beruhen auf Beobachtungen von Trends innerhalb der digitalen Ökonomie und Entwicklungstendenzen von KI-Märkten. Dabei ist festzuhalten, dass der Eintritt in den Markt der KI-Entwicklung im Wesentlichen von drei Faktoren bestimmt wird: i) den Zugriff auf eine große und diverse Menge an Daten, (ii) die Verfügbarkeit von hoher Rechenleistung, (iii) den Zugang zu Know-how zur Entwicklung von Algorithmen (Hall & Pesenti 2017; Vollhardt et al. 2021). Daraus ergeben sich potenzielle und zum Teil bereits zu beobachtende ökonomische Verteilungseffekte – sowohl auf Märkten der KI-Entwicklung als auch in Zielmärkten, also für KI-anwendende (oder nicht-anwendende) Organisationen – die als Risiken für eine nachhaltige Ausgestaltung der Ökonomie zu bewerten sind. Auch der Einsatz von KI-Systemen in Organisationen geht potenziell mit (im Sinne ökonomischer Nachhaltigkeit) risikobehafteten Einkommensverteilungen innerhalb der Belegschaft aber auch generell zwischen den Produktionsfaktoren Arbeit und Kapital sowie veränderten Arbeitsbedingungen einher (Burghin 2018; Norton 2017; Tratjenberg 2018) Die im Folgenden dargelegten Kriterien adressieren diese Risiken.

### 4.4.1 Marktvielfalt & Ausschöpfung des Innovationspotenzials

Unternehmen, die KI-Systeme entwickeln, kommt eine Schlüsselrolle in der Neugestaltung wirtschaftlicher Strukturen zu. Für eine sozial-ökologische Transformation bedarf es nachhaltiger Innovationen in allen wirtschaftlichen Sektoren. Hierbei können KI-basierte Verfahren einen positiven Beitrag leisten, denn große, meist passiv generierte Datenpools kombiniert mit Vorhersagealgorithmen haben das Potenzial, routinebasierte, arbeitsintensive Forschung in Innovationsprozessen zu substituieren und diese somit zu beschleunigen. Aus Nachhaltigkeitsperspektive ist es zu bevorzugen, dass möglichst viele Akteure aus allen wirtschaftlichen Sektoren Zugang zu den benötigten Datenpools haben, um dieses Innovationspotenzial auszuschöpfen.

Die drei eingangs genannten Schlüsselvoraussetzungen für die Entwicklung und das Trainieren von KI-Systemen (Daten, Rechenleistung, Know-how) stellen Eintrittsbarrieren zu KI-Märkten dar. Insbesondere mittels großer und geschlossener Datenpools konnten sich große Digitalunternehmen wie GAFAM und BATX[20], deren Geschäftsmodelle auf die Bindung möglichst vieler Nutzer*innen und dem Sammeln und Verarbeiten ihrer Daten basiert, bereits als KI-Marktführer hervortun (Simon et al 2019). Es sind Marktkonzentrationstendenzen zu beobachten (Bughin et al. 2018).

Marktkonzentrationen auf der Seite der KI-entwickelnden Unternehmen führen dazu, dass die Anwendungszwecke (d. h. Zielfunktionen der Algorithmen) für die KI-basierte Technologien entwickelt werden, von einigen wenigen Akteuren festgelegt werden. Dies kann zu einer Einschränkung der

[20] GAFAM dient als Akronym für die „amerikanischen Riesen“ Google, Apple, Facebook, Amazon und Microsoft, während BATX für die „chinesischen Riesen“ Baidu, Alibaba, Tencent und Xiaomi steht.

Entwicklung von KI-Systemen hinsichtlich der Diversität ihrer Zielfunktionen führen. Auf der Anwendungsebene können Marktkonzentrationen auf der Seite der KI-entwickelnden Unternehmen dazu führen, dass es zu einer Beschränkung der vermarkteten KI-Systeme hinsichtlich bestimmter (z. B. besonders profitträchtiger) Anwendungskontexte und Bedürfnisfelder (siehe z. B. Prause 2021; Carolan 2020; Rotz et al. 2019 im Kontext von Landwirtschaft) kommt oder Anbieter ihre Marktmacht nutzen, um eigene Gewinne auf Kosten von Konsument*innenrechten zu maximieren. Kann die Schlüsselrolle in der auf KI-Anwendungen basierten Neugestaltung wirtschaftlicher Strukturen aufgrund von großer Marktkonzentration nur von einigen wenigen Akteure wahrgenommen werden, kommt es zu einer Machtkonzentration in diesem Transformationsprozess. Das kann wiederum negative Auswirkungen auf notwendige sozial-ökologische Transformationsprozesse haben.

Aus Wettbewerbsperspektive setzen die Aussichten auf durch KI-Systeme beschleunigte Innovationsprozesse starke Anreize für Unternehmen frühzeitig für unterschiedliche Anwendungsbereiche geschlossene Datenpools anzulegen und auszuweiten, um Wettbewerbsvorteile durch Innovationsvorsprünge zu erlangen. Als Folge kann eine starke Fragmentierung von Datensets innerhalb einzelner Sektoren entstehen, was nicht nur die Innovationsproduktivität innerhalb des Sektors verringert, sondern auch die Spillover-Effekte zurück in den KI-entwickelnden Sektor und in andere Anwendungssektoren reduziert (Cockburn et al. 2019).

Die starken Anreize dafür, große, geschlossene Datenpools zu kontrollieren, um Marktanteile zu sichern, kann bei Unternehmen zu wettbewerbsverzerrenden Praktiken bei der Datenbeschaffung und Nutzeneinbußen für Konsument*innen führen: Etwa wenn große Digitalkonzerne durch Übernahme anderer Unternehmen mit datengetriebenen Geschäftsmodellen monopolartige Strukturen erzeugen, sie Nutzer*innen ihrer Anwendungen über ihre Datensammelpraktiken im Unklaren lassen, wenn sie ihre Marktposition nutzen, um Lock-In-Effekte für Nutzer*innen zu erzeugen und die Datenextraktion zu maximieren oder sie Daten von einzelnen Nutzer*innen von verschiedenen Plattformen zusammenzutragen und zu einem detaillierten Profil zusammenfügen.

### Indikatoren für Marktvielfalt & Ausschöpfung des Innovationspotenzials

- Zentrale Voraussetzung einer geringen Marktkonzentration, also einer hohen Marktvielfalt und damit einer hohen Ausschöpfung des Innovationspotenzials durch KI-Systeme, sind **faire Zugangsmöglichkeiten** zur KI-Entwicklung sowie zu Tools und Daten (insbesondere für Startups und gemeinnützige Organisationen). Das KI-entwickelnde Unternehmen macht hierfür neue KI-Technologien und Algorithmen (Wissenschaft und Quellcode) der Öffentlichkeit frei oder zu fairen Konditionen zugänglich (Bostrom 2017). Eine negative Bewertung erfolgt, sofern der Zugang nicht oder nur zu schweren Bedingungen ermöglicht wird. Des Weiteren setzt sich das KI-entwickelnde Unternehmen für offen zugängliche Datenpools bzw. Datentreuhand-Modelle ein (Cockburn et al. 2019; Borgogno & Colangelo 2019). Durch das KI-System generierte Datensätze fließen wiederum (anonymisiert und aggregiert) in offene Datenpools ein und stehen anderen Organisationen zur Verfügung. Insbesondere für kleine, neue entwickelnde Unternehmen (Startups) werden KI-Tools und grundliegende Frameworks und Bibliotheken zu günstigeren Preisen oder kostenlos zur Verfügung gestellt und somit eine höhere Adoption erzielt (Watney 2018). Gleichzeitig müssen Fragen nach Datenzugang auf regulatorischer Ebene adressiert werden.
- KI-Technologien und -Methoden stellen in der Regel nicht in sich geschlossene Systeme dar, sondern können als Plattformen interpretiert werden, auf denen verschiedene Methoden, Tools und Teilsysteme zusammenfließen. Damit vielfältige (sowohl KI-entwickelnde als auch KI-anwendende) Akteure, die Plattformen nutzen und durch externe Services erweitern kön-

nen, muss das KI-System oder die KI-Plattform **Schnittstellen** (APIs = Application Programming Interface) für Drittanbieter und andere KI-entwickelnde Unternehmen bereitstellen (Borgogno & Colangelo 2019). Diese Schnittstellen können sich zum Beispiel auf Daten und Abfragen sowie Funktionen beziehen. Im besten Fall werden die Schnittstellen offen (*public*) oder zu fairen Preisen als sogenanntes Partner-Modell angeboten.

- Eine weitere KI- und Plattform-spezifische Herausforderung sind Lock-In-Effekte innerhalb sogenannter digitaler Ökosysteme. Die Ermöglichung von **Multihoming** wirkt dem entgegen. Nutzer*innen und Entwickler*innen erhalten die Möglichkeit Plattformen (ohne großen Aufwand und Fallstricke) zu wechseln, parallel zu betreiben oder zu entwickeln (Hyrynsalmi et al. 2016). Der Plattformbetreiber stellt zudem die Kompatibilität zu anderen Plattformen und Tools sicher. Migrationen und Anbieterwechsel werden nicht unnötig erschwert.

### 4.4.2 Verteilungswirkung in Zielmärkten

Unternehmen kann der Zugang zu KI-basierten Anwendungen Marktvorteile gegenüber Konkurrenten verschaffen, die aus unterschiedlichen Gründen keinen Zugriff auf diese Technologien haben (Bughin et al. 2018). Neben fehlenden finanziellen Mitteln und fehlendem Know-how kann auch der Zuschnitt der Technologien auf die (Unternehmens-)eigenen Bedürfnisse ein Grund dafür sein, dass die Nutzung KI-basierter Technologien für bestimmte Akteure nicht in Frage kommt. Dies ist zum Beispiel der Fall, wenn Marktkonzentrationen auf der Seite der KI-entwickelnden Unternehmen dazu führen, dass es zu einer Beschränkung der vermarkteten KI-Systeme hinsichtlich bestimmter (z. B. besonders profitträchtiger) Anwendungskontexte und Bedürfnisfelder kommt (siehe 4.4.1) und marginalisierte Akteure nicht als Nutzer*innen adressiert werden. Solch ein strukturell bedingter fehlender Zugang zu KI-basierten Technologien kann zu Wettbewerbsverzerrungen auf den entsprechenden Zielmärkten und schließlich auch dort zu Marktkonzentrationen führen. Dies wiederum kann zu Nutzeneinbußen für Endkonsument*innen führen sowie die Innovationsfähigkeit des gesamten Sektors einschränken.

Der Zuschnitt KI-basierter Technologien auf Bedürfnisse von KI-anwendenden Unternehmen und Akteuren ist eng verknüpft mit der für die Entwicklung des KI-Modells verfügbaren Datenbasis. Bildet diese nur die Realität einiger dominanter Akteure des Zielmarktes ab, kann dies insbesondere in Märkten, die sich bereits aus diversen anderen politökonomischen Gründen durch große Ungleichheiten auszeichnen, zu einer weiteren Polarisierung und Konzentration auf den Zielmärkten führen (Rotz et al. 2019).

#### Indikatoren für Verteilungswirkung in Zielmärkten

- Damit auch kleine und marginalisierte Marktteilnehmer*innen von der Nutzung von KI-Systemen profitieren können, müssen entwickelnde Unternehmen **die Inklusivität der KI-Systeme in der Anwendung** sicherstellen. Das KI-System wird damit potenziell allen Marktakteuren und nicht nur dominanten Akteuren wirksam zur Verfügung gestellt. Die Zielfunktion der entwickelten algorithmischen Systeme sowie die der Entwicklung zugrundeliegende Datenbasis berücksichtigen die Bedürfnisse und Handlungsbedingungen marginalisierter Marktakteure des Zielmarktes und stärken nicht per se die Position dominanter Marktakteure (Watney 2018; Babina et al. 2020). Bewertungskriterien sind die Anpassungsfähigkeit an Datenmengen, *Accuracy*-Unterschiede zwischen Anwendungen für große und marginalisierte Marktakteure sowie die Diversität der Kunden, die die KI-Systeme einsetzen.
- Nicht nur die technischen Bedingungen müssen gegeben sein, damit marginalisierte Marktakteure KI anwenden können. Eine wichtige Voraussetzung sind die organisatorischen und finanziellen Rahmenbedingungen – sprich eine **Förderung von bspw. KMUs und NGOs**

(Watney 2018; Babina et al. 2020). Kleinen und mittelgroße Unternehmen sowie NGOs werden KI-Technologien vergünstigt und/oder als Open Source Produkt angeboten, um somit eine Einführung von KI-Technologien bei oftmals marginalisierten Marktakteuren zu fördern. Als Bewertung kann also eine Preisdifferenzierung für Kund*innen anhand ihrer Marktmacht sowie Angebote für zusätzlichen Support und Coaching herangezogen werden.

### 4.4.3 Arbeitsbedingungen und Arbeitsplätze

Der Einsatz von KI-basierten Technologien ist mit den Risiken verbunden, Arbeitsbedingungen für Arbeitnehmer*innen zu verschlechtern und negative verteilungsökonomische Auswirkungen für Arbeitnehmer*innen nach sich zu ziehen (Bughin et al. 2018; Tratjenberg 2018). KI-basierte Technologien ermöglichen die Automatisierung von einzelnen Arbeitsschritten und ganzen Prozessen. Sie können zum einen unterstützend für die menschliche Arbeitskraft wirken (*human-enhancing-innovations*, Tratjenberg 2018). Mit sensorischen, motorischen oder interpretativen Fähigkeiten können sie z. B. im Medizinbereich bei der Auswertung visueller Informationen von CT-Scans oder Röntgenaufnahmen assistieren (ebd.). Zum anderen können sie menschliche Arbeitskraft substituieren (*human-replacing-innovations*, ebd.). Gleichzeitig besteht aber auch die Möglichkeit, dass neue Arbeitsplätze geschaffen werden, die tendenziell höhere Qualifizierungen hinsichtlich technologischer Kompetenzen voraussetzen, Entfaltungsmöglichkeiten zur persönlichen Entwicklung der Arbeiter*innen bieten und potenziell höhere Einkommen erzielen (Kohl et al. 2020; Bughin et al. 2018). Andererseits kann es Fälle geben, in denen bestimmte Aufgaben besser oder billiger von Menschen ausgeführt werden – tendenziell betrifft dies monotone Arbeiten im Niedriglohnsegment. Das heißt, die Schaffung neuer Arbeitsplätze muss nicht zwingend mit der Schaffung menschenwürdiger Arbeit einhergehen (Sætra 2021). Diese mögliche Polarisierung des Arbeitsmarktes – hoch anspruchsvolle, hoch entlohnte Stellen auf der einen und monotone, niedrig entlohnte Stellen auf der anderen Seite – könnte mit einer Reduktion der Arbeitsplätze der „Mittelklasse“ einhergehen und soziale Unsicherheit verstärken (Kohl et al. 2020). Welcher Effekt – Substituierung oder Neuschaffung von Arbeitsplätzen – überwiegt, ist noch nicht abzusehen und unklar zu bewerten. Darüber hinaus ist davon auszugehen, dass Arbeitsmarktwirkungen in unterschiedlichen Sektoren variieren werden (Furman & Seamans 2019). Die Europäische Kommission (2018) legt nah, dass sichergestellt werden sollte, dass die sozialen Auswirkungen des KI-Systems gut verstanden werden. Das beinhaltet beispielsweise eine Überprüfung, ob die Gefahr des Verlustes von Arbeitsplätzen oder eine Dequalifizierung der Belegschaft besteht (Europäische Kommission 2018).

Neben den Auswirkungen der Anwendung von KI-Technologien auf den Arbeitsmarkt sind aber gleichermaßen die Arbeitsbedingungen in der KI-Entwicklung während des gesamten Lebenszyklus zu berücksichtigen. Vor allem für Crowdworker bzw. Clickworker in den Bereichen der Datenaufbereitung und -klassifizierung (bspw. Labeling) mangelt es häufig neben einer fairen Entlohnung auch an Aufstiegs- und Weiterbildungsmöglichkeit (Kittur 2013; Benner 2015; Salehi 2015; Schmidt 2019).

### Indikatoren für Arbeitsbedingungen und Arbeitsplätze

- Durch den verstärkten Einsatz von KI im Arbeitsumfeld besteht die Gefahr, dass sich die Arbeitsbedingungen beispielsweise durch verstärkte Überwachung der Mitarbeitenden verschlechtern oder Menschen ihren Arbeitsplatz verlieren. Die anwendende Organisation führt daher vor Einführung von KI-Systemen in Produktions- und Arbeitsprozessen eine **Abschätzung zu Auswirkungen für die Mitarbeiterschaft hinsichtlich Verschlechterungen der Arbeitsbedingungen sowie eines möglichen Stellenabbaus durch.** Sie strebt einen Interessensausgleich an und führt solche KI-basierten Verfahren ein, die menschliche Arbeit unterstützen und Arbeitsbedingungen verbessern. Wo menschliche Arbeit teilsubstituiert wird, werden Arbeitsprozesse so strukturiert, dass die von Menschen auszuführenden Tätigkeiten nicht monoton sind, sondern den Anforderungsprofilen der zuvor ausgeführten Tätigkeiten entsprechen. Unter anderem können Aspekte wie übermäßiges Monitoring (Moore 2019), Datenschutz & Privatsphäre, Einfluss auf HR- und Management-Entscheidungen, Transparenz, Fairness und Eingriffsmöglichkeiten als Bewertungsindikatoren für Arbeitsbedingungen herangezogen werden (Lane & Saint-Martin 2021). Wo menschliche Arbeit vollsubstituiert wird, erfolgt der Stellenabbau sozialverträglich nach einem zuvor mit der Mitarbeiterschaft vereinbarten Sozialplan.
- Die Sicherstellung **fairer Löhne entlang der Wertschöpfungskette** der KI-Entwicklung stellt einen weiteren Indikator für das Kriterium Arbeitsbedingungen und Arbeitsplätze dar. Das entwickelnde Unternehmen aber auch KI-anwendende Unternehmen evaluieren und verbessern die Arbeitsbedingungen entlang der gesamten Wertschöpfungskette der KI-Entwicklung (bspw. auch von digitalen Crowdworkern bzw. Clickworkern, die zum Labeln und Klassifizierung von Datensätzen beschäftigt werden).

## 4.5 Entwurf eines Indikatorensets für nachhaltige KI

Basierend auf der Analyse von bestehenden Diskursen, Konzepten und Ansatzpunkten, die sich mit den sozialen, ökologischen und ökonomischen Auswirkungen von KI-Systemen beschäftigen, haben wir für die dreizehn Nachhaltigkeitskriterien und fünf Querschnittsindikatoren jeweils Subindikatoren entwickelt oder abgeleitet, um so eine Bewertung dieser Indikatoren zu ermöglichen. Wir haben unser Kriterien- und Indikatorenset in Tabelle 2 dargestellt und nennen jeweils die relevanten Akteure, bei denen wir die Verantwortung für die Erfüllung dieses Kriteriums sehen. Die Kriterienentwicklung wurde in einem ersten Expert*innenworkshop diskutiert und die Operationalisierung basiert auf einem projektinternen interdisziplinären Austauschprozess. Wir haben in der letzten Spalte außerdem den Bezug unserer Indikatoren zu den SDGs angegeben, um aufzuzeigen, dass unsere Indikatoren selbstverständlich einen Bezug zu den Zielen für nachhaltige Entwicklung aufweisen. Die Erfüllung unserer Nachhaltigkeitsindikatoren kann somit auf der übergeordneten gesellschaftlichen Ebene einen Beitrag zu den SDGs leisten. Es geht dabei jedoch in erster Linie um die gesellschaftliche Einbettung dieser Systeme und „the profound and dynamic positive and negative impacts of artificial intelligence (AI) on societies, environment, ecosystems and human lives, including the human mind (…)“ (UNESCO 2021: 12). Wie genau diese Auswirkungen aussehen und zu bewerten sind, wird Gegenstand der weiteren Forschung und auch gesellschaftlicher Aushandlungsprozesse sein müssen. Wir gehen mit unserem Indikatorenset einen ersten Schritt zur Systematisierung dieser Auswirkungen entlang des KI-Lebenszyklus und möchten alle beteiligten Akteure ermutigen Nachhaltigkeitsbetrachtungen systematisch zu berücksichtigen.

## 4.6 Übersicht Kriterien- und Indikatorenset für nachhaltige KI

**Tabelle 2: Übersicht der Nachhaltigkeitskriterien und Indikatoren**

| Kriterium | Indikatoren | Subindikatoren (Operationalisierung) | Relevante Akteure | Bezug zu SDGs |
|---|---|---|---|---|
| **Organisatorische Verankerung** | | | | |
| **Querschnitts-kriterium** | Festgelegte Verantwortlichkeiten | Es existieren ein oder mehrere Ansprechpartner*innen für soziale und ethische Belange. | Einsetzende & entwickelnde Org. | |
| | | Die Zuteilung von Verantwortung für produzierte Ergebnisse ist klar und transparent geregelt, dokumentiert und kommuniziert. | Einsetzende & entwickelnde Org. | |
| | | Es existieren Regelungen zu Haftungsaspekten in Schadensfällen. | Einsetzende & entwickelnde Org. | |
| | Code of Conduct | Vorliegen eines Code of Conducts, der die Werte und Normen für die Implementierung und Nutzung des KI-Systems festlegt. | Einsetzende & entwickelnde Org. | |
| | Stakeholder-Analyse & -Beteiligung | Es wurden betroffene Stakeholder identifiziert und klassifiziert. | Einsetzende & entwickelnde Org. | |
| | | Die Stakeholder werden/wurden in den Designprozess, die Testphase sowie während neuer Releases einbezogen. Anzahl der durchgeführ-ten oder geplanten Stakeholder-Meetings. | Einsetzende & entwickelnde Org. | |
| | Dokumentation des KI-Systems | Aufzeichnung von Informationen hinsichtlich der Zielsetzung, Domain, Nutzer, Daten, Model, Feature-Selektion, Inputs, Tests, Metriken etc. (Google Model Card). | Entwickelnde Org. | |
| | Risikomanagement | Es findet ein standardisiertes und messbares Risk Assessment, Risk Monitoring und Risk Management statt. | Entwickelnde Org. | |
| | Beschwerde-mechanismus | Nutzer haben die Möglichkeit Fehler, unfaire und diskriminierende Ent-scheidungen, Privatsphäreeingriffe etc. dem KI-betreibenden Unterneh-men zu melden. | Einsetzende Org. | |

| Soziale Kriterien | | | | |
|---|---|---|---|---|
| **Transparenz & Verantwortungs-übernahme** | Transparenz, Erklärbarkeit und Prüfbarkeit des Modells | (1) Anzahl der Parameter im Modell. | Entwickelnde Org. | |
| | | (2) Nutzung von Deep Learning. | Entwickelnde Org. | |
| | | (3) Einsatz von Methoden zur Erhöhung der Transparenz und Erklärbarkeit. | Entwickelnde Org. | |
| | Informations-möglichkeiten zur Funktionsweise des Systems | (4) Nutzende/Betroffene werden über Einsatz des Systems informiert. | Einsetzende Org. | |
| | | (5) Es gibt öffentlich zugängliche Informationen zur Funktionsweise des Systems. | Einsetzende Org. | |
| **Nicht-Diskriminierung und Fairness** | Erfassung, Bewusstsein & Sensibilisierung für Fairness und Bias | (6) Höhe des Diskriminierungspotenzials anhand eines Impact-Assessments. | Einsetzende & entw. Org. |  *Geschlechtergleichstellung*  *Weniger Ungleichheiten*  *Gerechtigkeit* |
| | | (7) Anteil der KI-Systeme im Unternehmen, in denen Methoden zur Messung von Fairness und Bias zum Einsatz kommen. | Entwickelnde Org. | |
| | | (8) Benachteiligte Gruppen werden anhand von geschützten Attributen vorab definiert. | Einsetzende & entwickelnde Org. | |
| | Maßnahmen zur Verbesserung der Fairness | (9) Bei hohem Diskriminierungspotenzial werden Maßnahmen zur Beseitigung von Diskriminierung ergriffen. | Entwickelnde Org. | |
| **Technische Verlässlichkeit & Menschliche Aufsicht** | Performance-Kontrolle und Eingriffsmöglichkeit | (10) Es existieren Mechanismen zur Performancekontrolle. | Einsetzende & entw. Org. | |
| | | (11) Es existieren Maßnahmen, die menschlichen Eingriff ermöglichen. | Einsetzende & entw. Org | |

| | | | | |
|---|---|---|---|---|
| | Prüfung und Sicherstellung der Datenqualität | (12) Es existieren Maßnahmen zur Sicherstellung der Datenqualität (bspw. Audits oder kuratierte Datensätze). | Entwickelnde Org. | |
| **Selbstbestimmung & Datenschutz** | Sicherstellung der informationellen Selbstbestimmung | (13) Es wird ein Privacy-by-Design Ansatz konsequent umgesetzt. | Entwickelnde Org. |  *Gesundheit und Wohlergehen* |
| | | *(14)* Die Nutzer*innen besitzen die Kontrolle über ihre Daten. Datennutzung ist den Betroffenen transparent (z. B. Visualisierung), Nutzer*innen können über *opt-in* und *opt-out* über die Datennutzung bestimmen. Zweckgebundene Datennutzung durch Dritte. | Einsetzende Org. | |
| | | (15) Es existieren Mechanismen zur Benachrichtigung bezüglich der Sammlung, Verarbeitung und Verwendung von Daten. | Einsetzende Org. | |
| | Nutzungsautonomie | (16) Die Nutzung der KI ist selbstmotiviert und frei von Zwängen. | Einsetzende Org. | |
| | | (17) Verzicht auf suchtfördernde Mechanismen wie *Nudging* oder negativ beeinflussende Mechanismen wie *Dark Patterns*. | Einsetzende Org. | |
| **Inklusives und partizipatives Design** | Anwendung von Co-Design-Prinzipien | (18) Die Ausgestaltung des KI-System wurde mit Anwender*innen und/oder Endnutzer*innen und/oder Betroffenen reflektiert. | Einsetzende Org. | |
| | Sicherstellung von Barrierefreiheit und Zugänglichkeit | (19) Es werden unterschiedliche Fähigkeiten, Bedürfnisse und Präferenzen von Nutzenden berücksichtigt | Einsetzende Org. | |
| **Kulturelle Sensibilität** | Team Diversität und Berücksichtigung lokaler Wissensbestände | (20) Prozentualer Anteil der Mitarbeitenden je Geschlecht, Altersgruppe und Ethnizität. | Entwickelnde Org. |  *Weniger Ungleichheiten*  |
| | | (21) Es werden lokale Expert*innen in den Entwicklungsprozess integriert. | Entwickelnde Org. | |

| | | | | |
|---|---|---|---|---|
| | Anpassungsfähigkeit (*Retrain*-Option) | (22) Das KI-System lässt sich umschulen (*retrain*), bzw. an lokale und neue Anwendungskontexte, Normen und Werte (Parameter) anpassen. | Entwickelnde Org. | *Gerechtigkeit* |
| **Ökonomische Kriterien** | | | | |
| **Marktvielfalt & Ausschöpfung des Innovationspotenzials** | Zugangsmöglichkeiten zur KI-Entwicklung | (23) KI-Technologien und Algorithmen (Wissenschaft und Quellcode) werden der Öffentlichkeit frei oder zu fairen Konditionen zugänglich gemacht. | Entwickelnde Org. |  *Wirtschaftswachstum*  *Industrie, Innovation und Infrastruktur*  *Partnerschaften* |
| | | (24) Generierte Datensätze fließen in offene Datenpools und stehen anderen Organisationen zur Verfügung. | Entwickelnde Org. | |
| | | (25) Kleinen, neuen entwickelnden Unternehmen werden KI-Tools und Framework zu günstigeren Preisen oder kostenlos zur Verfügung gestellt. | Entwickelnde Org. | |
| | Schnittstellen | (26) Es werden Schnittstellen (APIs) für Drittanbieter und andere KI-entwickelnde Unternehmen bereitgestellt. | Einsetzende & entwickelnde Org. | |
| | Mulithoming und Kompatibilität | (27) Unternehmen, Nutzende und *developer communities* erhalten die Möglichkeit Plattformen (ohne großen Aufwand) zu wechseln, parallel zu betreiben oder zu entwickeln. Das Unternehmen stellt zudem die Kompatibilität zu anderen Plattformen und Tools sicher. | Entwickelnde Org. | |
| **Verteilungswirkung in Zielmärkten** | Inklusivität in der Anwendung | (28) Sicherstellung der Anpassungsfähigkeit an Datenmengen und Handlungsbedingungen. | Entwickelnde Org |  *Wirtschaftswachstum*  *Industrie, Innovation und Infrastruktur* |
| | | (29) Keine *Accuracy*-Unterschiede zwischen großen und marginalisierten Marktakteuren. | Entwickelnde Org. | |
| | | (30) Diversität der einsetzenden Kunden (Aufteilung nach HGB in kleine, mittelgroße und große Unternehmen). | Entwickelnde Org. | |
| | Förderung für KMUs und NGOs | (31) Kleinen Unternehmen werden KI-Technologien vergünstigt oder als Open Source-Produkt angeboten. | Entwickelnde Org. | |

| | | | | |
|---|---|---|---|---|
| **Arbeitsbedingungen und Arbeitsplätze** | Evaluation und Optimierung der Arbeitsbedingungen | (32) Eine Abschätzung zu Auswirkungen (Dequalifikation, Monotonisierung, Monitoring) für die Mitarbeitenden hinsichtlich Verschlechterungen der Arbeitsbedingungen wird durchgeführt. Bei Bedarf findet ein Interessensausgleich statt. | Einsetzende Org. |  *Menschenwürdige Arbeit* |
| | Faire Löhne entlang der Wertschöpfungskette | (33) Die Arbeitsbedingungen entlang der gesamten Wertschöpfungskette der KI-Entwicklung werden evaluiert und verbessert. | Entwickelnde Org. | |
| **Ökologische Kriterien** | | | | |
| **Energieverbrauch** | Berücksichtigung der Energieeffizienz | (34) Der Energieverbrauch wird in der Forschungs- und Entwicklungsphase des Algorithmus sowie in der Auswahl des Modells oder Systems berücksichtigt. | Einsetzende & entwickelnde Org. |  *Nachhaltige Produktion*  *Klimaschutz* |
| | | (35) Modelle mit niedrigerer Komplexität werden bei der Modellauswahl bevorzugt. | Einsetzende & entwickelnde Org. | |
| | | (36) Es werden vortrainierte Modelle und *Transfer Learning* zur Reduktion der Trainingszeit genutzt. | Einsetzende & entwickelnde Org. | |
| | Erfassung von Effizienzmaßen | (37) Während Modellentwicklung/-training/-einsatz werden Parameter zur Erfassung der Modelleffizienz gemessen:<br>- Energieverbrauch<br>- Laufzeiten<br>- Anzahl der Parameter<br>- Ausnutzung der Parameter<br>- Fließkommaoperationen<br>- Genauigkeit/Fehleranfälligkeit | Einsetzende & entwickelnde Org. | |
| | Optimierung der Energieeffizienz | (38) Es werden Methoden zur Komprimierung des Modells eingesetzt (z. B. Quantisierung, *Pruning*). | Einsetzende & entwickelnde Org. | |

| | | | | |
|---|---|---|---|---|
| | | (39) Es werden Methoden für ein effizientes Trainieren der Modelle angewendet (z. B. *model distillation*). | Entwickelnde & einsetzende Org. | |
| | | (40) Es werden Maßnahmen zur Reduktion der Datenmengen eingesetzt. | Entwickelnde & einsetzende Org. | |
| **$CO_2$- und Treibhausgasemissionen** | $CO_2$-Fußabdruck | (41) Direkte Emissionen, die durch das KI-System verursacht werden, werden quantifiziert. | Entwickelnde & einsetzende Org. |  *Klimaschutz* |
| | $CO_2$-Effizienz | (42) Verhältnis der verbrauchten Energie zu den dadurch erzeugten Emissionen. | Entwickelnde & einsetzende Org. | |
| | Emissionsausgleich | (43) Anteil der ausgeglichenen Emissionen an dem $CO_2$-Fußabdruck. | Entwickelnde & einsetzende Org. | |
| **Nachhaltigkeitspotenziale in der Anwendung** | Nachhaltige Zielfunktion in der Anwendung | (44) Das eingesetzte KI-System zielt primär darauf ab, Ressourcenverbräuche durch effizienteren Ressourceneinsatz zu reduzieren. | Entwickelnde & einsetzende Org. |  Nachhaltiger Konsum und Produktion  Klimaschutz |
| | Berücksichtigung von Nachhaltigkeitskriterien | (45) Empfehlungssysteme oder Entscheidungssysteme berücksichtigen in ihrer Entscheidungsfindung Nachhaltigkeitskriterien wie $CO_2$-Emissionen, Herstellungsbedingungen etc. | Entwickelnde & einsetzende Org. | |
| | Förderung nachhaltiger Produkte | (46) Das System stärkt die Position von nachhaltigen Produkten (zeigt bspw. nachhaltige Alternativen). | Entwickelnde & einsetzende Org. | |
| | Förderung eines nachhaltigen Konsums bzw. nachhaltiger Konsummuster | (47) Das KI-System fördert einen nachhaltigen Konsum der Nutzenden. Bei Verbrauchsgütern wie Strom oder Wasser wird eine Verschwendung von Gütern und Ressourcen verhindert. Die Nutzenden werden dazu angeregt weniger aber qualitativ höherwertige Produkte zu erwerben. | Entwickelnde & einsetzende Org. | |
| | Reduzierung des Ressourcen-verbrauchs | (48) Das KI-System optimiert Produktions-, Geschäfts- oder Verwaltungsprozesse und reduziert dadurch den relativen Ressourcenverbrauch. | Entwickelnde & einsetzende Org. | |
| | Auswirkung des KI-Systems auf die Produktqualität und -lebensdauer | (49) Softwarebedingte Hardware-Obsoleszenz wird vermieden (z. B. Modelle sind anpassbar für ältere Hardware). | Einsetzende Org. | |
| | | (50) Die Anwendung des KI-Systems zielt darauf ab, die Produktqualität zu erhöhen. | Einsetzende Org | |

| | | | | |
|---|---|---|---|---|
| **Indirekter Ressourcenverbrauch** | Nachhaltige Hardwareherstellung und -beschaffung | (51) Prozentualer Anteil der beschafften Hardware, die zertifiziert ist. | Einsetzende Org. |  Nachhaltige Produktion  Klimaschutz |
| | Nachhaltiger Rechenzentrums- und Hardwarebetrieb | (52) Das Rechenzentrum besitzt eine Umwelt/Energieeffizienz-zertifizierung. | Einsetzende Org. | |
| | | (53) Wichtige Effizienzmetriken wie *Power Usage Effectiveness* (PUE), *Carbon Usage Effectiveness* (CUE) oder *Water Usage Effectiveness* (WUE) werden berücksichtigt. | Einsetzende Org. | |
| | Kennzahlen des Verwertungsbetriebs | (54) Recyclingquote: Gewichtsmäßiger Anteil von recycelten Materialien an der entsorgten Hardware. | Einsetzende Org.<br>Verwertungsbetrieb | |
| | | (55) Wiederverwendungsquote:<br>Gewichtsmäßiger Anteil der Hardware, der aufbereitet oder wiederverwendet wird. | Einsetzende Org.<br>Verwertungsbetrieb | |
| | | (56) Gewichtsmäßiger Anteil von energetisch verwerteten Materialien an der entsorgten Hardware. | Einsetzende Org.<br>Verwertungsbetrieb | |
| | Entsorgungsszenarien für Server | (57) Einstufung des Entsorgungsprozesses in Szenarien. | Einsetzende Org.<br>Verwertungsbetrieb | |

# 4.7 Mapping der Nachhaltigkeitsindikatoren entlang des KI-Lebenszyklus

| | Organisationelle Einbettung | Konzeption-alisierung | Daten-Management | Modell-entwicklung | Modell-implementierung | Modellnutzung & Entscheidungsfindung |
|---|---|---|---|---|---|---|
| **Querschnitts-indikatoren** | • Stakeholdermanagement<br>• Dokumentation<br>• Risikomanagement<br>• Code of Conduct<br>• Verantwortlichkeiten | • Stakeholdermanagement<br>• Dokumentation<br>• Risikomanagement | • Dokumentation | • Stakeholdermanagement<br>• Dokumentation<br>• Risikomanagement | • Stakeholdermanagement<br>• Dokumentation | • Stakeholdermanagement<br>• Dokumentation<br>• Risikomanagement |
| **Transparenz & Verantwortung** | • Informations-möglichkeiten | | • Erklärbarkeit & Prüfbarkeit | | • Erklärbarkeit & Prüfbarkeit | • Informations-möglichkeiten |
| **Nicht-Diskriminierung & Fairness** | • Fairness Definition und Bewusstsein | | • Messung von Fairness & Bias | • Messung von Fairness & Bias | • Messung von Fairness & Bias<br>• Fairness-Maßnahmen | • Messung von Fairness & Bias |
| **Technische Verlässlichkeit & Menschliche Aufsicht** | | | • Datenqualität | | • Performance-Kontrolle | • Menschlicher Eingriff |
| **Selbstbestimmung & Datenschutz** | | • Nutzungsautonomie<br>• Verzicht auf Nudging | • Benachrichtigungen<br>• Privacy-by-Design | | | • Kontrolle der Daten<br>• Verzicht auf Nudging |
| **Inklusives & partizipatives Design** | | • Co-Design<br>• Barrierefreiheit & Zugänglichkeit | | | | |
| **Kulturelle Sensibilität** | • Team Diversität<br>• Lokale Experten | • Lokale Experten | | | • Retrain-Option | |
| **Energieverbrauch** | | • Berücksichtigung der Energieeffizienz | | • Berücksichtigung & Optimierung<br>• Erfassung von Effizienzmetriken | | • Berücksichtigung & Optimierung<br>• Erfassung von Effizienzmetriken |
| **$CO_2$- & THG-Emissionen** | • $CO_2$-Fußabruck<br>• $CO_2$-Effizienz<br>• Emissionsausgleich | | • $CO_2$-Fußabruck | • $CO_2$-Fußabruck<br>• $CO_2$-Effizienz | • $CO_2$-Fußabruck | • $CO_2$-Fußabruck |
| **Nachhaltigkeitspotenziale in der Anwendung** | • Nachhaltige Zielfunktion | • Nachhaltige Zielfunktion<br>• Berücksichtigung von Nachhaltigkeitskriterien<br>• Förderung nachhaltiger Produkte<br>• Nachhatliger Konsum | • Berücksichtigung von Nachhaltigkeitskriterien | | | • Reduzierung des Ressourcenverbrauchs<br>• Auswirkungen auf die Produktqualität |
| **Indirekter Ressourcen-verbrauch** | • Entsorgungsszenarien<br>• Hardwarebeschaffung<br>• Verwertungskennzahlen | • Rechenzentrum | | | | |
| **Marktkonzentration & Innovationspotenzial** | • Open Innovation Ansatz<br>• Anreize für KI-Startups | | • Schnittstellen<br>• Offene Datenpools | • Schnittstellen | • Multihoming & Kompatibilität | • Multihoming & Kompatibilität |
| **Verteilungswirkung in den Zielmärkten** | •Diversität der Kunden<br>•Anreize für KMUs | • Inklusivität in der Anwendung | | | • Inklusivität in der Anwendung<br>• Accuracy-Unterschiede | • Inklusivität in der Anwendung<br>• Accuracy-Unterschiede |
| **Arbeitsbedingungen und Arbeitsplätze** | • Faire Löhne entlang der Wertschöpfungskette | | | | | • Evaluation & Optimierung |

**Abbildung 10: Kriterien und Indikatoren entlang des KI-Lebenszyklus**

Quelle: Eigene Darstellung

Unsere Kriterien und Indikatoren adressieren verschiedene Phasen des Lebenszyklus von KI-Systemen, die sich in Konzeptualisierung, Datenmanagement, Modellentwicklung, Modellimplementierung sowie Modellnutzung und Entscheidungsfindung unterteilen lassen. Darüber hinaus haben wir die organisationale Einbettung den Lebenszyklusphasen hinzugefügt. Wir möchten verdeutlichen, dass viele Aspekte, die wir mit unseren Nachhaltigkeitskriterien berücksichtigen, in der Unternehmensstruktur und der Organisationskultur durch das Management verankert sein sollten. Denn die Entwicklung der sozio-technischen Systeme für künstliche Intelligenz findet im organisationalen Kontext statt und ist somit mit Verantwortungsstrukturen und Arbeitskulturen auf organisationaler Ebene verbunden.

Bei der **organisationalen Einbettung** sind dementsprechend alle Querschnittsindikatoren verortet, da sie sich vor allem darauf beziehen, unter welchen Bedingungen in Unternehmen oder anderen Organisationen (z. B. Forschung und Entwicklung, öffentlicher Sektor) die Entwicklung und Implementation von KI-Systemen stattfindet. Aspekte wie die Definition von Fairness, das Bewusstsein für faire Entscheidungen und die Diversität der Entwicklungsteams können ebenfalls im organisationalen Kontext beeinflusst werden, auch wenn die Wertemuster der Entwickler*innen freilich über den Organisationskontext hinausgehen. Insbesondere die ökonomischen Kriterien sind hier von Bedeutung, weil die Entscheidung über die grundsätzliche Ausgestaltung des Systems (Open Source, Offene Datenpools, Multihoming) der Organisation obliegt, die entweder selbst KI entwickelt oder die Entwicklung in Auftrag gibt. Im Hinblick auf die ökologischen Kriterien wie $CO_2$-Fussabdruck, $CO_2$-Effizienz oder auch den indirekten Ressourcenverbrauch geht es vor allem um das Engagement des Unternehmens oder der Forschungseinrichtung. Die ökologischen Auswirkungen der KI-Systeme müssen von Beginn mitgedacht werden und im Zusammenhang mit dem Umweltverbrauch der KI-Systeme und der Ausnutzung der digitalen Infrastruktur gesehen werden.

Die **Konzeptualisierung** ist ein besonders wichtiger Schritt für die Gestaltung von nachhaltigen KI-Systemen. Hier werden wesentliche Entscheidungen getroffen, wie die Frage, ob Stakeholder oder lokale Expert*innen involviert werden oder ob das System die Nutzungsautonomie respektiert und zugänglich ist. Der Aspekt des Co-Designs ist insofern von herausgehobener Bedeutung, als dass in diesem Prozessschritt viele Probleme, die im weiteren Verlauf der Modellentwicklung und -implementierung auftauchen können, adressiert werden können. Konsequent auf Co-Design und die Beteiligung betroffener Akteure zu setzen, kann viele ethisch-soziale Risiken in der KI-Entwicklung ex-ante vorbeugen oder minimieren. Auch ganz zentrale Fragen der Nachhaltigkeitswirkungen der Anwendung sind in dieser Phase ausschlaggebend. Denn welche Ziele mit dem System überhaupt verfolgt werden, sind hier festgelegt: Sollen beispielsweise besonders umweltfreundliche Konsummuster gefördert, Energie gespart oder die Nutzenden zum Mehrkonsum angeregt werden? Dabei bietet es sich zum Beispiel an, Nachhaltigkeitsaspekte – sofern diese verfolgt werden – in die *User Stories* zu integrieren. Im weiteren Projektverlauf werden dazu konkrete Vorschläge erarbeitet. Auch die Wahl des Rechenzentrums hat erheblichen Einfluss auf die Frage, wie nachhaltig die Nutzung von KI-Systemen erfolgen kann. Idealerweise liegen Kennzahlen zum ökologischen Betrieb des Rechenzentrums vor und Provider, die innovative Konzepte wie Abwärmenutzung, Wasserkühlung und eine möglichst hohe Energieeffizienz aufweisen, werden bevorzugt.

Im **Datenmanagement** liegt ein zentraler Ansatzpunkt für viele soziale Nachhaltigkeitsaspekte. Die Frage, wie gut die Datenqualität und wie repräsentativ der zugrundeliegende Datensatz ist, werden in dieser Phase entschieden. Möglichkeiten wie sogenannte kuratierte Datensets oder die Benachrichtigung der Nutzenden über die genutzten Daten sowie ein konsequenter Privacy-by-Design Ansatz, können gesellschaftliche Risiken der KI-Nutzung minimieren und die informationelle Selbstbestimmung fördern. Offene Datenpools und Schnittstellen können sowohl ökonomische als auch

ökologische Vorteile mit sich bringen, weil sie Datenströme optimieren, Austausch von Daten ermöglichen und somit das Innovationspotenzial sowie den Wettbewerb fördern können.

Die Phase der **Modellentwicklung** hat nicht nur auf eine gesellschaftlich vertretbare, sondern auch ökologisch sinnvolle Gestaltung der KI-Systeme Einfluss. Die Dokumentation des Modellentwicklungsprozesses sowie die Art und Weise der genutzten Daten und die Einbettung des Entwicklungsprozesses in ein geeignetes Risikomanagement sind wichtige Ausgangsvoraussetzungen zur Gestaltung nachhaltiger KI-Systeme. Nur so kann sichergestellt werden, dass die Nachvollziehbarkeit gewährleistet ist sowie die Modellarchitektur und Trainingsprozesse im Nachhinein auditierbar sind. In der Modellentwicklung, insbesondere im Training, können zur Verbesserung der ökologischen Nachhaltigkeit Messverfahren zur Erfassung des Energieverbrauches eingeführt werden, um möglicherweise jene Modellarchitekturen zu wählen, die bei gleichem Ergebnis mit weniger Energieverbrauch auskommen. Aktuelle Entwicklungen hin zu *Tiny AI* (Hao 2020*)*, also der mobilen, daten- und ressourcensparsamen KI, könnten dabei ein vielversprechender Ansatz sein.

Bei der **Modellimplementierung** geht es vor allem darum, kritisch zu überwachen und zu überprüfen, ob das KI-System verlässlich agiert. Im Zweifel geht es dann darum, Maßnahmen zu ergreifen, die eine *Performance*-Kontrolle oder das Nachjustieren des Trainingsprozesses (*Retrain*) ermöglichen. Gleichzeitig sollten Tools zur Messung bzw. Bewertung von Fairness eingesetzt werden. Die Erklärbarkeit und Prüfbarkeit von KI-Systemen ist dabei ein ganz entscheidender Aspekt, der in unabhängiger Form durch geeignete Methoden gewährleistet sein sollte. Im Hinblick auf die ökologischen Aspekte ist hier vor allem der $CO_2$-Fussabdruck entscheidend, der durch den Energieverbrauch in der Inferenz (also der Nutzung des KI-Systems) zustande kommt. Ob auch kleinere Marktakteure die Potenziale, die mit der KI-Entwicklung verbunden sind, nutzen können, hängt entscheidend von der Anpassung an Datenmengen und den Zugangsmöglichkeiten für kleinere und mittlere Akteure ab.

Wie in Kapitel 3 bereits erläutert, können die Nachhaltigkeitswirkungen von KI-Systemen nicht losgelöst von der **Modellnutzung und Entscheidungsfindung** betrachtet werden. Denn auch wenn im Design- und Entwicklungsprozess viele Stellschrauben für nachhaltige KI liegen, so stellt sich in dieser Phase die wichtige Frage, was mit dem KI-System letztendlich optimiert wird. Die Möglichkeiten und Anwendungsfelder haben wir in Tabelle 1 aufgezeigt und verdeutlicht, dass es unzählige Anwendungsfelder gibt. Allerdings können nur wenigen Anwendungen ein direkt positiver Beitrag zur Nachhaltigkeit unterstellt werden. Dennoch kann bei allen Systemen immer auch gefragt werden, welche Auswirkungen die Nutzung in dem jeweiligen Anwendungsfeld hat. Begünstigen sie beispielsweise nachhaltigen oder einen überflüssigen Konsum (z. B. Kingaby 2021) oder erhöhen sich durch vergangenheitsorientierte Daten oder Verzerrungen Finanzmarktrisiken (Vöpel 2020)? KI-Systeme können aber auch zur Aufdeckung von Geldwäsche, Menschenhandel oder zur nachhaltigen Waldbewirtschaftung eingesetzt werden. Ob und inwieweit die einzelnen Anwendungsfälle zur Nachhaltigkeit beitragen, muss daher stets im konkreten Fall geprüft werden. Unternehmen, die ihre wirtschaftlichen Aktivitäten an Nachhaltigkeit bzw. den SDGs ausrichten, könnten einen Nachhaltigkeitscheck für KI-Systeme einführen oder Nachhaltigkeitspotenziale der Anwendung in die User Stories integrieren. Ziel wäre zu prüfen, wie sinnvoll der konkrete Anwendungsfall aus Nachhaltigkeitssicht ist.

Über den gesamten Lebenszyklus können bei allen KI-Systemen Vorkehrungen getroffen werden, um die Nachhaltigkeit entlang des KI-Lebenszyklus zu verbessern sowie menschen- und umweltverträgliche Systeme zu entwickeln, die eine nachhaltige ökonomische Entwicklung zum Wohle des Planeten und für Frieden und Wohlstand sicherstellen.

# 5 Herausforderungen und Grenzen der KI-Nachhaltigkeitsbewertung

## 5.1 Nutzen und Grenzen eines indikatorenbasierten Ansatzes

Mit den Nachhaltigkeitskriterien für KI sollen mögliche Auswirkungen von KI-Systemen in einer übergreifenden Nachhaltigkeitsperspektive zusammengefasst und bewertbar gemacht werden. Gleichzeitig soll die Komplexität der entwickelten Kriterien und Indikatoren überschaubar bleiben, um ihre Anwendung in der Praxis zu ermöglichen und verschiedenen Akteuren Gestaltungsspielräume aufzuzeigen. Vor diesem Hintergrund ist unser Versuch, die vielfältigen Wirkungsebenen von KI-Systemen in einer umfassenden Nachhaltigkeitsperspektive zusammenzuführen, zwangsläufig mit Verkürzungen und Vereinfachungen verbunden. Die in diesem Vorhaben entwickelten Nachhaltigkeitskriterien und -indikatoren sind daher als ein Auftakt zur Diskussion darüber zu verstehen, wie man nachhaltige KI systematisieren und bewertbar machen kann.

Indikatoren sind dabei mit verschiedenen Vor- und Nachteilen verbunden, weil sie vielschichtige Sachverhalte und Problemlagen auf eine konkrete Dimension reduzieren sollen. Sie müssen gleichzeitig theoretischen, methodischen, praktischen und politischen Anforderungen genügen (SECO 2001). Im Idealfall können Indikatoren komplizierte Zusammenhänge einfach darstellen und auf nicht erkannte Steuerungsmöglichkeiten hinweisen (SECO 2001).

Die Wahl eines indikatorenbasierten Ansatzes soll die praktische Umsetzung nachhaltiger KI-Systeme ermöglichen, und dabei helfen konkrete Instrumente aufzubauen. Damit greifen wir die Kritik auf, dass es wünschenswert ist von einer bloßen Darstellung von Prinzipien zur praktischen Umsetzung zu kommen (z. B. AI Ethics Group 2020). Stattdessen solle es darum gehen, konkrete Indikatoren zu entwickeln und zu operationalisieren, damit die entsprechenden Akteure auch handeln können. Die Entwicklung guter Indikatoren sollte deshalb von Akteuren aus unterschiedlichen Disziplinen und mit unterschiedlichen Perspektiven vorgenommen werden. Im Projekt SustAIn erfolgte die Entwicklung in einem interdisziplinären Team aus den Bereichen Informatik, Ökonomie, Soziologie, Kommunikations- und Medienwissenschaft, Software-Entwicklung sowie Betriebswirtschaft und Politikwissenschaft. Darüber hinaus wurden die Kriterien und Indikatoren im Rahmen eines Stakeholderworkshops mit verschiedenen Akteuren aus der KI-Entwicklung und -Forschung diskutiert.

Indikatoren sollen über einen festgelegten, nicht oder nur sehr schwer messbaren Tatbestand Auskunft geben. Wir haben uns daher zunächst an den Nachhaltigkeitsdimensionen sowie dem KI-Lebenzyklus orientiert, um ein praktikables Indikatorenset zu schaffen. Viele der Indikatoren basieren auf aktuellen Diskussionen und Vorschlägen für die Bewertung von KI-Systemen. Aufgrund eines sehr umfassenden Diskurses über die ethisch-sozialen Aspekte konnte bei den sozialen Kriterien auf umfangreiche Literatur zurückgegriffen werden. Die Indikatoren sollen einen ersten Überblick ermöglichen, welche Aspekte aus einer umfassenden Perspektive auf nachhaltige KI von Bedeutung sind. Viele der Indikatoren knüpfen an andere Ansätze und Indikatorensysteme an (z. B. AI Ethics Group) und greifen bereits bestehende Bewertungskonzepte auf. Gerade wenn es um eine Bewertung der ökologischen Auswirkungen der digitalen Infrastrukturen geht, ist eine große Dynamik zu beobachten. Viele Kennzahlensysteme und Bilanzierungsmethoden werden erprobt und

weiterentwickelt (für den Bereich Datencenter siehe z. B. UBA 2018). Im Bereich der ökonomischen Indikatoren für nachhaltige KI-Systeme konnte nur auf wenige Vorarbeiten aufgebaut werden. Es mussten teilweise neue Kriterien und Indikatoren entwickelt werden. Grenzen sind diesem Ansatz insofern gesetzt als dass auch damit Wechselwirkungen zwischen verschiedenen Nachhaltigkeitsdimensionen nur unzureichend adressiert werden können. In Fallstudien sollen die Verbindungen zwischen verschiedenen Dimensionen im Projektverlauf etwas genauer betrachtet werden.

Der Umfang des Indikatorensets zeigt die Komplexität des Unterfanges, KI-Systeme in ihrer gesellschaftlichen und ökonomischen Einbettung aus Nachhaltigkeitsperspektive zu bewerten. Die Art der Indikatoren ist dabei höchst unterschiedlich und es sind sowohl qualitative als auch quantitative Indikatoren enthalten. Einige Indikatoren lassen sich standardisiert erfassen, andere müssen offener erhoben werden, um anschließend in Bewertungssysteme überführt zu werden. Zunächst dienen die Indikatoren dazu, die Bandbreite möglicher Auswirkungen zu operationalisieren. Ziel ist es, so den jeweiligen Akteuren überhaupt erst einen Zugang zu dieser Diskussion zu ermöglichen. Wie hoch der jeweilige Aufwand ist, über die Nachhaltigkeitsindikatoren Auskunft zu geben, werden wir in Fallstudien sowie in Kooperation mit weiteren Praxisakteuren im Projektverlauf erproben.

Wenngleich wir möglichst viele der bestehenden Konzepte und Analysen in die Nachhaltigkeitskriterien und -indikatoren haben einfließen lassen, können wir nicht alle Konzepte in jeglicher Detailtiefe berücksichtigen. In dem Kriterien- und Indikatorenset sind die wesentlichen Aspekte aus den aktuellen Diskursen um die ethisch-sozialen, ökologischen und ökonomischen Aspekte von Systemen künstlicher Intelligenz enthalten.

## 5.2 Bewertungsinstrumente zur Anwendung in der Praxis

Das hier vorgestellte Kriterien- und Indikatorenset fasst auf konzeptioneller Ebene zusammen, wie sich die Nachhaltigkeit von künstlicher Intelligenz messbar machen lässt. Das Set allein ist jedoch noch nicht praktisch anwendbar, sondern muss zunächst in Bewertungsinstrumente übersetzt werden. Dies erfolgte induktiv auf Basis einer systematischen Zuordnung und Gruppierung der Indikatoren. Einige Indikatoren beziehen sich ausschließlich auf den Anwendungsfall, andere treffen nur für die Entwicklungsphase zu. Manche Indikatoren beziehen sich auf den Umgang mit KI-Systemen im Allgemeinen in einer Organisation. Dies trägt auch dem Umstand Rechnung, dass Organisationen oft keine Aussage über individuelle KI-Systeme treffen können, sondern Prozesse und Kennzahlen für alle KI-Systeme in der Organisation definiert sind. Wiederum andere Indikatoren in unserem Katalog setzen normative Standards für nachhaltige KI-Systeme. Andere erfragen Informationen zu einem bestimmten Indikator, die erst später in eine Bewertung einfließen. Einige beziehen sich auf den Prozess, wie einzelne Systeme nachhaltig geplant und in Anwendung gebracht werden können. Andere schauen sich die *Performance* an. Zuletzt gehen weitere Indikatoren weit über den Handlungs- und Verantwortungsspielraum von einzelnen Organisationen hinaus und reflektieren stattdessen umfassende politische Regulierungsansätze. Um dieser Vielfalt an Zielsetzungen, Anwendungsebenen und adressierten Akteuren Rechnung zu tragen, schlagen wir vier Bewertungsansätze vor:

1. Einen Fragebogen zur Selbstauskunft für Organisationen, die Systeme künstlicher Intelligenz anwenden.

2. Einen Fragebogen zur Selbstauskunft für Organisationen, die Systeme künstlicher Intelligenz entwickeln.

3. Eine Richtlinie, die Schritte und Prozesse definiert, wie sich Systeme künstlicher Intelligenz nachhaltig entwickeln und anwenden lassen.

4. Ein politisches Bewertungsinstrument, das skizziert, wie regulatorische Ansätze auf politischer Ebene zur Förderung einer nachhaltigen Entwicklung und Anwendung von künstlicher Intelligenz beitragen.

Erste Gespräche mit Praxisakteuren deuten darauf hin, dass einige der metrischen Indikatoren aus den Fragebögen zur Selbstauskunft, bisher nur in geringem Maße vorliegen. Angaben zum Energieverbrauch, zur Energieeffizienz, zum Anteil der recycelten Hardwarekomponenten in der Entsorgung etc. werden häufig nicht standardisiert erfasst. Unternehmen können vielleicht den Energieverbrauch ihrer IT-Infrastruktur benennen, aber nur in seltenen Fällen für ihre KI-Systeme insgesamt geschweige denn für einzelne KI-Systeme. Die Messung des Energieverbrauchs müsste im Entwicklungsprozess dementsprechend mitberücksichtigt werden. Wir müssen also davon ausgehen, dass viele metrische Abfragen zunächst nicht beantwortet werden. Dennoch sind diese Indikatoren entscheidend für eine Bewertung der Nachhaltigkeit von KI. Unsere Instrumente müssen daher auch als Mittel verstanden werden, um Aufmerksamkeit für relevante Messkriterien zu generieren, die KI-entwickelnde und KI-anwendende Organisationen in Zukunft erfassen sollten. Viele der in den Fragebögen abgefragten Indikatoren beziehen sich allerdings nicht auf metrische Angaben. Sie sind daher in offenen Statements in Bezug auf Organisationsprozesse einfacher zu beantworten.

Gleichzeitig könnte eine politische Empfehlung sein, dass Unternehmen in Berichterstattungspflichten bestimmte Kennzahlen vorweisen müssen, so wie es im Bereich der Nachhaltigkeitsberichterstattung für viele Unternehmen zur Pflicht geworden ist. Da es solche Transparenzpflichten aber bisher nicht gibt, setzen wir aktuell auf Selbstauskünfte. Mit Blick auf mögliche Berichterstattungspflichten müsste gleichzeitig sichergestellt werden, dass diese Pflichten für kleine Akteure nicht zu überbordend gestaltet werden.

Als wichtiges ergänzendes Instrument zu den Selbstauskünften sind daher die Richtlinien zur Entwicklung nachhaltiger KI-Systeme zu sehen. Hier nehmen wir eine normative Setzung vor und skizzieren Entscheidungsprozesse in der Entwicklung eines nachhaltigen KI-Systems. Die Richtlinien können daher als unterstützendes Tool in der Entwicklung und Anwendung eines KI-Systems betrachtet werden. Sie können aber gleichzeitig auch als Tool fungieren, um bisherige Entwicklungs- und Anwendungsprozesse von KI-Systemen auf ihre Nachhaltigkeit zu prüfen. Die Bewertung der Nachhaltigkeit erfolgt in einem stufenbasierten System. So lässt sich ebenfalls skizzieren, wo weiteres Nachhaltigkeitspotenzial besteht.

# 6 Fazit

Vor dem Hintergrund gesellschaftlicher Folgen des Einsatzes von KI-Systemen hat die Diskussion über verantwortungsvolle und ethische KI eine große Dynamik entfaltet und auch deren Umwelteffekte finden zunehmend Beachtung. Gleichzeitig werden immer mehr Einsatzbereiche diskutiert. Zunehmend wird die Rolle dieser Technologie für die Erreichung der Ziele der nachhaltigen Entwicklung (SDGs) betont. Denn prinzipiell können KI-Systeme für sehr vielfältige Anwendungsbereiche eingesetzt werden, auch für sozial und ökologisch gewünschte Ziele (z. B. Armutsbekämpfung, Bildung, Energiewende, Mobilitätswende oder Agrarwende). Diese nachhaltigkeitsbezogenen Anwendungen spielen bislang in der Praxis aber nur eine untergeordnete Rolle. Es ist Vorsicht geboten, gesellschaftliche und ökologische Gerechtigkeits- und Verteilungsfragen an eine Technologie auszulagern, die Zielkonflikte nur unzureichend adressieren kann und deren Einsatz selbst mit problematischen Auswirkungen verbunden ist.

Eine Diskussion darüber, wie die Potenziale dieser Technologie genutzt werden können ohne negative soziale, ökologische und ökonomische Entwicklungen zu verstärken, ist daher dringend geboten. Wir schlagen ein Konzept für die Nachhaltigkeitsbewertung aller KI-Systeme vor. Die Nachhaltigkeitskriterien und -indikatoren ermöglichen es, alle drei Nachhaltigkeitsdimensionen entlang des KI-Lebenszyklus systematisch zu berücksichtigen. Damit knüpfen wir an international geführte Debatten an, die sich unter anderem in den UNESCO Empfehlungen zu ethischer KI wiederfinden. Denn diese Leitlinien beinhalten vielversprechende Ansätze, die sowohl ethisch-soziale Aspekte als auch ökologische Auswirkungen des Technologieeinsatzes berücksichtigen. Gerade in einer Zeit, in der mit dem Digitale-Dienste-Gesetz (*Digital Services Act*) oder der geplanten europäische KI-Verordnung wichtige politische Rahmensetzungen vorgenommen werden, müssen die Auswirkungen dieser Technologie umfassend adressiert werden. Der Koalitionsvertrag der neuen Ampel-Koalition setzt beispielsweise lediglich darauf nachhaltige Digitalisierung über nachhaltige Rechenzentren und zertifizierte IT-Hardware zu gestalten. Das Ansetzen bei Energiekosten und Hardwareinfrastrukturen ist zwar ein guter erster Ansatz, aber er greift viel zu kurz und deckt nur die offensichtlichsten Nachhaltigkeitsaspekte von Künstlicher Intelligenz ab. Unser Kriterien- und Indikatorenset setzt an diesen Punkten an und bietet konkrete Indikatoren für eine umfassende Nachhaltigkeitsbewertung an.

Das hier vorgestellte Kriterien- und Indikatorenset umfasst – unseres Wissens nach – den ersten systematischen und übergreifenden Vorschlag zur Operationalisierung der Nachhaltigkeit von KI-Systemen. Basierend auf einer Analyse der bestehenden Literatur und Ansatzpunkten in diesem Themenfeld machen wir einen theoretisch und praktisch fundierten Vorschlag, der die aktuelle Diskussion zusammenfasst, kondensiert, synthetisiert und in konkrete Indikatoren überführt. Wir möchten das Indikatorenset als Auftakt für eine Diskussion mit allen relevanten Akteuren und Stakeholdern verstehen, um die Diskussion zur nachhaltiger KI voranzubringen. Die Indikatoren sollen eine Orientierung bieten für Akteure, die diese Systeme entwickeln, einsetzen oder ihre Entwicklung innovationspolitisch vorantreiben.

Diese Diskussion wird allerdings kein Selbstläufer sein. Viele Organisationen sind daran interessiert, ihre KI-Systeme nachhaltiger zu gestalten. Bei vielen anderen Organisationen steht die Nachhaltigkeit von KI hingegen nicht weit oben auf der Prioritätenliste. Es braucht daher auch politische Weichenstellungen. Neben mehr Aufmerksamkeit für das Thema, sind mehr Informationen zu den Nachhaltigkeitsimplikationen von KI-Systemen notwendig. Die Verantwortungsübernahme aller beteiligten Akteure und die Befähigung zur Ausgestaltung nachhaltiger KI-Systeme muss vorangetrieben werden: Wer ist für die Nachhaltigkeitsauswirkungen von KI-Systemen verantwortlich? Wer

muss welche Rechte und Pflichten übernehmen, damit KI-Systeme nachhaltig – also zum Wohle der Allgemeinheit, von Mensch und Umwelt – eingesetzt werden und dennoch die mit ihr verbundenen Potenziale nutzbar gemacht werden können? Wir haben mit den Nachhaltigkeitskriterien für Künstliche Intelligenz einen ersten Schritt gemacht und eine Diskussionsgrundlage geschaffen.

# 7 Literaturverzeichnis

AI Ethics Impact Group (2019): From Principles to Practice - An interdisciplinary framework to operationalize AI ethics. Abrufbar: https://www.ai-ethics-impact.org/en (abgerufen am 29.11.2021).

AI for Good (2021): How AI is helping uncover modern slavery. AI for Good Blog. Abrufbar: https://aiforgood.itu.int/how-ai-is-helping-uncover-modern-slavery (Abgerufen am 29.11.2021).

AI HLEG (2019): High-level expert group on artificial intelligence. Ethics guidelines for trustworthy AI.

Algorithms to live by. Nat Mach Intell 2, 487 (2020). https://doi.org/10.1038/s42256-020-00230-w.

AlgorithmWatch (2021): Draft AI Act: EU needs to live up to its own ambitions in terms of governance and enforcement. 08/2021. Abrufbar: https://algorithmwatch.org/en/eu-ai-act-consultation-submission-2021 (Abgerufen am 29.11.2021).

Altenried, M. (2020): The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. Capital & Class, 44(2), 145-158.

Amodei, D., Hernandez, D., Sastry, G., Clark, J., Brockman, G., & Sutskever, I. (2018): AI and Compute. Abrufbar: https://blog.openai.com/aiand-compute. (abgerufen am 29.11.2021).

Andrews, D., & Whitehead, B. (2019): Data Centres in 2030: Comparative Case Studies that Illustrate the Potential of the Design for the Circular Economy as an Enabler of Sustainability. In Sustainable Innovation 2019: 22nd International Conference Road to 2030: Sustainability, Business Models, Innovation and Design.

Andrews, D., Newton, E. J., Adibi, N., Chenadec, J., & Bienge, K. (2021): A Circular Economy for the Data Centre Industry: Using Design Methods to Address the Challenge of Whole System Sustainability in a Unique Industrial Sector. Sustainability, 13(11), 6319.

Anthony, L. F. W., Kanding, B., & Selvan, R. (2020): Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. arXiv preprint arXiv:2007.03051.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020): Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82-115.

Association for Computing Machinery US Public Policy Coucil (2017): Statement on algorithmic transparency and accountability. Commun. ACM.

Babina, T., Fedyk, A., He, A. X., & Hodson, J. (2020): Artificial intelligence, firm growth, and industry concentration. Firm Growth, and Industry Concentration (November 22, 2020).

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021): On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623).

Benner, Christiane (Hg.) (2015): Crowdwork – zurück in die Zukunft? Perspektiven digitaler Arbeit. Frankfurt am Main: Bund-Verlag.

Birhane, A. (2020): Algorithmic colonization of Africa. SCRIPTed, 17, 389.

Birhane, A. (2021): Algorithmic injustice: a relational ethics approach. Patterns, 2(2), 100205.

Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., & Bao, M. (2021): The values encoded in machine learning research. arXiv preprint arXiv:2106.15590.

Bitkom (2019): Industrie 4.0 – jetzt mit KI. Hannover, 04/2019. Abrufbar: https://www.bitkom.org/sites/default/files/2019-04/bitkom-pressekonferenz_industrie_4.0_01_04_2019_prasentation_0.pdf (abgerufen am 29.11.2021).

Bitkom (2020): Unternehmen tun sich noch schwer mit Künstlicher Intelligenz. 06/2020. Abrufbar: https://www.bitkom.org/Presse/Presseinformation/Unternehmen-tun-sich-noch-schwer-mit-Kuenstlicher-Intelligenz (abgerufen am 29.11.2021).

BMWi (2019): Einsatz von Künstlicher Intelligenz in der Deutschen Wirtschaft. Stand der KI-Nutzung im Jahr 2019. 03/2020. Abrufbar: https://www.bmwi.de/Redaktion/DE/Publikationen/Wirtschaft/einsatz-von-ki-deutsche-wirtschaft.pdf?__blob=publicationFile&v=8 (abgerufen am 29.11.2021).

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Wang, W. (2021): On the Opportunities and Risks of Foundation Models. arXiv preprint arXiv:2108.07258.

Borgogno, O., & Colangelo, G. (2019): Data sharing and interoperability: Fostering innovation and competition through APIs. Computer Law & Security Review, 35(5), 105314.

Borderstep (2020): [Borderstep Institut für Innovation und Nachhaltigkeit] Videostreaming: Energiebedarf und CO2-Emissionen. Hintergrundpapier. Abrufbar: https://www.borderstep.de/wp-content/uploads/2020/06/Videostreaming-2020.pdf) (abgerufen am 05.12.2021)

Bostrom, N. (2017): Strategic implications of openness in AI development. Global policy, 8(2), 135-148.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020): Language models are few-shot learners. arXiv preprint arXiv:2005.14165.

Brundtland, G. H. (1987). Report of the World Commission on environment and development:" our common future.". UN.

Brynjolfsson, E.; Mcafee, A. (2017): Artificial intelligence, for real. Harvard Business Review.

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018): Notes from the AI frontier: Modeling the impact of AI on the world economy. McKinsey Global Institute.

Bünte C. (2019): Studie Künstliche Intelligenz – Die Zukunft des Marketings.

Bünte, C. , Pabst-Neuenfels E., Ngu c., Coleman M. (2018): Künstliche Intelligenz im Marketing.

Cai, E., Juan, D. C., Stamoulis, D., & Marculescu, D. (2017): Neuralpower: Predict and deploy energy-efficient convolutional neural networks. In Asian Conference on Machine Learning (pp. 622-637). PMLR.

Carnau, Peter (2011): Nachhaltigkeitsethik. Normativer Gestaltungsansatz für eine global zukunftsfähige Entwicklung in Theorie und Praxis. München: Rainer Hampp Verlag, 2011

Carolan, M. (2020): Acting like an algorithm: Digital farming platforms and the trajectories they (need not) lock-in. Agriculture and Human Values. https ://doi.org/10.1007/s1046 0-020-10032 -w.

Calvo, R. A., Peters, D., Vold, K., & Ryan, R. M. (2020): Supporting human autonomy in AI systems: A framework for ethical enquiry. In Ethics of Digital Well-Being (pp. 31-54). Springer, Cham.

Campolo, A., Sanfilippo, M. R., Whittaker, M., & Crawford, K. (2017): AI Now 2017 report.

Canziani, A., Culurciello, E., & Paszke, A. (2017): Evaluation of neural network architectures for embedded systems. In 2017 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 1-4). IEEE.

Cheatham, B., Javanmardian, K., & Samandari, H. (2019): Confronting the risks of artificial intelligence. McKinsey Quarterly, 1-9.

Chiusi, Fabio/Fischer, Sarah/Kayser-Bril, Nicolas/Spielkamp, Matthias (Hrsg.) (2020): Automating Society Report 2020. Algorithm-Watch; Bertelsmann Stiftung. Berlin, Gütersloh. Online verfügbar unter https://automatingsociety.algorithmwatch.org/ (abgerufen am 29.11.2021).

Cockburn, I. M., Henderson, R., & Stern, S. (2019): 4. The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis (pp. 115-148). University of Chicago Press.

Coeckelbergh, M. (2020): Artificial intelligence, responsibility attribution, and a relational justification of explainability. Science and engineering ethics, 26(4), 2051-2068.

Coleman, C., Kang, D., Narayanan, D., Nardi, L., Zhao, T., Zhang, J., ... & Zaharia, M. (2019): Analysis of dawnbench, a time-to-accuracy machine learning performance benchmark. ACM SIGOPS Operating Systems Review, 53(1), 14-25.

Delft University of Technology.

Deloitte (2017): Bullish on the business value of cognitive. Leaders in cognitive and AI weigh in on what's working and what's next. The 2017 Deloitte State of Cognitive Survey. Abrufbar: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-da-2017-deloitte-state-of-cognitive-survey.pdf (abgerufen am 29.11.2021).

Deloitte (2019): Transparency and Responsibility in Artificial Intelligence - A call for explainable AI. Abrufbar: https://www2.deloitte.com/nl/nl/pages/innovatie/artikelen/bringing-transparency-and-ethics-into-ai.html (abgerufen am 29.11.2021).

Deloitte (2020): KI-Studie 2020: Wie nutzen Unternehmen Künstliche Intelligenz?. Abrufbar: https://www2.deloitte.com/de/de/pages/technology-media-and-telecommunications/articles/ki-studie-2020.html (abgerufen am 29.11.2021).

Dena (2019): Künstliche Intelligenz als Chance für die Energiewirtschaft. Abrufbar: https://www.dena.de/newsroom/publikationsdetailansicht/pub/dena-umfrage-kuenstliche-intelligenz-als-chance-fuer-die-energiewirtschaft/ (abgerufen am 29.11.2021).

Dhar, P. (2020): The carbon impact of artificial intelligence. Nature Machine Intelligence, 2(8), 423-425.

Dignum, V. (2018): Responsible Artificial Intelligence. Social Artificial Intelligence Lab & Delft Institute Design for Values.

Dignum, V. (2019): Responsible artificial intelligence: how to develop and use AI in a responsible way. Springer Nature.

Dodge, J., Gururangan, S., Card, D., Schwartz, R., & Smith, N. A. (2019): Show your work: Improved reporting of experimental results. arXiv preprint arXiv:1909.03004.

Europäische Kommission (2019): Ethics guidelines for trustworthy AI. Abrufbar: https://ec.europa.eu/news-room/dae/document.cfm?doc_id=60425 (abgerufen am 29.11.2021).

Europäische Kommission (2020): White Paper on Artificial Intelligence: A European approach to excellence and trust. Abrufbar: com-mission-white-paper-artificial-intelligence-feb2020_en.pdf (europa.eu) (abgerufen am 29.11.2021).

Europäische Kommission (2020): The assessment list for trustworthy AI (ALTAI), Abrufbar: https://ec.europa.eu/newsroom/dae/docu-ment.cfm?doc_id=68342 (Abgerufen am 29.11.2021).

Europäische Kommission (2021): Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) un zur Änderung bestimmter Rechtsakte der Union. Abrufbar: https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52021PC0206. (Abgerufen am 29.11.2021).

Falco, G., Shneiderman, B., Badger, J. et al. (2021): Governing AI safety through independent audits. Nat Mach Intell 3, 566–571 https://doi.org/10.1038/s42256-021-00370-7.

Felländer-Tsai, L. (2020): AI ethics, accountability, and sustainability: revisiting the Hippocratic oath.

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019): Transparency you can trust: Transparency requirements for artifi-cial intelligence between legal norms and contextual concerns. Big Data & Society, 6(1), 2053951719860542.

Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., & Criado, N. (2021): Bias and Discrimination in AI: a cross-disciplinary perspective. *IEEE Technology and Society Magazine*, *40*(2), 72-80.

Forti, V., Balde, C. P., Kuehr, R., & Bel, G. (2020): The Global E-waste Monitor 2020: Quantities, flows and the circular economy poten-tial.

Frick, V., & Santarius, T. (2019): Smarte Konsumwende? Chancen und Grenzen der Digitalisierung für den nachhaltigen Konsum. In Das transformative Potenzial von Konsum zwischen Nachhaltigkeit und Digitalisierung (pp. 37-57). Springer VS, Wiesba-den.

Furman, J., & Seamans, R. (2019): AI and the Economy. Innovation policy and the economy, 19(1), 161-191.

García-Martín, E., Rodrigues, C. F., Riley, G., & Grahn, H. (2019): Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing*, *134*, 75-88.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2018): Datasheets for datasets. arXiv preprint arXiv:1803.09010.

Goertzel, B. (2014): Artificial general intelligence: concept, state of the art, and future prospects. Journal of Artificial General Intelli-gence, 5(1), 1.

Government of Canada (2019): Directive on Automated Decision-Making. (Online]) verfügbar unter: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 (abgerufen am 05.02.2021).

Gupta, V., Choudhary, D., Tang, P. T. P., Wei, X., Wang, X., Huang, Y., ... & Mahoney, M. W. (2020): Fast distributed training of deep neural networks: Dynamic communication thresholding for model and data parallelism. arXiv preprint arXiv:2010.08899.

Hagendorff, T., & Wezel, K. (2020): 15 challenges for AI: or what AI (currently) can’t do. AI & SOCIETY, 35(2), 355-365.

Hall, W., & Pesenti, J. (2017): Growing the artificial intelligence industry in the UK. Department for Digital, Culture, Media & Sport and Department for Business, Energy & Industrial Strategy. Part of the Industrial Strategy UK and the Commonwealth.

Hao, Karen (2020): Tiny AI. In: MIT Technology Review. Abrufbar: https://www.technologyreview.com/technology/tiny-ai/ (Abgerufen am 09.12.2021).

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020): Towards the systematic reporting of the energy and carbon footprints of machine learning. Journal of Machine Learning Research, 21(248), 1-43.

Hermann, E. (2021): Artificial intelligence in marketing: friend or foe of sustainable consumption?. AI & SOCIETY, 1-2.

Hoos, H. & Kersting, K. (2020): Die dritte Welle der Künstlichen Intelligenz. FAZ, Abrufbar: https://www.faz.net/aktuell/wirtschaft/digitec/die-dritte-welle-der-kuenstlichen-intelligenz-17100377.html (Abgerufen am 29.11.2021).

Homan, Karl (1996): Sustainability: Politikvorgabe oder regulative Idee? In: Homan, Karl (1996): Ordnungspolitische Grundfragen einer Politik der Nachhaltigkeit, S. 33-47.

Hyrynsalmi, S., Suominen, A., & Mäntymäki, M. (2016): The influence of developer multi-homing on competition between software ecosystems. Journal of Systems and Software, 111, 119-127.

Initiative “Konzernmacht beschränken” (2018): Konzernmacht in der digitalen Welt. Abrufbar: https://www.forumue.de/wp-content/uploads/2018/12/Konzernmacht_digitale-Welt.pdf (Abgerufen am 29.11.2021).

iRights.Lab & Bertelsmann Stiftung (2019): Algo.Rules: Regeln für die Gestaltung algorithmischer Systeme. 03/2019. Abrufbar: https://algorules.org/typo3conf/ext/rsmbstalgorules/Resources/Public/assets/pdf/Algo.Rules_DE.pdf (Abgerufen am 29.11.2021).

Jafari, S., Mtenzi, F., O’Driscoll, C., Fitzpatrick, R., & O’Shea, B. (2011): Measuring privacy in ubiquitous computing applications. Int. J. Digit. Soc, 2(3), 547-550.

Jain, S., Luthra, M., Sharma, S., & Fatima, M. (2020): Trustworthiness of Artificial Intelligence. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 907-912). IEEE.

Jansen, F. (2021): Resisting Surveillance in Public Spaces. Mozilla Foundation. Mozialla Festival. 02/2021. Abrufbar: https://foundation.mozilla.org/de/blog/resisting-surveillance-in-public-spaces/ (Abgerufen am 29.11.2021).

Jobin, A., Lenca, M. & Vayena, E. (2019): The global landscape of AI ethics guidelines. Nature Machine Intelligence, 2. September, p. 389–399.

Jungblut, S. I. (2021): Artificial Intelligence for Environmental and Climate Protection. Ökologisches Wirtschaften-Fachzeitschrift, 36(O1), 41-43.

Kaack, L., Donti, P., Strubell, E., & Rolnick, D. (2020): Artificial Intelligence and Climate Change: Opportunities, considerations, and policy levers to align AI with climate change goals. Abrufbar: https://us.boell.org/sites/default/files/2020-12/Artificial%20Intelligence%20and%20Climate%20Change_FINAL.pdf (Abgerufen am 29.11.2021).

Kaack, L., Donti, P., Strubell, E., Kamiya, G., Creutzig, F., et al (2021): Aligning artificial intelligence with climate change mitigation. ffhal-03368037 Abrufbar: https://hal.archives-ouvertes.fr/hal-03368037/document (Abgerufen am 29.11.2021).

Khakurel, J., Penzenstadler, B., Porras, J., Knutas, A., & Zhang, W. (2018): The rise of artificial intelligence under the lens of sustainability. Technologies, 6(4), 100.

Kingaby, H. (2021): Promises and Environmental Risks of Digital Advertising. Ökologisches Wirtschaften-Fachzeitschrift, 36(O1), 15-19.

Kittur, Aniket, et al. (2013): The Future of Crowd Work. In: CSCW ’13 Proceedings of the 2013 conference on Computer supported cooperative work, San Antonio, Texas, USA – 23.–27.02.2013. New York: ACM, 1301–1318.

Klinova K. & Korinek, A. (2021): AI and Shared Prosperity. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES)-

Kohl, J. L., van der Schoor, M. J., Syré, A. M., & Göhlich, D. (2020): Social sustainability in the development of service robots. In Proceedings of the Design Society: DESIGN Conference (Vol. 1, pp. 1949-1958). Cambridge University Press.

Kreps, D., & Fors, P. (2020): A Resource Perspective on E-Waste: A Global Problem with Local Solutions?. In Unimagined Futures–ICT Opportunities and Challenges (pp. 129-141). Springer, Cham.

Kroll, C., Warchold, A., & Pradhan, P. (2019): Sustainable Development Goals (SDGs): Are we successful in turning trade-offs into synergies?. Palgrave Communications, 5(1), 1-11.

Kwet, M. (2019): Digital colonialism: US empire and the new imperialism in the Global South. Race & Class, 60(4), 3-26.

Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019): Quantifying the carbon emissions of machine learning. arXiv preprint arXiv:1910.09700.

Lane, M., & Saint-Martin, A. (2021): The impact of Artificial Intelligence on the labour market: What do we know so far?.

Littig, B., & Griessler, E. (2005): Social sustainability: a catchword between political pragmatism and social theory. International journal of sustainable development, 8(1-2), 65-79.

Lottick, K., Susai, S., Friedler, S. A., & Wilson, J. P. (2019): Energy Usage Reports: Environmental awareness as part of algorithmic accountability. arXiv preprint arXiv:1911.08354.

Lu, Y., & Zhou, Y. (2019): A short review on the economics of artificial intelligence. CAMA Working Paper 54/2019.

Malone, C. G., & Belady, C. L. (2008): Optimizing Data Center TCO: Efficiency Metrics and an Infrastructure Cost Model. Ashrae Transactions, 114(1).

Manyika, J., Chui, M., Miremadi, M., Bughin, J., & George, K. (2017): A future that works: AI, automation, employment, and productivity. McKinsey Global Institute Research, Tech. Rep, 60, 1-135.

Matutinović, I. (2001): The aspects and the role of diversity in socioeconomic systems: an evolutionary perspective. Ecological Economics, 39(2), 239-256.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021): A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6), 1-35.

Menghani, G. (2021): Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better. arXiv preprint arXiv:2106.08962.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019): Model cards for model reporting. In Proceedings of the conference on fairness, accountability, and transparency (pp. 220-229).

Moore, P. (2019): OSH and the Future of Work: benefits and risks of artificial intelligence tools in workplaces, EU-OSHA Discussion papers, EU-OSHA, Bilbao. Abrufbar: https://osha.europa.eu/en/publications/osh-and-future-work-benefits-and-risks-artificial-intelligence-tools-workplaces/view (Abgerufen am 29.11.2021).

Münchener Kreis (2020): Leben, Arbeit, Bildung 2035+.Zukunftsstudie Müchener Kreis. Band VIII. Abrufbar: https://www.muenchner-kreis.de/fileadmin/user_upload/2020_Zukunftsstudie_MK_Band_VIII_Publikation.pdf (Abgerufen am 29.11.2021).

Nash, K. L., Blythe, J. L., Cvitanovic, C., Fulton, E. A., Halpern, B. S., Milner-Gulland, E. J., ... & Blanchard, J. L. (2020): To achieve a sustainable blue future, progress assessments must include interdependencies between the sustainable development goals. One Earth, 2(2), 161-173.

Naumov, M., Kim, J., Mudigere, D., Sridharan, S., Wang, X., Zhao, W., ... & Smelyanskiy, M. (2020): Deep learning training in facebook data centers: Design of scale-up and scale-out systems. arXiv preprint arXiv:2003.09518.

Norton, A. (2017): Automation and inequality: The changing world of work in the global South. Issue Paper. International Institute for Environment and Development.

Nussbaum, M. C. (2006): Education and democratic citizenship: Capabilities and quality education. Journal of human development, 7(3), 385-395.

Ogolla, S. & Gupta, A. (2018): Inclusive Design – Methods To Ensure A High Degree Of Participation in Artificial Intelligence (AI) Systems. University of Oxford Connected Life 2018 Link: https://files.persona.co/74670/Ogolla-Gupta-2018.pdf (Abgerufen am 29.11.2021).

Patterson, M. K., Tschudi, W., VanGeet, O., & Azevedo, D. (2011): Towards the Net-Zero Data Center: Development and Application of an Energy Reuse Metric. ASHRAE Transactions, 117(2).

Park, J., Naumov, M., Basu, P., Deng, S., Kalaiah, A., Khudia, D., ... & Smelyanskiy, M. (2018): Deep learning inference in facebook data centers: Characterization, performance optimizations and hardware implications. arXiv preprint arXiv:1811.09886.

Peter Clutton-Brock, David Rolnick, Priya L. Donti, Lynn H. Kaack, et al. (2021): Climate Change and AI Recommendations for Government Action. Abrufbar: https://www.gpai.ai/projects/climate-change-and-ai.pdf (Abgerufen am 29.11.2021).

Petschow, U., Lange, S., Hofmann, D., Pissarskoi, E., Moore, N., Korfhage, T., & Ott, D. (2018). Gesellschaftliches Wohlergehen innerhalb planetarer Grenzen. Der Ansatz einer vorsorgeorientierten Postwachstumsposition. Zwischenbericht des Projektes Ansätze zur Ressourcenschonung im Kontext von Postwachstumskonzepten.

Prause, L., Hackfort, S., & Lindgren, M. (2021). Digitalization and the third food regime. Agriculture and human values, 38(3), 641-655.

PWC (2021): Künstliche intelligenz in der Gesundheitswirtschaft – Wie KI zu einer besseren und günstigeren Gesundheitsversorgung beitragen kann. Abrufbar: https://www.pwc.de/de/gesundheitswesen-und-pharma/wie-kuenstliche-intelligenz-das-gesundheitssystem-revolutioniert.html (Abgerufen am 29.11.2021).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019): Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.

Raworth, K. (2017): Doughnut economics: seven ways to think like a 21st-century economist. Chelsea Green Publishing.

Reddy, V. D., Setz, B., Rao, G. S. V., Gangadharan, G. R., & Aiello, M. (2017): Metrics for sustainable data centers. IEEE Transactions on Sustainable Computing, 2(3), 290-303.

Regneri, M. (2021): Datenwert und Datenminimalismus: Wege zu nachhaltiger künstlicher Intelligenz. In CSR und Künstliche Intelligenz (pp. 189-207). Springer Gabler, Berlin, Heidelberg.

Reset.org (2020): Greenbook (1): Künstliche Intelligenz - Können wir mit Rechenleistung unseren Planeten retten?, 09/2020, Abrufbar: https://reset.org/greenbook_01_kuenstliche-intelligenz/ (Abgerufen am 29.11.2021).

Resseguier, A., & Rodrigues, R. (2020): AI ethics should not remain toothless! A call to bring back the teeth of ethics. Big Data & Society, 7(2), 2053951720942541.

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020): Designing fair AI for managing employees in organizations: a review, critique, and design agenda. Human–Computer Interaction, 35(5-6), 545-575.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., ... & Foley, J. A. (2009): A safe operating space for humanity. nature, 461(7263), 472-475.

Rohde, F., Gossen, M., Wagner, J., & Santarius, T. (2021): Sustainability challenges of Artificial Intelligence and Policy Implications. Ökologisches Wirtschaften-Fachzeitschrift, 36(O1), 36-40.

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., … & Bengio, Y. (2019): Tackling climate change with machine learning. arXiv preprint arXiv:1906.05433.

Rotz, S., Duncan, E., Small, M., Botschner, J., Dara, R., Mosby, I., ... & Fraser, E. D. (2019): The politics of digital agricultural technologies: a preliminary review. Sociologia Ruralis, 59(2), 203-229.

Salehi, N., Irani, L. C., Bernstein, M. S., Alkhatib, A., Ogbe, E., & Milland, K. (2015): We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In Proceedings of the 33rd annual ACM conference on human factors in computing systems, 1621-1630.

Sætra, H. S. (2021): A Framework for Evaluating and Disclosing the ESG Related Impacts of AI with the SDGs. Sustainability, 13(15), 8503.

Sadovoi, M. (2021). Tiny AI–a great step for artificial intelligence development. Technical-Scientific Conference of Undergraduate, Master and PhD Students

Schmidt, F. A. (2019). Crowdproduktion von Trainingsdaten: Zur Rolle von Online-Arbeit beim Trainieren autonomer Fahrzeuge (No. 417). Studie der Hans-Böckler-Stiftung.

Schneider, J., & Ziyal, L. K. (2019): We Need to Talk, AI. Dr. Julia Schneider.

Schödwell, B. (2018): Kennzahlen und Indikatoren für die Beurteilung der Ressourceneffizienz von Rechenzentren und Prüfung der praktischen Anwendbarkeit. 02/2018. Umweltbundesamt.

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020): Green AI. Communications of the ACM, 63(12), 54-63.

SECO (2001): Handbuch Indikatorenbildung für die wirtschaftliche Zusammenarbeit mit Entwicklungs- und Transitionsländer. Version 1.0- 11/2001

Sen, A. (2000): Social exclusion: Concept, application, and scrutiny.

Sevilla, J, Villalobos P., Cerón J.P., Burtell, M., Heim, L., Nanjajjar A.B. (2021): Parameter counts in Machine Learning. Towards Data Science. Abrufbar: https://towardsdatascience.com/parameter-counts-in-machine-learning-a312dc4753d0 (Abgerufen am 29.11.2021).

Siggiqui, S., 2018. The Guardian. Online: https://www.theguardian.com/uk-news/2018/mar/26/cambridge-analytica-trump-campaign-us-election-laws (abgerufen am 06.01. 2021).

Simon, J. P. (2019). Artificial intelligence: scope, players, markets and geography. Digital Policy, Regulation and Governance. 21 (3), 208-237.

Smuha, N. A. (2019): The eu approach to ethics guidelines for trustworthy artificial intelligence. Computer Law Review International, 20(4), 97-106.

Spiekermann, S. (2021): Value-based Engineering: Prinzipien und Motivation für bessere IT-Systeme. Informatik Spektrum, 44(4), 247-256.

Statista (2021): Smart Home – Prognose zur Penetrationsrate in Deutschland für die Jahre 2017 bis 2025. Veröffentlichung 06/2021. Abrufbar: https://de.statista.com/prognosen/885655/smart-home-penetrationsrate-in-deutschland (Abgerufen am 29.11.2021).

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., ... & Sörlin, S. (2015): Planetary boundaries: Guiding human development on a changing planet. Science, 347(6223).

Strubell, E., Ganesh, A., & McCallum, A. (2019): Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645-3650).

Su, Z., Togay, G., & Côté, A. M. (2020): Artificial intelligence: a destructive and yet creative force in the skilled labour market. Human Resource Development International, 1-12.

Talens Peiró, L., & Ardente, F. (2015): Environmental footprint and material efficiency support for product policy—Analysis of material efficiency requirements of enterprise servers, JRC-EC (Joint Research Centre—European Commission). *Publications Office of the European Union*.

The Internet Society (2017): Artificial Intelligence and Machine Learning: Policy Paper. 05/2017. Abrufbar: https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/ (Abgerufen am 29.11.2021).

Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2021): Deep Learning's Diminishing Returns: The Cost of Improvement is Becoming Unsustainable. IEEE Spectrum, 58(10), 50-55.

Trajtenberg, M. (2018): AI as the next GPT: a Political-Economy Perspective (No. w24245). National Bureau of Economic Research.

UBA (2019): Künstliche Intelligenz im Umweltbereich. Anwendungsbeispiele und Zukunftsperspektiven im Sinne der Nachhaltigkeit. Kurzstudie. Abrufbar: https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2019-06-04_texte_56-2019_uba_ki_fin.pdf (Abgerufen am 29.11.2021).

Uddin, M., & Rahman, A. A. (2012): Energy efficiency and low carbon enabler green IT framework for data centers considering green metrics. Renewable and Sustainable Energy Reviews, 16(6), 4078-4094.

UNESCO (2021): Draft Text on the Recommendations on Ethics of Artificial Intelligence with Track changes, 18.06.2021.

Uptime Institute (2021): 2021 Uptime Institute Global Data Center Survey. Abrufbar: https://uptimeinstitute.com/2021-data-center-industry-survey-results (Abgerufen am 29.11.2021).

Vallance, S., Perkins, H. C., & Dixon, J. E. (2011): What is social sustainability? A clarification of concepts. Geoforum, 42(3), 342-348.

van Wynsberghe, A. (2021): Sustainable AI: AI for sustainability and the sustainability of AI. AI and Ethics, 1-6.

Vavik, T., & Keitsch, M. M. (2010): Exploring relationships between universal design and social sustainable development: some methodological aspects to the debate on the sciences of sustainability. Sustainable development, 18(5), 295-305.

Vinuesa, R., Azizpour, H., Leite, I. et al. (2020): The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun 11, 233 https://doi.org/10.1038/s41467-019-14108-y.

Vöpel, Henning (2020): Daten und KI revolutionieren Banken, aber nicht das Banking. Zum Einfluss Künstlicher Intelligenz auf den Finanzsektor. Der Bank Blog, Beitrag vom 08.05.2020, abrufbar: https://www.der-bank-blog.de/daten-ki-banken/digital-banking/37662715/ (abgerufen am 06.12.2020)

Vollhardt, S., Schmidt, K., Kask, S., & Noga, M. (2021): Das intelligente Unternehmen: Effiziente Prozesse mit Künstlicher Intelligenz von SAP–Wie Unternehmen die hohen Erwartungen an die KI erfüllen können. In: Künstliche Intelligenz. 119-137. Springer Gabler, Berlin, Heidelberg.

Waltersmann, L., Kiemel, S., Stuhlsatz, J., Sauer, A., & Miehe, R. (2021): Artificial Intelligence Applications for Increasing Resource Efficiency in Manufacturing Companies—A Comprehensive Review. Sustainability, 13(12), 6689.

Wang, R. Y., & Strong, D. M. (1996): Beyond accuracy: What data quality means to data consumers. Journal of management information systems, 12(4), 5-33.

Watney, C. (2018): Reducing entry barriers in the development and application of AI. 10/2018. Abrufbar: https://www.rstreet.org/2018/10/09/reducing-entry-barriers-in-the-development-and-application-of-ai/ (Abgerufen am 29.11.2021).

Whitehead, B., Andrews, D., Shah, A., & Maidment, G. (2014): Assessing the environmental impact of data centres part 1: Background, energy use and metrics. Building and Environment, 82, 151-159.

Wiedmann, T., Lenzen, M., Keyßer, L. T., & Steinberger, J. K. (2020): Scientists' warning on affluence. Nature communications, 11(1), 1-10.

WBGU [Wissenschaftlicher Beirat globale Umweltveränderung] (2009): Kassensturz für den Weltklimavertrag." Der Budgetansatz, Berlin.

GESCHÄFTSSTELLE BERLIN
MAIN OFFICE
Potsdamer Straße 105
10785 Berlin
Telefon: + 49 – 30 – 884 594-0
Fax: + 49 – 30 – 882 54 39

BÜRO HEIDELBERG
HEIDELBERG OFFICE
Bergstraße 7
69120 Heidelberg
Telefon: + 49 – 6221 – 649 16-0

mailbox@ioew.de
www.ioew.de