TEXTE
29/2022

Digitalisierung und Gemeinwohl – Transformationsnarrative zwischen Planetaren Grenzen und Künstlicher Intelligenz

Abschlussbericht

Für Mensch & Umwelt

Umwelt Bundesamt

TEXTE 29/2022

Ressortforschungsplan des Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz

Forschungskennzahl 3718 11 105 0

FB000807

Digitalisierung und Gemeinwohl – Transformationsnarrative zwischen Planetaren Grenzen und Künstlicher Intelligenz

Abschlussbericht

von

Lorenz Erdmann, Kerstin Cuhls, Philine Warnke unter Mitarbeit von Bärbel Hüsing, Meent Mangels, Lia Meißner, Svetlana Meissner, Andreas Röß und Jan Simon
Fraunhofer-Institut für System- und Innovationsforschung ISI, Karlsruhe

Thomas Potthast, Leonie Bossert, Cordula Brand
Ethikzentrum IZEW, Universität Tübingen

Stefanie Saghri, Animation & Illustration Stefanie Saghri

Im Auftrag des Umweltbundesamtes

Impressum

Herausgeber
Umweltbundesamt
Wörlitzer Platz 1
06844 Dessau-Roßlau
Tel: +49 340-2103-0
Fax: +49 340-2103-2285
info@umweltbundesamt.de
Internet: www.umweltbundesamt.de

/umweltbundesamt.de
/umweltbundesamt

Durchführung der Studie:
Fraunhofer-Institut für System- und Innovationsforschung ISI
Breslauer Str. 48
76139 Karlsruhe

Abschlussdatum:
November 2021

Redaktion:
Fachgebiet I 1.1
Martina Eick

Publikationen als pdf:
http://www.umweltbundesamt.de/publikationen

ISSN 1862-4804

Dessau-Roßlau, März 2022

Kurzbeschreibung: Digitalisierung und Gemeinwohl

Im Projekt „Gemeinwohlorientierung im Zeitalter der Digitalisierung“ haben das Fraunhofer-Institut für System- und Innovationsforschung (ISI) und das Ethikzentrum der Universität Tübingen (IZEW) anthropologische und ethische Konzepte mit Fokus auf Künstliche Intelligenz (KI) analysiert und entwickelt, sowie darauf aufbauend neue Einstiege in Narrative für gesellschaftliche Transformationen geschaffen. Zehn Anwendungsfelder für KI, die mit grundlegenden Veränderungen der Mensch-Technik-Umwelt-Beziehungen einhergehen können, wurden identifiziert und beschrieben. Dabei stand die Differenzierung zwischen empirisch beobachtbaren Entwicklungen und spekulativen Visionen im Zentrum der Analyse. Für zwei Themen wurden ethische Analysen durchgeführt: Affective Computing und Autonome Systeme zur Erschließung von für Menschen (bislang) unverfügbaren Räumen. Ethisch relevante zu berücksichtigende Punkte sind (1) KI-basierte Manipulation und Täuschung von Emotionen, (2) Beteiligung und Zugangsgerechtigkeit inklusive der Rolle privater Unternehmen, (3) Extended Reality als besseres Lernen versus Entfremdung durch Entpersonalisierung und Entkörperlichung, (4) Privatheit persönlicher Daten, (5) Natürlichkeit versus Künstlichkeit und Authentizität hinsichtlich Emotionen, (6) Autonome Systeme als Teil der Frontier-Ideologie der nutzenorientierten Technisierung der Natur versus Forderungen nach (partieller) Unverfügbarkeit von Räumen, (7) rein symptomorientierte Techno-Fix-Zugänge versus Suffizienzförderung durch KI, (8) Modifikation von Verantwortungsrelationen durch KI, (9) Dual use beziehungsweise starke militärische Beteiligung und Interessen an KI-Forschung. Anknüpfungspunkte für die ethischen Befunde sind die Arenen Bildung für Nachhaltige Entwicklung und nachhaltiges Verhalten, die Unverfügbarkeit von Räumen im Anthropozän sowie Governance der Technikentwicklung. Mit Storyboards wurde ein Einstieg für neue Narrative zur Verfügung gestellt, der Umweltinteressierte und Medienschaffende dazu anregen soll, sich mit normativen Fragen der KI zu befassen. Der Abschlussbericht schließt mit der Formulierung von aktuellem Forschungsbedarf.

Abstract: Digitisation and Common Good

In the project "Orientation towards the common good in the age of digitalisation", Fraunhofer Institute for Systems and Innovation Research (ISI) and the Ethics Centre of the University of Tübingen (IZEW) analysed and developed anthropological and ethical concepts with a focus on Artificial Intelligence (AI). Based on this, new entry points into narratives for social transformations were created. Ten fields of application for AI, which bring about fundamental changes in human-technology-environment relations, were identified and described. The differentiation between empirically observable developments and speculative visions was at the centre of this analysis. Ethical analyses were carried out for two topics: Affective Computing and Autonomous Systems to access spaces not at humans’ disposition so far. Ethically relevant points to consider are (1) AI-based manipulation and deception of emotions, (2) participation and equity of access including the role of private companies, (3) Extended Reality as better learning versus alienation through depersonalisation and disembodiment, (4) privacy of personal data, (5) naturalness versus artificiality, and authenticity regarding emotions, (6) Autonomous Systems as part of the frontier ideology of a benefit-oriented technification of nature versus demands for (partial) unavailability of spaces, (7) purely symptom-oriented techno-fix approaches versus sufficiency promotion through AI, (8) modification of responsibility relations through AI, (9) dual use or strong military involvement and interests in AI research. Points of contact for the ethical findings are the arenas of education for sustainable development and sustainable behaviour, the normative dimension of indisposable spaces in the Anthropocene, and governance of technology development. Storyboards were used to provide an entry point for new narratives that should also encourage environmentally-concerned people

and journalists to engage with normative questions of AI. The final report concludes with the formulation of current research needs.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

3D	Dreidimensional
5G	Fünfte Mobilfunkgeneration
AC	Affective Computing
ACII	International Conference on Affective Computing and Intelligent Interaction
AGI	Artificial General Intelligence
AI	Artificial Intelligence
AR	Augmented Reality
AS	Autonome Systeme
ASEuR	Autonome Systeme zur Erschließung von (bislang) unverfügbaren Räumen
AUV	Autonome Unterwasservehikel
AV	Augmented Virtuality
BCI	Brain-Computer-Interface
BNE	Bildung für Nachhaltige Entwicklung
DL	Deep Learning
EEG	Elektroenzephalographie
ETF	Exchange-Traded Fund
HFT	High Frequency Trade
HPE	Human Performance Enhancement
IoE	Internet of Everything
IoT	Internet of Things
KI	Künstliche Intelligenz
MISST	Mobile, Imaging, pervasive Sensing, Social media and location Tracking
ML	Maschinelles Lernen / Machine Learning
MTU	Mensch-Technik-Umwelt
NE	Nachhaltige Entwicklung
NLG	Natural Language Generation
NLP	Natural Language Processing
ppm	Parts per million
SCS	Social Credit System
SDGs	Sustainable Development Goals
SIFF	Semi-intelligente Informationsfilter
UAV	Unmanned Aerial Vehicle
VR	Virtual Reality
XAI	Explainable Artificial Intelligence
XR	Extended Reality

Zusammenfassung

Im Projekt „Gemeinwohlorientierung im Zeitalter der Digitalisierung: Transformationsnarrative zwischen Planetaren Grenzen und Künstlicher Intelligenz“ haben das Fraunhofer-Institut für System- und Innovationsforschung (ISI) und das Ethikzentrum der Universität Tübingen (IZEW) anthropologische und ethische Konzepte mit Fokus auf Künstliche Intelligenz analysiert und entwickelt, sowie darauf aufbauend neue Einstiege in sinnstiftende Erzählungen (Narrative) für gesellschaftliche Veränderungsprozesse (Transformationen) geschaffen. Grundlage bildete eine kritische Bestandsaufnahme der Wissensbestände zu Entwicklungen und Wirkungen von Künstlicher Intelligenz. Im Fokus standen ausgewählte Anwendungsfelder von Künstlicher Intelligenz, die derzeitige Beziehungen zwischen Menschen, Technik und Umwelt grundlegend verändern können (disruptive Anwendungen). Bei diesem Vorhaben handelt es sich um ein Projekt der Vorlaufforschung des Umweltbundesamtes.

Teilziele und Forschungsinhalte waren:

- die Identifizierung und Charakterisierung von Digitalisierungsfeldern, die grundlegende Veränderungen der Mensch-Technik-Umwelt-Beziehungen bewirken können (Screening),
- die kritische Reflexion der Narrative über disruptive digitale Technologien sowie die Analyse und Erweiterung der ethischen Fragen hinsichtlich Nachhaltiger Entwicklung und Gemeinwohl in Bezug auf Künstliche Intelligenz (ethische Analyse),
- Einstiege in neue Transformationsnarrative unter Berücksichtigung disruptiver digitaler Technologien (Storyboards),
- Identifizierung von Forschungsbedarf.

Ein inter- und transdisziplinär zusammengesetzter Beirat unterstützte das Vorhaben. Das Projekt lief von November 2018 bis November 2021.

Der Wissenschaftliche Beirat der Bundesregierung Globale Umweltveränderungen sieht in seinem Gutachten *Unsere gemeinsame digitale Zukunft* die Digitalisierung als einen „Brandbeschleuniger bestehender nicht-nachhaltiger Trends“, von ihrem Potenzial her aber als einen Hebel und Unterstützer für die große Transformation. Einschlägige Transformationsnarrative einer nachhaltigen Entwicklung blenden jedoch aus, dass sich die Prämissen von der Verfasstheit der Menschheit (*conditio humana*) in den betrachteten Zeiträumen stark ändern können.

Entwicklungen der Digitalisierung, die die Mensch-Technik-Umwelt-Beziehungen grundlegend verändern können, bezeichnen wir als potenziell disruptiv. Hierunter verstehen wir die Erschütterung von überlieferten Selbstverständnissen in der Gesellschaft, so dass folgende Fragen neu aufgeworfen werden müssen:

1. Was verbindet uns Menschen mit (und trennt uns von) unseren Artefakten, einschließlich Technik?
2. Was macht das Menschsein und Zusammenleben aus?
3. Wo stehen wir in unserer natürlichen Umwelt bzw. in der Natur?

Das Vorhaben fokussiert inhaltlich auf Künstliche Intelligenz, für die es keine einheitliche Definition und zudem unterschiedliche epistemische Auffassungen gibt. Unter Künstlicher Intelligenz wird faktisch eine Sammlung von Technologien gefasst, welche selbständig Aufgaben erledigen können, die normalerweise menschliche Intelligenz erfordern. Künstliche Intelligenz

hat sich im Laufe der Zeit dahingehend entwickelt, dass sie selbstständig auch Operationen ausführen kann, die nicht (mehr) mit menschlicher Intelligenz allein zu bewältigen sind, wie zum Beispiel Big Data Analysen.

Anwendungsfelder von Künstlicher Intelligenz, die grundlegenden Veränderungen der Mensch-Technik-Umwelt-Beziehungen bewirken können

Im Hinblick auf den Aufgabenbereich des Umweltbundesamtes liegt der Fokus dieser Studie nicht auf Künstlicher Intelligenz allgemein oder um deren konkrete Umweltbilanz, sondern um Anwendungsfelder der Künstlichen Intelligenz, die grundlegende Veränderungen von Mensch-Technik-Umwelt-Beziehungen bewirken können. Aufbauend auf einer breiten und systematischen Analyse von Quellen wurden anhand der drei Leitfragen zu grundlegenden Veränderungen der Mensch-Technik-Umwelt-Beziehungen zehn potenziell disruptive Anwendungsfelder von Künstlicher Intelligenz identifiziert und konturiert:

- Affective Computing: Durch computerbasiertes Erkennen und Interpretieren, Auslösen und Erwidern von menschlichen Affekten und Emotionen können (para)soziale Mensch-Maschinen-Beziehungen entstehen, die unsere Vorstellungen vom Menschsein und Zusammenleben grundlegend in Frage stellen und verändern können.
- Simulation natürlicher Sprache: Durch die computergestützte Analyse von natürlicher Sprache und die Inszenierung von technischer Sprache als natürlich wird die Urheberschaft sprachlicher Artefakte, sowohl in mündlicher als auch in schriftlicher Form, kaschiert und öffnet damit Fehlzuschreibungen Tür und Tor (bis hin zu sogenanntem Deep Fake).
- Extended Reality: Die Umgebung der Menschen wird durch Extended Reality um virtuelle Aspekte erweitert (Augmented Reality) oder ersetzt (Virtual Reality). Wahrnehmen und Agieren des Menschen in seiner Umwelt erfolgen damit nicht unmittelbar, sondern medial vermittelt in digital überlagerten räumlichen Umgebungen.
- Digitales Enhancement: Durch die physische Verschmelzung des Menschen mit digitalen Elementen wird Technik zu einem integralen Bestandteil des Verhaltens, wodurch die Grenzen des Körpers und die Zuschreibungen eines Verhaltens als menschlichen oder technischen Ursprungs verschwimmen.
- Autonome Systeme zur Erschließung von für Menschen bislang (weitgehend) unverfügbaren Räumen, beispielsweise der Tiefsee und des Weltraums: Durch diese Erschließung überschreitet der Mensch seine lebensräumlichen Begrenzungen, die er als ein von anderen Lebensformen abhängiges ‚Landtier' bislang hatte.
- Big Data Gesellschaft: In einer Big Data Gesellschaft werden Entscheidungen in zahlreichen gesellschaftlichen Teilsystemen (Recht, Finanzen, etc.) massenhaft anhand von kombinierten sozioökonomischen Daten getroffen, wodurch das damit einhergehende Kalkül und die datengetriebene Vorhersage eine eigenständige normative Kraft entwickeln.
- Hyperkonnektivität: Hyperkonnektivität bezeichnet die ubiquitäre Vernetzung von Menschen, Artefakten und Bestandteilen der Natur in Echtzeit, wodurch die menschliche Fähigkeit zur differenzierten und vielseitigen Kommunikation technisch weitergeführt wird – bis hin zu einer für den Menschen nicht mehr durchschaubaren Komplexität.

- Autonome Systeme im Alltag: Autonome Systeme im Alltag werden selbst aktiv, anstatt dass der Mensch wie bislang Technik jedes Mal dazu veranlasst, ihn bedarfsabhängig zu unterstützen. Dadurch kann sich die autonome Technik schrittweise aus unserem Bewusstsein ‚schleichen', agiert dabei aber im Hintergrund weiter.
- Entschlüsselung des Lebens durch digitale Werkzeuge: Von der Isolierung der Faktoren für die Erklärung des Lebens durch Künstliche Intelligenz (insbesondere Analyse des Genoms, von phänotypischen Eigenschaften und Lebensstilen) versprechen sich ihre Protagonist*innen die Vorhersagbarkeit menschlicher Eigenschaften, Fähigkeiten und Lebenschancen.
- Schwarmintelligenz: Schwarmintelligenz bezeichnet die emergente kollektive Intelligenz einer per Selbstorganisation koordinierten Gruppe lebender Agenten wie Ameisen, Vögel, Fische oder auch Menschen, und – mit Künstlicher Intelligenz – auch technischer Agenten. Dadurch relativieren technische Systeme dieses vormalige Alleinstellungsmerkmal des Lebendigen.

Diese zehn potenziell disruptiven Digitalisierungsfelder zeichnen sich dadurch aus, dass sie wichtige Anwendungsfelder für Künstliche Intelligenz sind, zu grundlegenden Veränderungen der Mensch-Technik-Umwelt-Beziehungen führen können und hinsichtlich der Nachhaltigkeitstransformationen und mit dieser Verknüpfung verbundener ethischer Aspekte einen gewissen neuen Bedeutsamkeitsgrad vermuten lassen. Eine ethische Auseinandersetzung fehlt bislang insbesondere zu Modifikationen im Mensch-Technik-Umwelt-Verhältnis durch Anwendungen der Künstlichen Intelligenz.

Künstliche Intelligenz allgemein sowie diese zehn Anwendungsfelder wurden in einer einheitlichen Art und Weise mit Hilfe gezielter Recherchen und Interviews zu Profilen ausgearbeitet, indem Schlüsselbegriffe erläutert und ein Untersuchungsrahmen abgesteckt, zwischen empirisch beobachtbaren Entwicklungen und spekulativen Visionen unterschieden und die mögliche grundlegende Veränderung von Mensch-Technik-Umwelt-Beziehungen ergänzt wurden. Die Profile der zehn potenziell disruptiven Anwendungsfelder der Künstlichen Intelligenz dienten als Grundlage für die Auswahl von Schwerpunkten in der ethischen Analyse.

Tabelle 1: Entwicklungen und Visionen für Künstliche Intelligenz allgemein und für zehn potenziell disruptive Anwendungsfelder

Digitalisierungsfeld	Entwicklungen	Visionen
Künstliche Intelligenz allgemein	Technische Konjunkturen Normative Anforderungen	Superintelligenz Technischer Posthumanismus
Affective Computing	Erkennen von Affekten und Emotionen Erzeugen von Affekten und Emotionen	Intime Beziehungen mit Robotern Fortleben Verstorbener
Simulation natürlicher Sprache	Kognitive Assistenten Computational Creativity	Autonome Avatare Transformative Kreativität
Extended Reality	Führen im Raum Auslösen von Impulsen im Raum	Nahtlose Augmented Reality
Digitales Enhancement	Digitale Selbstvermessung Brain-Computer-Interfaces	Cyborgs Transhumanismus
Autonome Systeme in bislang unverfügbaren Räumen	Autonome Systeme für den Tiefseebergbau Extraterrestrische Produktion	Blue Economy and Society Space Economy and Society
Big Data Gesellschaft	Scoring und Microtargeting Finanztechnologie	Datengetriebene Gesellschaft Voll ausgebildeter digitaler Überwachungskapitalismus
Hyperkonnektivität	Internet der Menschen & Dinge Internet der Natur	Internet of Everything Dashboard für die Erde
Autonome Systeme im Alltag	Informatorische Systeme Physische Systeme	Verschwinden der Computer aus dem Bewusstsein
Entschlüsselung des Lebens durch digitale Werkzeuge	Genomanalyse Zusammenhang Genotyp und Phänotyp	Entschlüsselung des Alterns Vorhersage der Entfaltung des Lebens von Menschen
Schwarmintelligenz	Drohnenverbünde zur Indoor-Pflanzenbestäubung Drohnenverbünde für Sicherheit und Angriffe	Autonome Drohnenschwärme zur Outdoor-Pflanzenbestäubung Kriegsführung mit autonomen Drohnenschwärmen

Eine anwendungsfeldübergreifende Synthese der sich verändernden Mensch-Technik-Umwelt-Beziehungen legte drei übergreifende Wirkungskomplexe zutage, die teilweise bereits in Philosophie, Technikfolgenabschätzung und Innovationsforschung thematisiert werden, aber im Umweltsektor bislang kaum aufgegriffen worden sind:

- Künstliche Intelligenz verändert die Agency im Mensch-Technik-Umwelt-Gefüge grundlegend.
- Künstliche Intelligenz macht die Natur und ihre Wahrnehmung zunehmend ‚künstlicher'.
- Künstliche Intelligenz wird von Erwartungen an eine Megamaschine flankiert, deren Realisierung zwar unsicher ist, aber die dennoch Weltbilder wirksam verändern.

Für die ethische Analyse wurden zwei Themenkomplexe ausgewählt:

Affective Computing: Dieser Themenkomplex steht auch stellvertretend für die Veränderung der Alltagsumgebung und des Alltagsverhaltens durch Künstliche Intelligenz. Das Feld Affective Computing schließt das Feld Simulation natürlicher Sprache und Teilaspekte aus anderen Digitalisierungsfeldern wie Brain-Computer-Interfaces (Digital Enhancement) mit ein.

Autonome Systeme zur Erschließung von für Menschen bislang (weitgehend) unverfügbaren Räumen: Dieser Themenkomplex steht für Veränderungen der Handlungsspielräume von Menschen in Bezug auf ihre Umwelt im Anthropozän; der Begriff der Unverfügbarkeit hat dabei primär eine empirische aber zugleich eine normative Dimension möglicher bzw. erlaubter Zugänglichkeit. Aspekte aus dem Feld Extended Reality werden hier zum Teil ebenfalls angesprochen.

Ethische Aspekte

Die Frage nach einer „Künstlichen Intelligenz-Ethik" als Bindestrich-Ethik analog zu Umwelt- oder Wirtschaftsethik ist in Narrative der wissenschaftlich-technisch geprägten Welt eingewoben. Unter einem Narrativ kann eine erzählende, moralisch orientierende, vor allem aber sinnstiftende Struktur verstanden werden, die die Welt und darin bestimmte Entwicklungen zu verstehen und einzuordnen hilft. In den letzten Jahrzehnten wurde mit Bezug auf Fortschritte in Wissenschaft und Technik oft davon gesprochen, dass sich ‚ganz neue' ethische Fragen stellen. Auch die Entwicklungen der Digitalisierung und Künstlichen lntelligenz legen solche Narrativ-Momente nahe, denn eine radikal veränderte – gar: posthumane – *conditio humana* scheint eben auch eine neue Ethik zu verlangen.

Wir vertreten allerdings eine andere Position. Der „Tübinger Ansatz" einer Ethik in den Wissenschaften geht davon aus, dass sich in neuen gesellschaftlichen und technischen Konstellationen zwar grundlegende ethische Herausforderungen stellen und es dafür keine einfachen und/oder oft auch keine bewährten Umgangsweisen gibt. Im Falle der Künstlichen Intelligenz sind diese Kontexte neu. Nötig ist hierfür aber keine ‚ganz neue' Ethik, um Künstliche Intelligenz ethisch zu analysieren. Die anwendungsbezogene Ethik hält Methoden, Prinzipien, Normen, Wert oder Tugenden bereit, die auch zur Bewertung von Künstlicher Intelligenz herangezogen werden können.

Wissenschaft und Gesellschaft stehen allerdings vor der Aufgabe, die bestehenden grundlegenden ethischen Maßstäbe in einem neuen Handlungsfeld zu verorten und gleichzeitig bei der Erwägung ethischer Argumente den Akteurskonstellationen und der Geschwindigkeit der Änderungen in diesem neuen Handlungsfeld gerecht zu werden. Dies gilt umso mehr, als dass Ethik eigentlich prospektiv und umfassend problembezogen statt reaktiv und technologieinduziert vorgehen sollte. Es sind die technisch vermittelte Macht und das damit verbundene Gefahrenpotenzial planetaren Ausmaßes beziehungsweise der technischen Eingriffstiefe in Lebensprozesse und -strukturen selbst, die den neuen Kontext bilden. Damit sind die hier vorgestellten Überlegungen auch kompatibel mit aktuellen Arbeiten zur Verantwortung im Anthropozän und bezüglich der Ziele Nachhaltiger Entwicklung der Vereinten Nationen (SDGs).

Für die ethische Analyse wurden (1) grundsätzliche gerechtigkeitstheoretische Erwägungen vorgenommen. Zugleich fanden (2) die in vielen ethischen Debatten angesprochenen Besonderheiten der Abwägung bei Handlungen unter Unsicherheit Berücksichtigung. Für diese müssen Prinzipien wie das *Vorsorgeprinzip* herangezogen werden und im Speziellen auf Künstliche Intelligenz zugeschnitten werden. Weiterhin wurden (3) die in verschiedenen Bereichen der anwendungsbezogenen Ethik diskutierte Themenfelder, wie z. B. Fragen des Konsums, medienethische Herausforderungen, intensiver Ressourcenabbau und Bildungsgerechtigkeit zusammengeführt. Bei aller Antizipation möglicher heute noch

unrealistisch erscheinender Zukünfte muss eine anwendungsbezogene Ethik sich (4) auf einen seriösen Stand des Wissens und der Technik und der Ungewissheit(en) beziehen, um angemessene Beurteilungen entwickeln zu können.

Utopische und Dystopische Narrative

Grundsätzlich finden sich im gesamten Bereich der Künstlichen Intelligenz zwei Standard-Narrative, ein utopisches und ein dystopisches, so auch zu den beiden ausgewählten Themenkomplexen Affective Computing und Autonome Systeme zur Erschließung bislang unverfügbarer Räume.

Zum Thema Affective Computing wird auf der *utopischen* Seite (Entwicklung, Industrie) der Mehrwert für die Nutzer*innen betont. Argumentiert wird hier vor allem so, dass es Menschen glücklicher macht, wenn künstliche Agenten nicht nur nützliche, sondern auch freundliche und respektvolle Gegenüber sind. Kommunikation mit einem Bot, so die Argumentation, wird einfacher, wenn sie uns mehr an zwischenmenschliche Kommunikation erinnert. Weniger anstrengende und enervierende Auseinandersetzungen mit virtuellen Agenten tragen dazu bei, dass wir unsere Ziele schneller und entspannter erreichen, was uns im Alltag zufriedener macht. Geworben wird damit, dass Affective Computing Diskriminierungen minimieren kann, die aufgrund unterschiedlicher kultureller, die Emotionalität beeinflussender Praxen entstehen, indem sie spezifisch auf diese Unterschiede reagiert. Die *dystopische* Seite proklamiert durch Affective Computing den weiteren Verlust eines angenommenen Alleinstellungsmerkmals des Menschen. Die emotionale Verfasstheit von Menschen und ihre besondere Verknüpfung mit ihren rationalen Fähigkeiten, die zu einer einmaligen Urteilsbildungs- und Reflexionsfähigkeit führt, steht, so die Befürchtung, dann nicht mehr nur Menschen, sondern auch nichtmenschlichen Wesen zur Verfügung. In diesem Falle wären Maschinen demnach empathie- und sogar leidensfähig, was als eine Voraussetzung dafür gilt, in die Gruppe der moralisch zu berücksichtigenden Wesen aufgenommen zu werden und entsprechend Rechte in Anspruch nehmen zu können. Weiterhin wird die Gefahr betont, von ‚den Maschinen' nicht nur (emotional) abhängig zu werden – sondern ihnen vollständig unterlegen und unterworfen zu sein.

Auch die Rekapitulation von existierenden Narrativen zur Erschließung von für Menschen bislang unverfügbaren Räumen ergibt ein zweigeteiltes Bild. Auf der einen Seite kursiert das optimistische Narrativ, dass Aktivitäten mit Hilfe von auf Künstlicher Intelligenz basierenden Autonomen Systemen in der Tiefsee zu einem besseren Verständnis und leichterer Erschließung der Tiefsee führen werden, ohne dabei Menschen zu gefährden. Daraus leiten Befürworter*innen die Möglichkeit eines besseren Schutzes der Ozeane und speziell der Tiefsee mitsamt ihrer Lebenswelt ab. Auf der anderen Seite gibt es das pessimistische Narrativ, dass diese Aktivitäten zu einem intensiven Raubbau an der Tiefsee führen und die Tiefsee-Ökosysteme irreversibel, auch aufgrund des nicht mehr kontrollierbaren ‚Eigenlebens' der Autonomen Systeme, zerstören werden.

Die Spannung zwischen ethischen Argumenten und strategisch ausgerichteten Narrativen ist dabei offensichtlich, weil letztere eben nicht im Modus ‚zwanglosen Zwang des (besseren) Arguments wirken. Nicht immer lassen sich beide so sauber auseinanderhalten, wie in dieser idealtypischen Trennung, weil auch ethische Argumentationen narrative Elemente enthalten können. Gleichwohl, dies sei ausdrücklich betont, ist die Unterscheidung von systematischen Argumenten und interessengeleiteter Manipulation durch Narrative weiterhin möglich und nötig.

Offene Diskussionspunkte hinsichtlich Mensch-Technik-Umwelt-Beziehungen sowie Ethik

Hinsichtlich möglicher Veränderungen der Mensch-Technik-Umwelt-Beziehungen lässt sich festhalten, dass diese eher allgemein bleiben müssen, solange über Künstliche Intelligenz im Allgemeinen gesprochen wird. Die beiden vertiefenden Betrachtungen des *Affective Computing* und der *Autonomen System zur Erschließung bislang unverfügbarer Räume* haben aber exemplarisch gezeigt, dass verschiedene Technologien manche Aspekte (z. B. Mensch-Technik im Bereich Affective Computing und Mensch-Umwelt im Bereich der bislang unverfügbaren Räume) besonders stark beeinflussen können.

Für ‚KI im Allgemeinen' lässt sich in Bezug auf die Veränderung der *Mensch-Technik-Beziehung* ausmachen, dass sich das Kompetenzen-Spektrum von Menschen verändern wird, wenn und weil bestimmte bislang nötige Kompetenzen durch Künstliche Intelligenz übernommen und vom Menschen nicht mehr genutzt, dafür aber andere gefördert werden. Ferner lässt sich ein Überschreiten der bisherigen Grenzziehung zwischen Menschen als einzigen Gestaltenden der Technik konstatieren, wenn aus der Technik selbst eine ‚kreative Kraft' wird. In Bezug auf das *menschliche Zusammenleben* ermöglichen KI-Technologien eine Annäherung an globale Gerechtigkeit, wenn sie so gestaltet und zur Verfügung gestellt werden, dass sie die Repräsentation legitimer Anspruchsgruppen fördern und die Inklusion bislang benachteiligter Gruppen gewährleisten können. *Mensch-Umwelt-Beziehungen* ändern sich hinsichtlich der realen und gedachten Unverfügbarkeit von Natur für Menschen, die immer weiter zum Verschwinden gebracht wird, was sich in entsprechende Anthropozän-Narrative einfügen lässt. Es folgt außerdem eine Verschiebung der Verhältnisbestimmung von Virtualität und Realität des Mensch-Umwelt-Bezugs. Die Verschiebung verläuft nicht zwingend in Richtung Entfremdung, sondern es können auch neue, virtuell vermittelte leiblich erfahrbare Bezüge ermöglicht werden. In dieser neuen Welt schwindet die vertraute ‚Einheit in der Verschiedenheit', also eine Welt, in der nicht mehr eindeutig identifizierbar ist, was dem Menschen, der Technik oder der natürlichen Umwelt zuzuschreiben ist.

All die genannten Aspekte müssen in Wissenschaft, Politik und der öffentlichen Debatte intensiv diskutiert werden. Für jeden einzelnen Aspekt steht es aus, auszuloten, welche Mittel- und Langzeitfolgen sie für verschiedene Gesellschaften mit sich bringen, welche Änderungen gewünscht und gewinnbringend sein werden/können, in welchen Zusammenhängen auf die Anwendung bestimmter Künstliche Intelligenz Technologien aus guten Gründen besser verzichtet werden sollte und in welchen Zusammenhängen die Anwendung bestimmter Technologien Künstlicher Intelligenz ethisch unproblematisch oder gar geboten ist.

Zu berücksichtigende Aspekte (Points to Consider)

Aus der ethischen Analyse von Künstlicher Intelligenz und ihrer Auswirkungen auf Mensch-Technik-Umwelt-Beziehungen ergeben sich folgende *Points to Consider* für eine auf Künstliche Intelligenz bezogene anthropologisch informierte Ethik:

1. Große Möglichkeiten bietet die „Fähigkeit" von Technologie, sich selbst durch Lernen an veränderte Situationen und andere Umgebungen anzupassen. Dies ermöglicht unter anderem kosteneffizientere und qualitativ hochwertigere Arbeit in vielen Bereichen. Hierbei gilt es, konkrete Arbeitsbereiche zu identifizieren, für die davon ausgegangen werden kann, dass die Arbeit mit Künstlicher Intelligenz qualitativ hochwertiger als mit Menschen geleistet werden kann, ohne dass dabei potenziell sinnstiftende Arbeit für den Menschen verloren geht.
2. Kritische Fragen nach der Notwendigkeit einer ausgeweiteten Ressourcenprospektion oder räumlichen Ausdehnung durch Autonome Systeme müssen vor dem Hintergrund der Suffizienz gestellt werden. Was ist materiell wirklich nötig für ein gelingendes Leben, was

würde einen problemtischen rein extraktivistischen Naturzugang aufrechterhalten? Es ist zu klären, wie Suffizienz-Maßnahmen – entgegen verbreiteter individueller konsumbetonter Lebensansprüche – politisch angeschoben, ausgestaltet und umgesetzt werden können, so dass nachhaltige Lebensstile vermehrt Eingang finden und welche Rolle Künstliche Intelligenz dabei spielen kann – anstatt mithilfe von Künstlicher Intelligenz nichtnachhaltige Lebensweisen zu ‚reproduzieren' und zu verstärken.

3. Künstliche Intelligenz bietet neue Möglichkeiten für symptomorientierte Problemlösungen, gleichzeitig besteht das Risiko, einen Techno-Fix zu reproduzieren, durch den wiederum von nötigen anderen, an den Ursachen ansetzenden, Maßnahmen abgelenkt wird. Beispielsweise ist die Möglichkeit, mit Hilfe Autonomer Systeme den Abfall nach seiner Einbringung in die Weltmeere und andere Gewässer stark zu reduzieren aus Perspektive der Nachhaltigen Entwicklung zunächst positiv zu bewerten. Es besteht allerdings das Risiko, die Ursachen der Abfallentstehung zu ignorieren und damit nichtnachhaltiges Produzieren und Konsumieren grundsätzlich aufrechtzuerhalten.
4. Eine „*KI made in Europe*" demokratisch und partizipativ zu entwickeln wäre im Sinne Nachhaltiger Entwicklung, wenn die unterstellten ‚europäischen Werte', die implementiert werden sollen, an intra- und intergenerationeller Gerechtigkeit ausgerichtet werden. In diesem Kontext ist zu fragen, inwiefern die Art und Weise der Entwicklung und Regulierung von Künstlicher Intelligenz Entwicklungen in Richtung solcher normativen Desiderata verstärkt oder hemmt. Hierbei gilt es, die *inhärente Struktur neuer Künstliche Intelligenz* Anwendungen so auszugestalten, dass die Technik der Förderung von Fähigkeiten und Gemeinwohl eher dienen können, unter anderem durch Transparenz der Technikentwicklung.
5. Um intra- und intergenerationelle Gerechtigkeit zu erreichen, muss ein ausreichender Zugang zu Anwendungen Künstlicher Intelligenz und anderer digitaler Techniken für alle ermöglicht werden. Gelingt dies, und werden diskriminierende Biases in der Programmierung der Künstlichen Intelligenz vermieden bzw. minimiert – und in jedem Falle transparent gemacht –, würde eine verstärkte Inklusion bislang benachteiligter sowie vulnerabler Gruppen ermöglicht. Dabei bezieht sich der Aspekt des gerechten Zugangs auch auf die Entwicklung der Techniken selbst. Denn umfassende Teilhabe schließt die Teilhabe an Forschung, Entwicklung und ökonomischem Nutzen mit ein. Hierbei gilt es, einen solchen gleichberechtigten Zugang global zu Entwicklung und Nutzung von Künstlicher Intelligenz zu schaffen.
6. Damit zusammenhängend muss der doppelte Bildungsauftrag ernst genommen werden: Es besteht sowohl die Notwendigkeit, alle Menschen weltweit in die Lage zu versetzen, die Künstliche Intelligenz für Bildungszwecke zu nutzen, als auch die Fähigkeit und das Wissen zu vermitteln, Künstliche Intelligenz zu hinterfragen. Zudem besteht die Notwendigkeit, Entwickler*innen zu befähigen, gesellschaftliche und moralische Fragen der Technikentwicklung selbst kritisch reflektieren zu können sowie kommunikationsfähig zu werden, um in einen Austausch mit verschiedenen gesellschaftlichen Gruppen über KI und ihre Entwicklung und Implementierung treten zu können.
7. Hinsichtlich einer Bildung für Nachhaltige Entwicklung, auch speziell für die Natur- und Umweltbildung, ist die Frage nach der Möglichkeit oder dem Verlust leiblich-sinnlicher Naturerfahrung und Praxis, die von Mensch zu Mensch vermittelt wird, von entscheidender Bedeutung. Es gilt weiterhin mit Bezug auf Künstliche Intelligenz auszuloten, welche Verluste leiblich-sinnlicher Naturerfahrung von besonders großer Bedeutung für Psyche, aber auch ‚Geisteshaltung' in Bezug auf Natur der Menschen sind.
8. Hinsichtlich des Inhalts und der Strukturen der öffentlichen Kommunikation werden durch den vermehrten Einsatz von Künstliche Intelligenz Anwendungen qualitative Änderungen

erwartet beziehungsweise zum Teil bereits sichtbar. Diese führen zu Modifikationen von Verantwortungsrelationen, welchen bislang nicht die auf Grund ihrer gesellschaftlichen „Durchdringungstiefe" notwendige Aufmerksamkeit in der wissenschaftlichen Debatte zukommt. Hierbei gilt es, die konkreten Änderungen in der öffentlichen Kommunikation durch Künstliche Intelligenz zu beobachten und zu verstehen, welche Verantwortungswahrnehmungen erodieren und wo eingegriffen werden sollte.

Zusammenfassend lässt sich sagen, dass die vielleicht wichtigsten Themen einer auf Künstliche Intelligenz bezogenen Ethik, die auf Prinzipien der Nachhaltigen Entwicklung und des Gemeinwohls und damit Gerechtigkeitsprinzipien und dem guten Leben beruht, diejenigen Bereiche betreffen, in denen ethische Fragen eng mit philosophisch-anthropologischen Fragen der conditio humana und der Mensch-Technik-Umwelt-Beziehungen verbunden sind:

- *Unverfügbarkeit*: Wie stark wird die lebensweltlich vertraute – aber nicht fixiert und absolut zu denkende – Trennung von Mensch-Technik-Umwelt mittels Künstlicher Intelligenz unterlaufen, so dass sich neue anthropologische Orientierungsfragen nach dem spezifisch Menschlichen und seiner Mitwelt in der technischen Zivilisation stellen, wenn die Bereiche mehr und mehr zusammenfallen? Und – davon zumindest analytisch zu trennen – wo ist dies (hinsichtlich Unverfügbarkeit des Menschlichen oder der Natur) aus welchen Gründen (nicht) wünschenswert? Diese Frage lässt sich nicht beantworten, wenn allein die Folgen von Künstlicher Intelligenz betrachtet werden: Der Zusammenhang von Zielen (Problemstellung und Lösungsoptionen), den aufgewählten Mitteln (und möglichen Alternativen) und von möglichen Folgen und Nebenfolgen muss stets integriert beachtet werden.

- *Bildung als/und kritisches Denken*: Anwendungen der Künstlichen Intelligenz können den Lebenswelt- und Umweltbezug des Lernens radikal modifizieren und virtualisieren. Auch hier stellt sich die Frage danach, was sich im technisch vermittelten Weltbezug substituieren lässt und was dabei verloren geht. Die zweite Frage besteht darin, wie sich in der Konsequenz Formen des kritischen Denkens und ihre Förderung durch Bildung verändern. Und drittens ist die ethische Frage zu stellen, welche Bildungsprozesse vermittelt durch Künstliche Intelligenz wünschenswert sind und welche nicht. Auch hier sind wiederum Ziele, Mittel und Folgen gesamtheitlich zu betrachten.

Nicht zuletzt für die Umweltpolitik sind die beiden Punkte der (Un)Verfügbarkeit und der Bildung für Nachhaltige Entwicklung in *globaler intra- und intergenerationeller* Perspektive zentrale Kriterien für die weitere Gestaltung von KI. Sie kommen in der Frage nach Grenzen der Virtualisierung und Grenzüberschreitung zwischen Menschen, Technik und Umwelt zusammen und schließen so auch an Konzepte einer starken Nachhaltigkeit an. Zusätzlich zur Perspektive einer Substitutionsfähigkeit von Natur als kritische Ressourcenfrage ist die Frage der leiblichen Orientierung und Präsenz von Menschen in ihrer Mitwelt, die sich nicht leicht durch Künstliche Intelligenz ersetzen und virtualisieren lässt, – und/oder lassen sollte –, aufgeworfen.

Anknüpfungspunkte für die Aushandlung der praktischen Umsetzung ethischer Desiderata im Innovationssystem

Aufbauend auf der ethischen Analyse war es das Ziel, mögliche Anknüpfungspunkte für die Denklinien der nachhaltigkeitsethisch informierten, auf Künstliche Intelligenz bezogenen Ethik innerhalb des vorhandenen Regime- und Institutionengeflechts zu identifizieren. Dazu wurden aktuelle Medienbeiträge zu den beiden Vertiefungsthemen *Affective Computing* und *Autonome*

Systeme für die Erschließung von für Menschen bislang unverfügbaren Räumen unter besonderer Berücksichtigung der Ozeane zusammengetragen und auf Themen und Akteure hin analysiert. Ergänzend wurden wissenschaftliche Konferenzen und Reports herangezogen. Auf dieser Basis werden Vorschläge für mögliche Verknüpfungsarenen für die Befunde aus diesem Projekt unterbreitet. Die beiden Bereiche *Affective Computing* und *Autonome Systeme für die Erschließung von für Menschen bislang unverfügbaren Räumen unter besonderer Berücksichtigung der Ozeane* weisen jeweils sehr spezifische Anknüpfungs-Arenen auf. Zugleich lassen sich einige übergreifende Aspekte mit konkreteren Unterthemen herausstellen:

1. Arena Bildung für Nachhaltige Entwicklung und nachhaltigeres Verhalten

- Künstliche Intelligenz als Mittel und Gegenstand von Bildung
- Manipulation durch Künstliche Intelligenz als Herausforderung für Bildung
- Erkenntnisse aus der Bildungsforschung für eine auf Künstliche Intelligenz bezogene Ethik

2. Arena Bislang unverfügbare Räume im Anthropozän

- Rolle unverfügbarer Räume für die menschliche Identitätsbildung
- Rolle von Künstlicher Intelligenz bei der Erschließung bislang unverfügbarer Räume (in und um den Menschen)
- Rolle von Künstlicher Intelligenz für den Naturschutz im Anthropozän

3. Arena Governance der Künstlichen Intelligenz (inclusive Ethik und Standardentwicklung)

- Spezifische ethische Herausforderungen von Affective Computing/Emotional Artificial Intelligence
- Rolle von Künstlicher Intelligenz bei der Erschließung von für Menschen bislang unverfügbaren Räumen und ethische Implikationen hinsichtlich normativ verstandener Unverfügbarkeit

Storyboards zum Einstieg in die Aushandlung neuer Transformationsnarrative

Für die Kommunikation der Befunde aus dem Projekt wurde ein methodisch innovativer Weg beschritten. Ziel war es, mit einer graphischen Repräsentation einen Einstieg in die Formulierung neuer Transformationsnarrative zu schaffen, die aktuelle Entwicklungen aufgreifen, Dilemmata und Interessenkonflikte beleuchten soll, es aber offenlässt, wie die Erzählung weitergeht. Dafür wurden Einstiegserzählungen geschrieben.

Am Anfang stand das Ausprobieren unterschiedlicher Erzählweisen (im Sinne von Abläufen bzw. Drehbüchern) auf der Basis des generierten Wissens und der normativen Debatten zu den beiden Themenkomplexen *Affective Computing in der Bildung* und *Autonome Systeme für die Erschließung der Tiefsee*. Für die Ausarbeitung dieser beiden Einstiege in Erzählungen – hier Storybords genannt – wurden zentrale Entwicklungen der Künstlichen Intelligenz und damit verbundene ethisch relevante Themen gesichtet und einige davon für die Storyboards ausgewählt.

Die Storyboards sollten Alltagsbezug haben und emotional „packend“ sein, sie sollten allgemeinverständlich und für unterschiedliche Zielgruppen, insbesondere Umweltinteressierte und Medienschaffende, einsetzbar sein. Die Storyboards sind Einstiege in „neue“ Erzählungen,

die zwar in der Zukunft spielen und unterschiedliche Probleme anreißen oder Fragen stellen, aber eben deshalb noch nicht zu Ende erzählt sind. Es geht dabei nicht um „richtig“ oder „falsch“ bzw. „positive“ oder „negative“ Zukünfte, sondern darum, sich zu vergegenwärtigen, welche Fragen auf die Menschen zukommen können und wie sie ausgehandelt werden können.

Die beiden Storyboards führen zunächst das Fachthema ein und entwickeln dann den Einstieg in eine alltagsnahe Geschichte. Am Ende stehen heute schon existierende reale Beispiele für die thematisierten Anwendungen Künstlicher Intelligenz sowie offene Fragen, die zur Reflexion und Diskussion einladen, ohne eine normative Position vorwegzunehmen.

Dementsprechend wurden zwei Einstiege in „neue“ Erzählungen geschrieben, bebildert und über ein sogenanntes Scrollytelling umgesetzt. Das Scrollytelling erlaubt es, die Storyboards in der eigenen Geschwindigkeit zu rezipieren. Die Geschichten können als Scrollytelling im Internet ab Januar 2022 angesehen werden: uba-ki-storyboard.de. Mit der Entwicklung von Storyboards wurde ein medialer Einstieg in die Schaffung neuer Narrative für gesellschaftliche Transformationen zur Verfügung gestellt, der auch interessierte Laien dazu anregen soll, sich mit normativen Fragen der Künstlichen Intelligenz zu befassen.

Forschungsbedarf

Das Vorhaben „Gemeinwohlorientierung im Zeitalter der Digitalisierung: Transformationsnarrative zwischen Planetaren Grenzen und Künstlicher Intelligenz“ ist ein Projekt der Vorlaufforschung. Das Themenfeld wurde initial erschlossen, sondiert und zahlreiche ‚offene Enden‘ wurden identifiziert. Hieraus lassen sich wiederum Forschungsbedarfe ableiten, die im Falle transformativer Forschung vom Handlungsbedarf nicht sinnvoll zu trennen sind. Drei übergreifende Forschungsstränge wurden identifiziert:

- die Konzeption einer umfassenden Forschungsprogrammatik im Zeitalter der Digitalisierung und des Anthropozäns unter ausdrücklichem Einbezug anthropologischer und ethischer Aspekte;
- detaillierte inhaltliche Ausgestaltung einer umfassenderen Ethik für das Zeitalter der Digitalisierung und des Anthropozäns unter Einbezug der Umweltperspektive als Teil einer umfassenden Perspektive von Nachhaltiger Entwicklung und Ethik sowie des Gemeinwohls;
- die weitere Zusammenführung von Digitalisierung, Ethik und Umwelt und Nachhaltiger Entwicklung, Gemeinwohl und Zukunftsgestaltung in Form von partizipativen Prozessen, Diskursen und Narrativen.

Summary

In the project ”Orientation towards the Common Good in the Age of Digitalisation: Transformation Narratives between Planetary Boundaries and Artificial Intelligence” the Fraunhofer Institute for Systems and Innovation Research (ISI) and the Ethics Centre of the University of Tübingen (IZEW) analysed and developed anthropological and ethical concepts with a focus on Artificial Intelligence and, building on this, created new entry points into narratives that create meaning for processes of social change (transformations). The basis was a critical stocktaking of the body of knowledge on the developments and effects of artificial intelligence. The focus was on selected fields of application of artificial intelligence that can fundamentally change current relationships between humans, technology and the environment (disruptive applications). This project is a preliminary research project of the German Environment Agency.

Specific objectives and research contents were:

- the identification and characterisation of fields of digitalisation that can bring about fundamental changes in human-technology-environment relations (screening),
- the critical reflection of narratives about disruptive digital technologies as well as the analysis and expansion of ethical questions regarding sustainable development and the common good in relation to artificial intelligence (ethical analysis),
- entry into new transformation narratives thereby considering disruptive digital technologies (storyboards), and
- identification of research needs.

An inter- and transdisciplinary advisory board supported the project. The project ran from November 2018 to November 2021.

In its report *Our Common Digital Future*, the German Advisory Council on Global Change sees digitalisation as an "accelerant of existing unsustainable trends", but in terms of its potenzial as a lever and supporter of the Great Transformation. However, relevant transformation narratives of sustainable development ignore the fact that the premises of the constitution of humankind (*conditio humana*) can change significantly in the time periods under consideration.

We describe developments in digitalisation that can fundamentally change human-technology-environment relations as potenzially disruptive. By this, we mean the unsettling of traditional self-understandings in society, so that the following questions must be raised anew:

1. What connects us humans to (and separates us from) our artefacts, including technology?
2. What constitutes being human and living together?
3. Where do we stand in our natural environment or in nature?

In terms of content, the project focuses on artificial intelligence, for which there is no uniform definition and, moreover, different epistemic understandings. Artificial intelligence is in fact a collection of technologies that can independently perform tasks that normally require human intelligence. Artificial intelligence has developed over time to the extent that it can also independently perform operations that can no longer be handled by human intelligence (alone), such as big data analyses.

Fields of application of artificial intelligence that can bring about fundamental changes in human-technology-environment relations.

In view of the remit of the German Environment Agency, the focus here is not on artificial intelligence in general or on specific environmental ressource aspects, but on fields of application of artificial intelligence that can bring about fundamental changes in human-technology-environment relations. Building on a broad and systematic analysis of sources, and based on the three questions on fundamental changes in human-technology-environment relations, ten potenzially disruptive fields of application of AI were identified and contoured:

- Affective Computing: Through computer-based recognition and interpretation, triggering and responding to human affects and emotions, parasocial human-machine relationships can emerge that can fundamentally question and change our ideas of being human and living together.
- Simulation of Natural Language: Through the computer-assisted analysis of natural language and the staging of technical language as natural, the authorship of linguistic artefacts, both in oral and written form, is concealed and thus opens the door to false attributions (up to and including Deep Fake).
- Extended Reality: The environment of people is extended by virtual aspects (Augmented Reality) or replaced (Virtual Reality). People here no longer perceive and act in their environment directly but mediated by digitally superimposed spatial environments.
- Digital Enhancement: Through the physical fusion of humans with digital elements, technology becomes an integral part of behaviour, blurring the boundaries of the body and the attribution of behaviour as being of human or technical origin.
- Autonomous Systems to access spaces that have been indisposable to humans so far, for example the deep sea and outer space: Through this opening up, humans transcend the limitations of their living space, which they previously had as a "land animal" dependent on other life forms.
- Big Data Society: In a Big Data society, decisions in numerous social subsystems (law, finance, etc.) are made en masse based on combined socio-economic data, whereby the the accompanying calculation and data-driven prediction develop a normative force of their own.
- Hyperconnectivity: Hyperconnectivity refers to the ubiquitous networking of people, artefacts, and components of nature in real time, whereby the human capacity for differentiated and multifaceted communication is technically moved forward - to the point of complexity that is no longer transparent to humans.
- Autonomous Systems in everyday life: Autonomous systems in everyday life become active themselves, instead of humans causing technology to support them each time depending on an actual need, as it has been the case in the past. This allows autonomous technology to gradually ‘creep’ away from our consciousness, while continuing to act in the background.

- Decoding life through digital tools: By isolating the factors for explaining life through artificial intelligence (especially analysis of the genome, phenotypical characteristics, and lifestyles), its protagonists hope to make human characteristics, abilities and life chances more predictable.
- Swarm Intelligence: Swarm intelligence refers to the emergent collective intelligence of a self-organizd group of living agents such as ants, birds, fish or even humans, and – with artificial intelligence – also technical agents. Thereby technical systems further relativize this former characteristic of the living.

These ten potenzially disruptive fields of digitalisation are characterised by the fact that they are important fields of application for artificial intelligence, can lead to fundamental changes in the human-technology-environment relationship, and suggest a certain degree of novel implications with regard to sustainability transformations and ethical aspects. An ethical debate especially on modifications in the human-technology-environment relationship through applications of artificial intelligence has been lacking so far.

Artificial Intelligence in general and these ten fields of application were elaborated as profiles by a shared methodological approach with the help of targeted research and interviews, thereby explaining key terms and defining a framework of investigation, distinguishing between empirically observable developments and speculative visions, and indentifying further possible fundamental changes in human-technology-environment relations. The profiles of the ten potentially disruptive fields of application of AI served as a basis for the selection of focal points in the ethical analysis.

Tabelle 2: Developments and Visions for Artificial Intelligence in general and for ten potentielly disruptive application fields

Digitalisation field	Developments	Visions
Artificial Intelligence in general	Technical ups and down Normative requirements	Superintelligence Technical Posthumanism
Affective Computing	Recognition of affects and emotions Generation of affects and emotions	Intimate relationships with robots Continued life of the deceased
Simulation of Natural Language	Cognitive Assistants Computational Creativity	Autonomous Avatars Transformative Creativity
Extended Reality	Guidance in space Triggering impulses in space	Seamless Augmented Reality
Digital Enhancement	Digital self-measurement Brain-Computer-Interfaces	Cyborgs Transhumanism
Autonomous Systems in hitherto indisposable spaces	Autonomous Systems for deep-sea mining Extraterrestrial production	Blue Economy and Society Space Economy and Society
Big Data Society	Scoring and Microtargeting Financial technology	Data-driven society Full-formed digital surveillance capitalism
Hyperconnektivity	Internet of people & things Internet of nature	Internet of Everything Dashboard for the Earth
Autonomous systems in everyday life	Informational systems Physical systems	Disappearance of computers from consciousness
Decoding of life through digital tools	Genome analysis Relationship between genotype and phenotype	Decoding of ageing Predicting the unfolding of human life
Swarm intelligence	Drone fleet networks for indoor plant pollination Drone fleet networks for security and attack	Autonomous drone swarms for outdoor plant pollination Warfare with autonomous drone swarms

A cross-application synthesis of the changing human-technology-environment relationships revealed three overarching impact complexes, some of which are addressed in philosophy, technology assessment and innovation research, but which have hardly been taken up by the environmental sector to date:

- Artificial intelligence fundamentally changes the agency in the human-technology-environment relationship.
- Artificial intelligence is making nature and its perception increasingly ”artificial”.
- Artificial intelligence is flanked by expectations of a mega-machine, the realisation of which is uncertain, but which nonetheless effectively change worldviews.
- Two thematic complexes were selected for the ethical analysis:

Affective Computing: This complex of topics stands for changes in the everyday environment and everyday behaviour through artificial intelligence. The field of *Affective Computing* includes the field of *Simulation of Natural Language* and partial aspects of other digitalisation fields such as brain-computer interfaces (*Digital Enhancement*).

Autonomous Systems to access spaces that have been indisposable to humans so far: This complex of topics represents changes in the scope of action of humans in relation to their environment in the Anthropocene. Aspects from the field of *Extended Reality* are also addressed here in part.

Ethical aspects

The question of an "artificial intelligence ethics" analogue to, e.g., environmental ethics or business ethics is itself woven into narratives of the world shaped by science and technology. A narrative can be understood as a morally orientating story, but above all a sense-making structure that helps to understand and classify the world and certain developments therein. In recent decades, with reference to advances in science and technology, there has often been talk of "entirely new' ethical questions arising. The developments in digitalisation and artificial intelligence also suggest such narratives, because a radically changed – even posthuman – human condition also seems to demand a new ethics.

However, we take a different position. The "Tübingen approach" to ethics in the sciences and humanities assumes that new social and technical constellations pose fundamental ethical challenges and that often there are no simple and/or established ways of dealing with them. In the case of artificial intelligence, it is precisely these contexts that are new. However, there is no need for a 'completely new ethics' to analyse artificial intelligence, because application-related ethics provides methods, principles, norms, values or virtues that can also be used to evaluate this field of science and technology.

However, science and society are faced with the task to adopt and specify existing, fundamental ethical standards for a new field of practice. The challenge is to adequately take into account the constellations of actors and the speed of change in this new field when considering ethical arguments. This is all the more important because ethics ideally should proceed prospectively and comprehensively in relation to the problem instead of only in a reactive and technology-induced mode. It is the technically mediated power and the associated potential danger of planetary implications, and the technical depth of intervention in life processes and structures themselves that form the new context. Thus, the considerations presented here are also compatible with current work on responsibility in the Anthropocene and especially regarding the Sustainable Delevopment Goals (SDGs) of the United Nations.

For the ethical analysis, (1) fundamental issues of justice theory were considered. At the same time, (2) the particularities of weighing actions under uncertainty, which are addressed in many ethical debates, were considered. For this, principles such as the precautionary principle must be adopted to and specified for artificial intelligence. Furthermore, (3) the topics discussed in various areas of application-oriented ethics, such as consumption, media-ethical challenges, resource depletion, and educational justice, were brought together. With all due anticipation of possible futures that seem unrealistic today, application-related ethics must refer (4) to a serious state of knowledge and technology, and uncertainty(ies) to allow for developing appropriate assessments.

Utopian and Dystopian Narratives

Basically, two standard narratives can be found in the entire field of artificial intelligence, a utopian and a dystopian one, as is also the case with the two selected topics of affective computing and autonomous systems for accessing hitherto indisposbale spaces.

About affective computing, the utopian side (developers, industry) emphasises the added value for users. The argument is that people are happier when artificial agents are not only useful but also friendly and respectful counterparts. Communication with a bot, the argument continues, becomes easier when it reminds us more of interpersonal communication. Less stressful and enervating arguments with virtual agents help us to achieve our goals faster and in a more relaxed way, which makes us more satisfied in everyday life. It is advertised that affective computing can minimise discriminations that arise due to different cultural practices affecting emotionality by specifically responding to these differences. The dystopian side ciritises the further loss of an alleged unique human characteristic through affective computing. The emotional constitution of humans and its special link with their rational abilities, which leads to a unique ability to form judgements and reflect, will then, so the fear, no longer be available only to humans, but also to non-human beings. In this case, machines would be capable of empathy and even suffering, which is considered a prerequisite for being included in the group of beings to be considered morally and to be able to claim rights accordingly. Furthermore, the danger of not only becoming (emotionally) dependent on “the machines” but of being completely inferior to and hence dominated by them is emphasised.

The recapitulation of existing narratives on the access of spaces hitherto indisposable to humans also yields a dichotomous picture. On the one hand, there is the optimistic narrative that activities carried out in the deep sea with the help of autonomous systems based on artificial intelligence will lead to a better understanding and accessing of the deep sea without endangering people. As a result, proponents suggest better protection of the oceans and especially the deep sea and its life forms. On the other hand, there is the pessimistic narrative that these activities will lead to intensive overexploitation of the deep sea and irreversibly destroy deep-sea ecosystems, also because of unstoppable Autonomous Systems.

The tension between ethical arguments and strategically oriented narratives is obvious because the latter do not work in the mode of 'unconstrained compulsion of the (better) argument'. It is not always possible to distinguish between the two as clearly as in this ideal-typical separation, because ethical argumentation can also contain narrative elements. Nevertheless, it should be explicitly emphasised that the distinction between systematic arguments and interest-driven manipulation through narratives is still possible and necessary.

Open discussion points regarding human-technology-environment relations and ethics

Regarding possible changes in human-technology-environment relations, of course their identification remains on a rather general level as long as artificial intelligence is being discussed in general. However, the two in-depth considerations of affective computing and autonomous systems for opening up indisposable spaces have shown exemplary that different technologies can particularly influence certain aspects (e.g., human-technology in the area of affective computing and human-environment in the area of hitherto indisposable spaces).

For AI in general and the change in the human-technology relationship, it can be expected that the spectrum of practiced human competencies will change if and because certain competencies are taken over by artificial intelligence and are no longer executed by humans, but others are promoted instead. Furthermore, the previous boundary between humans as the sole creators of technology will be crossed when technology itself becomes a 'creative force'. In terms of human coexistence, AI technologies enable an approach to global justice if they are designed and made available in such a way that they can promote the representation of legitimate stakeholders and ensure the inclusion of previously disadvantaged groups. Human-environment relations are changing in terms of the real and imagined unavailability of nature to humans, which is increasingly being made to disappear, fitting into corresponding Anthropocene narratives. It

also follows a shift in the determination of the relationship between virtuality and reality of the human-environment relationship. The shift is not necessarily only in the direction of alienation, but can also enable new, virtually mediated, bodily experienceable relationships. In this new world, the familiar 'unity in diversity' is fading, i.e., a world in which it is no longer possible to clearly identify what can be attributed to humans, technology, or the natural environment.

All the aspects mentioned need to be discussed thoroughly in science, politics, and public debate. For each individual aspect, it remains to be sounded out which medium and long-term consequences they entail for different societies, which changes will/could be desired and profitable, in which contexts the application of certain AI technologies would be better dispensed with for good reasons, and in which contexts the application of certain AI technologies is ethically unproblematic or even demanded.

Points to Consider

From the ethical analysis of artificial intelligence and its impact on human-technology-environment relations, the following Points to Consider arise for an anthropogically informed ethics related to artificial intelligence:

1. Great opportunities are offered by the 'ability' of technology to adapt itself to changing situations and different environments through learning. This enables, among other things, more cost-efficient and higher quality work in many areas. Here, it is important to identify the specific areas of work for which it can be assumed that work by AI can be done in a higher quality than by humans, without losing meaningful work for humans.
2. Critical questions about the necessity of expanded resource prospecting or spatial expansion by autonomous systems must be asked against the background of sufficiency. What is, not least regarding resources, really necessary for a good life, what would merely support unsustainable extractivist lifestyles? It must be elaborated how sufficiency measures can be politically initiated, designed, and implemented against the widespread consumerist lifestyle demands, so that sustainable ways of life are increasingly adopted – instead of 'reproducing' and increasing unsustainable lifestyles with the help of artificial intelligence – and what role artificial intelligence can play in this.
3. Artificial intelligence offers new opportunities for symptom-oriented problem solving, but at the same time there is a risk of reproducing a techno-fix in some cases, which in turn distracts from other measures directed to the causes of the problem. For example, the possibility of using autonomous systems to greatly reduce waste after it has been discharged into the world's oceans and other bodies of water is at first glance positive from the perspective of sustainable development. However, there is a risk of ignoring the causes of waste generation and thus fundamentally perpetuating unsustainable production and consumption.
4. Developing an "AI made in Europe" in democratic and participaroty ways would be in the spirit of sustainable development if the assumed 'European values' to be implemented were aligned with intra- and intergenerational justice. It must be asked to what extent the way AI is developed and regulated: whether it strengthens or inhibits developments towards such normative desiderata. In this context, it is important to design the inherent structure of new artificial intelligence applications in such a way that the technology is more likely to serve the promotion of capabilities and the common good (among other things, through transparency in and of technology development).
5. To achieve intra- and intergenerational equity, sufficient access to AI applications and other digital technologies must be made possible for all. If this succeded, and if discriminatory biases in the programming of AI are avoided or minimised – and in any case made transparent – this would enable greater inclusion of previously disadvantaged and

vulnerable groups. The aspect of equitable access also relates to the development of the technologies themselves. For comprehensive participation includes participation in research, development, and economic benefitsharing. In this context, it is imperative to create such equal access to the development and use of artificial intelligence on a global scale.

6. In connection with this, the dual educational task must be taken seriously: There is both the need to enable all people wordwide to use AI for educational purposes and the need to impart the ability and knowledge to question AI. In addition, there is a need to enable developers to critically reflect on social and moral issues of technology development themselves and to become able to communicate for engaging with various societal stakeholders on the development and implantation of AI.
7. Regarding Education for Sustainable Development, also specifically for nature and environmental education, the question of the possibility or loss of bodily-sensual experience of nature and practice, which is conveyed in shared practices from person to person, is of decisive importance. It is still necessary to explore which losses of bodily-sensual experience of nature are of particularly great significance for the psyche, but also 'mindset' in relation to people's nature regarding AI.
8. Concerning content and structures of public communication, qualitative changes are expected or already partly visible through the increased use of artificial intelligence applications. These lead to modifications of responsibility relations, which have not yet received the necessary attention in the scientific debate due to their social 'penetration depth'. In this context, it is important to observe the concrete changes in public communication through artificial intelligence and to understand which perceptions of responsibility are eroding and where intervention should take place.

Summing up, perhaps the most important issues of an AI-related ethics based on principles of sustainable development and the common good, i.e., principles of justice and the good life, concern areas where ethical questions are closely linked to philosophical-anthropological questions of the human condition and of human-technology-environment relations:

- *Indisposability*: To what extent is the separation of human-technology-environment – which is familiar in lifeworld but cannot be thought of as fixed and absolute – undermined by (applications of) artificial intelligence? Do new anthropological questions of orientation arise about the specifically human and its co-environment in technical civilisations when the separation of the areas is increasingly blurred? And – to be separated from this at least analytically – where is the indisposability, e.g., of the human sphere or the natural sphere (not) desirable and for what reasons? This question cannot be answered if only the consequences of artificial intelligence are considered: The context of goals (problem definition and solution options), the choice of means (and possible alternatives) and of possible consequences and side-effects must be considered in an integrated way.

- *Education as/and critical thinking*: Artificial intelligence applications can radically modify and virtualise the lifeworld and environmental reference of learning. Here, too, the question arises as to what can be substituted in the purely technically mediated reference to the world and what is lost in the process. The second question is how forms of critical thinking and their promotion through education will change as a consequence. And thirdly, there is the ethical question of which educational processes mediated by artificial intelligence are

desirable and which are not. Here, too, the goals, means and consequences must be considered in an integrated way.

Not least for environmental policy, the two points of indisposability and education for sustainable development in a global intra- and intragenerational perspective are central criteria for the further design of Artificial Intelligence. They come together in the question of limits to virtualisation and boundary crossing between people, technology and the environment and thus also connect to concepts of strong sustainability. In addition to the limited substitutability of nature as a critical resource issue, also the question of bodily orientation and integration of people in their living environment is at stake, which cannot – and maybe may not – easily be replaced and virtualised by artificial intelligence.

Starting points for negotiating the practical implementation of ethical desiderata in the innovation system

Building on the ethical analysis, the aim was to identify possible points of contact for the lines of thought of an ethics related to artificial intelligence informed by sustainability within the existing network of regimes and institutions. To this end, current media contributions on the two subjects of *Affective Computing* and *Autonomous Systems to access spaces that have been indisposable to humans so far*, with a special focus on the oceans, were compiled and analysed in terms of topics and actors. In addition, scientific conferences and reports were consulted. On this basis, suggestions are made for possible linking arenas for the findings from this project. The two selected subjects have very specific connection arenas. At the same time, some overarching aspects with more concrete sub-themes can be identified:

1. Arena Education for Sustainable Development and Sustainable Behaviour

Addressing artificial intelligence as a means and subject of education

- Addressing manipulation through artificial intelligence as a challenge for education
- Insights from educational research for an ethics related to artificial intelligence

2. Arena Indisposable spaces in the Anthropocene

- The role of indisposable spaces for human identity formation
- The role of artificial intelligence in the development of indisposable spaces (in and around humans)
- Role of artificial intelligence for nature conservation in the Anthropocene

3. Arena *Governance of Artificial Intelligence* (including ethics and standards development)

- Specific ethical challenges of Affective Computing/Emotional Artificial Intelligence
- Role of artificial intelligence for opening up spaces previously indisposable to humans and ethical implications with regard to normatively understood unavailability

Storyboards to start negotiating new transformation narratives

A methodologically innovative approach was taken to communicate the findings of the project. The aim was to create an entry point into the formulation of new transformation narratives with a graphic representation that would pick up on current developments, illuminate dilemmas and

conflicts of interest, but leave it open how the narrative would continue. New entry narratives were written for this purpose.

The first step was to try out different types of narrating (in the sense of sequences or scripts) based on the knowledge generated and the normative debates on the two thematic complexes of *Affective Computing in education* and *Autonomous Systems to access spaces that have been indisposable to humans so far*, using the example of the deep sea. For the elaboration of these two entries into narratives –called "storyboards" here – central developments in Artificial Intelligence and related ethically relevant topics were screened and some of them were selected for the storyboards.

The storyboards should have relevance to everyday life and be emotionally 'enthralling'; they should be generally understandable and applicable for different target groups, especially for the environmental sector, and for environmentally-concerned people and journalists. The storyboards are introductions to 'new' narratives that are set in the future and touch on different problems or pose questions but are therefore not yet finished. It is not about "right" or "wrong" or "positive" or "negative" futures, but about illustrating what questions people are likely to face and how they can be negotiated.

The two storyboards first introduce the subject matter and then develop the entry into a story that is close to everyday life. At the end, there are real examples that already exist today for the thematised applications of Artificial Intelligence as well as open questions that invite reflection and discussion without pre-empting a normative position.

Accordingly, two entries into 'new' narratives were written, illustrated, and implemented via so-called "scrollytelling". Scrollytelling allows the storyboards to be received at one's own speed. The stories can be viewed as scrollytelling on the internet from January 2022 on: uba-ki-storyboard.de. With the development of storyboards, a media entry point into the creation of new narratives for social transformations has been made available, which should also encourage interested laypeople to engage with normative questions of Artificial Intelligence.

Need for research

The project "Orientation towards the common good in the age of digitalisation: transformation narratives between planetary boundaries and artificial intelligence" is a project of "forward planning research". The topic area was 'unlocked', explored, and numerous open issues were identified. From this, in turn, research needs can be derived, which in the case of transformative research cannot be meaningfully separated from the practical need for action. Three overarching research strands were identified:

- the conception of a comprehensive research agenda in the Age of Digitalisation and the Anthropocene, explicitly including anthropological and ethical aspects;
- a detailed unfolding of an encompassing ethics for the Age of Digitalisation and the Anthropocene, incorporating the environmental and sustainable development as well as the common good perspectives;
- the further merging of digitalisation, ethics, environmental/sustainability/common good perspectives for contributing options of shaping the future in the form of participatory processes, discourses, and narratives.

1 Einleitung

1.1 Ausgangslage

Unter dem Digitalisierungsprozess werden im Wirtschaftsleben auf der Mikroebene die **Digitalisierung** von Geschäftsmodellen und Geschäftsprozessen gefasst (Bitkom 2016), zum anderen aber die tiefgreifende Umgestaltung der Gesellschaft als – je nach Lesart emergente oder gestaltete – Transformation, die sich aus dem Zusammenwirken von gesellschaftlichen und Digitalisierungsprozessen auf der Makroebene ergibt (WBGU 2019b). Die Digitalisierung ist auf dem Weg, nahezu alle Lebensbereiche des Menschen zu durchdringen, so dass sie als ein wesentlicher Motor des technisch-sozialen Wandels in der jüngeren Moderne gelten kann (Bundesregierung 2018a). Aktuell stehen digitale Technologiefelder wie Blockchain, Cloud-Computing, Extended Reality, das Internet der Dinge, Robotik, Big Data und Künstliche Intelligenz in der öffentlichen Debatte (Gotsch und Erdmann 2018). Die künstliche Intelligenz ist weder mit der Digitalisierung identisch noch für jede Anwendung wie zum Beispiel Blockchain oder Robotik zwingend erforderlich.

Für **Künstliche Intelligenz** gibt es keine einheitliche Definition und zudem unterschiedliche epistemische Auffassungen.[1] Unter Künstlicher Intelligenz (KI) wird faktisch eine Sammlung von Technologien gefasst, welche selbständig Aufgaben erledigen können, die normalerweise menschliche Intelligenz erfordern (vgl. u. a. (Lenzen 2002b) (Mainzer 2016b), (Grunwald 2019)). Künstliche Intelligenz wurde im Laufe der Zeit dahingehend entwickelt, dass sie auch selbstständig Operationen ausführen kann, die nicht (mehr) mit menschlicher Intelligenz zu allein zu bewältigen sind, wie zum Beispiel Big Data-Analysen. Wesentliche Fortschritte der KI-Entwicklung wurden in den letzten Jahren auf dem Gebiet des Maschinellen Lernens erzielt. Durch Steigerung der Leistungsfähigkeit der Computer-Hardware und die Verfügbarkeit großer Datenbestände konnten Analyse und Erkennung von Mustern in Datensätzen mittels tiefer neuronaler Netzwerke entscheidend verbessert werden (Hao 2019). Beim Maschinellen Lernen (ML) verändern sich die digitalen neuronalen Netze als Folge der Interaktion mit der Umgebung ohne weitere menschliche Intervention. Das Potenzial von ML zur Entscheidungsunterstützung ist von digitalen Plattformunternehmen erkannt und in ‚maßgeschneiderte' Internet-Inhalte und -Services umgesetzt worden, darunter Sprachassistenten für Smartphones und in Empfehlungssystemen von digitalen Handelsplattformen. Internet-User und Zivilgesellschaft haben hierauf uneinheitlich reagiert: von der Begrüßung der neuen Services bis hin zur kritischen Beobachtung von algorithmischer Entscheidungsfindung (vgl. u. a. AlgorithmWatch).

Der Beginn des Projektes „Gemeinwohlorientierung im Zeitalter der Digitalisierung: Transformationsnarrative zwischen Planetaren Grenzen und Künstlicher Intelligenz" fiel im November 2018 zeitlich nahezu mit der Veröffentlichung der *Strategie Künstliche Intelligenz* der Bundsregierung (Bundesregierung 2018b) zusammen. Die Bundesregierung fasst KI darin als Basisinnovation auf, die zum Treiber der Digitalisierung und autonomer Systeme in allen Lebensbereichen wird (Bundesregierung 2018a, 2018b). Die Bundesregierung will mit ihrer *Strategie Künstliche Intelligenz* „die Voraussetzungen zur Nutzung der Chancen und des Potenzials von KI schaffen", wobei sie jedoch bestrebt sei, „die KI in sämtlichen Politikfeldern aktiv im Sinne einer menschenzentrierten, gemeinwohlorientierten Nutzung für Wirtschaft und Gesellschaft auf Basis der demokratischen Grundordnung mitzugestalten" (Bundesregierung 2018b). Die öffentliche Debatte stand eine Weile ausschließlich im Zeichen einer **Technikeuphorie**, die sich auf die großen Chancen von KI kapriziert. Beispielsweise widmete

[1] Der Begriff der Intelligenz deckt ein weites Feld ab und ist zudem umstritten (Stephan und Walter 2013). Ungeachtet dessen schreiben Menschen trainierten Algorithmen Intelligenz zu (Daum 2019b, S. 27).

sich der Digital-Gipfel 2018 in Nürnberg dem Schwerpunkt *Künstliche Intelligenz – ein Schlüssel für Wachstum und Wohlstand*[2], ohne dass der kritischen Zivilgesellschaft und entsprechenden Forschungsbeiträgen dort eine Stimme eingeräumt wurde.

Die Regierungen moderner Gesellschaften und innovative Unternehmen sind unter den Bedingungen des Wettbewerbs grundsätzlich daran interessiert, mächtige Werkzeuge wie KI für ihre Ziele und Zwecke einzusetzen. Laut Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen (WBGU) handelt es sich beim Methodenarsenal der KI „möglicherweise um die mächtigsten Werkzeuge, die jemals von unserer Zivilisation angefertigt wurden (WBGU 2019b). Der Werkzeuggebrauch durch bestimmte Menschen und Menschengruppen in bestimmter Art und Weise wirft normative Fragen auf, wer davon profitiert beziehungsweise davon betroffen ist oder welche Chancen und Risiken damit einhergehen (Grunwald 2019). Technikgestaltung und Hoheit über den Technikeinsatz sind also von wesentlicher Bedeutung dafür, wem oder was der Technikeinsatz zukünftig nützt oder schadet. Damit ist der KI notwendig eine **politische Dimension** eingeschrieben (Müller-Mall 2020; Ramge 2020). Auch vor diesem Hintergrund wurde von der Bundesregierung die Enquete Kommission Künstliche Intelligenz ins Leben gerufen, die nach zweijähriger Arbeit im Oktober 2020 ihren Abschlussbericht präsentierte (Enquete-Kommission Künstliche Intelligenz 2021). Auch der Digitalgipfel 2021 (Online) thematisierte neben den Chancen von KI auch deren Risiken und Gestaltungspotenziale.[3]

Zweifelsohne gibt es eine Reihe von gemeinwohlorientierten Einsatzszenarien für KI, beispielsweise den Einsatz autonomer Systeme im Katastrophenfall, die Vorhersage der Ausbreitung von Epidemien, das Erkennen von Risiken für die Stabilität des Finanzsystems und konkrete KI-Anwendungen für den Umweltschutz. Dennoch wird KI in Deutschland bislang in weitaus dominierendem Maße zur Förderung des ‚Privatwohls' von Unternehmen und Konsument*innen eingesetzt, so zum Beispiel in Form von KI-generierten Kaufempfehlungen auf digitalen Plattformen. Die Digitalisierung ist für die sich formierenden Akteure ein **Gelegenheitsfenster**, um Machtpositionen neu abzustecken und Einfluss auf Regulierungen zu nehmen. Dies bedeutet nicht, dass die vorherige analoge Welt gemeinwohlorientiert gewesen ist. Die Digitalisierung trifft aber auf fruchtbaren Boden, indem sie strukturelle Muster der vorhersehbaren Interaktion von Menschen aufgreift, wie zum Beispiel Individualismus oder die Trennung von Produktion und Konsum (Nassehi 2019).

Auch extreme, unsichere oder widersinnige Technikvisionen können mit ihren Versprechen **Transformationsdiskurse** bestimmen, selbst wenn sie – wie möglicherweise die sogenannte „Superintelligenz" (Bostrom 2017; Harari 2016) – nicht realisierbar sein sollten; (Mieth 2002, 2002) spricht bei solchen Konstellationen von der normativen Kraft des Fiktionalen. Dessen ungeachtet blenden die einschlägigen Transformationsnarrative einer nachhaltigen Entwicklung (WBGU 2011; Brohmann und David 2015; Schnurr et al. 2018) jedoch aus, dass sich die Prämissen von der Verfasstheit der Menschheit in den betrachteten Zeiträumen deutlich verändern können (WBGU 2019b). Angesichts dieser bislang wenig aufeinander bezogen Diskursrichtungen stellt sich die Frage nach Transformationsverständnissen, die erstens Digitalisierungsdynamiken aufgreifen und zweitens im Hinblick auf nachhaltige Entwicklung fruchtbar machen ((Muraca 2020), (WBGU 2019b)). Der WBGU sieht die Digitalisierung in seinem Gutachten *Unsere gemeinsame digitale Zukunft* als einen „Brandbeschleuniger bestehender nicht-nachhaltiger Trends", von ihrem Potenzial her aber als einen Hebel und Unterstützer für die große Transformation ((WBGU 2019a), S. 1). Die große Transformation zur

[2] vgl. https://www.de.digital/DIGITAL/Redaktion/DE/Dossier/digital-gipfel-2018.html

[3] vgl. https://www.de.digital/DIGITAL/Navigation/DE/Konferenzen/konferenzen.html

Nachhaltigkeit könne nur gelingen, wenn sie unter den sich wandelnden Bedingungen des digitalen Zeitalters – Vernetzung, Kognition, Autonomie, Virtualität und Wissensexplosion – stattfindet (WBGU 2019a), S. 9 f.).

Der WBGU unterscheidet drei Dynamiken des digitalen Zeitalters (WBGU 2019a):

1. In der ersten Dynamik *Digitalisierung für Nachhaltigkeit* liegt der Blick auf der Umsetzung der Sustainable Development Goals bis 2030.
2. In der zweiten Dynamik *Nachhaltige digitalisierte Gesellschaften* geht es um die Nutzbarmachung und Einhegung der digitalen Kräfte durch Gesellschaften.
3. In der dritten Dynamik *Die Zukunft des Homo Sapiens* wird der Mensch selbst durch die digitale Revolution verändert.

Sind die erste und die zweite Dynamik bereits Gegenstand intensiver gesellschaftlicher Debatten, so wird die dritte Dynamik nur in begrenzten, meist akademischen Zirkeln verhandelt.

An dieser Ausgangslage setzt das Projekt „Gemeinwohlorientierung im Zeitalter der Digitalisierung: Transformationsnarrative zwischen Planetaren Grenzen und Künstlicher Intelligenz“ an.

1.2 Projektziele und -hintergrund

Das **Projekt** „Gemeinwohlorientierung im Zeitalter der Digitalisierung: Transformationsnarrative zwischen Planetaren Grenzen und Künstlicher Intelligenz“ analysiert und entwickelt ethische Konzepte und sinnstiftende Erzählungen (Narrative) für gesellschaftliche Veränderungsprozesse (Transformationen) unter ausdrücklicher Berücksichtigung von solchen digitalen Technologien, die die derzeitigen Beziehungen zwischen Mensch, Technik und Umwelt grundlegend verändern können (disruptive digitale Technologien), insbesondere Künstlicher Intelligenz (KI).

Teilziele und Forschungsinhalte waren:

- die Identifizierung und Charakterisierung von Digitalisierungsfeldern, die grundlegende Veränderungen der Mensch-Technik-Umwelt-Beziehungen bewirken können (Screening),
- die kritische Reflexion der Narrative über disruptive digitale Technologien sowie die Analyse und Erweiterung der ethischen Fragen hinsichtlich Nachhaltiger Entwicklung und Gemeinwohl in Bezug auf Künstliche Intelligenz (ethische Analyse),
- Einstiege in neue Transformationsnarrative unter Berücksichtigung disruptiver digitaler Technologien (Storyboards),
- Identifizierung von Forschungsbedarf.

Ein inter- und transdisziplinär zusammengesetzter Beirat unterstützte das Vorhaben. Das Projekt lief von November 2018 bis November 2021.

Chancen und Risiken von KI werden inzwischen öffentlich ausgiebig diskutiert. Hierbei nehmen einzelne Themen wie die Auswirkungen von KI auf den Arbeitsmarkt, Datenschutzaspekte und die Macht einzelner Akteure in der Datenökonomie einen großen Raum ein. Einige KI-Anwendungen haben das Potenzial, Stoffströme und Eingriffe in Ökosysteme so zu verändern, dass ein Wirtschaften innerhalb der planetaren Grenzen erleichtert werden könnte. Auch die Bereitstellung, der Betrieb und die Entsorgung der KI-Software und der KI-Hardware selbst

verursachen Stoffströme und Eingriffe in Ökosysteme. Die Umweltbilanz dieser gegenläufigen Effekte ist unbekannt und auch **nicht Gegenstand dieses Vorhabens**.

In diesem Vorhaben geht es schwerpunktmäßig um Entwicklungen der Digitalisierung, die die **Mensch-Technik-Umwelt-Beziehungen** grundlegend verändern können. Solche Entwicklungen bezeichnen wir als potenziell disruptiv.[4] Hierunter verstehen wir das Rütteln an den überlieferten Selbstverständnissen in der Gesellschaft, so dass folgende Fragen neu aufgeworfen werden müssen:

- Was verbindet uns mit (und trennt uns von) unseren Artefakten, einschließlich Technik?
- Was macht das Menschsein und Zusammenleben aus?
- Wo stehen wir in unserer natürlichen Umwelt bzw. in der Natur?[5]

Solche Fragen werden in der nachhaltigkeitsbezogenen **Transformationsforschung** nicht explizit adressiert (vgl. Feola 2015), auch wenn sie die dort benannten Erfolgsfaktoren für Transformation maßgeblich beeinflussen können (u. a. Wertewandel). Gerade im Hinblick auf die Formulierung von Narrativen für die Umweltpolitik (vgl. u. a. (Marscheider-Weidemann et al. 2016); (Espinosa et al. 2017)) scheint ein Aufgreifen insbesondere von KI deshalb überzeugend zu sein, weil es sich erstens um eine potenziell disruptive Technologie handelt, die für zukünftige Transformationen als zentral angesehen wird und zweitens, weil sie sich – entgegen technikdeterministischer Vorstellungen – sehr wohl gestalten lässt.

Im Hinblick auf globale Umweltveränderungen wird meistens auf das Konzept der planetaren Grenzen rekurriert (Rockström et al. 2009). Im Hinblick auf die grundlegenden Veränderungen der Mensch-Technik-Umwelt-Beziehungen durch KI reicht das Set an Indikatoren für die planetaren Grenzen nicht zur Qualifizierung der wesentlichen Transformationsdynamiken aus, sondern es ist – darin besteht weitgehende Einigkeit in der wissenschaftlichen Debatte um Nachhaltige Entwicklung – Teil eines um gesellschaftliche und kulturelle Dimensionen erweiterten normativen Bezugsrahmens für die mit diesen Veränderungen einhergehenden **ethischen Aspekte**, insbesondere unter Berücksichtigung des Gemeinwohls und des Kerns der nachhaltigen Entwicklung, der intra- und intergenerationalen Gerechtigkeit. Solch eine Erweiterung des Planetare Grenzen-Konzepts findet sich beispielsweise im *Safe and Just Operating Space* (z. B. (Raworth 2018)). Weitere Elemente eines normativen Bezugsrahmens im vorliegenden Projekt werden an entsprechender Stelle identifiziert.

In der **populärwissenschaftlich aufgemachten Literatur** zu KI dominieren die weitreichenden Konsequenzen extremer Zuspitzungen. Zum Beispiel behauptet (Harari 2016), dass die Koevolution von Biotechnologie und Informationstechnologie nur zu drei möglichen Zukunftsszenarien führen kann: (1) die technisch-medizinische Schaffung des Übermenschen durch Bioengineering, (2) das Verschmelzen des Menschen mit künstlichen Neuronalen Netzen durch Cyborgisierung oder (3) die Abschaffung des Menschen durch die Eingliederung der

[4] Die grundlegenden Eigenschaften von Mensch-Technik-Umwelt-Beziehungen und die Einzigartigkeit des Menschen sind Gegenstand der Anthropologie. Wichtige Strömungen sind unter anderem die philosophische und die evolutionäre Anthropologie. Hellmut Plessner als einer der zentralen Vertreter der philosophischen Anthropologie hat drei „Gesetze" der conditio humana ausgemacht: (1) Das Gesetz der natürlichen Künstlichkeit (Schöpfung künstlicher Welten), (2) das Gesetz der vermittelten Unmittelbarkeit (Schaffung symbolischer Ordnungen) und (3) das Gesetz des utopischen Standpunkts (Die Reflexion seiner Vergänglichkeit lässt den Menschen Bezugspunkte außerhalb seiner Selbst wählen) (Plessner 2003). Die evolutionäre Anthropologie befasst sich mit Themen wie kollektiver Intentionalität und Anpassung des Menschen an die Umwelt (vgl. Hartung 2018).

[5] Wir sind uns der Schwierigkeit der Begrifflichkeiten Natur und natürliche Umwelt durchaus bewusst. Einerseits ist der Mensch zugleich Naturwesen und Kulturwesen (Plessner 2003). Andererseits ist die natürliche Umwelt des Menschen immer auch anthropogen beeinflusst. Generisch wird im Folgenden der Terminus „natürliche Umwelt" verwandt, es sei denn der Begriff passt nicht in diesem Sinne.

Menschheit in das Internet der Dinge (IoT) und damit die Reduzierung des Menschen zu einem "Ding" in einer "Daten-Religion". Eine differenzierte Einschätzung der zugrundeliegenden Entwicklungslinien unterbleibt jedoch, so dass normativ aufgeladene Visionen als Zukunftsfakten ausgegeben werden.

In der **naturwissenschaftlich-technischen und in der geistes- und sozialwissenschaftlichen Fachliteratur** wird KI – wie auch andere Digitalisierungstechnologien – häufig auf einem hohen Abstraktionsniveau behandelt, so dass oft allgemeine ethische Probleme und Prinzipien für die Einschätzung der Digitalisierung bemüht werden. Wenn auf Anwendungsfeldebene argumentiert wird, dann treten immer wieder die Beispiele des Entscheidens beim Autonomen Fahren oder beim Einsatz bewaffneter Drohnen auf (vgl. u. a. (Misselhorn 2018).

Die Besonderheiten des Projektes:

Mit dem Projekt soll ein Beitrag zur Demythifizierung der KI-Entwicklungen, der Versprechen und vorherrschenden Erzählungen geleistet werden, wobei die damit einhergehende Aufklärung selbst als ein Beitrag zur Gemeinwohlorientierung verstanden werden kann:

- Im Mittelpunkt des Erkenntnisinteresses stehen grundlegende Veränderungen von Mensch-Technik-Umwelt-Beziehungen durch spezifische Anwendungen von KI. Es wird zwischen heute konkret feststellbaren Entwicklungslinien und abstrakten und ggf. hypothetischen langfristigen Visionen unterschieden.
- Vertiefte ethische Analysen werden anhand von KI-Themenkomplexen auf einem mittleren Abstraktionsebene vorgenommen (zwischen KI allgemein und KI-Einzelfällen).
- Es werden Einstiege in Transformationsnarrative entwickelt, die heutige Entwicklungen aufgreifen, dabei Konflikte und Dilemmata aufwerfen und die Weitererzählung aber offenlassen. Dieses experimentelle Kommunikationsformat soll auch bislang nicht involvierte, aber möglicherweise von KI betroffene Akteursgruppen dazu anregen, sich mit normativen Fragen der KI zu befassen.

Angesichts der Vielschichtigkeit und Dynamik des Untersuchungsgegenstandes geht es im Sinne des Projekttyps „Vorlaufforschung“ darum das Thema aufzubohren, zu sondieren und im Projektrahmen nicht sinnvoll adressierbare Forschungsfragen zu identifizieren („offene Enden“).

1.3 Vorgehen und Aufbau des Berichtes

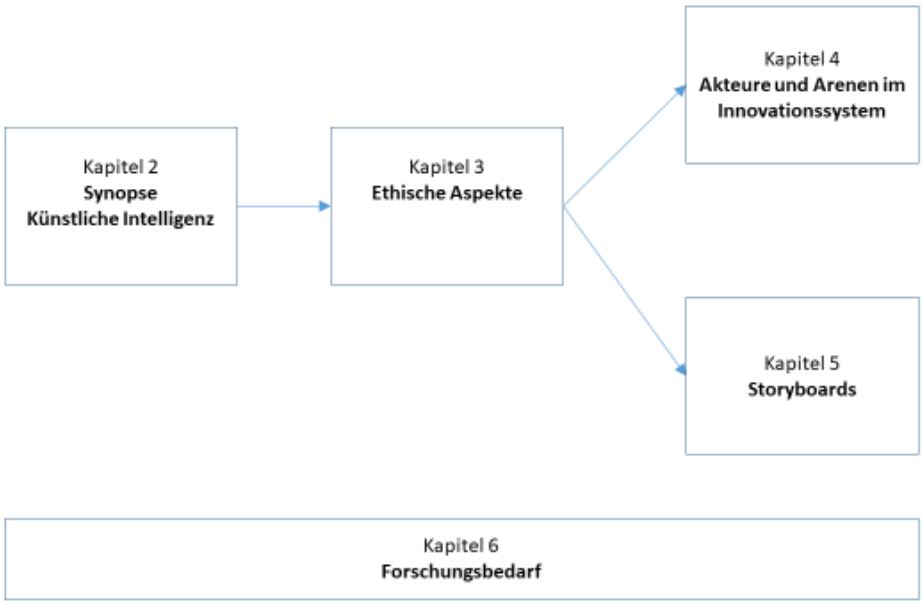
Das Projekt „Gemeinwohlorientierung im Zeitalter der Digitalisierung: Transformationsnarrative zwischen Planetaren Grenzen und Künstlicher Intelligenz“ umfasste vier inhaltliche Arbeitspakete, an der sich auch der Aufbau dieses Abschlussberichtes orientiert,

- als analytisches Fundament das Screening von KI und ihren disruptiven Anwendungsfeldern in einer synoptischen Darstellung (Kapitel 2) und die ethische Analyse mit der Ausweisung wichtiger ethischer Aspekte (Kapitel 3) und
- als handlungsorientiert Elemente die Anknüpfung im Innovationssystem mit der Benennung wichtiger Akteure und möglicher Aushandlungsarenen für die ethischen Aspekte (Kapitel 4) und die Storyboards als Einstieg in neue Transformationsnarrative (Kapitel 5).

Der Forschungsbedarf (im Kapitel 5) wurde anhand eines internen Reviews des Projektes ausgeführt.

Das Projekt wurde von einem wissenschaftlichen Beirat begleitet (vgl. Anhang). Dieser tagte drei Mal, am 02. Oktober 2019 zum Screening, am 08. Mai 2020 zur ethischen Analyse und am 09. Juli 2021 zu den Storyboards. Die Hinweise aus diesen Beiratssitzungen wurden rezipiert und, soweit im Rahmen dieses Projektes möglich, umgesetzt.

Abbildung 1: Vorgehen und Aufbau des Berichtes



Quelle: eigene Abbildung

Lesehinweise

Im Zeitraum des Projektes von November 2018 bis November 2021 hat sich die Quellenlage zu KI dynamisch entwickelt. Es ist nicht möglich, den Fundus berücksichtigter Quellen tagesaktuell zu halten. Punktuell konnten neue Quellen (z.B. (Heßler und Liggieri 2020)) auch für länger abgeschlossene Projektaktivitäten berücksichtigt werden. Die wesentlichen Strukturierungsleistungen dieses Projektes sind aus unserer Sicht aktuell und werden es voraussichtlich auch noch eine Zeit lang bleiben.

Dieser Abschlussbericht wurde mittels eines pragmatischen Vorgehens unter Einbezug vielschichtiger Quellen erstellt. Dementsprechend finden sich populärwissenschaftliche neben wissenschaftlichen Quellen und Experteneinschätzungen neben eigenen Einschätzungen. Es wird an der jeweiligen Stelle kenntlich gemacht, was Materialbefunde und was eigene Wertungen sind.

Die Kapitel 2-5 gliedern sich in ein Unterkapitel zu Zielen und konzeptionellem Ansatz, weiteren inhaltlich verschieden gelagerten Unterkapiteln und abschließenden Schlussfolgerungen.

In diesem Bericht wird, wenn möglich, die gender-neutrale Schreibweise verwendet (z. B. Lehrkräfte). Wo dies nicht sinnvoll möglich ist, wird die gendergerechte Schreibweise verwendet (z. B. Lehrer*innen). Hiervon abweichend wird in den Storyboards aufgrund besonders hoher Anforderungen an Verständlichkeit der in Massenmedien wie Tagesschau und Süddeutsche Zeitung verwendete Standard (z. B. Lehrerinnen und Lehrer) verwendet.

2 Synopse: Künstliche Intelligenz und ihre potenziell disruptiven Anwendungen

Aufgabe des Arbeitspaketes 2 im Gesamtprojekt war es, (1) die Informationslage zu KI-Entwicklungen zu sichten, (2) Anwendungsfelder von KI im Hinblick auf grundlegende Veränderungen von Mensch-Technik-Umwelt-Beziehungen zu identifizieren, was dann der Auswahl von Themenkomplexen für die ethische Analyse in Arbeitspaket 3 Orientierung bot. Einschränkend ist zu betonen, dass eine Gesamtschau erwünschter Zukünfte und der spezifische Beitrag von KI zur Annäherung an diese Zukünfte nicht beabsichtigt wurde. Ausgangspunkt sind die heutigen Begriffe und empirisch beobachtbare Entwicklungslinien der KI.

2.1 Ziele und konzeptioneller Ansatz

Angesichts des pervasiven und ubiquitären Charakters der Digitalisierung/KI kann keine vollständige Systematik der Digitalisierungs-/KI-Anwendungsfelder entwickelt werden. In diesem Vorhaben wurde deshalb ein Prozess aufgesetzt, der eine große Quellenbreite berücksichtigt (Scoping) und anhand definierter Kriterien bewertet (Screening).

Die Identifizierung und Charakterisierung von potenziell disruptiven Digitalisierungsfeldern erfolgte anhand folgender mit dem Umweltbundesamt abgestimmter Kriterien:

- Digitalisierungsfeld mit potenziell hoher Anwendungsrelevanz für KI
- Möglichkeit grundlegender Veränderungen von Mensch-Technik-Umwelt-Beziehungen durch dieses Digitalisierungsfeld (potenziell disruptiv)
- Vermutete Neuigkeit des Digitalisierungsfeldes hinsichtlich der Implikationen für Nachhaltigkeitstransformationen und Gemeinwohlorientierung

Folgende Hauptschritte wurden ausgeführt:

1. Schritt: Identifizierung von potenziell disruptiven Digitalisierungsfeldern

Ein multidisziplinäres Screening-Team am Fraunhofer ISI stellte einen Quellenpool für das Scoping und Screening möglichst breit und vielschichtig zusammen.[6] Diese Basisquellen wurden ausgewertet und analysiert.[7] Anschließend wurden die rund 120 Einträge verschlagwortet und im Hinblick auf grundlegende Veränderungen von Mensch-Technik-Umwelt-Beziehungen durch KI geclustert.[8] In Abstimmung mit dem Umweltbundesamt hat das Screening-Team dann zehn Digitalisierungsfelder für die detaillierte Charakterisierung ausgewählt.

[6] Zu den wesentlichen Quellen gehörten aktuelle Fraunhofer Projekte, Veröffentlichungen von Industrieverbänden (u. a. Bitkom, ZVEI), staatlichen Einrichtungen insbesondere zur Technikfolgenabschätzung (u. a. TAB, TA Swiss, Ethikrat), KI-Konferenzen (u. a. QCon.ai The Applied AI Software Conference for Developers), Studien von Consultern (u. a. Gartner, Pricewaterhouse), Online Magazine wie Telepolis (Heise-Verlag), The Economist und das Kulturmagazin Perlentaucher, Technology Review des MIT, Springer-Science, peer-reviewte Fachzeitschriften wie Futures und TED-Talks. Zudem waren Einzelpersonen beim Auftakt des Wissenschaftsjahres 2019 KI, BMBF, beim Stakeholder-Gespräch KI Enquete, Stuttgart und beim BMU-Kamingespräch KI-Anwendungen für den Umweltschutz präsent.

[7] In einer Tabelle wurden die Einträge hinsichtlich einer Bezeichnung und Kurzbeschreibung der Digitalisierungsfelder vorgenommen. Dann wurden unter Zuhilfenahme der Quellen Einschätzungen in Bezug auf die oben angeführten drei Kriterien (Anwendungsrelevanz für KI, grundlegende Veränderungen von Mensch-Technik-Umwelt-Beziehungen und Neuigkeitsgrad) vorgenommen.

[8] Am 15. März 2019 erarbeitete das Screening-Team ein Gesamtbild der Befunde, nahm Abgrenzungen, Fokussierungen und Differenzierungen vor und benannte Themenpaten für die Formulierung von Kurbeschreibungen zur projektinternen Nutzung.

2. Schritt: Charakterisierung der zehn potenziell disruptiven Digitalisierungsfelder

Die zehn potenziell disruptiven Digitalisierungsfelder wurden in einer einheitlichen Form (Profil) dargestellt. Zu den Beschreibungskategorien gehören:

- das Festlegen des Untersuchungsrahmens (Definition, Differenzierungen, Abgrenzungen, spezifische Rolle von KI) und
- die Einschätzung von Entwicklungen, differenziert nach heute konkret feststellbaren Entwicklungslinien und langfristigen Visionen.

Befunde zu grundlegenden Veränderungen von Mensch-Technik-Umwelt-Beziehungen wurden für jedes Profil separat aufgeführt,[9] dann aber am Schluss verdichtet (vgl. Unterkapitel 2.5).

Die Digitalisierungsfelder wurden anhand dieser Kategorien durch gezielte Recherche themenspezifischer Quellen und eigene Einschätzungen beschrieben und anhand von vertiefenden Interviews (vgl. Anhang) und auf Projekttreffen konsolidiert.[10]

In analoger Weise wurde ein Ankerpapier zu KI für die zehn Digitalisierungsfelder verfasst. Zudem wurde in einem Exkurs das Konzept der planetaren Grenzen hinsichtlich seiner Reichweite, Einschränkungen und Erweiterungsmöglichkeiten eingeschätzt[11], in diesem Abschlussbericht aufgrund der detaillierten Bezugnahme in Arbeitspaket 3 nicht dargestellt.

3. Schritt: Diskussion auf der Ersten Beiratssitzung

Auf der Beiratssitzung am 02. Oktober 2019 wurden Hinweise für inhaltliche Schwerpunkte und die ethische Vertiefung gegeben.

Im Folgenden erfolgt ein kompakter Überblick zu Künstlicher Intelligenz. Darauf werden die Profile für die zehn potenziell disruptiven KI-Anwendungsfelder nacheinander vorgestellt, einschließlich einer Einschätzung der Stabilität der Entwicklungslinien. Das Kapitel schließt mit einer Extraktion von drei übergeordneten grundlegenden Veränderungsdynamiken von Mensch-Technik-Umwelt-Beziehungen durch KI und verortet darin die zehn potenziell disruptiven KI-Anwendungsfelder.

2.2 Überblick zu Künstlicher Intelligenz

Der Branchenverband (Bitkom 2018) versteht unter KI einen Werkzeugkasten mit den Kategorien (1) *Assess* (messen, erkennen, identifizieren): z. B. Sprach- und Bilderkennung, (2) *Infer* (simulieren, vorhersagen): z. B. Problemlösen und Entscheidungsfindung und (3) *Respond* (agieren, handeln): z. B. Wissensmodifikation und Steuerung von autonomen Systemen. Die einzelnen Elemente dieses Werkzeugkastens entwickeln sich fort und lassen sich in immer neuer Weise kombinieren. Die Bundesregierung hat im November 2018 ihre *Strategie Künstliche Intelligenz* vorgelegt (Bundesregierung 2018c), in der neben induktiver Musteranalyse und Mustererkennung (1) auch Deduktionssysteme/maschinelles Beweisen (2), wissensbasierte Systeme/Expertensysteme (3), Robotik/autonome Systeme (4) und intelligente multi-modale Mensch-Maschine-Interaktionen (5) adressiert werden. Einige dieser Teilgebiete wurden zuvor unter Bezeichnungen wie „Big Data“, „Smart Systems“ oder „Industrie 4.0“ gefördert, weshalb Kritiker im Zusammenhang mit KI von einem neuen Marketing-Begriff sprechen (Daum 2019a;

[9] vgl. Darstellung im internen Projektbericht zu AP2 Screening

[10] Auf einem internen Treffen am 29. Mai 2019 am Fraunhofer ISI wurden die Entwürfe der Profile vergleichend analysiert und anschließend modifiziert. Auf einem gemeinsamen Projektreffen von Fraunhofer ISI und IZEW am 19. Juli 2019 wurden die Befeunde weiter konsolidiert.

[11] vgl. Darstellung im internen Projektbericht zu AP2 Screening

Schröter 2019). Anstelle von ‚Künstlicher Intelligenz' wird auch von ‚kleiner Intelligenz' gesprochen, denn KI erledigt bislang nur Aufgaben, die beim Menschen als einzelne Kompetenzen bezeichnet werden; KI verfügt aber nicht über das breite menschliche Problemlösungsrepertoire (Schröter 2019). Für die der KI übertragenen Aufgaben ist menschliche Intelligenz oft nicht notwendig oder geeignet. In diesen gesellschaftlichen Auseinandersetzungen geht es um die Deutung des Wesens, der Reichweite und der Ausgestaltung von KI im weitesten Sinne.

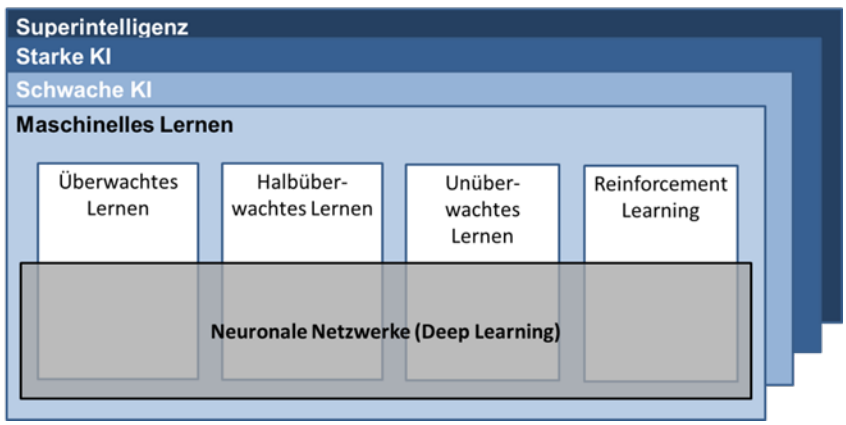
Wesentliche Treiber für die Entwicklung der KI in den letzten Jahren sind der ständig wachsende Strom neuartiger Daten, immer größere Rechenkapazitäten zu moderaten Kosten und immer mehr konkrete Problemlösungen durch KI, die entsprechende weitere Investitionen legitimieren (Bitkom 2018). Schlüsselakteure sind die großen global agierenden digitalen Plattformunternehmen aus den USA und China, teilweise in enger Kooperation mit staatlichen Einrichtungen. Auch die Bundesregierung hat als Antwort auf die Französische KI-Strategie im November 2018 eine eigene KI-Strategie verabschiedet (Bundesregierung 2018b). Auch supranationale Organisationen wie die EU (European Commission 2018)und die OECD (OECD 2018) befassen sich mit KI und ihrer Rolle für die Gesellschaft.

KI wird in Wirtschaft und Gesellschaft bereits breit eingesetzt. KI-Anwendungen können sich grundsätzlich überall dort finden, wo Daten verarbeitet werden. Sie unterstützen Unternehmen in der Gestaltung ihrer Geschäftsmodelle und -prozesse sowie datengetriebene Entscheidungsfindungen (Bitkom 2018). KI wird von Milliarden von Menschen bewusst oder unbewusst genutzt, insbesondere in Form von Sprach- und Bildverarbeitung durch Mobiltelefone und in Form von Algorithmen, die das Nutzerverhalten auf digitalen Plattformen leiten sollen (u. a. Empfehlungssysteme bei Amazon, Netflix, Spotify oder im Browser). Wir werden also schon heute durch KI gezielt und oft subtil beeinflusst.

Untersuchungsrahmen

Für Künstliche Intelligenz (KI) – *Artificial Intelligence* (AI) im Englischen – gibt es keine einheitliche **Definition** und zudem unterschiedliche **epistemische Auffassungen**. Abbildung 2 zeigt einige wesentliche Kategorisierungen von Künstlicher Intelligenz.

Abbildung 2: Framework zu KI und Maschinellem Lernen (vereinfacht)



Quelle: Fraunhofer ISI/IZEW basierend auf (Ackermann 2018)

Glossar zu Abbildung 2

Die **Schwache KI** (*Weak AI, Narrow AI*) zielt auf die Schaffung und Nachahmung menschlichen Verhaltens und Denkens sowie auf die automatische und autonome Erledigung klar definierter und abgegrenzter Aufgaben (Goram o. J.). Solche KI-Anwendungen lösen bereits heute spezifisch definierte Probleme, wie die Erkennung bestimmter Muster in Computertomographischen Bildern. Die **Starke KI** (*Strong AI, Artificial General Intelligence (AGI)*) soll über die Fähigkeiten eines normalen erwachsenen Menschen verfügen (Nicholson 2019). Von anderer Seite wird die starke KI als eine Illusion bezeichnet (Nadin 2019). Die **Superintelligenz** bezeichnet eine Intelligenz, die menschlicher Intelligenz überlegen ist ((Mainzer 2016b), S. 206).

Unter **Maschinellem Lernen** (*machine learning (ML)*) werden algorithmische, maschinelle Repräsentationen des menschlichen Lernprozesses gefasst. Anhand von Beispieldaten oder Erfahrungen aus großen Datenmengen kann der Computer definierte Aufgaben ausführen oder vordefinierte Entscheidungen treffen. ***Deep Learning*** ist eine ML-Methode, bei welcher die Daten durch sich selbstständig an den Dateninput anpassende neuronale Netzwerke verarbeitet werden. Die Anpassung erfolgt in Form von Veränderungen der Gewichtungen zwischen den Verbindungen in einem künstlichen neuronalen Netzwerk, so dass relevante Muster in Datensätzen erfasst werden können.[12] In den letzten Jahren sorgte die *Deep-Learning*-Methode für Durchbrüche in den Bereichen Erkennen, Verarbeiten und Generieren von Text, Bild, Ton und Code, Erkennen von Zusammenhängen zwischen verschiedenen Merkmalstypen, Handeln und Entscheiden beeinflussen/unterstützen sowie Robotik und Autonome Systeme.

Beim **überwachten Lernen** (*supervised learning*) lernt der Algorithmus anhand vieler mit Metadaten (z. B. Label für Kategorie) versehener Daten (z. B. Fotos), um anschließend neue Daten richtig zuzuordnen oder ähnliche Entscheidungen zu treffen (Klassifikation, Regression). Beim **unüberwachten Lernen** (*unsupervised learning*) erfolgt das Lernen ohne kommentierte Beispiele, nur aus Erfahrung mit großen Datenmengen oder aus realweltlichen Echtzeitdaten (Clusterung, Anomaliedetektion, Assoziation). Beim **halbüberwachten Lernen** werden ungelabelten Daten Label zugewiesen. **Reinforcement Learning** bezeichnet in Analogie zur Bestrafung oder Belohnung von Tieren einen anreizzentrierten Lernmechanismus (modellbasiert, wertebasiert, politikbasiert).

Für die Innovationsforschung und -politik ist es hilfreich, wesentliche Entwicklungsparadigmen zu unterscheiden, um gezielt und sachbezogen agieren zu können. Auf einem höheren Abstraktionsniveau unterscheiden (Erdmann und Röß 2020) deshalb drei Paradigmen für die Beschreibung und Deutung von KI: (1) das Problemlösungsparadigma (das selbstständige und effiziente Lösen von Problemen (Mainzer 2016a)), (2) Das Mensch-als-Maßstab Paradigma (das selbständige und effiziente Erledigen von Aufgaben, die normalerweise menschliche Intelligenz erfordern bzw. diese übersteigen[13] (Lenzen 2002a)) und (3) das Evolutionsparadigma (durch Variation und Selektion von Varianten werden den veränderten Umweltbedingungen angepasste Kompetenzen etabliert (Dennett 2018).[14]). Grundsätzlich sind emergente Phänomene einer Koevolution von *Seed-AI* und Kultur zwar denkbar, aber mit ihren Wirkungen und Nebenwirkungen kaum antizipierbar ((Horn und Bergthaller 2019), S. 90 - 92).

[12] In der Informatik bezieht sich der Begriff neuronale Netze auf Systeme, die aus Schichten relativ einfacher Rechenelemente bestehen, die als Neuronen bezeichnet werden.

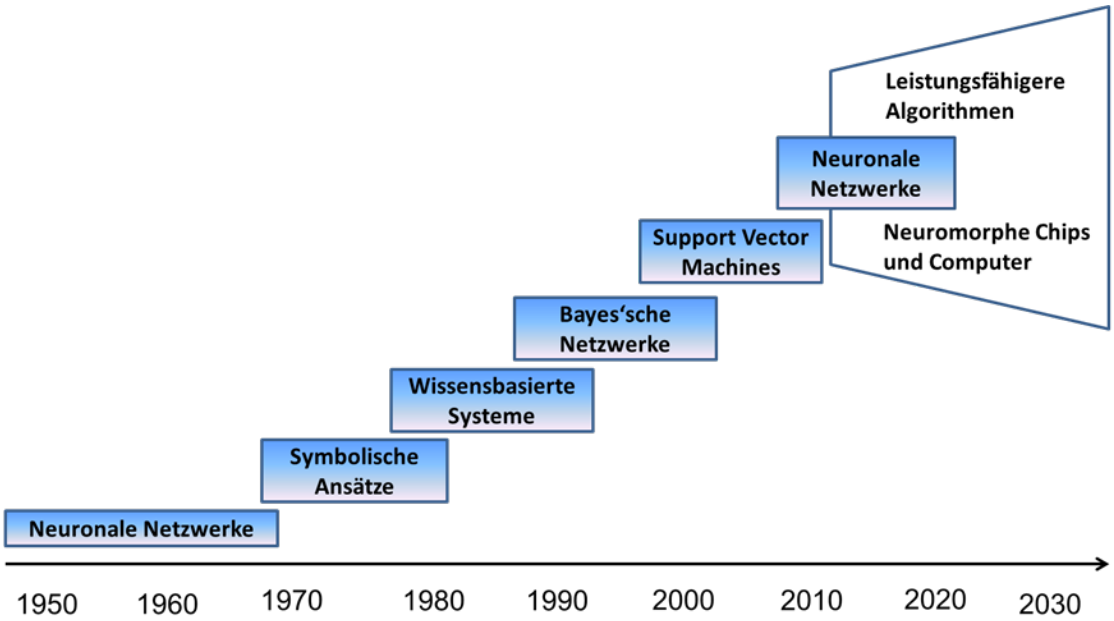
[13] „Starke KI" steht für KI mit gleichen oder dem Menschen überlegenen intellektuelle Fertigkeiten (Bundesregierung 2018c), Artificial General Intelligence für KI, die die Bandbreite des menschlichen Problemlösungsrepertoires überschreitet und Superintelligenz allgemein für eine KI, die menschlicher Intelligenz in verschiedener Hinsicht überlegen ist (Mainzer 2016a).

[14] *Seed-AI* bezeichnet ein „KI-System mit der Fähigkeit, sich selbst zu verbessern" (Mainzer 2016a)

Entwicklungen

Eine Analyse der Abstracts von über 16.000 Forschungsartikeln zu KI unterscheidet verschiedene **KI-Konjunkturen** (Hao 2019). Ein Blick zurück zeigt (vgl. Abbildung 3), dass die KI-Forschung durch den plötzlichen und raschen Aufstieg und Fall von einzelnen KI-Technologien charakterisiert ist.

Abbildung 3: Entwicklungslinien der KI (vereinfacht)



Quelle: Fraunhofer ISI/IZEW basierend auf (Hao 2019) und (Warnke et al. 2019)

Ende der 1990er erfolgte eine Abkehr von regelbasierten Wissenssystemen und eine Hinwendung zum Maschinellen Lernen, indem nicht mehr Tausende von Regeln codiert werden mussten, sondern die Regeln automatisch aus großen Datenbeständen extrahiert wurden. Neuronale Netze als Kernelement des *Deep Learning* setzten sich gegenüber anderen Techniken ab 2014 durch. In den meisten praktischen Anwendungen dominiert heute das überwachte Lernen. Seit 2014 steigt die Zahl der Publikationen zum *Reinforcement Learning* rapide an. Welche Ansätze in Zukunft dominieren werden ist unsicher (Hao 2019).

Eine wichtige Grundlage für die Entwicklung von KI sind das Verständnis und die Simulation von kognitiven Prozessen. Seit 2013 wird das "**Human Brain Project**" mit einem Volumen von über einer Milliarde Euro gefördert.[15] Ziel ist die Simulation des gesamten menschlichen Gehirns innerhalb von zehn Jahren. Mit dem Human Brain Projekt sind Hoffnungen und Versprechen verknüpft, die Funktionsweise kognitiver Prozesse aufzuklären, diese formal zu beschreiben und technisch zu simulieren (u. a. Brain Functional Mapping). KI-Algorithmen unterstützen wiederum dabei, neue Erkenntnisse über menschliche Kognition zu erlangen.

Zentrale aktuelle **technische KI-Entwicklungslinien** beziehen sich einerseits auf die Steigerung der Leistungsfähigkeit der Algorithmen und andererseits auf die Nachbildung intelligenter biologischer Prinzipien in der Hardware. Mögliche Durchbrüche in der Entwicklung von intelligenten KI-Algorithmen werden für *Duelling Networks* („schneller Lernen")[16], *Capsule*

[15] https://www.humanbrainproject.eu/en/

[16] *Duelling Networks* basieren auf dem Prinzip, dass eines der beiden neuronalen Netzwerke realistische Artefakte wie Fotos generiert (Generator) und dass das andere Netzwerk zwischen Fake und Fakt unterscheidet (Diskriminator). Durch Feedbacks lernen Generator und Diskriminator sehr schnell.

Networks („menschenähnlicher Lernen“)[17] und *Few-Shot-Image Recognition* („Lernen anhand weniger Daten“)[18] erwartet (Warnke et al. 2019). Im Bereich der KI-geleiteten Hardwareentwicklung versprechen neuromorphe Chips und Computer eine unmittelbare Steigerung der Leistungsfähigkeit durch KI (Warnke et al. 2019), aber auch die allgemeinen Forschungsanstrengungen zur Erhöhung der Leistungsfähigkeit der Hardware wie Quantencomputing können der KI zugutekommen.

Über die technischen KI-Entwicklungslinien hinaus gibt es Entwicklungsparadigmen, die nicht so sehr die Funktionalität als vielmehr bestimmte andere **Anforderungen an die Technikentwicklung** hervorheben. Die Entwicklungslinie *Explainable AI* (XAI) zielt darauf, die Mechanismen für KI-basierte Entscheidungsempfehlungen offenzulegen, *Human-centered AI* darauf, diese am Menschen und *Beneficial AI* darauf, diese am Gemeinwohl auszurichten.

Anwendungsfeldübergreifende langfristige Visionen für die KI sind die Superintelligenz und der technische Posthumanismus:

Langfristige Visionen

Es können drei Formen der **Superintelligenz** unterschieden werden (Mainzer 2016b): (1) Schnelle Superintelligenz "kann alles, was der Mensch kann, nur schneller"; (2) Kollektive Superintelligenz "besteht aus vielen Teilsystemen, die weniger leisten können als Menschen. Das kollektive KI-System ist dem Menschen aber überlegen"; (3) Neue Superintelligenz "hat neue intellektuelle Fähigkeiten, die Menschen nicht im Ansatz besitzen". Es ist unklar, ob es möglich sein wird, die Anfangsbedingungen für KI so zu konstruieren, dass eine Intelligenz jenseits der menschlichen Existenz selbst überlebensfähig sein kann (vgl. u. a. (Bostrom 2017); (Kurzweil 2013)).

Der **technische Posthumanismus** soll eine neue evolutionäre Stufe des Menschen begründen, auf der der Mensch nur noch in digitalen Bewusstseinsformen, d.h. ohne Körper, existiert (z. B. Upload des menschlichen Bewusstseins in einen externen Datenspeicher). Bei diesen Visionen geht es um eine Unsterblichmachung (Loh 2018), Ergänzung, radikale Umwandlung oder gar Ersetzung der menschlichen Natur, indem der Mensch durch von ihm geschaffene „perfekte“ Entitäten abgelöst wird. Dabei bleibt offen, ob dies als Vollendung oder Überwindung des Projekts Mensch aufgefasst werden soll (Röcke 2017; Bundesministerium für Bildung und Forschung 2018).

Stabilität der Entwicklungslinien: Die Investitionen in KI-Technikentwicklung und KI-Anwendungsentwicklung sind beträchtlich, ohne vorhersagen zu können, welche Einzellösungen sich schließlich durchsetzen werden. „Superintelligenz“ und technischer Posthumanismus sind langfristige Visionen, deren Realisierung hochspekulativ ist.

2.3 Synopse potenziell disruptiver Digitalisierungsfelder

Im Folgenden werden die zehn Digitalisierungsfelder, die grundlegende Veränderungen der Mensch-Technik-Umwelt-Beziehungen bewirken können, charakterisiert. Wesentliche grundlegende Veränderungen der Mensch-Technik-Umwelt-Beziehungen werden in Abschnitt 2.5 integriert dargestellt. Jedes Profil gliedert sich in die Teilabschnitte Untersuchungsrahmen und Entwicklungen, einschließlich Visionen, gefolgt von einer Einschätzung der Stabilität der Entwicklungslinien seitens des Fraunhofer ISI. Es gibt zahlreiche Querbezüge zwischen den einzelnen KI-Anwendungsfeldern und KI-Entwicklungslinien. Die Trennung zwischen den

[17] *Capsule Networks* sind hierarchische neuronale Netze die eher der Art entsprechen, wie Menschen Informationen verarbeiten.

[18] *Few-Shot Learning* bezweckt, aus einem sehr kleinen Datensatz die Klassifizierung von Daten zu lernen; ebenso wie das menschliche Gehirn Objekte nach ein- bis zweimaligem Sehen wiedererkennen kann.

Digitalisierungsfeldern ist praktischer Natur, um Schwerpunkte für die ethische Analyse auszuwählen, in der dann Querbezüge zwischen den Digitalisierungsfeldern elaboriert werden. An dieser Stelle wird auf die Darstellung von Querbezügen verzichtet.

2.3.1 Affective Computing

Untersuchungsrahmen

Das **Affective Computing** setzt an der technischen Nachbildung und Stimulation der emotionalen Dimension menschlichen Lebens an. Konkret geht es um das Wahrnehmen, Interpretieren und Erwidern menschlicher Affekte und Emotionen. Zu den wichtigsten Affektphänomenen für den Menschen gehören situationsgebundene episodische Emotionen (z. B. Ärger oder Freude), dispositionale Emotionen (z. B. anhaltende Schuldgefühle), Stimmungen (z. B. Ausgelassenheit), existenzielle Gefühle (z. B. überwältigt sein), atmosphärische Gefühle (z. B. eisige Atmosphäre im Gespräch), Bauchgefühle (z. B. Bewertungen komplexer Entscheidungsabsichten) und affektive Stile (z. B. unterstützend oder arrogant) (Colombetti und Stephan 2013). Der Ansatz der situierten Affektivität sieht in Bezug auf affektive Prozesse die Notwendigkeit, über kognitive Prozesse hinaus Körper, Umwelt und deren Interaktion einzubeziehen (Wilutzky et al. 2013).

Analytisch lassen sich zwei Anwendungsformen des Affective Computing unterscheiden ((Kehl und Coenen 2016), S. 147): Die erste Anwendungsform sind emotional sensitive Systeme bzw. Roboter, denen in erster Linie eine Kontroll- oder Assistenzfunktion zukommt. Über die Detektion und das **Erkennen von Emotionen** (z. B. über Mimik und Stimmlage etc.) **und körperlichen Affekten** (z. B. Pulsschlag, Blutdruck oder Hautfeuchtigkeit) sollen sie den adressierten Menschen ein bestimmtes Verhalten nahelegen bzw. selbst automatisch Entscheidungen treffen oder den Menschen in Kontrollkontexten hinsichtlich ihrer Persönlichkeit und ihrem tatsächlichen Empfinden besser einschätzen. Solche Systeme sind auf dem Niveau von technischen Hilfsmitteln angesiedelt. Hiervon können solche Systeme bzw. Roboter unterschieden werden, die einen Selbstzweck darstellen. Sie sollen auf emotional intelligente Weise auf Gefühle des menschlichen Gegenübers reagieren, Emotionen situationsspezifisch produzieren bzw. simulieren und damit beim menschlichen Gegenüber wiederum **Gefühle hervorrufen**. Systeme dieser Art können dadurch als Interaktionspartner*innen auftreten, zu denen Menschen nicht mehr nur ein instrumentelles Verhältnis besitzen, sondern eine Art „parasoziale Beziehung“ (Gutmann 2011), die bisherige zwischenmenschliche Beziehungen und die damit verbundenen Bedürfnisse nach Nähe und emotionalem Austausch ergänzen oder gar ersetzen kann.

Was verändert sich durch KI beim Affective Computing?

KI erfüllt beim Affective Computing mit Mustererkennung in Gesten, Mimik und anderen körperlichen Manifestationen von Emotionen, Maschinellem Lernen und Generierung von responsiven Reaktionen auf Gefühle wichtige Grundfunktionen.

Entwicklungen

Grundsätzlich lassen sich die Entwicklungslinien „Erkennen von Emotionen“ und „Erzeugen von Emotionen“ unterscheiden.

Schlüsseltechnik für das **Erkennen von Emotionen** ist die Gesichtserkennung. Inzwischen werden Emotionen auch mittels Textanalyse, Stimmlage, Herzschlag- und Atemmustern, etc. gemessen. In Entwicklung sind derzeit temporäre elektronische Tattoos, die die Aktivität von Muskeln oder Nervenzellen messen (tragbare Hautelektroden), Sentiment Analysis (u. a.

Textanalyse, Wearables zur Erkennung der Stimmlage bzw. Bestimmung von Herzschlag- und Atemmustern) und die berührungslose Detektion von Mimik und Gesten (Warnke et al. 2019).

Anwendungen des Affective Computing in ihrer Kontroll- oder Assistenzfunktion werden heute in verschiedenen Branchen benutzt. Insbesondere in der Werbebranche und der Filmindustrie kommen Systeme, die die emotionalen Reaktionen des Publikums auf bestimmte Stimuli erfassen und interpretieren (u. a. Kameras, Eye Tracking) großflächig zum Einsatz (Schnabel 2016). Die Detektion von Emotionen zu Kontrollzwecken durch KI-Systeme befindet sich bereits, wenn auch in überschaubarem Maße, in der Anwendungsphase (z. B. Lügendetektoren).

Die Schlüsseltechniken für das **Erzeugen von Emotionen** haben sich noch nicht deutlich herauskristallisiert. Das elektronische Spielzeug Tamagotchi und seine Nachfolger stimulieren seit langem das Kümmern um digitale Wesen. Die Entwicklung humanoider Roboter (Warnke et al. 2019) vermag mittelfristig die Mensch-Maschine-Beziehung der heutigen Mensch-Mensch-Beziehung in Teilbereichen noch ähnlicher machen.

Für eine positive Resonanz von Seiten der Nutzer werden emotional sensitive Robotersysteme auch in ihrer Materialität ansprechend gestaltet. So ist das Aussehen von Assistenzrobotern häufig einem Kindchenschema nachempfunden oder die Geräte werden aus biegsamem, weichem und damit haptisch ansprechendem Material hergestellt (Schnabel 2016). Roboter bzw. Systeme, die selbst Emotionen simulieren und damit menschenähnliche soziale Beziehungen eingehen können, befinden sich derzeit in der Entwicklungs- oder Pilotphase. Die möglichen Anwendungen von emotional sensitiven bzw. intelligenten Systemen sind vielfältig (u. a. Pflegesektor, Psychotherapie, Sexbots).

Emotional intelligente Roboter bzw. Systeme können nicht mehr als reine technische Hilfsmittel gelten, vielmehr sollen sie autonom agieren und ein möglichst gleiches Interaktionsniveau wie menschliche Beziehungen erreichen (Schnabel 2016). Emotional intelligente Roboter als Beziehungspartner und als lebensüberdauernde Gefährten stellen Entwicklungen dar, die in dem sich ausbreitenden Gefühl der Einsamkeit in Ländern wie Japan, England und Deutschland einen wesentlichen Treiber finden könnte (Schnabel 2016). Die Treiber und Versprechen von autonomen Beziehungsrobotern variieren in unterschiedlichen gesellschaftlichen und kulturellen Kontexten. Wichtige Faktoren sind beispielsweise der Alterung von Gesellschaften, Kontroll- und Sicherheitsdiskurse sowie allgemeine Technikakzeptanz (Schnabel 2016).

Langfristige Visionen

Folgt man der Analyse des niederländischen Rathenau Instituts, lässt sich die Vision intimer emotionaler Beziehungen zwischen Mensch und Roboter als "*intimate-technological Revolution*" ((Kehl und Coenen 2016), S. 145) fassen, bei der emotional intelligente **Roboter als Beziehungspartner** das langfristige übergreifende Entwicklungsziel darstellen.

Bei der Entwicklung von lebensüberdauernden emotional intelligenten Gefährten (Chatbots bzw. digitale Avatare) werden die Gefühle, Erinnerungen und das Kommunikationsverhalten einer lebenden Person durch KI-Systeme analysiert und über deren Tod hinaus „konserviert“. Für Nahestehende stehen diese „Gefährten“ dann **lebensüberdauernd** zeitlich unbegrenzt als Interaktionspartner*in zur Verfügung und für Dritte kann die Interaktion mit lebensüberdauernden Avataren bedeuten, dass ihnen kein gesichertes Wissen über den Tod bzw. die Lebendigkeit der leiblichen Person zur Verfügung steht (Schneider 2020).

Stabilität der Entwicklungslinien: Das Erkennen von Emotionen durch KI-basierte Verfahren ist etabliert und zieht durch große Marktpotenziale weitere Entwicklungsanstrengungen nach sich.

Für das Erzeugen von Emotionen durch humanoide Roboter ist derzeit in Deutschland und Europa die Entstehung eines Massenmarktes nicht erkennbar.

2.3.2 Simulation natürlicher Sprache

Untersuchungsrahmen

Sprache ist ein sich fortwährend entwickelndes, komplexes System von Lauten und Zeichen zum Zwecke der Kommunikation. Neben ihrer Herkunft (z. B. Arabisch, Deutsch, Italienisch,) werden Sprachen unter anderem auch hinsichtlich ihrer Funktionalität (u. a. Alltagssprache, Fachsprache, Programmiersprache) differenziert. Während die Syntax sich auf die formalen Beziehungen zwischen den Zeichen bezieht, geht es bei der Semantik um die Beziehung zwischen den Zeichen und den Dingen, auf die sich die Zeichen beziehen (Kompa et al. 2013). Ein Code ist eine Regel für die Transformation einer Nachricht aus einer symbolischen Form in eine andere (Deutsche Stiftung Weltbevölkerung 2018; Giménez et al. 2018).

Dieses Profil umfasst das Erkennen, Verarbeiten, Erzeugen und Simulieren von Sprache, einschließlich Übersetzungen von einer Sprache in eine andere. Die **computergestützte Analyse und Generierung natürlicher Sprache** orientiert sich an der von Menschen verwendeten schriftlichen und mündlichen Ausdrucksweise (Bitkom 2018). Das Erkennen und die Verarbeitung bzw. die Erzeugung und Simulation **gesprochener Sprache** durch Computer dienen in erster Linie der Mensch-Maschine-Konversation, im Falle des Bezugs auf **schriftlich kodierte Sprache** wird in erster Linie das Wissensmanagement adressiert.

Wie verändert KI die Simulation natürlicher Sprache?

Das automatische Einfangen von Texten und Sprache sowie die Verarbeitung und Extraktion relevanter Informationen dienen als Basis für viele intelligente Anwendungen und Dienste. **KI** erfüllt mit Mustererkennung, Maschinellem Lernen und Generierung natürlicher Sprache (*Natural Language Generation* (NLG)) wichtige Grundfunktionen. Die Kontextabhängigkeit der Semantik ist eine wichtige Randbedingung für eine effektive maschinelle Sprachverarbeitung (Kompa et al. 2013). Generell wird auch über digitale Artefakte wie Bilder, Töne oder Objekte kommuniziert, zu deren Erkennen und Erzeugen vergleichbare KI-Technologien eingesetzt bzw. entwickelt werden.

Entwicklungen

Die Verarbeitung und Generierung natürlicher Sprache wird in den Entwicklungslinien Kognitive Assistenten und Computational Creativity praktisch angewendet

Kognitive Assistenten sind Assistenzsysteme, „die Menschen bei der Ausführung kognitiver Aufgaben und Entscheidungsfindungen unterstützen oder gar ersetzen sollen“ ((Hecker et al. 2017), S. 36). Teilweise werden die Begriffe kognitive Assistenten, virtuelle Assistenten, Smarte Assistenz, intelligente und virtuelle Agenten, kommunizierende Agenten, *Conversational Agents*, *(Chat)Bots*, *Cognitive Computing* oder *Companion Systems* synonym verwendet. Insgesamt steigen sowohl der Anteil der kognitiven Assistenten an der menschlichen Gesamtkommunikation als auch die Qualität der Sprachverarbeitung (Specia et al. 2017).

Sprachassistenten (einschließlich *Chatbots*) basieren auf der Verarbeitung natürlicher Sprache (*Natural Language Processing*, NLP) mit derselben Technologie, die Spracherkennung ermöglicht und die auch die Basis virtueller Assistenten wie Amazon Alexa, Google Now, Siri von Apple oder Cortana von Microsoft bildet. So kann Amazon Alexa beispielsweise Flüstern als solches erkennen und zurückflüstern oder mit verschiedenen Akzenten sprechen (Prüfer 2019). Solche Sprachassistenten verarbeiten die sprachlichen Inhalte der Sprechenden, prognostizieren

Intentionen und reagieren darauf in sprachlicher Form. Durch Sprachassistenten können auch emotionale Dispositionen verstärkt oder modifiziert werden.

Es werden persönliche und Service-Assistenten voneinander unterschieden. Persönliche Assistenten (*Virtual Personal Assistants*) können mit Nutzer*innen via Audio/Text kommunizieren und verschiedene weitere Funktionen ausführen ((Hecker et al. 2017), S. 37). Service-Assistenten (*Virtual Customer Assistants*) werden für kundenorientierte Dienstleistungen eingesetzt ((Hecker et al. 2017), S. 40). Sie liefern dem Kunden im Rahmen eines Gesprächs die gesuchten Informationen verdichtet und kontextrelevant und können in seinem Auftrag handeln. Textbasierte Dialogsysteme erlauben das Chatten mit einem technischen System. *Chatbots* nehmen Anrufe entgegen, unterhalten und positionieren uns in Warteschleifen, nehmen Bestellungen auf oder beraten bei der Suche nach einzelnen Produkten.[19]

Eine wesentliche Barriere in der Kommunikation zwischen Menschen sind ihre verschiedenen Sprachen. Sprachassistenten bieten hierbei Möglichkeiten, die sich mit Menschen in der analogen Welt praktisch nicht umsetzen lassen (Burchardt und Uszkoreit 2018a). Hierzu gehören insbesondere das Angebot nutzergerechter, einfacher Sprache sowie die Echtzeitübersetzung von einer Sprache in eine andere Sprache. Die Qualität von Übersetzungen hat durch KI-basierte Services wie DeepL und Google Translate deutlich zugenommen (Stieler 2018). Die KI-basierte Echtzeitübersetzung in verschiedene Sprachen (Mustererkennung und Maschinelles Lernen) steht somit vermutlich kurz vor dem ökonomischen Durchbruch (vgl. (Gartner 2018)).

Computational Creativity (auch *Artificial Creativity, Creative Computing*) ist ein multidisziplinäres KI-Forschungsfeld, das darauf abzielt, Computerprogramme zu entwickeln, die zu Kreativität auf menschlichem Niveau fähig sind (Association for Computational Creativity 2016). Kreativität bezeichnet den schöpferischen Prozess des Hervorbringens von Neuem, wobei die fünf Phasen Vorbereitung, Inkubation, Einsicht, Bewertung und Ausarbeitung unterschieden werden können (Schmid und Funke 2013). Für Computational Creativity wird auf jüngere Erkenntnisse der Neurowissenschaft der Kreativität zurückgegriffen ((Warnke et al. 2019), S. 53). In Abgrenzung zu gängigen KI Ansätzen, werden Computational Creativity-Ansätze meist nicht in einem Problemlösungs-Paradigma[20] entwickelt und getestet, sondern einem Artefaktgenerierungs-Paradigma zugeordnet, das darauf ausgerichtet ist, etwas von kulturellem Wert zu produzieren (Colton und Wiggins 2012).

Computational Creativity ist von den meisten Menschen unbemerkt in den Alltag eingezogen: Beispiele für sprachbezogene Computational Creativity sind *Narrative Science* (automatische Sport-und Wetternachrichten, Quartalszahlen von Investmentbanken), *Mexica* (Kurzgeschichten, die durch KI geschrieben wurden), *The Joking Computer* (computergenerierte Witze), *Poem Machine* (interaktives Gedichtschreibe-Werkzeug) und *Wibbitz* Software (Generation von Nachrichtenvideos aus Texten, Videosequenzen, Fotos und einer automatisch generierten Stimme). Mehr und mehr journalistische Texte werden automatisiert generiert und verbreitet. Computational Creativity ist dabei gleichzeitig Ursache und Folge der Vervielfachung der Menge an Texten und Sprechakten in Sekundenschnelle. In Verwaltung, Rechtsprechung und Kundenberatung (z. B. in der Finanzbranche) werden Routinetexte generiert, die durch intelligente Schreibprogramme erstellt werden könnten (Mainzer 2016b). In der Wissenschaft

[19] Beispielsweise unterhalten sie sich mit einsamen Menschen, scannen Stimmen auf depressive Verstimmungen oder bieten Seelsorge (vgl. auch die Abschnitte 2.3.1 und 3.3 zu Affective Computing).

[20] Eine Aufgabe, die automatisiert werden soll, wird als eine bestimmte Art von Problem formuliert. Je nach Art der Verarbeitung (z. B. Deduktion, Verallgemeinerung, die erforderlich ist, um eine Lösung zu finden, kommen andere Ansätze zum Einsatz wie Theoremprüfung oder maschinelles Lernen.

könnten Bots die von einem Wissenschaftler stammenden Daten, Argumente und Ergebnisse einlesen und automatisch in dem Schreibstil des Wissenschaftlers angepassten Texte verwandeln (Mainzer 2016b). Darüber hinaus gibt es eine Reihe anderer nicht-textbasierter Anwendungen, die den Möglichkeitsraum von Computational Creativity illustrieren.[21]

Langfristige Visionen

Individual AI, auch **Persönliche KI**, soll an die Person angepasst sein, diese Person am besten kennen, bei der Erledigung täglicher Aufgaben helfen und gegen mögliche Angriffe fremder KI schützen. Die heutigen persönlichen Assistenten avancieren zukünftig unter Umständen zu **Avataren**, die Menschen simulieren (Gheorghiu et al. 2017).

Es werden Computational Creativity Programme angestrebt, die in der Lage sind, neue unvorhergesehene Modalitäten zu erschaffen, die über menschliche Fähigkeiten hinausgehen. Diese **transformative Kreativität** (Boden 2004) besteht darin, dass ein System in der Lage ist, neue Möglichkeiten in einem neuen konzeptionellen Raum zu finden, indem es den alten konzeptionellen Raum transformiert – anstatt vertraute Dinge neu zu kombinieren oder neue Möglichkeiten innerhalb desselben konzeptionellen Raumes zu erkunden.

Stabilität der Entwicklungslinien: Die Entwicklungen zum Erkennen und Generieren gesprochener und geschriebener Sprache sind weit vorangeschritten und ziehen große Investitionsvolumina an sich. Die Musikband ABBA gab Ende des Jahres 2021 ihr virtuelles (nicht ihr leibliches) „Comeback" als alterslose Avatare, die anhand von aktuellen Bewegungsprofilen und realer Auftritte im Jahr 1979 realisiert worden sind (Ergin und Ammer 2021). Die zukünftige Verbreitung von Avataren und die Realisierungschancen von transformativer Kreativität lassen sich derzeit kaum seriös einschätzen.

2.3.3 Extended Reality

Untersuchungsrahmen

Unter dem Oberbegriff **Extended Reality** (XR) werden Augmented Reality (AR, dt. erweiterte Realität), Augmented Virtuality (AV, dt. erweiterte Virtualität) und Virtual Reality (VR, dt. virtuelle Realität) sowie deren Mischform Mixed Reality (MR, dt. gemischte Realität) zusammengefasst. Voraussetzung ist die digitale Repräsentation einer räumlichen Umgebung.

Im Falle der **Augmented Reality** wird die reale räumliche Umgebung durch seinen digitalen Zwilling überlagert, wodurch eine Erweiterung der natürlichen Wahrnehmung der Umgebung durch digitale Informationen und Objekte ermöglicht wird. Durch Computertechnologie und Wearables (Handy, Brille, Handschuh) werden digitale Elemente möglichst nahtlos in die Realität eingefügt und es entstehen neue real-virtuelle Umgebungen. Die Idee besteht im Kern darin, Informationen zu jeder Zeit an jedem Ort abzurufen und in Echtzeit über die Realität zu legen, um Nutzer*innen zu informieren, sie bei Aufgaben zu unterstützen oder sie zu unterhalten.

Aus technologischer Sicht ist AR eine große Herausforderung, da die Technologie auf einem komplexen Zusammenspiel von Sensoren beruht, die die Position und die Aufmerksamkeit der Nutzer*innen verfolgen können, ein gutes Verständnis der Position im Raum (3D) ermöglichen, mit dem Internet verbunden und möglichst leicht und gut handhabbar sein müssen (Greenfield 2018). AR-Anwendungen sind auf die Verfügbarkeit von Echtzeitdaten aus der Umgebung und

[21] beispielsweise das *Painting Fool Project* (Automatisierung physischer und kognitiver Aspekte von Malerei), der *Iamus* Kompositionscomputer (Komposition von Songs aller Genres) und der *IBM Computer Watson* (Erfindung von Kochrezepten)

der Erweiterungsdatenschicht angewiesen. Digitale Zwillinge der realen räumlichen Umwelt sind die Basis für AR (Warnke et al. 2019). Die digitale Umgebung lernt aus unseren Reaktionen und passt sich unseren digital aufgezeichneten Verhaltensweisen und Handlungen im Raum an. Die überblendeten Informationen sind selektiv und können vom Anbieter zu eher individualisierten oder zu eher generalisierten Umgebungswahrnehmungen eingesetzt werden.

Virtual Reality bezeichnet die computertechnischen Simulationen einer realistisch erscheinenden Umgebung mit 3D-Bild (und meist auch Ton). In diese synthetische Umgebung können Betrachtende vollends eintauchen (Immersion). Benutzereingaben über Datenhelme und -Handschuhe werden unmittelbar in Steuerbefehle umgesetzt, die sich direkt auf die virtuelle Umgebung auswirken (ITWissen.info 2017).

Wie verändert KI die Extended Reality?

KI-unterstützte Datenanalysen und KI-basierte Schnittstellen vermitteln zwischen der digitalen und der menschlichen Welt (DFKI 18.04.2018).

Entwicklungen

Für AR können zwei grundlegende Entwicklungslinien unterschieden werden, die Führung des Menschen und die Auslösung digitaler Interaktionen durch den Menschen.

Zur **Führung von Menschen** werden **AR-Anwendungen** in vielen Bereichen bereits eingesetzt, u. a. in Bildung, Training, Unterhaltung, Navigation, Produktions-und Wartungsarbeiten in der Industrie, Medizin, Marketing und Werbung (u. a. *Eye Tracking*) sowie der Erinnerungsarbeit.

In der Industrie sind AR und VR weitestgehend angekommen. Interaktive Handbücher liefern Informationen in Echtzeit und geben Live-Anweisungen an Personen, die direkt mit Maschinen arbeiten. Durch das Abspielen der Informationen auf der AR-Brille haben sie die Hände frei und können die für ihre Arbeit relevanten Informationen ohne zeitliche Verzögerung verwenden. So nutzt beispielsweise DHL seit 2017 *Google Glass Enterprise*, um Mitarbeiter*innen Arbeitsanweisungen und Informationen live anzuzeigen (DHL 08.02.2017). VR ist besonders im Bereich der Weiterbildung beliebt, denn Schulungen und Ausbildungen können direkt an Objekten durchgeführt, interessant gestaltet und verkürzt werden (Baumgartner 2018).

Der Verkauf von *Google Glass* auf dem Konsummarkt wurde 2015 allerdings aufgrund mangelnder Akzeptanz eingestellt (Lindner 2015). 2019 brachte das Start-up *North* seine Brille *Focals* raus, welche einer normalen Brille sehr ähnelt. Die Brille ist mit dem Telefon verbunden und über Sprache und einen Handring steuerbar. *Focals* verfügt über folgende Funktionen: Einblendung von Nachrichten aus sozialen Medien, Stimme zu Text-Übertragung zum Versenden von Nachrichten, live-Navigation und Suchfunktion, Kommunikation mit Alexa (Musik abspielen, Nachrichten hören, Wetteranzeige, Smart Home steuern usw.) sowie Kalender- und Wetterfunktionen.

Auch die Identifikation von Personen hat bereits marktreife Anwendungen. Menschen mit Gesichtsblindheit, die Schwierigkeiten haben, Menschen wiederzuerkennen, könnte AR im Alltag ein nützlicher Begleiter sein. In China tragen Polizist*innen AR-Brillen, um Daten mit der Nationalen Polizeidatenbank in Echtzeit abzugleichen und somit Verdächtige zu identifizieren (Bauer 2018b).

In der Automobilindustrie sollen *AR Head-up Displays* in einigen Jahren serienreif für viele Automobile angeboten werden. Durch AR können Informationen, wie Geschwindigkeit, Navigation, Spurassistent und Abstand vor dem Auto angezeigt werden (Continental 2019).

Zum **Auslösen von digitalen Impulsen** in AR gibt es zahlreiche nicht räumlich repräsentierte digitale Schnittstellen am Markt. Beim Blick durch die Smartphone-Kamera, durch AR-Brillen wie *Focals* oder durch spezielle Headsets wie *Microsofts HoloLens* können Nutzer*innen per Klick-, Sprach- oder Gestensteuerung mit virtuellen Objekten interagieren.

Haptische Hologramme und volumetrische Displays befinden sich erst in der Entwicklung. Solche Schnittstellen eröffnen neue Möglichkeiten der Interaktion von Menschen mit digitalen Geräten im Raum (z. B. Auslösung durch Berührung des Hologramms) und damit auch neuer, sinnlicher Erfahrungen der virtuellen Realität ((Warnke et al. 2019), S. 48).

Langfristige Visionen

Das Ziel der aktuell miteinander konkurrierenden Technologien ist ein möglichst unauffälliges Gerät, das die virtuelle Welt nahtlos über die natürlichen Sinne der Nutzer*innen legt. Dabei können Smartphones, Brillen, Kontaktlinsen oder sogar kybernetische Implantate, die die digitalen Objekte direkt an menschliche Nerven übermitteln, zum Einsatz kommen (Warnke et al. 2019). Mit **nahtloser AR** würde jeder Mensch seine oder ihre personalisierte Version der Welt sehen. Alles, was wir in unserem Blickfeld sehen, wird in die jeweilige Sprache der Nutzer*in übersetzt. Im Flugzeug oder Theater gibt es personalisierte Schilder, die den Sitzplatz anzeigen oder die Farbe des Himmels passt sich der aktuellen Stimmung der Nutzer*in an. Es gibt virtuelle Haustiere, die einem überallhin folgen (Warnke et al. 2019).

Stabilität der Entwicklungslinien: XR ist im professionellen Bereich bereits weit verbreitet, im privaten Bereich hängt ihre Einführung stark von der bislang nur in einzelnen Gruppen ausgeprägten Nutzer*innenakzeptanz ab. Die Markteinführung innovativer räumlicher Schnittstellen wie haptische Hologramme ist aus heutiger Sicht nicht absehbar.

2.3.4 Digitales Enhancement

Untersuchungsrahmen

Unter ***Enhancement*** werden Eingriffe in die körperliche und geistige Konstitution des Menschen verstanden, die über die Wiederherstellung und Bewahrung seiner natürlichen Fähigkeiten hinausgehen (Fuchs et al. 2002). Sie können sich auf physische Eigenschaften (z. B. Bewegungsabläufe, Sinneswahrnehmungen, Körperkräfte, Ausdauer, Attraktivität), kognitive Fähigkeiten (z. B. Gedächtnisleistungen, Konzentrationsfähigkeit), Verhaltensweisen (z. B. Aggressivität, sozial erwünschtes Verhalten) sowie Emotionen (z. B. Hervorrufen positiver Gefühle) beziehen.

Zur Steigerung ihrer Leistungsfähigkeit benutzen Menschen zunehmend technische Artefakte. In diesem Profil liegt der Schwerpunkt auf Enhancement durch die Nutzung technischer Artefakte, bei denen KI bzw. Digitalisierung essenziell für ihre Funktionalität sind. Einbezogen wird ein breites **Spektrum von Artefakten**, die sich darin unterscheiden, (1) wie eng und reversibel sie mit dem menschlichen Körper integriert sind, (2) ob körperliche, kognitive oder emotionale Eigenschaften und Fähigkeiten des Menschen damit beeinflusst werden sollen, (3) inwieweit der Mensch, der die technischen Artefakte verwendet, deren Funktionen selbst steuert bzw. inwieweit seine Aktivitäten gesteuert werden, (4) inwieweit es sich um die Optimierung der eigenen Person oder einer anderen handelt.

Wie verändert KI das digitale Enhancement?

Schlüsseltechnologien für das Digitale Enhancement wie Gehirn-Computer-Schnittstellen (BCIs) kombinieren Wissen und Techniken aus den Neurowissenschaften, der Signalverarbeitung und des

maschinellen Lernens (Coleman 2018). Die bereits etablierte digitale Selbstvermessung durch am oder im Körper getragene Geräte profitiert in der Datenauswertung und Kombination mit anderen Datensätzen von den Fortschritten beim maschinellen Lernen.

Entwicklungen

Auf dem Gebiet des Digitalen Enhancement haben insbesondere zwei Entwicklungslinien das Potenzial, grundlegende Veränderungen von Mensch-Technik-Umwelt-Beziehungen anzustoßen: Die Selbstvermessung und Brain Computer Interfaces (BCIs).

In bestimmten Gruppen der Gesellschaft ist die **Selbstvermessung** fest etabliert: Eine Person misst sich aktiv mit Geräten und Applikationen, um auf Basis der Analyseresultate Lebensstil und Verhalten in den Bereichen Fitness/Leistung, Wellness und Gesundheit zu optimieren und die Motivation dafür zu erhalten (Scheermesser et al. 2018, S. 57; Meidert et al. 2018, S. 44).

Relevante Geräte sind insbesondere Wearables, d.h. reversibel mit dem Körper verbundene kleine Computer (wie z. B. Smart Watches, Fitnessarmbänder und digitale Brillen), die i.d.R. Körperparameter, aber auch Verhalten, Gefühlszustände und gesundheitsrelevante Symptome messen; im Falle der Neurotechnologie auch Kappen, mit denen Hirnströme gemessen bzw. Nerven stimuliert werden können (Ienca et al. 2018). Online-Plattformen dienen dem Datenaustausch, der Datenauswertung und der Interaktion mit der Selbstvermessungs-Community (Albrecht et al. 2016; Evers-Wölk et al. 2018). Die Datensätze aus der Selbstvermessung können mit Metadaten (z. B. Datum, Aufenthaltsort) und weiteren Datensätzen (z. B. Lebensstil, Einkäufe) verknüpft werden.

Etwa 75 % der kommerziell erhältlichen Artefakte und Apps sind dem Lifestyle-Bereich zuzuordnen. Die Betreiber der digitalen Plattformen fördern durch Bereitstellung kostenloser Apps die Verbreitung von Selbstvermessungspraktiken. Ihre Geschäftsmodelle beruhen auf der ökonomischen Verwertung der durch die Selbstvermessenden erhobenen Daten, indem sie mit Daten aus anderen Quellen zusammengeführt und Nutzerprofile erstellt werden. Etwa 25 % der Angebote sind für Krankheitsmanagement und -behandlung bestimmter Erkrankungen (z. B. Diabetes, Depressionen und Angststörungen, Multiple Sklerose, Parkinson etc.) gedacht und dienen der Kontrolle und Dokumentation bestimmter Körperparameter, dem Monitoring und der Förderung der Therapieadhärenz (Meidert et al. 2018; Scheermesser et al. 2018; Evers-Wölk et al. 2018; Albrecht et al. 2016; Mück et al. 2019).

Durch die digitale Selbstvermessung wird eine Datengrundlage zur Identifizierung und Charakterisierung von Bevölkerungsgruppen gelegt, und der Übergang von Wearables zu dauerhaft in den Körper integrierten intelligenten Implantaten und BCIs ist fließend.

Brain-Computer-Interfaces (BCI) umfassen eine große Gruppe von Artefakten, die eine direkte Verbindung zwischen Hirnaktivitäten und Computer ermöglichen, ohne dass Dateneingaben z. B. über Tastaturen oder Mikrofone erforderlich sind (Coleman 2018). Messgrößen sind i.d.R. elektrische Signale von Nervenzellaktivitäten oder Indikatoren für Hirnstoffwechselaktivitäten. BCI können reversibel mit dem Körper verbunden oder implantiert sein. Entsprechende Geräte sind in der Forschung und der Medizin etabliert, teilweise auch im kommerziellen Direct-to-Consumer-Bereich. Zu unterscheiden sind (Roelfsema et al. 2018):

- „nur lesende" BCI, die die Hirnsignale des Nutzers an einen Computer weiterleiten, wo sie verarbeitet und in Aktionen (z. B. Steuerung von künstlichen Gliedmaßen) umgesetzt oder ausgelesen (z. B. "Lügendetektor", „Gedankenlesen“) werden.[22]
- „nur schreibende" BCI, die die Gehirnaktivität beeinflussen, z. B. durch Stimulation bestimmter Hirnregionen. Hierzu zählen beispielsweise Retinaimplantate, die Lichtreize in elektrische Impulse umwandeln, die über den Sehnerv ans Gehirn weitergeleitet werden, um dort ein Bild zu erzeugen.[23]
- Closed-Loop BCI, die sowohl „lesen und schreiben" können und i.d.R. bestimmte Hirnregionen nur in Abhängigkeit vom zuvor gemessenen Zustand stimulieren.

Wurden BCI-Technologien ursprünglich für eng begrenzte Fragestellungen in spezifisch regulierten Bereichen in spezialisierten Einrichtungen im Forschungs-, Medizin- und Rehabilitationsbereich entwickelt und eingesetzt, so ermöglichen immer leistungsfähigere, miniaturisierte und portable Geräte mit wachsendem Funktionsumfang die Nutzung für eine steigende Anzahl von Zwecken bis hin zu einer Alltagsanwendung „für jedermann".

Künftig realisierbar erscheinen Anwendungen des **Kognitiven Enhancement** bzw. der *Augmented Cognition*, bei denen in den BCI vorhersagende, beratende und automatisierte Funktionalitäten einschließlich des permanenten Monitorings der Hirnaktivität in Echtzeit integriert und kombiniert werden.[24] Entsprechende Assistenzsysteme sind im militärischen Bereich von großem Interesse, da Streitkräfte bei Einsätzen in Gefechten, in verschiedenen Klimazonen und Umgebungen hohen physischen und psychische Belastungen ausgesetzt sind, deren Bewältigung von großer Bedeutung für den Gefechtserfolg ist. Daher werden **Human Performance Enhancement** (kurz: HPE)-Systeme entwickelt und erprobt, um Ausdauer und Resilienz der Soldatinnen und Soldaten zu erhöhen.[25]

[22] Elektroenzephalografie-Kappen (EEG-Headsets) zur Messung und zum Monitoring von Hirnaktivitäten werden Privatpersonen zu Preisen von wenigen hundert Euro angeboten. Diese Geräte zeichnen Hirnstromsignale auf, verbunden mit Software, die eine Auswertung der Rohdaten für verschiedene, buchbare Zwecke ermöglicht. Diese Headsets sollen dem Trainieren der kognitiven Fähigkeiten oder der Verbesserung von Konzentration, Meditation, Entspannung/Schlaf dienen, oder aber der Steuerung von Geräten, z. B. beim Gaming (Ienca et al. 2018).

[23] Geräte zur Beeinflussung von Nerven- und Hirnaktivitäten werden kommerziell für wenige hundert Euro angeboten, die nicht-invasiv bestimmte Hirnregionen mit magnetischen, elektrischen oder Ultraschall-Impulsen stimulieren, was z. B. der Verbesserung der Konzentration, dem Gehirntraining, der Behandlung und Vorbeugung von Migräne oder der Behandlung von Depressionen, Angstzuständen und Schlaflosigkeit dienen soll (Ienca et al. 2018). Im Tierversuch konnten nicht nur Bewegungen des Tiers von extern gesteuert werden, sondern auch Verhalten wie Fressen, Trinken und Fortpflanzung. Möglich ist auch das Auslösen von Gefühlen oder das Hervorrufen von Erinnerungen (Roelfsema et al. 2018). Die invasive Neurostimulation ließe sich prinzipiell auch zur gezielten Manipulation von Personen einsetzen.

[24] Hierzu zählen beispielsweise (Roelfsema et al. 2018):

- BCI, die der Erweiterung der Sinneswahrnehmungen dienen: Beispielsweise könnten Cochlear- und Retinaimplantate so gestaltet werden, dass sie z. B. auf Ultra- und Infraschall bzw. ultraviolette und infrarote Strahlung ansprechen.
- BCI, die zu Sinneseindrücken zusätzliche Informationen hinzuspielen (z. B. Hinzuspielen von Namen bei Gesichtserkennung) oder die Aufmerksamkeit auf die relevanten Bestandteile lenken (z. B. beim Autofahren auf potenzielle Gefahren, die der Aufmerksamkeit des Fahrenden entgehen („Ball rollt auf die Straße").
- BCI, die das Ausführen komplexer oder feinmotorischer Bewegungs- und Verhaltensmuster ermöglichen, die normalerweise nur mit langjähriger Übung und umfassendem Training beherrscht werden („Laie fliegt Flugzeug").
- BCI, die es ermöglichen, dass ein Teil der Datenverarbeitung im Gehirn erfolgt, ein Teil im externen Computer. Hierdurch könnten der „Arbeitsspeicher" des Kurzzeitgedächtnisses vergrößert, das Langzeitgedächtnis verbessert, die Assoziationen zwischen Konzepten (z. B. Gesichter und Namen) verbessert sowie neue Problemlösungsstrategien ermöglicht werden.
- BCIs, die Gedanken und Informationen (d.h. detaillierte Wahrnehmungen, Überlegungen, Absichten, Neigungen, Stimmungen, Gefühle) zwischen Personen mit höherer Effizienz als über Sprache oder Schrift übertragen.

[25] Beispielsweise hat sich die Bundeswehr auf ihrer 52. Sicherheitspolitischen Informationstagung der Clausewitz-Gesellschaft e.V. und der Führungsakademie der Bundeswehr (FüAkBw) 2018 mit dem Thema KI befasst, wobei auch HPE thematisiert wurde (Hoffmann und Scheffler 2018). Aktuell durchgeführt wird das Projekt Human Performance Enhancement: Smart Textiles und Augmented Reality (HPE-STAR).

Langfristige Visionen

Weitgehend (noch) im Bereich des Science Fiction angesiedelt sind Visionen von **Cyborgs** (Röcke 2017, S. 328 f.), d.h. Mensch-Maschine-Mischwesen mit einem hohen Grad der Technisierung des Körpers, die durch dauerhaft in den Körper integrierte intelligente Implantate übernatürliche Fähigkeiten erlangen (**Transhumanismus**). Bei diesen biovisionären Entwürfen geht es um eine Verbesserung der menschlichen Natur (schneller rennen, besser hören, neue Sinne, besser denken, etc.) bis hin zur Lebensverlängerung und Unsterblichkeit.

Stabilität der Entwicklungslinien: Die digitale Selbstvermessung schreitet stabil voran und etabliert sich außerhalb der speziellen Community der Selbstvermessenden zunehmend in breiteren Bevölkerungsgruppen, u. a. befördert beispielsweise von Krankenkassen oder der Gesundheitspolitik („App auf Rezept"). Dagegen ist unklar, inwieweit BCI den Bereich der Forschung, der Medizin sowie der behinderungskompensierenden Technik verlassen werden. Wahrscheinlich ist eine Umsetzung von BCIs für spezielle Aufgaben, nicht aber der Einsatz von BCIs als gesellschaftlich verbreitete Normalität. Die Machbarkeiten des Uploads und Downloads von Gedanken und ihre Repräsentationen im Computer sind zweifelhaft.

2.3.5 Autonome Systeme zur Erschließung von bislang für den Menschen unverfügbaren Räumen

Untersuchungsrahmen

Bislang für den Menschen weitgehend unverfügbare Räume sind in erster Linie lebensfeindliche Räume. Letztere sind dadurch gekennzeichnet, dass die Belastung und Gefährdung des Menschen mit der Aufenthaltsdauer wachsen, dass unzumutbar hohe Risiken für den Menschen bestehen oder dass Menschen die Umgebung nur mit spezieller Schutzausrüstung betreten können (AG Lebensfeindliche Umgebungen 2019). Weite für den Menschen lebensfeindliche Räume entziehen sich bislang einer Nutzung für **Wirtschafts- oder Siedlungszwecke**.

Hinsichtlich der durch die Digitalisierung geschaffenen Expansionsmöglichkeiten ragen zwei lebensfeindliche Umgebungen heraus, der Weltraum und die Ozeane.[26] Die Kármán-Linie in 100 Kilometern Höhe über dem Meeresspiegel grenzt den **Weltraum** und die Erdatmosphäre voneinander ab. Der Weltraum ist für den Menschen aufgrund von Schwerelosigkeit, Temperaturextremen und fehlender Atemluft lebensfeindlich. Es sind jedoch bestimmte energetische und stoffliche Ressourcen verfügbar und es herrschen physische Bedingungen, die für bestimmte Produktionsprozesse vorteilhaft sind. Das **Meer** bedeckt rund zwei Drittel der Erdoberfläche. Anrainerstaaten üben staatliche Hoheitsrechte bis zur ausschließlichen Wirtschaftszone (maximal 200 Seemeilen vom Festland entfernt) aus, dahinter schließen sich der Festlandssockel und die hohe See an (WBGU 2013). Das Meer ist für den Menschen unter anderem deshalb lebensfeindlich, weil dort wesentliche Ressourcen für das Überleben, wie z. B. Süßwasser, fehlen. Gleichzeitig bietet das Meer zahlreiche Ressourcen, die für eine wirtschaftliche Nutzung interessant sind. Der Weltraum, der maritime Festlandssockel und die hohe See zählen zu den globalen Gemeinschaftsgütern, für die besondere zwischenstaatliche Regelungen gelten. Einzelne Akteure versuchen sich, diese Gemeinschaftsgüter anzueignen.

[26] Darüber hinaus gibt es verschiedene terrestrische lebensfeindliche Umgebungen, wie z. B. Wüsten (Antarktis, Sahara) und die Wildnis (u. a. tropischer Regenwald), von Katastrophen betroffene Regionen (u. a. Brände, Erdbeben, Kontaminationen) oder Regionen, die durch Geoengineering umgestaltet werden (u. a. nukleare Endlager).

Wie verändert KI die Erschließung schwer zugänglicher Räume?

Lernende Systeme in lebensfeindlichen Umgebungen sind technische Systeme, „die einerseits über gewisse Autonomieeigenschaften und maschinelle Intelligenz verfügen und andererseits adaptiv und lernfähig sind“ ((AG Lebensfeindliche Umgebungen 2019), S. 6). Sie werden für Aktivitäten im Weltraum und in der Tiefsee zur Kostenreduzierung und zur Performanceerhöhung eingesetzt, zur Überwindung menschlicher und technischer Steuerungsbarrieren und in lokalen Entscheidungskontexten (Fong 2018). Hierdurch werden Risiken für das Personal verringert, die Reaktionszeit verkürzt und Fähigkeitslücken geschlossen.

Entwicklungen

Künstliche Intelligenz, Autonome Systeme und andere Digitalisierungstechnologien zur Erschließung lebensfeindlicher Räume sind nicht unbedingt neu, aber mit ihrer wachsenden Leistungsfähigkeit spielen sie die Rolle von Enablern. Die Entwicklungen werden differenziert nach Weltraum und Meer und den jeweiligen Zwecken der wirtschaftlichen Nutzung und Besiedlung eingeschätzt. Zudem werden zwei Anwendungen charakterisiert, die im Hinblick auf Nachhaltigkeitstransformationen als Chancen erscheinen, autonome Systeme zum Sammeln von Abfall aus Gewässern und autonome Systeme zur Aufforstung.

In Bezug auf die Nutzung des Weltraums wird in größerem Umfang in autonome Systeme für die **extraterrestrische Produktion** investiert. Hierunter fallen die Rohstoffgewinnung (*Space Mining*) und die Herstellung von Produkten (*Space Manufacturing*) im „leeren" Weltraum, auf dem Mond, auf Asteroiden/Kometen und anderen Planeten (Feireiss und Najjar 2018). Im Weltraum gibt es besondere physische Bedingungen (Vakuum, Mikroschwerkraft, Abwesenheit von Luftreibung, Temperaturextreme, Verfügbarkeit von Energie und Ressourcen etc.), die bestimmte Prozesse ausschließlich oder nur dort sehr effektiv ausführen lassen. Die Firma *Made In Space* nutzt bereits die einzigartigen physischen Bedingungen des Weltraums für die Entwicklung von Materialien mit veränderten Eigenschaften und neue Produktionsprozesse, wie z. B. die Herstellung von optischen Fasern unter Mikrogravitationsbedingungen. Der Asteroid-Bergbau könnte eine größere Vielfalt und Volumina an Materialen erbringen als der Tiefseebergbau (Gheorghiu et al. 2017). Durch Nutzung autonomer Systeme für die extraterrestrische Produktion kann weitgehend auf die kostspielige bemannte Raumfahrt verzichtet werden und die Raumfahrt trägt durch die kommerziellen Zwecke selbst zu ihrer Refinanzierung bei. Der **Weltraumtourismus** mit suborbitalen Raumfahrzeugen (*Spaceliner)* und Weltraumhotels ist eine stabile Entwicklungslinie der bemannten Raumfahrt; die Bezüge zu Digitalisierung sind jedoch gering.

Langfristige Visionen

Die NASA verfolgt eine dauerhafte Präsenz auf dem Mond unter Etablierung einer Mondwirtschaft (***Lunar Economy***) (NASA 2019). Hierzu sollen Robotiklabore in den Weltraum gebracht werden, die dort Raumfahrzeuge fertigen, Mondmaterial mittels 3D-Druck formen und daraus Infrastrukturen errichten.

Die NASA (NASA 2019) strebt auch bemannte Missionen auf den Mars an. Finanzstarke und mächtige Akteure (u. a. Elon Musk) propagieren die Vision, dass eine ausgesuchte Gruppe an Menschen **die Erde dauerhaft verlassen** und im Katastrophenfall das Überleben der Spezies Mensch im Weltraum mit Hilfe von autonomen, selbsterhaltenden Systemen sicherstellen soll.[27]

[27] Die sozialen Schwierigkeiten eines solchen selbsterhaltenden Systems sind anhand des Biosphäre 2 Experimentes deutlich geworden, was T.C. Boyle in seinem Roman "Die Terranauten" anschaulich beschrieben hat (Boyle 2017).

Stabilität der Entwicklungslinien: Die durch autonome Systeme ermöglichte extraterrestrische Produktion ist hinreichend weit entwickelt und zieht Investitionen an sich, wohingegen ein dauerhaftes Verlassen der Erde auch nur einer winzigen Gruppe an Menschen spekulativ ist.

Hinsichtlich des **Tiefseebergbaus** werden der Abbau von Methanhydrat für energetische Zwecke ((WBGU 2013), S. 205) und die Ausbeutung von Manganknollen, Kobaltkrusten und Massivsulfiden für die Gewinnung mineralischer Rohstoffe (WBGU 2013), S. 53f.)) anvisiert. Die These „Über 30 % der Ressourcen für Elektronik stammen aus der Tiefsee" halten 16 von 32 Antwortende einer Delphi-Befragung vor 2040 für realisierbar (Gheorghiu et al. 2017). Die Plattform „Lernende Systeme" hat ein Szenario „Unter Wasser autonom unterwegs" entwickelt: Hierbei setzt ein Schiff ein lernendes robotisches Unterwassersystem aus, das Unterwasseranlagen inspiziert und repariert und dessen Daten ausgelesen und mit KI analysiert werden (AG Lebensfeindliche Umgebungen 2019).

Es gibt derzeit bereits eine Reihe bewohnbarer **Unterwassereinrichtungen** ((Warnke et al. 2019), S. 171), darunter das Meeresfahrzeug *SeaOrbiter*, die sich selbsttragende Sub-Biosphäre 2 für bis zu 100 Menschen, die *Atlantica Expeditions* (erste Unterwasserkolonie) und submarine Strukturen von staatlichen US-Einrichtungen. UN-Habitat hat dazu aufgerufen, **schwimmende Städte** als eine brauchbare tragfähige Lösung zur Anpassung an die Erderhitzung und an urbane Herausforderungen zu entwickeln (Williams 2019; UN Habitat 2019; Sandoval 2019). Ähnlich wie bei der Errichtung künstlicher Inseln stellen sich auch bei der Errichtung von schwimmenden Städten Fragen nach der Bewertung der Flächeninanspruchnahme und der Raumordnung neu. Die Realisierung von schwimmenden Städten ist durch Investoren wie Peter Thiel (Cosgrave 2017; Garfield 2018) in Reichweite gelangt. Analog zu den heutigen Ölplattformen (Mueller und Massaron 2018) benötigen solche schwimmenden Städte auch aufwändige Sensorik und KI-Anwendungen, um sie an den jeweiligen Meereszustand sicher anzupassen.

Langfristige Visionen

In einer ***Blue Economy*** könnte der durch autonome Systeme technisch-ökonomisch realisierbare Tiefseebergbau langfristig einen wesentlichen Anteil an der Versorgung der globalen Wirtschaft mit Energie und mineralischen Rohstoffen beisteuern (Waterborne Technology Platform 2019).

Gemäß der Vision des ***Seasteading*** könnten sich dauerhaft besiedelte schwimmende Städte und Unterwasserhabitate als maritime Mikrostaaten mit eigenen Regeln des Zusammenlebens ausstatten und damit eine lebensweltliche Perspektive zum Leben an Land bieten.

Stabilität der Entwicklungslinien: Methanhydratabbau in der Tiefsee und meeresbezogenes *Climate Engineering* sind hinreichend weit entwickelt und ziehen Investitionen an sich, wohingegen das menschliche Leben in schwimmenden Städten oder Unterwasser, auch falls es über ein experimentelles Stadium hinausgehen sollte, in absehbarer Zukunft auf sehr kleine Gruppen von Menschen beschränkt bleiben würde.

Plastikmüll gelangt indirekt über die Flüsse und über direkten Eintrag in die **Meere**. Alleine im großen pazifischen Müllwirbel treiben 45–129 Tausend Tonnen Plastik in einem Gebiet von 1,6 Millionen Quadratkilometern (Lebreton et al. 2018). Verschiedene Plastikarten haben im Meer unterschiedliche Lebensdauern, die bis zu mehreren Hundert Jahren dauern kann. Zudem kann Plastik durch UV-Strahlung, mechanische und biologische Prozesse in mikroskopisches Plastik, das ein hohes Gefährdungspotenzial für die marine Umwelt und marine Organismen aufweist, umgewandelt werden (Ackermann 2018; WBGU 2013).

Zahlreiche Anstrengungen, den Plastikmüll aus den Meeren zu entfernen sind in den letzten Jahren publik geworden. Der *Ocean Cleanup Interceptor* soll in der Lage sein, autonom Kunststoff aus den Flüssen zu sammeln, bevor dieser das Meer erreicht. „Nach vierjähriger Entwicklungszeit stehen nun die ersten vier der Vorrichtungen bereit." (dw 2019). In England sammelt der sogenannte *WasteShark* bereits bis zu 60 Kilogramm Abfall pro Fahrt im küstennahen Bereich (Gabbatiss 2019). Im aktuellen FuE-Projekt „Mobiles autonomes Unterwassersystem" werden Unterwasserfahrzeuge entwickelt, die zukünftig im Meer autonom Müll sammeln oder Altmunition aufspüren können (Holbach 2019). Hierbei sollen neue Missionen durch KI von vorherigen Missionen lernen. Ein anderes Konzept zum Einfangen von Plastik aus dem Meer beruht auf langen linienförmigen Einheiten (z. B. aus Kork), an denen Sammeleinrichtungen befestigt sind. Solche Einrichtungen sind passiv, treiben auf dem Meer und benötigen keine Energie (Cockburn 2019).

Langfristige Vision

Autonome Roboterflotten im Meer könnten massenhaft Müll aus den Weltmeeren sammeln und an Land zur Entsorgung zurückbringen. Der pazifische und andere Müllwirbel würden in die Anthroposphäre zurückgeholt werden können. Solche Visionen treiben die Entwickler von autonomen Roboterflotten und anderer passiver Konzepte an (Cockburn 2019; dw 2019; Gabbatiss 2019; Holbach 2019).

Stabilität der Entwicklungslinien: Derzeit gibt es viele einzelne Projekte für autonome Roboterflotten zur Säuberung der Meere im Pilotstadium. Ein massenhafter Einsatz scheint technisch möglich zu sein, politisch-ökonomisch müssten jedoch entsprechende Mittel und Maßnahmen zur Hochskalierung der Einzelaktivitäten mobilisiert werden.

Die **Aufforstung** wird als eine der wichtigsten langfristigen Klimaschutzoptionen angesehen. Das globale Wiederaufforstungspotenzial wird auf 0,9 Milliarden Hektar zusammenhängende Waldflächen geschätzt (Bastin et al. 2019). Dies entspräche einer Ausweitung der heutigen globalen Waldfläche um 25 %. Die größten Potenziale finden sich in den sechs Ländern USA (103 Millionen Hektar); Kanada (78,4 Millionen Hektar); Australien (58 Millionen Hektar); Brasilien (49,7 Millionen Hektar); und China (40,2 Millionen Hektar) (Bastin et al. 2019). In anderen Weltregionen wird Wald in erster Linie zur Verbesserung des Mikroklimas, zur Verringerung der Bodendegradation und für die land- und forstwirtschaftliche Nutzung aufgeforstet. *The Great Green Wall of the Sahara and the Sahel* (GGW) ist eine im Jahr 2007 ins Leben gerufene multinationale Anstrengung, um durch Aufforstung von rund 100 Millionen Hektar Land quer über den Afrikanischen Kontinent vom Senegal bis Djibouti (8.000 Kilometer) die Landdegradation einzudämmen (Goffner et al. 2019). Laut Eigenauskunft sind 15 % der anvisierten Aufforstung bereits umgesetzt.[28]

Aufforstung ist eine harte, arbeitsintensive, oft auch gefährliche Tätigkeit für den Menschen. Es wird von zahlreichen Projekten, die Wiederaufforstung lebensfeindlicher Räume technisch zu unterstützen, berichtet. eine Der jüngst entwickelte autonome Baumpflanzroboter *TreeRover* könne viel schneller als Menschen Bäume pflanzen (Mielczarek 2018). Der *GrowBot* der Firma SkyGrow ist ein unbemanntes Landfahrzeug, das laut Firmenangabe Bäume zehnmal schneller pflanzen kann als ein trainierter Mensch (Simpson 2018). Drohnen von BioCarbon Engineering, die Setzlinge in den Boden schießen, können laut Eigenauskunft der Firma 100.000 Bäume an einem einzigen Tag pflanzen (Grossmann 2017).

[28] vgl. https://www.greatgreenwall.org/about-great-green-wall

Langfristige Vision

Die Firma SkyGrow verfolgt die Vision, Bäume schneller zu pflanzen als durch Landnutzungswandel und Abholzung verloren gehen. Das robotische Pflanzsystem soll hochgradig skalierbar sein (Simpson 2018).

Stabilität der Entwicklungslinie: Derzeit gibt es einige Projekte für Baumpflanzungen durch autonome Roboter oder Drohnen. Ökobilanzen autonomer Pflanzsysteme und eine Gegenüberstellung ihrer CO_2-Emissionen über den Lebenszyklus hinweg gegenüber dem durch sie erzielbaren CO_2-Senkungspotenzial liegen nicht vor. Ein massenhafter Einsatz scheint technisch möglich zu sein, politisch-ökonomisch müssten jedoch entsprechende Mittel und Maßnahmen zur Hochskalierung der Einzelaktivitäten mobilisiert werden.

2.3.6 Big Data Gesellschaft

Untersuchungsrahmen

Big Data ist durch fünf Merkmale charakterisiert (Salzig 2016): *Volume* (die enormen Datenmengen), *Variety* (die Vielfalt der Datentypen und -quellen), *Velocity* (die Geschwindigkeit, mit der Daten generiert, ausgewertet und weiterverarbeitet werden), *Validity/Veracity* (die Qualität bzw. Wahrhaftigkeit der Daten) sowie *Value* (der Mehrwert, der aus Daten generiert werden kann). Unter dem Begriff **Big Social Data** wird der Zusammenhang von Technologien zur Erfassung, Vorhersage, Bewertung, Kontrolle und der Steuerung des Verhaltens von Agenten auf der Grundlage von Datenauswertungen zusammengefasst. Agenten können Einzelpersonen, Personengruppen oder auch digitale Agenten wie z. B. Trading Bots sein. Unter dem Begriff der **Big Data Gesellschaft** werden die sich aus Big Social Data ergebenden gesamtgesellschaftlichen Folgen diskutiert. Hierzu gehört eine Vielzahl an grundlegenden Umwälzungen des gegenwärtigen sozialen Lebens, der Wirtschaft und des staatlichen Handelns (Kind und Ferdinand 2018). Von besonderem Interesse sind **Veränderungen von Entscheidungsprozessen** durch Big Social Data (Magrabi und Bach 2013). Grundsätzlich wird zwischen analytischem und intuitivem Entscheiden unterschieden, wobei Emotionen eine wesentliche Rolle spielen können. Weitläufig wird auch zwischen wertbasierter, sozialer und perzeptueller Entscheidungsfindung unterschieden.

Wie verändert KI die Big Data Gesellschaft?

Mittels **KI** können durch die Verknüpfung von Datensätzen (z. B. zu neuen Kennwerten) Muster erkannt, statistische Zusammenhänge konstruiert und Entscheidungsgrundlagen geschaffen werden (Kelleher und Tierney 2018). Das Treffen von Entscheidungen durchdringt die ganze Disziplin der KI, wobei unter anderem Markov-Entscheidungsprozesse (Aktionsketten, die einen Ausgangszustand mit einem Zielzustand verbinden) und agentenorientierte Ansätze eingesetzt werden (Magrabi und Bach 2013). Das von der steigenden Verfügbarkeit großer und heterogener Echtzeitdatensätze profitierende maschinelle Lernen, hat die Herausbildung einer Big Data Gesellschaft wesentlich befördert.

Entwicklungen

Haupttreiber für die Big Data-Gesellschaft sind die scheinbar objektiveren, genaueren und effizienteren Entscheidungsgrundlagen durch Big Data sowie die tatsächliche Verfügbarkeit von immer mehr und vielfältigeren Daten mit hoher zeitlicher und räumlicher Auflösung in Bezug auf Personen, Organisationen, Objekte und die Umwelt (Li 2018). KI-Anwendungen für Big-Data-Analysen sind durch große global agierende digitale Plattformunternehmen aus den USA (Amazon, Apple, Google, Microsoft) und von in China ansässigen Unternehmen (Baidu, Alibaba,

Tencent) in enger Kooperation mit staatlichen Einrichtungen vorangetrieben worden (OECD 2018). Blockchain und Quantencomputing könnten die Datensicherheit in einer Big Data-Gesellschaft zukünftig grundlegend verbessern und damit zu einem weiteren starken Treiber avancieren (Warnke et al. 2019).

Anwendung finden KI-basierte Big-Data-Analysen in den unterschiedlichsten sozialen Bereichen wie Konsum und Werbung, staatliche Kontroll- und Sicherheitspolitik, politische Prozesse wie Wahlen, Rechtsprechung und Finanzprodukthandel. Die Sammlung und Verwertung von Daten wird im Wesentlichen durch die digitalen Plattformunternehmen und staatliche Akteure gesteuert, wobei etwaige Gemeinwohlinteressen dabei oft nicht in der breiten Öffentlichkeit mitverhandelt werden. Einzelne Anwendungen wie das Microtargeting von Konsumenten beim Online-Handel (Kahlenborn et al. 2019)[29] oder digitale Medienkompetenz[30] werden bereits öffentlich hinsichtlich ihrer Konsequenzen diskutiert.

Im Hinblick auf grundsätzliche Veränderungen von Mensch-Technik-Umwelt-Beziehungen lohnt es sich, einen Blick auf zwei in der Transformationsforschung kaum rezipierte Entwicklungslinien zu werfen: Erstens, das Scoring und Microtargeting und, zweitens, Big Data für Finanzentscheidungen.

Das **Scoring und Microtargeting** von Personen im öffentlichen Sektor wird vor allem im Zusammenhang mit dem *Social Scoring System* in China thematisiert. Je nach Region gibt es dort unterschiedliche Ausprägungen beispielsweise für die Kreditvergabe (*Social Credit System, SCS*), zur Belohnung erwünschten Verhaltens (z. B. Nutzung des ÖPNV) oder zur Bestrafung unerwünschten Verhaltens (z. B. Littering). China plant bis 2020 die landesweite Einführung eines *SCS* (Kind und Ferdinand 2018). Dort kann das SCS auf privatwirtschaftliche wie auf staatliche Datenbanken zugreifen, die Auskunft über Zahlungsmoral, Strafregister, Einkaufsgewohnheiten, digitales Surf- und Kommunikationsverhalten geben. Mit Unterstützung von KI wird auf Basis dieser Daten eine Punktsumme ermittelt, die darüber entscheidet welche Aussichten die betreffende Person auf die Gewährung eines Kredits, eine berufliche Beförderung oder ins Ausland reisen zu dürfen hat (Kühnreich 2018).

In den USA wird das politische Microtargeting seit 2008 als Wahlkampfinstrument genutzt (Kind und Weide 2017). Die Parteien (Demokraten und Republikaner) nutzen Big Social Data zur Erstellung psychologischer Profile von Personen und Nutzergruppen, um Wahlwerbung effektiver zu platzieren. Im Rechtswesen einiger US-Bundesstaaten findet bereits die Software COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) Anwendung, die die Rückfallwahrscheinlichkeit von Straftätern berechnet und in die Urteilsverkündigung des Gerichts eingeht. Die Sammlung und Auswertung von Daten, die bei der Nutzung von digitalen Diensten, insbesondere der großen Internetplattformen wie Amazon, Google oder Facebook entstehen, zur Generierung von individuellen oder gruppenbezogenen Konsumprofilen, ist auch in Deutschland weithin bekannt. Diese Profile stellen für die Internetplattformen wiederum eine begehrte Handelsware dar, deren Kunden dem einzelnen Nutzer in der Regel nicht bekannt sind.

In Deutschland wie auch in Europa beschränkt sich die Nutzung von Big Data zum Großteil auf wirtschaftliche Kontexte, wie Bonitätsprüfung, personalisierte Werbung oder Personalrekrutierung für Unternehmen. Arbeitgeber und Personalagenturen sammeln, analysieren und verdichten im Internet zugängliche Datenbestände mit Hilfe von KI zu

[29] Erste Konzepte zur Ausrichtung beispielsweise von Empfehlungssystemen am Gemeinwohl statt am vermeintlichen Individualnutzen und maximalem Unternehmensgewinn verharren angesichts der Akteurs- und Machtkonstellationen bislang weitgehend im Gedankenstadium (Veenhoff 2019).

[30] Nutzer*innen bedürfen nach Auffassung einzelner Akteure einer kritischen Befähigung zur Nutzung digitaler Plattformen. Kritische Befähigung bedeutet auch Manipulationen (z. B. Confirmation Biases und Filterblasen über soziale Medien) entgegenzuwirken (vgl. u. a. Remotti et al. 2016).

Personenprofilen, die es erlauben sollen, die Eignung von potenziellen Arbeitnehmern für eine Arbeitsstelle besser beurteilen zu können. Ein weiteres potenzielles Einfalltor für das sozioökonomische Scoring und Microtargeting in Deutschland sind Systeme zur Beurteilung der finanziellen Bonität. An solchen Informationen haben insbesondere Finanzdienstleister und die SCHUFA (u. a. Kreditwürdigkeit), aber auch beispielweise Vermieter ein grundsätzliches Interesse. In einer kürzlich abgehaltenen repräsentativen Befragung befürwortete immerhin jeder Sechste in Deutschland ein *Social Scoring System* nach chinesischem Vorbild (YouGov 04.02.2019). Zusammengenommen lassen diese Entwicklungen die schleichende und verdeckte Entstehung von Social Scoring Systemen zum Microtargeting auch in Deutschland als nicht ausgeschlossen erscheinen.

Das Finanzsystem mit seinen Preisbildungsmechanismen auf den Märkten und an den Börsen wird als gigantische Informationsverarbeitungsmaschine gesehen, die dezentrale Signale aggregiert, in monetäre Einheiten transformiert und damit erst vergleichbar macht (Harari 2016). Viele Investoren arbeiten mit deterministischen Handelsrobotern (**Trading Bots**), die Handelsgeschäfte mit Aktien, Anleihen, Wertpapieren und Fonds automatisch platzieren. Die gebräuchlichsten algorithmischen Handelsstrategien folgen Trends in gleitenden Mittelwerten, Ausbrüchen aus einem Korridor und Preisniveauänderungen (DeCleene 2019). Beim automatisierten Hochfrequenzhandel (HFT) wird eine große Zahl kleiner Aufträge in hoher Geschwindigkeit in den Markt eingespeist.

Wurden der Handel mit Finanzprodukten bislang vor allem als automatisierter Hochfrequenzhandel abgewickelt, so wird zunehmend KI zum Erkennen und Bewerten von Signalen für Änderungen der Güte von Finanzanlagen (Börsendaten, Unternehmensmeldungen, Nachrichten, soziale Medien, etc.) eingesetzt (Harari 2016). KI-Handelsroboter können aus Erfahrung und Fehlern lernen und selbst den Algorithmus, den sie nutzen, auswählen. Diese erhöhte Fähigkeit zum Einschätzen von neuen Marktsituationen macht smarte KI-Algorithmen gegenüber den bisherigen Trading Bots überlegen (Krauss et al. 2016). Hiermit geht eine Verschiebung von deterministischen hin zu lernenden Modellen einher.

Der *Equbot* listete im Jahr 2018 einen mit KI betriebenen Equity Exchange-Traded Fund (ETF) an der Börse. Er analysiert Unternehmensberichte von 6000 Firmen, 1 Mio. Archivierungen täglich und Daten zur Marktstimmung und wählt danach 30-70 Aktien aus. Der *Equbot* war damit dem Standard & Poors 500 Firmen Standard in seiner Anlagestrategie überlegen. Forscher haben die Leistungsfähigkeit von verschiedenen KI-Typen getestet, die den Aktienmarkt von 1992-2015 simulierten (Krauss et al. 2016). Alle hatten größere Einkünfte als der Markt, wobei Random Forest besser als tiefe neuronale Netzwerke und Gradient-boosted Forest abschnitt. Es gibt immer weiterhin Entwicklungsbedarf hinsichtlich der grundlegenden Analyse (Verhältnis von Semantik und Syntax, Integration von Information in den größeren Kontext) (vgl. u. a. (Baron 2018); (ConsorsBank 2018)).

KI-gestützte Big Data-Analysen sind auch ein wichtiger Faktor für **Dynamic Pricing** (Lippert 2018), bei dem Preise in hoher zeitlicher Dynamik fluktuieren (z. B. wetterbedingt) oder an einzelne Nutzer*innen angepasst werden können (z. B. individualisierte Preise). Hierdurch werden grundlegende ökonomische Annahmen wie die des Gleichgewichtspreises oder die der Konsumentenrente (entfällt bei vollständiger Ausschöpfung der Zahlungsbereitschaft des einzelnen Kunden durch dynamisch angepasste Einzelpreisfestsetzung seitens des Verkäufers) außer Kraft gesetzt ((Hilty et al. 2003), S. 63 ; (Lippert 2018)). Es ist umstritten, ob die Signale und Muster, die aus Big Data extrahiert werden, Finanzentscheidungen effektiv auf ein breiteres und realweltlicheres Fundament stellen und damit zu verbesserten ökonomischen Einzelentscheidungen führen können (Frühauf 2019). Auch besteht jedoch kein empirisch

belastbarer Zusammenhang, dass zahllose optimierte Mikroentscheidungen dem volkswirtschaftlichen Gemeinwohl zuträglich sind.

Langfristige Visionen

Eine **datengetriebene Gesellschaft** kann viele mögliche Facetten aufweisen, darunter die subtile Identifizierung und Erkennung von Personen, ihrer Absichten und Bedarfe durch KI-Analysen und die mittelbare Erzwingung konformen Verhaltens durch Social Scoring und Nudging beim Konsum. Big Data-getriebene Finanzmärkte können neue Investment- und Divestment-Schwerpunkte hervorbringen und Dynamic Pricing kann eine neue Stufe des Kapitalismus fördern, die auch ein neues Verständnis wirtschaftlicher Vorgänge erforderlich macht. In der Politik wird KI nach Auffassung zahlreicher Experten zukünftig EU-weit zur Entscheidungsunterstützung genutzt (Gheorghiu et al. 2017). In Kombination von Big Data in Wirtschaft und Politik kann ein **totalitärer Überwachungskapitalismus** entstehen (Zuboff 2018).

Stabilität der Entwicklungslinien: Das Scoring und Microtargeting ist in begrenztem Ausmaß bereits wenig beachtete gesellschaftliche Praxis, während ein großskaliges staatliches Social Scoring System wie in China in Deutschland und den meisten europäischen Ländern nicht realisierbar erscheint. KI zur Auswertung großer und diverser Datenbestände für den Finanzprodukthandel ist hinsichtlich seines Veränderungspotenzials auf der bestehenden öffentlich zugänglichen Informationsgrundlage schwer einzuschätzen. Stabile Entwicklungslinien aus anderen Profilen mit Relevanz für Big Data sind das Erkennen von Emotionen und Gesten (Affective Computing), die massenhafte Analyse von Sprachdaten (Simulation natürlicher Sprache) und die digitale Selbstvermessung (Digitales Enhancement). Die Integration von Augmented Reality in die Big Data Gesellschaft ist wahrscheinlich, wohingegen ein umfangreiches Pooling von Daten zu Genotyp, Phänotyp, Lebensstil und Umwelt (Entschlüsselung des Lebens durch digitale Werkzeuge) aus heutiger Sicht in Deutschland und den meisten europäischen Ländern kaum realisierbar erscheint.

2.3.7 Hyperkonnektivität

Untersuchungsrahmen

Konnektivität ist eine Bezeichnung für die Verbundenheit von einzelnen Elementen. Der Begriff der Hyperkonnektivität steht für eine **Überfülle an Verbindungen** (Lauchenauer 2018). Analytisch lassen sich eine horizontale und eine vertikale Dimension der digitalen Konnektivität unterschieden (Arbeitskreis Systemaspekte 2018). Die **horizontale Dimension** bezieht sich darauf, welche physischen und virtuellen Elemente miteinander verbunden sind. Hierzu gehören beliebige Objekte wie zum Beispiel Alltagsgegenstände, Fahrzeuge, Geräte und Maschinen, Organisationen wie zum Beispiel Unternehmen und Nichtregierungsorganisationen, Menschen und andere Lebewesen wie zum Beispiel Weidevieh und Bäume. Die **vertikale Dimension** umfasst die Verbindung der physischen Welt mit Sensoren und digitalen Signalwandlern, über die Datenverarbeitung und Vernetzung bis hin zur Management- und Leitebene.

Wie verändert KI die Hyperkonnektivität?

KI kommt sowohl in der Nachrichtenübertragung (u. a. Mustererkennen und Maschinelles Lernen) als auch in der Übersetzung von Informationen auf Feldebene in Entscheidungen auf der Management- und Leitungsebene (u. a. Expertensysteme) zum Einsatz.

Entwicklungen

Wesentlicher Treiber für die Hyperkonnektivität ist der IP V6 Internetstandard, der eine Potenzierung der an das Internet anschließbaren digitalen Elemente ermöglicht. Eine zentrale Rolle spielen auch Leitbilder mit ihren Versprechungen, die Verbesserung und Ausweitung der Netzinfrastruktur (z. B. Einführung von 5G) und rechtliche Vorgaben (z. B. eCall – automatischer Notruf bei Autounfällen). Idealtypisch lassen sich drei zentrale Entwicklungslinien unterscheiden, das Internet der Menschen, das Internet der Dinge (*Internet of Things*, IoT) und das Internet der Natur. Als langfristige übergreifende Vision wird das *Internet of Everything* (IoE) in Aussicht gestellt.

Das **Internet der Menschen** ist bereits weitgehend Bestandteil unseres Alltagslebens. Weltweit gibt es derzeit über 8 Milliarden Mobilfunk- und über 4 Milliarden Internetanschlüsse (Tenzer 2019) für rund 7,6 Milliarden Menschen (Deutsche Stiftung Weltbevölkerung 2018). Menschen sind zunehmend online. Aktuelle Entwicklungsschwerpunkte sind die ubiquitäre Verbreitung des schnellen Internet (u. a. Satellitenprogramme von Space-X[31], Amazon[32]), die verbesserte Anbindung des Menschen an das Internet (insb. Schnittstellen)[33] sowie Tools zur Überwindung menschlicher Kommunikationsbarrieren (vgl. Abschnitt 3.2).

Das **Internet der Dinge** (*Internet of Things*, IoT) ist eine technische Vision, Objekte jeder Art in ein universales digitales Netz zu integrieren. Die darin befindlichen Objekte sind digital repräsentiert und miteinander verbunden. Der Mensch ist in die Interaktion zwischen den Maschinen nicht direkt eingebunden; sie interagieren autonom (Wissenschaftlicher Dienst Bundestag 2012). In Deutschland wird mit dem Konzept von Industrie 4.0 der Akzent konkret auf die Verzahnung der industriellen Produktion und Lieferketten mit Informations- und Kommunikationstechnik gelegt. „Ziel dieser neuen, vierten industriellen Revolution ist der Aufbau einer agilen, mit Künstlicher Intelligenz durchsetzten Planungs-und Fertigungsinfrastruktur, mit der Unternehmen schneller und kostengünstiger auf spezifische Kundenwünsche eingehen können. Am Ende dieses Prozesses soll das seriengefertigte Massenprodukt (das Ergebnis des klassischen industriellen Prozesses) abgelöst werden vom kundenindividuell angefertigten Gegenstand, der sich massenweise herstellen lässt“ (Pschera 2016). Eine besondere Rolle spielen herbei der 3D-Druck (Gheorghiu et al. 2017) und Cyber-physikalische Produktionssysteme (Gotsch und Erdmann 2018). Rund 25 % der Maschinen von produzierenden Unternehmen in Deutschland sind smart (Bitkom 23.04.2018). Im Konsumgüterbereich gab es 2017 bereits 5,2 Mrd. IoT Geräte (Reply AG o.J.). Das IoT zeigt sich gegenwärtig vorwiegend in der Vernetzung innerhalb einzelner Sektoren (z. B. Industrie 4.0, *Connected Cars*, *Smart Home*), in Ansätzen auch Sektor-übergreifend (z. B. Services in der *Smart City*, Kopplung von Stromnetz und Elektromobilität). Die Daten des IoT sind für Maschinelles Lernen sehr wertvoll, auch um die eigenen Algorithmen zu verbessern (Delgado 2017).

Der Terminus des **Internet der Natur** ist bislang nur wenig konturiert. In Analogie zum Verschwinden der Dinge hinter ihrem digitalen Abbild im IoT kann man auch von einem Verschwinden von Pflanzen, Tieren, biotischen und abiotischen Elementen von Ökosystemen hinter ihren digitalen Repräsentationen sprechen (Pschera 2016). Mustererkennung und

[31] Das Raumfahrtunternehmen Space-X (Elon Musk) will ein schnelles Internet auch in entlegene Gegenden bringen (Frankfurter Allgemeine Zeitung 2019). Das schnelle „Internet für Alle“ soll Einnahmen generieren, um Elon Musk’s Weltraummissionen (Entwicklung neuer Raketen und Raumschiffe für die Besiedlung des Mars) zu finanzieren (vgl. Abschnitt 3.6 Erschließung lebensfeindlicher Räume durch Digitalisierung). Die zunächst 60 Satelliten sind in 44 Kilometern Höhe abgesetzt worden, mit 720 Satelliten können man eine moderate Internetabdeckung um den Globus erreichen, in der vollen Ausbaustufe „Starlink“ sind 11.943 Satelliten vorgesehen.

[32] Auch Amazon arbeitet unter dem Namen „Project Kuiper“ an einem Satellitennetzwerk.

[33] Langfristig könnten leichte Antennen aus Metamaterialien Individuen direkt mit Satelliten verbinden, ohne dass sie sich über lokale Internetstrukturen einloggen müssten (Gheorghiu et al. 2017).

Maschinelles Lernen sind wesentliche Technologien, die die Voraussetzungen zur Ökosystemdiagnose und Entscheidungsunterstützung liefern. Digital etikettierte Pflanzen und Nutztiere sollen die Land- und Forstwirtschaft effizienter und natürliche Ressourcen weniger störungsanfällig machen (World Economic Forum et al. 2018).

Langfristige Visionen

Im ***Internet of Everything*** **(IoE)** sind die drei Bereiche Internet der Menschen, Internet der Dinge und Internet der Natur miteinander vernetzt. Grundsätzlich gibt es nichts, was von der Vernetzung ausgeschlossen ist, wodurch alles digital repräsentiert und vernetzt werden kann (Pschera 2016).

Die Vision eines **Dashboards für die Erde** (World Economic Forum et al. 2018), in dem relevante Datenbestände sowohl zu dem Ökosystemen als auch zum gesellschaftlichen Stoffwechsel miteinander verknüpft sind, verspricht mit Hilfe von KI ein effektives Erdsystemmanagement leisten zu können.

Stabilität der Entwicklungslinien: Die Ausweitung und Intensivierung des Internets der Menschen ist eine stabile Entwicklung. Ausmaß und Funktionalitäten des Internets der Dinge (IoT) in Zukunft sind schwer einschätzbar, es wird jedoch intensiv an einer weitgehenden Realisierung gearbeitet. Das Internet der Natur ist in beispielhafter Form als Realität greifbar, bislang jedoch nicht in ausgedehnter und vernetzter Form. Das Internet of Everything (IoE) und ein Erdsystemmanagement aus einem Guss sind aus heutiger Sicht Wiederbelebungen allumfassender Kontroll- und Steuerungsvorstellungen.

2.3.8 Autonome Systeme im Alltag

Untersuchungsrahmen

Autonome Systeme bestehen aus technischen Geräten und Maschinen, die ihre Funktionen ohne menschliches Eingreifen ausführen und dabei Merkmale intelligenten Verhaltens aufweisen können. Sie sind durch die Einbettung miniaturisierter elektronischer Komponenten in technische Geräte und Maschinen charakterisiert, die ihre Umgebung erkennen (Ort, Identität, Sensorik, etc.), miteinander vernetzt sind und Operationen als System selbstständig ausführen können (Bewegung, Aktorik etc.) (Hilty et al. 2003). Autonome Systeme ziehen schrittweise in den **Alltag** der Menschen ein und sind zunehmend allgegenwärtig. Durch die schleichende Gewöhnung an die selbstständig operierende Technik in alltäglichen Lebensbereichen wie Wohnen, Mobilität, Arbeit, Freizeit, Information und Kommunikation, Einkaufen und Inanspruchnahme von Dienstleistungen rückt die Technik in den Hintergrund und damit aus dem Bewusstsein. Autonome Systeme sollen von unangenehmen Arbeiten befreien, Routinen erleichtern, die Lebensqualität erhöhen, den Ressourcenverbrauch minimieren, Kosten senken und zur Sicherheit beitragen (Kadner et al. 2017; Stone et al. 2016; Fraunhofer IIS o.J.).

Wie verändert KI autonome Systeme im Alltag?

Die **Funktionsweise** Autonomer Systeme beruht in der Regel auf Automatisierung, auf Befehlssequenzierung (Ereignis-, Zeit- oder Hierarchie-getriggert) und teilweise auch auf KI (Maschinelles Lernen, Wahrnehmung von Objekten, etc.). Autonome Systeme haben die Fähigkeit, sich an ihre Umwelt anzupassen, dabei zu lernen und eigenständig zu agieren (Kadner et al. 2017). Dagegen sind automatisierte Systeme zumindest teilweise auf menschliche Überwachung und Eingriffe angewiesen (Deutscher Ethikrat 2017). Dieses Profil fokussiert deshalb auf die eigenständige technische Funktionalität Autonomer Systeme.

Entwicklungen

Die Entwicklung und Markteinführung Autonomer Systeme im Alltag ist bereits weit vorangeschritten. In fast jedem Land existieren entsprechende Förderprogramme, entweder unter der Überschrift Robotik, Künstliche Intelligenz, Digitalisierung oder Autonome Systeme.[34] Laut EFI-Studie (Dumitrescu et al. 2018) sind Systeme mit Assistenzfunktion, die vom Menschen bewusst aktiviert und gesteuert werden können, in allen Anwendungsfeldern des Alltags bereits heute verbreitet. Die bisherigen großen Märkte liegen in der Unterhaltung – Information und Kommunikation (Spracherkennung und folgende Suche im Internet, Musik-Boxen, die auf Zuruf funktionieren) und Alltagsunterstützung sowie die Ausführung physischer Operationen (Bestellsysteme, Navigationsgeräte usw.). Idealtypisch können physische und informatorische Operationen unterschieden werden, wobei diese Kategorisierung nicht trennscharf ist.

Informatorisch operierende Systeme:

Im Bereich der Information und Kommunikation für das Selbstmanagement bzw. Unterhaltung existieren am Markt bereits kognitive Assistenten mit Kalender- und Erinnerungsfunktion, Lern- und Abfragesysteme sowie Organizer für das tägliche Management, auch im Haushalt. Assistenten im Internet informieren und aktivieren die Nutzer*innen, bestimmte Handlungen auszuführen. Im Hinblick auf Unterhaltung nehmen Fernseher automatisch Verbindungen zum Internet auf und bieten personalisierte Angebote an. Übergreifend haben digitale (Sprach-)Assistenten Einzug in den Haushalt gehalten (vgl. Abschnitt 2.3.2 Simulation natürlicher Sprache), die auch zur Steuerung der Haustechnik oder von Fahrzeugen dienen können (Day 2019) (s.u. physisch operierende Systeme).

Erwartet wird für das Selbstmanagement eine Weiterentwicklung von Kerntechnologien zur Wahrnehmung, Selbstregulation und zum Lernen, Planen und Schlussfolgern (Dumitrescu et al. 2018). Im Hinblick auf die Unterhaltung liegen wesentliche Entwicklungsschwerpunkte auf der Entwicklung von Geräten zur Emotionserkennung (Day 2019) und von erweiterten Gefährten wie zum Beispiel eine holografische Gefährtin, die mit Männern kommuniziert oder Sexroboter (Day 2019) (vgl. Abschnitt 2.3.1 Affective Computing).

Physisch operierende Systeme:

Im Teilbereich Wohnen sind Autonome Systeme bereits weit verbreitet. Hierzu gehören selbstständige Steuerungen der Haustechnik, Sensoren (z. B. Kameras) und Messgeräte (z. B. Smart Meter) als Grundlage für autonome Prozesse, intelligente Sicherheitssysteme (z. B. Gesichtserkennung an der Haustür, Alarmsysteme, Brandschutz) und unterschiedliche Roboter (z. B. Staubsaugerroboter, Haustier- und Spielzeugroboter (Fischer 2018). Automatisierte Einkäufe können inzwischen über autonome Einkaufsassistenten, die Leerstände erkennen, Bedarfe vermuten und Angebote vorschlagen, abgewickelt werden (Bundesregierung 2018b).

Im Teilbereich Automatisiertes/Autonomes Fahren werden fünf Stufen unterschieden: Stufe 1 - Assistiertes Fahren, Stufe 2 - Teilautomatisiertes Fahren („Hands Off"), Stufe 3 - Hochautomatisiertes Fahren, Stufe 4 - Vollautomatisiertes Fahren und Stufe 5 - Fahrerloses (autonomes) Fahren („Brain Off"). Automatisiertes Fahren ist bereits heute mit entsprechend ausgestatteten Fahrzeugen im realen Verkehr bei bestimmten Rahmenbedingungen und Verkehrssituationen möglich, während Autonomes Fahren auf Stufe 5 noch in ferner Zukunft liegt oder nie erreicht werden wird (Deutscher Ethikrat 2016; Krail 2018).

Im *Gartner Hype Cycle* Modell für neue Technologien (Gartner 2018) sind zahlreiche Autonome Systeme verortet, darunter das Autonome Fahren auf Stufe 4 und Stufe 5, das *Connected Home,*

[34] Für einen Überblick öffentlicher Förderung in Deutschland siehe Dumitrescu et al. 2018.

autonome mobile Roboter und smarte Roboter sowie *Smarte Workspaces* (Panetta 2019). Nicht gelistet und dementsprechend unsicher ist der Einzug von humanoiden Service- und Pflegerobotern in die Haushalte, damit Menschen länger in den eigenen vier Wänden wohnen bleiben können (Bosch o.J.).

Langfristige Visionen

Joseph Weizenbaum formulierte bereits im Jahr 2001 folgende Vision: „Der Computer wird in den nächsten 5 bis 10 Jahren **aus unserem Bewusstsein verschwinden**. Wir werden einfach nicht mehr über ihn reden, wir werden nicht über ihn lesen, außer natürlich Fachleute." In einer Delphi-Befragung wurde folgende These zur Bewertung gestellt (Gheorghiu et al. 2017): „Die Hälfte der früher passiven Materialien und Dinge (Wände, Straßen, Möbel, Zeichen) wird interaktiv und reagiert auf ihre Umgebung mittels Sensoren, adaptiver Materialien und ubiquitärer Elektronik.“ 61 von 84 Antwortenden sind der Auffassung, dass dies vor 2040 der Fall sein wird.

Stabilität der Entwicklungslinien: Zwar ist der Computer allgegenwärtig und in eingebetteten Systemen nahezu unsichtbar geworden (Hilty et al. 2003), er nimmt in den gesellschaftlichen Debatten – entgegen der Vision von Joseph Weizenbaum – jedoch derzeit wichtige Rollen ein. Die bislang in den Alltag eingezogenen Autonomen Systeme beziehen sich vorwiegend auf vernetzte Einzelgeräte oder kleinere Systeme. Es wird zwar an der Verschmelzung von Kommunikationsassistenten mit Robotern und anderen dezenten Mensch-Maschine-Schnittstellen geforscht (Knight 2019), eine übergreifende Technikkonvergenz der einzelnen Autonomen Systeme in einem Internet der Dinge (vgl. Abschnitt 2.3.7) oder zu intelligenten Schwärmen (vgl. Abschnitt 2.3.10) ist derzeit aber nicht absehbar. Sollten die technisch unterstützenden Systeme im Alltag mehr und mehr wahrnehmen, auf unvorhergesehene Veränderungen in der Umwelt reagieren, lernen, Entscheidungen treffen und autonom handeln zu können, so könnte die Autonomisierung der Systeme im Alltag einen Schub in Richtung ihres Rückzuges in den Hintergrund (Unsichtbarkeit) auslösen.

2.3.9 Entschlüsselung des Lebens durch digitale Werkzeuge

Untersuchungsrahmen

Dieses Profil umfasst den Einsatz von KI in der **lebenswissenschaftlichen Forschung mit Relevanz für die Anwendung am Menschen** mit dem Versprechen der Entschlüsselung des Lebens - und damit implizit auch dem Versprechen, tiefere Erkenntnisse über das Wesen des Menschen, über sich selbst zu erfahren. Im Folgenden liegt der Schwerpunkt auf der **Genom- und Postgenomforschung**, insbesondere der Entschlüsselung des menschlichen Genoms und weiterer sogenannter „-omics-Ansätze".[35] Ziel dieser Ansätze ist die Ermittlung der erblichen und umweltbedingten Ursachen für Gesundheit und Krankheit, Verhalten und Wesen des Menschen, um damit die Grundlagen für das Verständnis, die Analyse sowie die Veränderung dieser Eigenschaften zu legen. In diesem Profil werden die Werkzeuge zur genetischen Veränderung des Menschen selbst aber nicht betrachtet, da sie nur mittelbare Bezüge zur Digitalisierung aufweisen.[36] Ein weiterer wichtiger Bereich ist die Hirnforschung, die in Abschnitt 2 unter dem „Human Brain Project“ behandelt wird.

[35] Transkriptom, Proteom, Metabolom, Epigenom, Metagenomanalysen des Mikrobioms etc.

[36] Die Anwendung von KI in der Genom- und Postgenomforschung birgt das Potenzial, Werkzeuge zur genetischen Veränderung des Menschen gezielt zu designen (z. B. Genomeditierung, Ansätze der synthetischen Biologie, Viren oder modifizierte Zellen als Vektoren im Rahmen fortgeschrittener Therapien wie Gen- und Zelltherapien oder Tissue Engineering). Transhumanisten propagieren, diese Werkzeuge über therapeutische Anwendungen hinaus für eine genetische ‚Verbesserung des Menschen‘ zu nutzen.

Wie verändert KI die Erkenntnisse in den Lebenswissenschaften?

KI wird in den Lebenswissenschaften mit dem Ziel eingesetzt, das menschliche Genom zu entschlüsseln, die Bedeutung einzelner Gene und die Zusammenhänge von Genotyp und Phänotyp aufzuklären und um genetische Faktoren, Lebensstil- und Umweltfaktoren voneinander zu isolieren. Durch Maschinelles Lernen aus großen kombinierten Datensätzen könnte Wissen generiert werden, mit dem möglicherweise besser erklärt werden kann, wie genetische Unterschiede zwischen Menschen komplexe Eigenschaften wie z. B. Disposition für Krankheiten, physische und kognitive Fähigkeiten oder Persönlichkeitsmerkmale beeinflussen. Die Genom- und Lebensforschung ist sehr datenintensiv, weshalb sich mit der Bioinformatik eine eigene Teildisziplin etabliert hat (Warnke et al. 2019).

Entwicklungen

Von 1990-2004 wurde das erste „Big Science-Projekt der Biologie", das „**Human Genome Project**" durchgeführt. Sein Ergebnis war die Bestimmung der genauen Abfolge der ca. 3 Milliarden Basenpaare, aus denen das menschliche Genom aufgebaut ist, durch Sequenzierung (Lander et al. 2001; Venter et al. 2001). Damit waren große Hoffnungen und Versprechen verknüpft, den Zusammenhang zwischen Gensequenz und der Entwicklung des menschlichen Organismus und seiner Merkmale (einschließlich Prädispositionen für Krankheiten, den Alterungsprozess oder kognitive Fähigkeiten) aufklären zu können. Daran anknüpfend können zwei Entwicklungslinien unterschieden werden, die Verbesserung der Genomanalyse und erweiterte Korrelationen der Genomdaten mit anderen Daten.

In den letzten Jahren wurde die **Genomanalyse** wesentlich verbessert. Insbesondere wurden enorme Leistungssteigerungen und Kostensenkungen bei Genomsequenzierungstechnologien erreicht (Goodwin et al. 2016). Derzeit ist die Sequenzierung vollständiger Genome in der medizinischen Versorgung bestimmter kleiner Patientengruppen (z. B. bei bestimmten Krebserkrankungen oder der Diagnose seltener Erkrankungen) in Universitätskliniken nicht unüblich (Rehm 2017). Es erscheint möglich, dass in wenigen Jahren die vollständige Genomsequenzierung als Routinediagnostik für breite Bevölkerungsgruppen in allen Lebensphasen etabliert sein wird. In der funktionellen Genomforschung werden v. a. überwachte Deep Learning-Algorithmen eingesetzt, um die funktionellen Konsequenzen individuell unterschiedlicher Genomsequenzen (Varianten) zu verstehen und vorherzusagen (Zhou et al. 2018).[37] Die aktuellen medizinischen und kommerziellen Aktivitäten zielen vor allem auf die Diagnose und Prognose von Krankheiten sowie auf Interventionen mit dem Ziel der Behandlung und Heilung ab. Einige wichtige ursprünglich gehegte Hoffnungen und Versprechen aus dem „Human Genome Project" sind unerfüllt geblieben. Die Ursache hierfür liegt zum einen in der Schwierigkeit, von Gensequenzen auf bestimmte phänotypische Eigenschaften zu schließen, da dies ein hochkomplexer Prozess von Wechselwirkungen und Rückkopplungen zwischen Genom, Zellbestandteilen und dem menschlichen Organismus ist und zudem einer Genomsequenz nicht ohne weiteres eine Funktion zugewiesen werden kann. Zum anderen tragen genetische Ursachen nur einen Teil zur Ausprägung von Eigenschaften bei, auch Umwelt und Lebensstil spielen eine wesentliche Rolle.

Die Berücksichtigung von Lebensstil- und Umweltfaktoren zur **Korrelation und Erklärung des Zusammenhangs von Genotyp und Phänotyp** steht erst am Anfang, wird aber in zunehmendem Maße in die Forschung integriert, da immer größere und lebensnähere MISST-

[37] Beispielsweise zur Klärung von solchen Fragen: Welche Genvarianten liegen in einem Genom vor? Welche Funktionen haben bestimmte Abschnitte im Genom (z. B. Bindung von Proteinen, Methylierung von DNA, Expressionsstärke)? Welche Funktionen haben Genvarianten in verschiedenen Bereichen des Genoms? Welche Genvarianten sind ursächlich am Krankheitsgeschehen beteiligt?

Datenmengen zur Verfügung stehen (Interview 7). MISST steht für *Mobile, Imaging, pervasive Sensing, Social media and location Tracking*. Somit handelt es sich um Daten, die über mobile Anwendungen, Bilder, Ortsbestimmung, Sensoren (z. B. Fitnesstracker) oder soziale Medien erhoben werden (Dunseath et al. 2018). Vor diesem Hintergrund wird der KI großes Potenzial zugemessen, die enormen Datenmengen im Hinblick auf die angestrebten Anwendungen zu analysieren und interpretieren zu helfen (Salter und Salter 2017; Eraslan et al. 2019; Topol 2019). Insbesondere das Deep Learning hat in den letzten Jahren verschiedene Forschungsansätze auf diesem Gebiet gefördert (Zou et al. 2019; Webb 2018; Leung et al. 2016; Eraslan et al. 2019). In der Genomforschung überwiegen bislang Modelle des Zusammenhangs zwischen Genomsequenz und Phänotyp, für die viel Vorwissen erforderlich ist. Durch nicht überwachtes Deep Learning können datengetriebene Modelle nunmehr automatisch konstruiert werden, indem große Datensätze[38] (Rehm 2017; Wainberg et al. 2018; Topol 2019) auf verborgene Muster hin analysiert werden, aus denen sich die zuvor unbekannten Mechanismen ableiten lassen. Man hofft, dass es auf diese Weise möglich wird, ursächliche Zusammenhänge zwischen Geno- und Phänotyp aufzudecken und über eine wissensbasierte Gruppeneinteilung von Menschen die Präzisionsmedizin[39] zu fördern. Zudem sollen Multiskalenmodelle von Zellen, Organen oder ganzen Individuen geschaffen werden („virtueller Zwilling"). Solche Modelle können dafür eingesetzt werden, die Wirkungen von Lebensstilen und Umwelteinflüssen auf die Gesundheit zu simulieren.

Langfristige Visionen

Die Isolierung des Faktors Gen von anderen Faktoren durch KI könnte wesentliche Impulse für Forschungsgebiete wie die **Entschlüsselung des Alterns** bis hin zu Faktoren für ein „ewiges Leben" liefern (Harari 2016).

Das Pooling von durch KI analysierbaren Daten aus der Genom- und Postgenomforschung, aus der Lebensstil- und Umweltforschung birgt das Potenzial, weitreichende **Vorhersagen** in Bezug auf die Fähigkeiten, Lebenschancen und Anfälligkeiten von einzelnen Personen und Personengruppen mit einer bestimmten Genauigkeit machen zu können.[40]

Stabilität der Entwicklungslinien: Aufgrund technologischer Fortschritte in der Genomsequenzierung wurden in den letzten Jahren erhebliche Reduktionen des Kosten- und Zeitaufwands erreicht, so dass die Totalgenomanalyse als Standardmethode in der Gesundheitsversorgung in realisierbar erscheinende Nähe rückt. Seit 2015 wird Deep Learning zunehmend auf -omics-Daten angewendet und diese zudem mit weiteren Daten (z. B. Patientenakten, Lifestyle und Umweltdaten) zusammengeführt. Nach Experteneinschätzung (vgl. (Interview 7), (Interview 1) ist dies der Ansatz der Wahl, um großskalig Variationen in der Bevölkerung zu beschreiben und zu analysieren. Die tatsächliche Wirksamkeit dieser Ansätze ist jedoch noch unsicher. Die Anwendungsentwicklungen im Bereich Medizin sind robust, während

[38] z. B. Daten aus -omics-Analysen und bildgebenden Verfahren, Patientenakten, wissenschaftliche Literatur, ethnische Zugehörigkeit, Daten zu Lebensstil und Lebensführung, Daten zu Umwelteinflüssen

[39] So wird beispielsweise innerhalb des Förderkonzepts Medizininformatik das Konsortium HighMed gefördert, das am Deutschen Krebsforschungszentrum und der Universität Heidelberg ein Datenintegrationszentrum für „Genomics"- und „Radiomics"-Daten (OmicsDIC) aufbauen und betreiben und seine Expertise in der klinischen Anwendung der Gesamtgenomsequenzierung erweitern wird (http://www.highmed.org/).

[40] In einer Delphi-Befragung wurde folgende These zur Bewertung gestellt (Gheorghiu et al. 2017): „Daten über die Interaktion zwischen Patienten und ihrer persönlichen Umgebung (einschließlich abiotischer und biotischer Faktoren) werden routinemäßig und systematisch gesammelt für die Diagnose und personalisierte Therapiepläne". 41 von 59 Antwortenden halten dies für vor 2040 realisiert. KI/Deep Learning kann bei der Kombination strukturierter Daten (Genotyp, Phänotyp, Genomics) mit halb- oder unstrukturierten Daten (Lebensstil, Umwelt, Gesundheitsökonomie) und entsprechender Mustererkennung helfen.

die Anwendungsentwicklung zur gezielten sozialen Stratifizierung (z. B. Bildungs- und Lebenschancen) unsicher ist.

2.3.10 Schwarmintelligenz

Untersuchungsrahmen

Die einschlägigste Definition von **Schwarmintelligenz** bezeichnet diese als „The emergent collective intelligence of groups of simple agents [die emergente kollektive Intelligenz einer Gruppe einfacher Agenten]" (Bonabeau et al. 2010). Agenten können dabei natürliche Spezies wie Ameisen, Vögel und Fische, Zellen, Menschen oder auch digitale Agenten wie Roboter, Drohnen und Software-Agenten sein, die Aufgaben im Verbund lösen (Monchalin et al. 2019). Statt von Schwarmintelligenz wird – meist im Zusammenhang mit Menschen – auch von kollektiver oder kollaborativer Intelligenz oder von Gruppenintelligenz gesprochen. Allerdings ist menschliche Gruppenintelligenz wahrlich kein Selbstläufer (Dueck 2018).

In diesem Kurzprofil geht es um selbstorganisierte intelligente **Roboter- und Drohnenschwärme**.[41] Schwarmrobotik ist definiert als „A group of non-intelligent robots forming, as a group, an intelligent robot [eine Gruppe nicht-intelligenter Roboter, die als eine Gruppe einen intelligenten Roboter formen]" (Beni 2005). Der Begriff Drohne wird für jegliches Fahrzeug verwendet, das über Autopilot-Algorithmen gesteuert wird (Warnke et al. 2019). Im militärischen Bereich spricht man auch von unbemannten Systemen, die zu Wasser, zu Lande oder in der Luft (unmanned aerial vehicle (UAV)) untereinander interagieren (Petermann und Grünwald 2011). Der Fokus liegt auf den emergenten Wirkungen vieler einfacher selbstorganisierter Roboter und Drohnen; die Wirkungen einzelner selbstoperierender Agenten oder generelle Multi-Roboter-Systeme werden nicht betrachtet.

Wie verändert KI die Scharmintelligenz?

Zu den Vorteilen von Schwarmlösungen gehört, dass die Zahl der eingesetzten Agenten an die Größe der Aufgabe angepasst werden kann (Skalierbarkeit) und, dass der Ausfall einzelner Agenten die Zielerreichung nicht in Frage stellt (Resilienz). Die Entscheidungsfindung findet dezentral statt, indem jeder der Agenten anhand von implementierten Regeln (computerbasierten Algorithmen) individuell dazu veranlasst wird, die nächste Aktivität auszuführen. Die Agenten haben lokalisierte Messfähigkeiten und kommunizieren untereinander (Monchalin et al. 2019). Zukünftig soll sich der Mensch auf die Führung des Gesamtsystems konzentrieren, während die Agenten die Aufgabe dann selbstständig ausführen (Huppertz 2016). Diese Eigenschaften sind für **offene Anwendungen** beispielsweise in Ökosystemen, auf Gefechtsfeldern, Agrarlandschaften oder unbestimmten Gegenden interessant.

Entwicklungen

Die Informationslage zur Differenzierung der Funktionalitäten von einzelnen bzw. massenhaft kooperierenden Robotern und Drohnen ist spärlich ((Warnke et al. 2019), S. 50f.).[42] Nach unserer Einschätzung kommen die Schwarmintelligenzeffekte von Robotern und Drohnen in

[41] Hierzu gehören auch Gruppen von Software-Agenten, die vordefinierte Aufgaben selbstständig im Austausch untereinander und im Austausch mit anderen digital codierten Informationsträgern lösen (Hilty et al. 2003). Sie werden jedoch hier nicht weiterverfolgt.

[42] Drohnen werden vor allem in der Fernerkundung (u. a. für Landwirtschaft, Infrastrukturüberwachung, Erdbeobachtung, Katastrophenschutz, territoriale Überwachung, Ausspähen von Angriffszielen und feindlichen Operationen, Erkennungsmissionen in urbanen Gebieten und in der Logistik im weitesten Sinne (Transport, Militär), eingesetzt (Marscheider-Weidemann et al. 2016). Roboter haben ein breites Einsatzspektrum von Industrierobotern über Serviceroboter und Agrarroboter bis hin zu humanoiden Pflegerobotern. Hinzu kommt der Sonderfall der Fahrzeug-zu-Fahrzeug-Kommunikation beim Autonomen Fahren. Sonderfälle sind Robotereinsätze für Katastrophenmissionen (Such- und Rettungsroboter o. ä.) und Nanobots (u. a. interne chirurgische Operationen, Lieferung von Wirkstoffen an gezielte Krebszellen).

vier ausreichend stabilen Entwicklungslinien besonders zur Geltung: im Verkehr/in der Logistik, in der Medizin, in der Landwirtschaft und im Bereich der militärischen Sicherheit. Die Visionen *Smart Dust*[43] und selbstreplizierende nanoskalige Maschinen sind aus heutiger Sicht hochspekulativ ((Gheorghiu et al. 2017), S. 172).

Grundlegende Veränderungen der Mensch-Technik-Umwelt-Beziehungen scheinen für bewaffnete Roboter-/Drohnenschwärme und für die Bestäubung von Pflanzen mit Bienen-Roboterschwärmen möglich zu sein. Schwarmintelligenzanwendungen im Bereich Verkehr/Logistik und ihre Umwelteffekte sind bereits seit langem Gegenstand der Forschung und Entwicklung[44] und werden deshalb hier nicht weiterverfolgt. Für die Landwirtschaft werden autonome Roboterflotten für die Feldbewirtschaftung entwickelt, die auf Einzelpflanzenebene säen, wässern, düngen, zurückschneiden und ernten. Die Ausbildung von Roboter-Pflanzen-Gesellschaften ist ein eigener Forschungszweig (Warnke et al. 2019). Von besonderem Interesse für grundlegende Veränderungen des Mensch-Technik-Umwelt-Verhältnisses ist die Entwicklungslinie Roboter-Insektenschwärme zur Bestäubung von Pflanzen.

Roboter-Insektenschwärme werden grundsätzlich für alle oben angeführten Einsatzgebiete entwickelt, hier geht es jedoch konkret um die Bestäubung von Agrarpflanzen, wodurch ein natürlicher Ökosystem-Service durch einen technischen Service ersetzt werden könnte. Der aktuelle Entwicklungsstand ist wie folgt: Eine Roboter-Biene (*RoboBee*) vom Wyss-Institut ist derzeit etwa halb so groß wie eine Büroklammer, wiegt weniger als ein zehntel Gramm und fliegt mit Hilfe von „künstlichen Muskeln", also Materialien die sich beim Anlegen einer Spannung zusammenziehen. Diese autonomen Mikro-Luftfahrzeuge sind in sich geschlossen, fliegen selbstgesteuert, und koordinierten ihr Verhalten in großen Gruppen. Zu den Einsatzgebieten gehört die Getreidebestäubung (Wyss Institute o.J.). Für Treibhäuser ist ein „*proof of concept*" für ein vollautonomes Mikro-Luftfahrzeug (APIS-Bestäuberdrohne der TU Delft) erstellt worden, wozu auch eine Marktanalyse und eine Untersuchung der technischen Herausforderungen gehört (Macovei 2017). Das US-Einzelhandelsunternehmen Walmart hat 2018 sechs Patente für Agrardrohnen angemeldet. Eines ist für eine Roboter-Biene, die alle Services von Monitoring über Bestäubung bis zum Versprühen von Pestiziden vornehmen können (Gohd 2018; Wholey 2018). Hierbei wird nicht nur damit argumentiert, dass Roboter-Bienen natürliche Bienen ersetzen können, sondern dass sie im Vergleich zur Insektenbestäubung die Erträge sogar erhöhen könnten (Cherney 2021). Es ist derzeit sehr unklar oder gar fraglich, wie weit die technische Entwicklung von Roboter-Bienen tatsächlich mit Blick auf einen Praxiseinsatz ist.

Langfristige Visionen

In einer aktuellen Delphi-Befragung wurde folgende These den Experten zur Bewertung gestellt (Gheorghiu et al. 2017): „Robot insects are deployed for 70 % of pollination to secure the EU's agricultural production" [**Roboter-Insekten werden für 70 % der Bestäubung eingesetzt, um die Agrarproduktion der EU zu sichern**]. Von den Antwortenden (n=36) ist die Hälfte der Ansicht, dass dies nach 2040 (n=11) oder nie (n=7) der Fall sein werde. Eine starke Minderheit hält die Realisierung der These bis 2040 für möglich (n=15). Zweifel bestehen im Hinblick auf die Großflächigkeit, Effektivität und – am wichtigsten – Sicherheit der Anwendung (n=24). Das

[43] *Smart Dust* wird ein System aus vielen kleinen mikroelektromechanischen Systemen (MEMS) wie Sensoren, Robotern und Geräten genannt, die z. B. Licht, Temperatur, Vibrationen, Magnetismus oder sogar Chemikalien erkennen können. In der Regel kommunizieren die energieautarken Elemente über *Radio Frequency Identification* (RFID). Smart Dust-Komponenten werden über ein kabelloses Computernetzwerk gesteuert und verteilen sich über eine größere Fläche, um alle Funktionen zu erfüllen.

[44] Die Logistik mit Drohnen hat sich bereits zu einem ökonomisch wichtigen Sektor entwickelt (Gheorghiu et al. 2017, S. 184 f.).

Bestäubungsproblem kann einfacher durch Fixierung entweder von Bienen oder Pflanzen gelöst werden (n=10). Als Treiber werden das Interesse des Militärs an Roboter-Insekten (n=14) und das ungeklärte Bienensterben (n=13) genannt. Eine Substitution der natürlichen Bestäubung durch eine technische Bestäubung wird aus verschiedenen Gründen sehr skeptisch gesehen, darunter technisch-ökonomische Hemmnisse und Ökosystemschäden durch Roboterbienen (Potts et al. 2018).

Stabilität der Entwicklungslinien: Das Bienensterben hat aktuell in Deutschland ein beträchtliches Mobilisierungspotenzial. Die vor allem großmaßstäbliche Realisierung von Roboter-Bienenschwärmen ist jedoch alles andere als klar, sie wäre zudem mit erheblichen Unsicherheiten und Risiken behaftet (vgl. (Potts et al. 2018)). Zudem würden digitale Roboter-Bienenschwärme bei der künstlichen Bestäubung mit künstlich gezüchteten Bienenvölkern konkurrieren, die gegenüber bestimmten Insektiziden resistent sind. Künftige Machbarkeit und möglicher Markterfolg von Roboter-Bienenschwärmen sind dementsprechend extrem unsicher.

Drohnenschwärme erweitern und transformieren die Möglichkeiten, technische Infrastrukturen und Menschen zu schädigen. Gleichzeitig werden zunehmend Anti-Drohnen-Systeme eingesetzt (The Economist 2019). China testet Flotten mit bereits 119 UAV, in einer Ausstellung des Chinesischen Militärmuseums wird ein UAV-Schwarm gezeigt, der auskundschaftet, blockiert und als Schwarmangriff ein Flugzeug angreift (Warnke et al. 2019). Auch die „Bundeswehr hat eine Studie ausgeschrieben, die die Nutzung von Drohnen zur Aufklärung des „gläsernen Gefechtsfeldes" untersuchen soll“ (Borchers 2019). Nahezu nicht erkennbare Mikrodrohnen/Roboterinsekten mit *Micro-Payload* (Nutzlast) werden für Aufklärungsmissionen in Städten entwickelt (Warnke et al. 2019). Zentrale Entwicklungsrichtungen sind die Verbesserung der Sensoren und das Verstehen (z. B. Gesichtserkennung und Diskriminierung), um mit der Umwelt und untereinander wie gewünscht interagieren zu können (Warnke et al. 2019). Schwärme haben in Bezug auf militärische Nutzungen einen entscheidenden Vorteil gegenüber einzelnen Agenten: „the biggest advantage of a „swarm“ is the ability of machines to work together in numbers. And when it comes to the battlefield, numbers matter (McMullan 2019).“

Langfristige Visionen

Autonome Drohnenschwärme erlauben es Militärs, Streitkraft aufzustellen, die zahlreicher, schneller, besser koordiniert sind, als es durch Menschen alleine der Fall wäre. Zudem sind sie ausbaufähig (Scharre 2018). In einer aktuellen Delphi-Befragung wurde folgende These den Experten zur Bewertung gestellt (Gheorghiu et al. 2017): „Autonomous weapons and robots are deployed to monitor and secure at least 50 % of the EU’s land sea, and air borders, sensitive places and infrastructures [Autonome Waffen und Roboter überwachen und sichern mindestens 50 % der Land-, See- und Luftgrenzen, empfindliche Plätze und Infrastrukturen)“. Von den Antwortenden (n=27) sind sieben (26 %) der Ansicht, dass autonome Verteidigungssysteme notwendig sind, um kritische Infrastrukturen vor Schwarmattacken von autonomen Drohnen zu schützen, weil menschliches Handeln zu langsam ist.

Stabilität der Entwicklungslinien: Die Entwicklungslinie Drohnenschwärme zur Schädigung von Infrastrukturen und Menschen bis hin zur Kriegsführung wird wesentlich von miteinander konkurrierenden militärischen Forschungseinrichtungen vorangetrieben. Die Algorithmen zur Koordination von Drohnen- und Roboterflotten verbessern sich stetig. Insgesamt ist das Militär wohl der stärkste Treiber der Entwicklung von selbstorganisierten unbemannten kleinen flug- und tauchfähigen Systemen.

2.4 Grundlegende Veränderungen von Mensch-Technik-Umwelt-Beziehungen

In der Debatte zum Verhältnis von Technosphäre und den planetaren Grenzen wird gemutmaßt, dass aus KI und dem Zeitalter des Anthropozän eine transformative Gemengelage entstehen könnte (Braidotti 2019), beziehungsweise dass wir uns an einem Sattelpunkt befänden, auf dem sich das Internet der Dinge (IoT), Smart Systems und Industrie 4.0 als Lock-In für die Zukunft erweisen könnten (Rosol 2019).

Im Folgenden stehen solche grundlegenden Veränderungen von Mensch-Technik-Umwelt-Beziehungen durch KI im Vordergrund, die teilweise in Philosophie und KI-Forschung selbst gesehen werden, in Technikfolgenabschätzung und Innovationsforschung aber kaum thematisiert werden. Aus der Analyse grundlegender Veränderungen der Mensch-Technik-Umwelt-Beziehungen haben wir drei zentrale Stränge identifiziert (Erdmann und Röß 2020):

2.4.1 KI verändert die Agency im Mensch-Technik-Gefüge

Der Akteur-Netzwerk-Theorie (ANT) zufolge war es schon immer eine Täuschung, Handlungsfähigkeit allein dem menschlichen Akteur zuzuschreiben. Die ANT versucht, den klassisch unterstellten Dualismus zwischen menschlichen Akteuren und nicht-menschlichen Akteuren aufzuheben und **Handlungen als Assoziationen zwischen Technik und Mensch** zu verstehen (Latour 2005). In den Diskursen rund um die Entwicklung von KI-Systemen wird die Vorstellung eines rein instrumentellen Charakters dieser Technologien unterlaufen:

Eine besondere Form der Radikalisierung dieses Prinzips ist die vollständige Delegation von Aktivitäten ohne menschliche Intervention. In diesem Falle könnten sich die Menschen aus der Mensch-Technik-Umwelt-Interaktion weitgehend zurückziehen und anderen Aufgaben widmen. Insbesondere KI-Anwendungen in Form von Autonomen Systemen bergen das Potenzial, Handlungsfähigkeit und Souveränität neu zu verteilen. Ob eine zunehmende Delegation von bestimmten Tätigkeiten an Autonome Systeme einen grundlegenden Wandel im Mensch-Technik Verhältnis auslöst, ist davon abhängig, wie stark solche Systeme Handlungsoptionen im Vorfeld bereits selektieren und strukturieren. In diesem Falle verschiebt sich die Agency zugunsten der KI-Systeme. Menschen werden durch KI im Alltag aktiviert und die Handlungsmöglichkeiten der menschlichen Akteure werden durch KI eingeschränkt. Die Folgen für den Menschen sind das Verlernen von bisherigen Praktiken und das Erlernen neuer Praktiken im Umgang mit autonomen Systemen im Alltag (eigene Einschätzung). Indem bisherige Selbstwirksamkeitserfahrungen durch Hilfserfahrungen und permanente Unterstützung ersetzt werden, verändert sich der Mensch selbst (Harari 2016). Dies gilt insbesondere für *Autonome Systeme im Alltag*, aber grundsätzlich auch für *Autonome Systeme zur Erschließung bislang für den Menschen unverfügbarer Räume* und die *Schwarmintelligenz*.

KI-Systeme können einzelne genuin menschliche Fähigkeiten simulieren. Die FuE-Gebiete des *Natural Language Processing* (NLP) und der *Computational Creativity* erzeugen u. a. gesprochene bzw. geschriebene sprachliche Artefakte. Für Hörer*innen in einem digital vermittelten Sprechakt oder der Leser*innen eines digital codierten Schriftstückes wird es in Zukunft zunehmend schwieriger zu unterscheiden, ob Sprache oder ein anderes Artefakt von einem Menschen oder einer Maschine generiert wurde und mit was für einer Art Gegenüber – Mensch oder Maschine – sie es in einer Kommunikationssituation zu tun haben. Diese – bereits seit langem diskutierte – Nicht-Unterscheidbarkeit zwischen technischer und menschlicher Urheberschaft kann das Vertrauen in Technik und Menschen grundlegend ändern. Gelingt es, mit KI auch neuen Problemen und ad-hoc Situationen auf für den Menschen bislang nicht mögliche Weise gerecht zu werden, so würde sich der Handlungsspielraum durch KI qualitativ

erweitern (Nassehi 2019). Eine (un)bewusste, weitreichende und massenhafte Verbreitung von Lügen und Falschinformationen kann zur Normalität werden (Fake News und Deep Fake), wobei eine Differenzierung zwischen Wahrheit und Fake ggf. nur noch für Expert*innen oder gar nicht mehr möglich ist, was ebenfalls nichts absolut Neues ist, aber in seiner Massivität erhebliche disruptive bzw. destruktive Implikationen haben würde.

Grundsätzlich neigen Menschen dazu, nicht nur Tiere und Pflanzen (Dennett 2018), sondern auch Technik zu anthropomorphisieren. „Computer" hießen früher die Menschen, die Rechenaufgaben manuell ausführten (Ceruzzi 2012). Heute sprechen die Menschen mit dem Lautsprecher Amazon Echo über seine Oberfläche Alexa, so als ob es sich dabei um einen Menschen handelte (Daum 2019a). Die Anthropomorphisierung lässt sich an einem breiten Spektrum von KI-Anwendungen illustrieren (Weber 2020). Insbesondere das *Affective Computing* steht für eine KI-Entwicklungslinie, die menschliche Emotionen erkennen sowie selbst auf emotionale Weise auf die User reagieren und damit – in ihrer ambitioniertesten Form – Bedürfnisse nach emotionalem Austausch durch Mensch-Technik-Interaktion simulieren oder gar ersetzen kann. Wenn sich Menschen aber auch ‚verstanden fühlen', können kognitive Assistenten für die Nutzer*innen auch anthropomorphen Charakter annehmen (eigene Einschätzung). Die durch das Affective Computing geförderte emotionale und parasoziale Mensch-Maschine-Beziehung vermag zwischenmenschliche Beziehungen zu ergänzen oder im Einzelfall gar zu ersetzen (Gutmann 2011). Die Ergänzung, Ersetzung und Veränderung zwischenmenschlicher Beziehungen durch Affective Computing bedeutet eine „Verdinglichung", deren Auswirkungen auf die natürliche Umwelt unbekannt sind (eigene Einschätzung). Auch die Allgegenwart von hörenden und sprechenden Sprachassistenten (*Simulation natürlicher Sprache*) könnte dazu führen, dass es perspektivisch keinen Bereich menschlicher Intimität mehr gibt (wiedergegeben von (Daum 2019b)).

2.4.2 KI macht die Natur und ihre Wahrnehmung zunehmend „künstlicher"

Hellmut Plessner hat „**Das Gesetz der natürlichen Künstlichkeit**" (Schöpfung künstlicher Welten) als ein zentrales Merkmal der *conditio humana* ausgemacht (Plessner 2003). Die immer schon schwierige Unterscheidung zwischen Natürlichkeit und Künstlichkeit wird durch KI noch komplexer. Im Zusammenspiel mit anderen digitalen Technologien sprechen einige KI-Anwendungen dafür, dass sich dieses Verhältnis zwischen Natürlichkeit und Künstlichkeit weiter in Richtung des Künstlichen verschieben und sich die menschliche Spezies schleichend an den fortschreitenden Verlust der tatsächlichen Natur gewöhnen wird (*Environmental Generational Amnesia)* (Kahn 2011):

Autonome Systeme ermöglichen es, bislang für den Menschen unverfügbare Räume zu erschließen. Hierzu zählen Roboterflotten zur Aufforstung von Wüsten und zur Reinigung der Meere von Müll ebenso wie autonome Systeme zur Rohstoffgewinnung in der Tiefsee und zur Produktion im Weltraum. Dadurch wird die *frontier* des menschlichen Handelns (in Sinne der Landnahme Nordamerikas durch europäische Kolonialisten) auf für den Menschen bislang unverfügbare Gebiete ausgedehnt, wobei die Wildnis als Gegenspielerin zur zivilisierten Welt verschwindet und sich die Erde und der Weltraum als reparierbare und quasi unerschöpfliche Produktionsfaktoren im Bewusstsein der Menschen verankern könnten. Autonome Habitate im Weltraum bzw. auf dem Meer ermöglichen dem Menschen neuartige Lebensweisen in lebensfeindlichen Umgebungen, bringen ihn jedoch auch in extremste Abhängigkeit von der zugrundeliegenden Technik, insbesondere auch von KI (Feireiss und Najjar 2018). Die Inanspruchnahme lebensfeindlicher Räume erfordert neue ökologische Bewertungen (teilweise auch neue Bewertungsinstrumente) und die lebensfeindlichen Räume werden damit der Logik der Raumplanung unterworfen (eigene Einschätzung). Eine Ausweitung der ‚*frontier*' wirft dabei

auch die Frage auf, welche Grenzen die Menschheit daran anschließend ins Visier nehmen wird. Einer möglichen massenhaften Erweiterung des Bewusstseins um das Meer und die Erde als Teil des Kosmos (Feireiss und Najjar 2018) steht eine mögliche Ablenkung von den heutigen großen Herausforderungen wie Erderhitzung und Verbrauch nichterneuerbarer Ressourcen gegenüber. Das Sammeln von Plastikmüll aus dem Meer und die Wiederaufforstung von Ödland können dazu führen, dass der Mensch vermehrt ein Bewusstsein dafür entwickelt, dass auch eine großskalige Schädigung von Naturräumen durch Technik reversibel wäre. Hierdurch könnten Weltbilder gefördert werden, die die Erde insgesamt als reparierbar ansehen und sich Bewertungen von Umwelträumen wandeln (eigene Einschätzung). Eine starke Änderung des Mensch-Umwelt-Verhältnisses wird angenommen, wenn Menschen dem Planeten Erde ‚entsteigen' können; dies impliziert das Risiko, dass Natur- und Umweltschutzbemühungen als hinfällig eingeschätzt werden.

Das Maschinelle Lernen aus realweltlichen Datensätzen ist eine Schlüsseltechnologie für die Anwendung von digitalen Technologien in und am menschlichen Körper wie zum Beispiel Gehirn-Computer-Schnittstellen und Wearables zur digitalen Selbstvermessung. Der digital durchdrungene Mensch gehört im Vergleich zum herkömmlichen Menschen stärker der künstlichen Welt an. Die neue Qualität speziell von BCI liegt darin, dass das *Enhancement* nicht mehr durch Erziehung, Bildung, Übung, Moral, Aufklärung, humanistische Kultur (Röcke 2017, S. 328) oder auch Medikamente erreicht werden soll, sondern durch Technologien, die grundsätzlich eine bidirektionale Kommunikation zwischen Mensch und Computer und gekoppelte Lernprozesse erlauben. Durch BCI werden perspektivisch als typisch menschlich geltende kognitive (Interview 3) und affektive Fähigkeiten der externen Beobachtung sowie der Manipulation zugänglich (Roelfsema et al. 2018). Unter anderem die Gedankenfreiheit als Merkmal des Menschen würde hierdurch eingeschränkt (eigene Einschätzung). Änderungen von Identitäts-Vorstellungen sind wahrscheinlich bei gleichzeitiger Gefahr des Verlusts von Authentizität sowie natürlicher Körpergefühle, die Technologie führt ggf. zur Modifikation des Selbstverständnisses als eigenverantwortliches Subjekt sowie zur Verstärkung diesbezüglich bestehender Stereotypen bzw. Herausbildung neuer Stereotype.

Durch *Extended Reality* (insbesondere Augmented Reality und Virtual Reality) verbringen Menschen immer mehr Zeit in technisch vermittelten Umgebungen. KI-unterstützte Datenanalysen vermitteln zwischen der echten Umgebung, ihrer digitalen Repräsentation und Erweiterung, und menschlicher Wahrnehmung. AR und VR bergen das Risiko einer emotionalen Distanzierung von der natürlichen Umwelt, indem diese digital mediatisiert, personalisiert oder standardisiert wahrgenommen wird**.** VR und AR können dafür genutzt werden, um unangenehme Aspekte der Wirklichkeit (wie z. B. Ungerechtigkeiten, soziale Not oder Naturzerstörung) auszublenden und um gänzlich neue Welten zu erschaffen, in denen als wünschenswert geltende Erfahrungen hervorgerufen und für eine breite Bevölkerung virtuell zugänglich gemacht werden (Greenfield 2018). Wenn Agenten realistisch in VR-Brillen eingeblendet werden können, werden diese vermehrt als menschliche Präsenz wahrgenommen (Interview 9). Die angezeigten Informationen der AR sind nur für den Träger und die Trägerin der AR-Devices sichtbar. Menschen in der Umgebung merken unter Umständen nicht einmal, dass ihr Gegenüber eine AR-Brille trägt (Opazität) (Carman 2019). AR vermag durch die Konstruktion digital erweiterter Umgebungen das überlieferte Selbstverständnis einer geteilten Wahrnehmung der Realität zu mindern bzw. zu fördern, je nachdem ob die AR eher individualisierte oder standardisierte Wahrnehmungen der Umgebung suggeriert (eigene Einschätzung). Der Versuch, Ambiguität in der Weltwahrnehmung zu unterdrücken, kann zu noch mehr Ambiguität führen (Bauer 2018a). Indem jegliche natürliche und künstliche Umgebung durch digitale Schnittstellen wahrnehmbar gemacht und selektiv erweitert wird,

wird „Das Gesetz der natürlichen Künstlichkeit“ radikal weitergeführt (Kahn, Severson, Ruckert 2009).

2.4.3 Die Versprechen der Megamaschine und ihre Einlösung

Unterstützt durch dezentral eingesetzte KI im IoE steigen Vielfalt und Ausmaß der Mensch-Technik-, der Technik-Technik- und der Technik-Umwelt-Kommunikationen. Die daraus entstehende **Megamaschine** wird mit großen Versprechungen versehen, wobei aber im Auge zu behalten ist, dass sich die **Aushandlungsarenen** für die Gestaltung der Lebenswelt erheblich verändern. Die Delegation der Entscheidungshoheit von der Techniknutzung an das Technikdesign, die Fantasie des KI-gestützten Erdsystem-Managements anstelle der politischen Aushandlung und die Verstetigung bestehender Praktiken im Zuge der Realweltreproduktion durch ML stellen die Technikgestaltung vor neue Aufgaben:[45]

Unter dem Signum der *Hyperkonnektivität* steigen Vielfalt und Ausmaß der Mensch-Mensch-, der Technik-Technik-, der Umwelt-Umwelt- und auch der Technik-Umwelt-Kommunikationen ohne Intervention des Menschen (eigene Einschätzung). Die unzähligen Verknüpfungen von Objekten, Organisationen, Menschen und anderen Lebewesen in einer hyperkonnektiven Welt sind für einen Großteil der Menschheit nicht durchschaubar ((Friedrich 2012; Krauss et al. 2016; Friedrich 2012). In jedem Gerät im IoE und auf jedem Server der digitalen Plattformökonomie kann ML stattfinden. Diese Undurchschaubarkeit vermag dazu führen, dass sich Menschen von den vernetzten Systemen abwenden und damit bisherige Mensch-Technik-Interaktionen ersetzen. Dies bedeutet, dass der Mensch immer seltener in Echtzeit in die Prozessausführung operativ eingreift. Anstelle des Abrufens einer definierten Technikfunktion steht bei der *Schwarmintelligenz* ein emergentes Verhalten, das grundsätzlich nicht erklärbar ist. Durch autonome Drohnen- und Roboterflotten können neue Formen von Kontrolle bzw. Kontrollverlust, einschließlich unbeabsichtigtem und unerklärlichem Verhalten des Schwarms, entstehen. Andere sehen eine autonome Drohnenarmee als eine rein hypothetische Gefahr (Warnke et al. 2019). Im Falle des Einsatzes von Roboter-Bienenschwärmen zur Bestäubung findet eine Substitution der Schlüsselfunktion eines natürlichen Organismus, hier: Bienen, in Ökosystemen durch Technik, statt (eigene Einschätzung). Die verschiedenen Varianten technisch auf das Bienensterben zu reagieren, wie z. B. mit Roboter-Bienenschwärmen, erschüttern unser überliefertes Bild der notwendigen natürlichen Bestäubung als Komponente für Ökosystemfunktionalität – unabhängig von der Realisierbarkeit solcher Ansätze.

Die Vision des IoE wird vom Versprechen eines KI-gestützten Erdsystem-Management-Systems flankiert (World Economic Forum et al. 2018), das sich entweder realisieren lässt oder aber als Illusion entpuppen wird. In ersterem Falle würden durch die erforderliche *Hyperkonnektivität* gezielte Transformationen grundsätzlich unterstützt werden können, im letzteren Falle würden die evolutionären Dynamiken überwiegen und den Gestaltungsanspruch der Menschheit für die Welt und die tatsächlichen Bemühungen in die Irre führen. Indem mit Deep Learning in immer größeren und komplexeren Datensätzen Muster analysiert und erkannt werden können, soll eine neue Qualitätsstufe des Wissens und der Erkenntnis über die Biologie des Menschen sowie über die ursächlichen und mechanistischen Zusammenhänge zwischen Genotyp und Phänotyp erreicht werden. Ein umfassender Pool könnte alle wesentlichen Daten und KI-Werkzeuge enthalten, die ein neues Ausfechten der „*nature-versus-nurture*“-Debatte stimulieren würde. Die *Entschlüsselung des Lebens* durch digitale Werkzeuge ermöglicht eine Neureflexion der Determinanten des Lebens. Je nach Ausprägung liegt ein neuer Gendeterminismus oder eine Entmystifizierung der Rolle von Genen im Bereich des Möglichen (eigene Einschätzung). Die

[45] vgl. die Projekte Ethical and Societal Implications of Data Science (https://e-sides.eu/) und EuDEco (http://data-reuse.eu/).

Frage, wer über die kombinierten Datenpools und die für ihre Analyse erforderlichen Werkzeuge zur Vorhersage von Fertigkeiten, Lebenschancen und Anfälligkeiten (vgl. u. a. (Interview 1)) verfügt und wer davon betroffen ist kann erhebliche Auswirkungen auf das eigene Selbstverständnis und auf das Zusammenleben haben. Die Totalität der Vorhersagbarkeit könnte zum einen verbesserte Einschätzungen und Gestaltungen des Lebensweges fördern, zum anderen aber auch einen steigenden Druck auf den einzelnen zum vorausschauenden Wissen über sich selbst und zu vorsorgendem Handeln erzeugen.

Auch unvollständige und widersprüchliche Daten gehen in die Erkennung von Mustern in Datensätzen durch KI mit ein und tragen zur Prognose künftiger Einzelereignisse bei. Ihre Eignung wird durch den Realweltabgleich beurteilt (World Economic Forum et al. 2018). Beim unüberwachten ML verändert sich die Software in Interaktion mit der Realwelt ohne situationsbedingtes menschliches Eingreifen, während beim Reinforcement Learning die Aufgabensetzung in einer Schleife von Verwerfen und Weitermachen immer wieder extern angepasst wird. Im Verborgenen etablieren und verfestigen sich auch für Personen in Deutschland schleichend Social Scoring Systeme (u. a. Arbeitsmarkt, Kreditwürdigkeit). In einer *Big Data-Gesellschaft* werden personenbezogene Entscheidungen anhand von Zuordnungen einzelner zu hochaggregierten statistischen Datenkollektiven getroffen. Die Art, wie in einer Big Data-Gesellschaft entschieden wird, unterscheidet sich grundlegend von den bisherigen Entscheidungsprinzipien. Das Verhandeln von Entscheidungen durch Menschen wird dabei durch starke Präformierung von Entscheidungen durch KI-basierte Datenanalysen ergänzt und teilweise ersetzt, und kann dadurch verhaltenssteuernd sowie normbildend wirken (eigene Einschätzung). Microtargeting verstellt den Blick auf die Welt, blendet die Heterogenität, Widersprüchlichkeit und die Unsicherheit der modernen Welt aus (Bauer 2018a). Die Datenkontrolle durch große Technologieunternehmen stellt eine „Gefahr" für die Menschheit dar ((OECD 2018), S. 26). Die Stellung der Plattformunternehmen bedeutet, dass wenige Akteure mit Big Data und KI tief in die Gesellschaft hineinreichende Macht ausüben können. Solche Gesellschaften sind dann von globalen Big Data Praktiken getrieben, ohne sich explizit für sie entschieden zu haben. Sollte China mit seinem Social Scoring System effektiv umweltgerechtes Verhalten fördern können, so könnten sich demokratische Gesellschaften angesichts der sich verschärfenden Umweltkrisen dazu gezwungen sehen, sich mit Social Scoring Systemen oder ähnlich wirkmächtigen Steuerungssystemen für das individuelle Umweltverhalten zu befassen bzw. diese Option grundsätzlich abzulehnen (eigene Einschätzung).

Exkurs: Jenseits der Planetaren Grenzen (vgl. auch Kapitel 3)

Das Leitplankenkonzept der **Planetaren Grenzen** soll einen sicheren Handlungsraum für die Menschheit gewährleisten (Rockström et al. 2009).[46] Das Überschreiten dieser Grenzen würde dagegen nichtlineare, abrupte Umweltveränderungen von Systemen im kontinentalen bzw. globalen Maßstab auslösen können.[47] Moderne Gesellschaften wirtschaften zunehmend vom

[46] In einschlägigen internationalen Foresight-Journals (*Futures, Foresight*) lassen sich Argumentationsstränge ausmachen, die unabhängig vom Konzept der Planetaren Grenzen einen sicheren Handlungsraum des Menschen thematisieren. Avin et al. 2018) verstehen unter einem kritischen System ein System oder einen Prozess, der bei einer Störung eine signifikante Reduzierung der menschlichen Fähigkeit in seiner jetzigen Form zu überleben auslösen könnte. Acht Risikoszenarien werden formuliert: (1) Asteroiden-Einschlag, (2) vulkanische Super-Eruption, (3) natürliche Pandemie, (4) Ökosystemkollaps, (5) Nuklearkrieg, (6) Krankheitserreger durch Bioengineering, (7) bewaffnete KI und (8) Katastrophen im Zuge des Geoengineerings. (Torres 2019) zählt zu den existenziellen Risiken die Umweltzerstörung (mit Referenzierung der planetaren Grenzen), die Demokratisierung von Wissenschaft und Technik (Biotechnologie, Nanotechnologie, KI) und maschinelle Superintelligenz.

[47] Die Konzepte, Befunde und Schlussfolgerungen wurden im Jahr 2015 modifiziert (Steffen et al. 2015). Die Festlegungen der Planetaren Grenzen sind für einige wenige Kategorien anerkannt (z. B. Stratosphärischer Ozonabbau), für die meisten Kategorien jedoch quantitativ umstritten (z. B. Schwellenwerte für Klimaveränderung) oder nicht formuliert (z. B. atmosphärische Aerosolbeladung). So könnten auch unterhalb der Planetaren Grenze von 350 ppm CO2 in der Atmosphäre durch Kipp-Effekte und daran anschließende Dominokaskaden abrupte globale Klimaveränderungen entstehen (Steffen et al. 2018). Für die Bestimmung anderer Planetarer Grenzen wie die Einführung neuer Entitäten fehlt bislang eine geeignete Metrik. Faktisch lebt die Menschheit in

technischen Kapital und verfolgen Strategien der Verringerung ihrer Abhängigkeit vom natürlichen Kapital (Costanza et al. 1997). Je nachdem, ob die Planetaren Grenzen unterschritten oder überschritten werden, kann die Menschheit vom natürlichen Kapital leben, oder muss vom technischen Kapital leben. Das Konzept der Planetaren Grenzen rekurriert seinem Anspruch nach auf die außerordentlich wichtigen globalen und regionalen Umweltdimensionen des Erdsystems. Es ist darüber hinaus aber eindeutig, dass auch andere Dynamiken Leitplanken für die menschliche Entwicklung darstellen, insbesondere die technologische Transformation und andere gesellschaftspolitische Entwicklungen. Am meisten diskutiert und vielleicht am plausibelsten sind diesbezüglich Nanotechnologie, Biotechnologie und KI, die ebenfalls miteinander kombiniert werden könnten (Baum et al. 2019).

2.5 Schlussfolgerungen

Mit der Identifizierung und Charakterisierung von zehn potenziell disruptiven Anwendungsfeldern der KI wurde in erster Linie eine Grundlage für die Auswahl von Themenkomplexen in Arbeitspaket 3 geschaffen. Durch die Strukturierung in Entwicklungen und Visionen sowie die Bündelung von grundlegenden Veränderungen der Mensch-Technik-Umwelt-Beziehungen durch dieses Set disruptiver KI-Anwendungen wurde gleichzeitig ein Wissensfundament für eine sachliche Befassung mit KI-Entwicklungsdynamiken und ihren potenziellen Auswirkungen geschaffen (Tabelle 3).

Teilbereichen bereits in einer Welt der überschrittenen Planetaren Grenzen. Die Planetaren Grenzen genetische Vielfalt (Teil der Kategorie Integrität der Biosphäre) und biochemische Flüsse gelten als definitiv überschritten. Die Erderhitzung („Klimawandel“) und die Reduzierung der Waldflächen (Kategorie Landnutzungsänderung) befinden sich in der Unsicherheitszone (Steffen et al. 2015). Auch findet sich die Position, dass die globale Frischwassernutzung die Planetaren Grenzen überschritten hat (Jaramillo und Destouni 2015).

Tabelle 3: Entwicklungen und Visionen für Künstliche Intelligenz allgemein und für zehn potenziell disruptive Anwendungsfelder

Digitalisierungsfeld	Entwicklungen	Visionen
Künstliche Intelligenz allgemein	Technische Konjunkturen Normative Anforderungen	Superintelligenz Technischer Posthumanismus
Affective Computing	Erkennen von Affekten und Emotionen Erzeugen von Affekten und Emotionen	Intime Beziehungen mit Robotern Fortleben Verstorbener
Simulation natürlicher Sprache	Kognitive Assistenten Computational Creativity	Autonome Avatare Transformative Kreativität
Extended Reality	Führen im Raum Auslösen von Impulsen im Raum	Nahtlose Augmented Reality
Digitales Enhancement	Digitale Selbstvermessung Brain-Computer-Interfaces	Cyborgs Transhumanismus
Autonome Systeme in bislang unverfügbaren Räumen	Autonome Systeme für den Tiefseebergbau Extraterrestrische Produktion	Blue Economy and Society Space Economy and Society
Big Data Gesellschaft	Scoring und Microtargeting Finanztechnologie	Datengetriebene Gesellschaft Voll ausgebildeter digitaler Überwachungskapitalismus
Hyperkonnektivität	Internet der Menschen & Dinge Internet der Natur	Internet of Everything Dashboard für die Erde
Autonome Systeme im Alltag	Informatorische Systeme Physische Systeme	Verschwinden der Computer aus dem Bewusstsein
Entschlüsselung des Lebens durch digitale Werkzeuge	Genomanalyse Zusammenhang Genotyp und Phänotyp	Entschlüsselung des Alterns Vorhersage der Entfaltung des Lebens von Menschen
Schwarmintelligenz	Drohnenverbünde zur Indoor-Pflanzenbestäubung Drohnenverbünde für Sicherheit und Angriffe	Autonome Drohnenschwärme zur Outdoor-Pflanzenbestäubung Kriegsführung mit autonomen Drohnenschwärmen

3 Ethische Aspekte

Arbeitspaket 3 hat im Rahmen des Gesamtprojekts die Funktion, exemplarisch sowohl ‚KI-ethisch' als auch umweltethisch relevante Felder aufzuzeigen, die für die Auftraggeber (Umweltbundesamt, ggf. das gesamte Umweltressort) im Auge behalten sollte. Dabei werden anthropologische Fragen und vor allem ethische Konfliktfelder, die durch den Einsatz von KI entstehen, analysiert, die für Nachhaltigkeitstransformationen bedeutsam sein können.

3.1 Ziele und konzeptioneller Ansatz

AP 3 nimmt eine vertiefende ethisch-philosophische Analyse vor. Dafür werden in diesem Unterkapitel zunächst einige zentrale Begriffe erläutert und deren Verwendung im vorliegenden Beitrag geklärt. Danach wird die normative Basis der ethischen Analyse begründet und skizziert. Daraufhin erfolgt ein Überblick zu anthropologischen und ethischen Fragen der KI (Kap. 3.2) sowie zwei spezifische thematische Erkundungen zu Affective Computing (Kap. 3.3) und Autonomen Systemen zur Erschließung für Menschen unverfügbarer bzw. schwer zugänglicher Räume (Kap. 3.4). Die Auswahl letzterer erfolgte sowohl (a) hinsichtlich des erwarteten Disruptionspotenzials im Hinblick auf grundlegende Veränderungen von Mensch-Technik-Umwelt-Beziehungen, als auch (b) hinsichtlich relevanter nachhaltigkeits- und umweltethischer Fragen sowie (c) einer aktuellen oder/und mittelfristig erwartbaren Realistik der Anwendungsfelder, um – aus heutiger Sicht – allzu hypothetische oder gar kontrafaktische Debatten möglicher Anwendungen zu vermeiden. Schließlich wurden (d) zwei auf den ersten Blick sehr unterschiedliche Anwendungsfelder ausgewählt.

Affective Computing: Dieses Feld bietet zahlreiche relevante Aspekte, wie Fragen der Möglichkeit und Zulässigkeit der Anthropomorphisierung von Maschinen und die Frage, was es für das menschliche Selbstverständnis bedeutet, wenn Maschinen kognitive und emotionale Fähigkeiten überzeugend simulieren können. In Bezug auf das Erkennen von Emotionen bietet dieses Feld die Möglichkeit, über den Aspekt der natürlichen Sprache hinaus mit einzubeziehen, dass Emotionen auch aus Gesichtszügen/Körperhaltungen gelesen werden können und entsprechend zur Informationsverarbeitung genutzt werden. Teilaspekte aus anderen Digitalisierungsfeldern wie die fortschreitende Entwicklung im Rahmen des Bioengineerings (vor allem bei Brain-Computer-Interfaces, BCIs) können im Kontext des Affective Computings ebenfalls mitgedacht werden.

Autonome System zur Erschließung von für Menschen bislang unverfügbaren (bzw. schwer zugänglichen) Räumen: Dieses Feld bietet zahlreiche Untersuchungspunkte, die in der bisherigen Debatte stark vernachlässigt werden und damit das größte Potenzial ein bislang unterrepräsentiertes Thema in Form von zukünftigem Forschungsbedarf zu erschließen. Zudem ergibt sich hier der stärkste direkte Bezug zu umweltbezogenen Fragestellungen. Damit kann dieses Feld paradigmatisch für eine sich ändernde Mensch-Natur-Beziehung stehen, was im Kontext der anderen Digitalisierungsfelder teilweise nur sehr mittelbar gegeben war. Dabei ist jedoch zu beachten, dass keine Engführung auf ökologische Aspekte stattfindet, weil zahlreiche Fragen globaler Gerechtigkeit aufgeworfen werden, die es zu berücksichtigen gilt. Aspekte aus dem Feld *Extended Reality* können hier zum Teil ebenfalls angesprochen werden.

Kapitel 3.5 nimmt dann eine Synthese vor und bündelt die Erkenntnisse, auch hinsichtlich von relevanten, zu berücksichtigenden Punkten (*points to consider*), die zugleich künftigen Forschungsbedarf anzeigen.

Während Kapitel 3.1 und 3.2 mit dem Thema Ethik beginnen und dann das Feld von Mensch-Technik-Umwelt in philosophisch-anthropologischer Hinsicht bearbeiten, wird in den beiden

thematischen Vertiefungsstudien zuerst letztere Dimension adressiert und danach die Ethik behandelt. Es zeigt sich, dass bei aller analytischen Unterscheidung und Trennung beide Aspekte aber eng miteinander verknüpft sind (Potthast 2015). Insgesamt orientiert sich das Vorgehen an einer interdisziplinären Ethik in den Wissenschaften, die in der sachgerechten Kombination empirischer Wissensgehalte und ausgewiesener Wert- und Normdimensionen zu sogenannten „gemischten Urteilen" oder zumindest Hinweisen auf ethisch relevante *points to consider* gelangt (vgl. (Ammicht Quinn et al. 2015)).

Erkenntnisziele

Die ethische Analyse der Anwendungsbeispiele legt einen Fokus darauf, wie die zu bewertende Technologie eine Transformation hin zu Nachhaltiger Entwicklung behindern oder vorantreiben kann. Damit wird der starke Bezug zu umfassender Gerechtigkeit, zur Gemeinwohlorientierung und dem Einhalten Planetarer Grenzen Rechnung getragen. Ein weiterer Fokus liegt auf dem Potenzial der jeweiligen KI-Technologie, Mensch-Technik-Umwelt-Beziehungen (in Teilen voraussichtlich stark) zu verändern sowie auf einer Bewertung dieses Potenzials. Hierbei ist zentral, dass die genannten Perspektiven nicht additiv sind und nicht auf derselben Ebene liegen. Dennoch bzw. genau dadurch ergibt sich aus der Vielfältigkeit der Blickwinkel eine umfassende Reflexion auf die ausgewählten Technologien, die ausreichend konkret fokussiert ist und zugleich die großen Fragen der Zukunft im Blick hält.

Normativer Rahmen: Nachhaltige Entwicklung und Gemeinwohl

Selbstverständlich gibt es nicht ‚die' einheitliche Ethik, sondern unterschiedliche Systeme der Moralphilosophie, von Tugendethiken über Pflichtenethiken und den Utilitarismus bis hin zu Vertragstheorien und der Diskursethik (vgl. (Ott et al. 2016)). Auch im Bereich der anwendungsbezogenen Ethik gibt es inhaltlich oder methodisch unterschiedliche Ansätze. In systematischer und methodischer Hinsicht sei darauf hingewiesen, dass uns diese Vielfalt bewusst ist, dass wir aber in pragmatischer Absicht das Gemeinsame all dieser Ansätze betonen, ohne die Unterschiede zu leugnen (vgl. (Ammicht Quinn et al. 2015)). Wir haben daher einen normativen Rahmen gewählt, der zumindest auf völkerrechtlicher Ebene von fast allen Staaten akzeptiert und ratifiziert wurde, und der vor allem die praktische Umweltpolitik unmittelbar als rechtliche Basis prägt. Dies ist zum ersten Nachhaltige und zum zweiten das Gemeinwohl.[48]

Ressourcen und Gerechtigkeit: Von Planetaren Grenzen zur Nachhaltigen Entwicklung inkl. der Frage nach dem Guten Leben

Das zentrale, auch durch das Umweltbundesamt motivierte, normative Framework, anhand dessen KI-Technologien ethisch analysiert werden, ist das der **Nachhaltigen Entwicklung (NE)** im Sinne der Vereinten Nationen (WCED 1987) sowie ein Verständnis von Gemeinwohl, welches als eng mit NE verknüpft angesehen wird. Eine Transformation hin zu NE sowie eine Orientierung an NE-relevanten Maßnahmen stellen geeignete Perspektiven dar, um Antworten auf die gegenwärtig vorherrschenden globalen anthropogen verursachten Herausforderungen, wie die Klima- und Biodiversitätskrise, Hunger und Mangelernährung, Migration, Zunahme von Populismus und Rassismus u. a., suchen und vorschlagen zu können.

Mit der normativen Rahmung durch NE wird das (im Projekttitel benannte) Konzept der Planetaren Grenzen (PB) erweitert. Dieses auf die Forschungsgruppe um Rockström zurückgehende (Rockström et al. 2009; WBGU 2011) und weltweit (u. a. vom WBGU (WBGU 2011), (WBGU 2013)) aufgenommene und weiterentwickelte Konzept beschreibt globale ökologische Grenzen, deren Überschreitung die Stabilität bzw. Resilienz der Ökosysteme gefährdet und damit die Lebensgrundlage für Menschen. Auch wenn dieses umweltpolitisch viel

[48] Zu den Limitationen des Ansatzes siehe Kap. 3.5.

beachtete Konzept wichtige Aspekte – wie die messbar bereits überschrittenen Grenzen bestimmter Ökosystem-Parameter in Hinblick auf die biochemischen Kreisläufe, die Biodiversität, die Landnutzung und das Klima – zum Ausdruck bringt, ist es zugleich Kritik ausgesetzt. Die Kritikpunkte erachten wir für so wichtig, dass eine Erweiterung dieses Konzepts um eine umfassende NE-Perspektive als unhintergehbar angesehen wird. So verbleiben (a) die rein physisch gedachten globalen Grenzen in gegebenen politischen, ökonomischen und sozialen Strukturen und benennen (b) keine genuin ethischen Grundlagen jenseits eines sehr allgemeinen Überlebens der Menschheit. Es werden (c) die Symptome globaler Umweltkrisen in den Blick genommen, anstatt deren Ursachen. Der Fokus auf naturwissenschaftlich messbare Daten im Sinne eines (d) „westlichen" Wissensverständnisses führt dazu, so ein weiterer Kritikpunkt, dass bestimmte – in der Regel nicht-westliche – Lebensweisen und Menschengruppen aus dem Blick geraten und damit marginalisiert oder ignoriert werden (können). Es hat zudem zur Konsequenz, dass Wissensbestände keinen Eingang finden, die nicht mit dem im Konzept zugrunde gelegten (westlichen) Naturwissenschafts-Verständnis übereinstimmen, aber dennoch wichtige Lösungsansätze für globale Krisen bieten könnten, wie zum Beispiel *indigenous knowledge* (vgl. (Senanayake, S. G. J. N. 2006) für die Bedeutung für NE). All diese Kritikpunkte sind weder neu oder spezifisch für die Planetaren Grenzen, sie müssen jedoch gleichwohl bedacht werden.

Das Konzept Nachhaltiger Entwicklung ist diesen Kritikpunkten nicht in so grundlegender Weise der zu engen Fokussierung auf physische Grenzen ausgesetzt. Wenn beispielsweise eine soziale Praxis dazu führt, dass eine planetare Grenze überschritten wird, so bietet das Planetare Grenzen-Konzept mit seinem Fokus auf rein ökologisch messbare Daten keine Methode und keine inhärenten normativen Kriterien, diese soziale Praxis genauer in den Blick zu nehmen. Das Konzept der Nachhaltigen Entwicklung, das Aspekte des Planetaren Grenzen-Konzepts explizit beinhaltet, bietet mehr Möglichkeiten: Durch den Fokus auf Gerechtigkeit bietet es normative Kriterien, um die umweltrelevanten Aspekte in ihrem Zusammenspiel mit sozialen Aspekten zu evaluieren. Die Debatte darum, welche Entwicklungen als nachhaltig gelten können und welche nicht, welche Ziele der Sustainable Development Goals priorisiert werden sollten und wie politische Umsetzungen gestaltet werden können, wird kontrovers geführt (dies betrifft besonders die sogenannten Entwicklungsdimensionen des NE-Konzepts u.a. hinsichtlich der Frage nachhaltigen Wirtschaftswachstums). Einig sind sich die Protagonist*innen dieser Debatte jedoch darüber, dass intra- und intergenerationelle Gerechtigkeit die zentrale ethische Grundlage Nachhaltiger Entwicklung (NE) ist (sogenannte Zieldimension des Konzepts). Diese Annahme wird international seit dem wegweisenden Brundtland-Report geteilt, in dem NE wie folgt definiert wird:

Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs. It contains within it two key concepts: the concept of 'needs', in particular the essential needs of the world's poor, to which overriding priority should be given; and the idea of limitations imposed by the state of technology and social organization on the environment's ability to meet present and future needs. ((WCED 1987), Chapter 2)

Hieraus ergibt sich in Bezug auf **intragenerationelle Gerechtigkeit** die Priorisierung derjenigen Menschen, die global am schlechtesten gestellt sind. Näher bestimmt werden muss, welche Bedürfnisse (*needs*) betrachtet werden sollten. Da der sogenannte *basic needs*-Ansatz, nach dem alle das bekommen sollten, was Menschen zum „nackten Überleben" benötigen, den Anforderungen an ein gutes Leben (oder, nach Hans Jonas ((Jonas 1979), S. 36): „Permanenz echten menschlichen Lebens auf Erden") nicht nachkommen kann, legen wir zur konkreteren Bestimmung der Bedürfnisse Martha Nussbaums (Nussbaum 2000, 2010) Fähigkeitenansatz

zugrunde. Für Nussbaum gilt es anzustreben, dass alle den *scope of justice* umfassenden Individuen bestimmte Fähigkeiten ausleben (bzw. Fähigkeiten in Tätigkeiten umsetzen) können, so (wie) sie das wollen. Welche Fähigkeiten das sind, legt Nussbaum in einer sogenannten Fähigkeiten-Liste fest, die sie selbst als offene Liste betrachtet. Die Liste umfasst 1. Leben, 2. Körperliche Gesundheit, 3. Körperliche Integrität, 4. Sinne, Vorstellungskraft und Denken, 5. Gefühle, 6. Praktische Vernunft, 7. Zugehörigkeit [zu einer Gemeinschaft, Sozialität], 8. Interaktion mit anderen Spezies, 9. Spiel und 10. Kontrolle über die eigene Umwelt [im Sinne von demokratischer politischer Teilhabe und Gestaltungsmöglichkeit] ((Nussbaum 2010), S. 112 - 114). Gemäß Nussbaum verlangt es die Gerechtigkeit, dass jeder Mensch (gegebenenfalls durch entsprechende Maßnahmen) in die Lage versetzt werden muss, alle Fähigkeiten der Liste in Tätigkeiten fördern und praktizieren zu können. In einer gerechten Gesellschaft sollten Fähigkeiten nicht gegeneinander abgewogen werden müssen. Ist dies gewährleistet, kann von einem guten menschlichen Leben gesprochen werden.

Der Fähigkeitenansatz ist ein produktiver Versuch, die Forderung nach inter- und intragenerationeller Gerechtigkeit und der damit verbundenen notwendigen Bedürfnisse jenseits des „nackten physischen Überlebens“ zu konkretisieren und dabei nicht alle Menschen als in ihren Bedürfnissen nach Realisierung ihrer Fähigkeit identisch anzusehen. Damit ist jedoch – um im Thema zu bleiben – kein normativer Algorithmus vorgegeben, sondern Hinsichten, die weiter diskursiv und deliberativ genauer zu bestimmen und auszuhandeln sind.

Vorliegend wird ein umfassendes Verständnis von Gerechtigkeit mit Bezug auf NE vertreten, gemäß dem soziale, ökologische, ökonomische sowie potenzielle andere Dimensionen der Gerechtigkeit so stark miteinander verwoben sind, dass sie nicht als separate Gerechtigkeits-Bereiche (und auch nicht als separate Nachhaltigkeits-Säulen, s.u.) aufgefasst werden können und daher auch nicht sollen. Ein solch umfassender Ansatz bringt jedoch auch praktische Schwierigkeiten wie die Handhabung der hohen Komplexität, erschwerte Kommunikation, Gefahr des Aufweichens durch Interessengruppen, fortwährende Relativierungs- und Rechtfertigungsdiskurse oder unklare Prioritäten in Abwägungsfragen mit sich. Diese Herausforderung besteht allerdings stets bei umfassenden „dichten Begriffen“ (Williams 1985), insbesondere in der politischen Arena, was die analogen Auslegungsdifferenzen von Freiheit oder Demokratie belegen. Ein wissenschaftlich anspruchsvolles Konzept muss die Komplexitätsherausforderungen annehmen und stets mitdenken, das umfassende Verständnis aber dennoch aufrechterhalten und möglichst kommunizierbar machen.[49]

Der Fokus auf **Fähigkeiten** gibt eine Antwort auf die Frage, wie tatsächlich gerechte Gesellschaften konstituiert sein sollten. Sie bzw. die in ihnen wichtigen Institutionen müssen es ermöglichen, dass alle Menschen, die für ein gutes menschliches Leben notwendigen Fähigkeiten in Tätigkeiten umsetzen können, so sie das wollen, und nicht durch Faktoren wie Armut, mangelnde Teilhabemöglichkeiten oder körperliche Beeinträchtigungen daran gehindert werden. Dadurch ist dieser Fokus sowohl für eine konkretere Bestimmung intra- und intergenerationellen Gerechtigkeit in der NE-Debatte, als auch zur Bestimmung dessen, was unter Gemeinwohl (siehe unten) gefasst werden sollte, zielführend. Wie diese Gesellschafts-Konstituierung umgesetzt werden kann, muss in politischen und gesellschaftlichen Aushandlungsarenen diskutiert und präzisiert sowie von entsprechend demokratisch legitimierten Institutionen entschieden werden. Das Konzept Nachhaltiger Entwicklung ist zudem plural angelegt: Es gibt nicht genau ‚den‘ einen richtigen Pfad, sondern es kann mehrere konkrete Entwicklungswege in Richtung mehr Nachhaltigkeit geben. In jedem Fall handelt es

[49] Siehe dazu die Ansätze des „Scrollytelling“ im Kap. 5 des vorliegenden Berichts.

sich um normative Entwürfe, die hinsichtlich der Legitimität und Wünschbarkeit ihrer Ziele, Mittel, Folgen und Nebenfolgen zu prüfen bzw. begründen sind.

Das Konzept Nachhaltiger Entwicklung, wie es hier verstanden wird, fokussiert entsprechend auf gerechte gesellschaftliche Verhältnisse und schließt alle relevanten Bereiche mit ein. Gerechte gesellschaftliche Naturverhältnisse stehen zwar im Fokus als Bedingung der Möglichkeit planetaren Lebens, jedoch sind diese nicht losgelöst von anderen Verhältnissen zu betrachten, sondern im Gegenteil sehr stark mit anderen Bereichen wie technischen oder ökonomischen verknüpft. Die – in der Debatte häufig anzutreffende – reduzierende Rede von „ökologische Nachhaltigkeit" ist daher unseres Erachtens falsch. „Ökologische Nachhaltigkeit" gibt es ebenso wenig wie „ökonomische" oder „soziale Nachhaltigkeit". Nachhaltigkeit bzw. Nachhaltige Entwicklung ist immer zugleich als ökologisch, ökonomisch und sozial zu denken bzw. – und das ist entscheidend – hat jede NE-relevante Herausforderung sowohl ökologische, ökonomische wie auch soziale Implikationen. Diese Implikationen können gegebenenfalls einzeln stärker fokussiert werden, ihre grundsätzliche Verwobenheit muss dabei jedoch immer mitgedacht werden. Wir verwenden die Termini Nachhaltigkeit bzw. Nachhaltige Entwicklung äquivalent, bevorzugen aber letzteren, weil er zum einen die globale politische Gerechtigkeitsdebatte ausdrücklich(er) unter Bezug auf die UN-Debatten adressiert, zum anderen auch weniger als Nachhaltigkeit eine bloße Fokussierung auf kluge Ressourcennutzung nahelegt.

Eine wichtige Erweiterung dieses im Kern anthropozentrischen Konzepts bietet die UN-Biodiversitätskonvention (Convention on Biological Diversity, CBD, beschlossen auf der UN-Weltkonferenz für Umwelt und Entwicklung in Rio den Janeiro, auf der auch in der „Agenda21" das Prinzip Nachhaltiger Entwicklung als verbindlich beschlossen wurde). Hier haben alle Signatarstaaten die Aussage der Präambel mit gezeichnet, dass die Biologische Vielfalt auch einen eigenen Wert (*intrinsic value*) aufweist und ein gemeinsames Erbe der Menschheit darstellt.[50] Zwar wird in der CBD nicht erläutert, ob damit ein relationaler eudaimonistischer Wert gemeint ist, der zwischen Menschen und Biodiversität entsteht, der über einen bloßen Nutzenwert hinausgeht, aber eben mindestens anthroporelational bleibt (Eigenwert; vgl. Eser / Potthast 1999), oder ob es sich um einen moralischen Wert der Biodiversität an und für sich – also einen Selbstwert völlig unabhängig von menschlicher Wertschätzung – handeln soll. Gleichwohl ist damit die Bedeutsamkeit der belebten Natur jenseits ihrer unmittelbaren Nützlichkeit als Ressource völkerrechtlich bekräftigt und damit das Konzept Nachhaltiger Entwicklung über eine enge Anthropozentrik hinausgerückt (Potthast 2014).

Ebenfalls im Kontext der ‚UN-Riokonferenz' beschlossen wurde das – bereits länger aus dem Umweltrecht bekannte – Vorsorgeprinzip oder *precautionary principle*:

Angesichts der Gefahr irreversibler Umweltschäden soll ein Mangel an vollständiger wissenschaftlicher Gewissheit nicht als Entschuldigung dafür dienen, Maßnahmen hinauszuzögern, die in sich selbst gerechtfertigt sind. Bei Maßnahmen, die sich auf komplexe Systeme beziehen, die noch nicht voll verstanden worden sind und bei denen die Folgewirkungen von Störungen noch nicht vorausgesagt werden können, könnte der Vorsorgeansatz als Ausgangsbasis dienen. (Agenda 21, Kap. 35, Abs. 3)[51]

Oder, mit Hans Jonas ((Jonas 1979), S. 70) formuliert: „Der schlechten Prognose den Vorrang zu geben gegenüber der guten, ist verantwortungsbewusstes Handeln im Hinblick auf zukünftige Generationen". Selbstverständlich gilt auch hier, dass die konkrete Auslegung des Prinzips

[50] Alle relevanten Dokumente unter www.cbd.int (letzter Zugriff 25.10.2021).

[51] Online verfügbar unter: http://www.agenda21-treffpunkt.de/archiv/ag21dok/kap35.htm, zuletzt geprüft am 15.11.2021

durchaus strittig ist, aber dessen grundlegende Bedeutsamkeit zur Gefahrenabwehr wird nicht bestritten, wo es um mögliche extrem schwerwiegende, ggf. globale, irreversible Effekte geht.

Gemeinwohl

Wie der NE-Begriff ist auch der Gemeinwohl-Begriff ein vielverwendeter Begriff, der im Rahmen zahlreicher strategisch-politischer und wissenschaftlicher (Politische Theorie, Ökonomik, Ethik) Kontexte verwendet wird. An dieser Stelle soll der Begriff etwas enger gefasst werden, um deutlich zu machen, was vorliegend darunter verstanden wird. Franz-Xaver Kaufmann ((Kaufmann 2002), S. 33) bezeichnet Gemeinwohl als die Maxime eines auf die politische Gemeinschaft bezogenen Handelns.[52] Abgesehen davon, dass eine Antwort auf die Frage gegeben werden muss, welche politische Gemeinschaft damit gemeint ist, sollte in pluralistischen Gesellschaften „korrekterweise nicht von Gemeinwohl im Singular, sondern von Gemeinwohlbelangen im Plural“ gesprochen werden (Hasenöhrl 2005). Vier Fragen sind zentral, die beantwortet werden müssen, um das Gemeinwohl-Verständnis zu klären ((Offe 2012), S. 673):

- Wer ist der „*social referent*“? (Welcher Gruppierung soll das Gemeinwohl dienen; Familie, lokale Gruppe, Nation, Europäische Ebene, Menschheit?)
- Welches ist der „*temporal scope*“? (Drei relevante temporäre Beziehungen: Gegenwart, nahe Zukunft, weiter entfernte Zukunft)
- Welches sind die „*substantive features*“? Darunter versteht z. B. Claus Offe (ebd.) „the goods and values that are supposed to be attained or realized through conduct oriented towards the common good”.
- Welche *Akteure* und welche *Verfahren* sollen Teil einer bindenden Antwort auf diese drei Fragen werden?

Die von Offe im Gemeinwohlkontext adressierten Arbeitsfragen lassen sich innerhalb des Konzepts der Nachhaltigen Entwicklung gut abbilden, da sie ebenso adressiert werden. Die vierte Frage verweist explizit auf die stets mit zu betrachtenden Akteur*innen innerhalb der verschiedenen Handlungsfelder, die durch die untersuchten Technologien angesprochen bzw. betroffen sind.[53]

Für die vorliegende Untersuchung sollte der *social referent* als gesamte Menschheit einschließlich zukünftiger Generationen gedacht werden. Die Autonomen Systeme (AS) finden global Einsatz und Auswirkungen des Technologieeinsatzes können ebenfalls globaler Natur sein. Ein solch weites Verständnis des sozialen Referenten erschwert allerdings sowohl den ethischen Analyserahmen als auch eine mögliche Konsens-Findung. Je größer die Gruppe ist, die mit dem Begriff umfasst wird, desto schwieriger wird es, sich auf ein gemeinsames Verständnis von Gemeinwohl zu einigen. Je kleiner die Gruppe ist, desto größer wird allerdings die Gefahr,

[52] Ähnlich führen Meyer et al. (Meyer et al. 2013, S. 14) eine von drei verschiedenen Begriffsverständnissen des Begriffs „Wohlfahrt“ auf, gemäß der Wohlfahrt im wissenschaftlichen Kontext als Gesamtnutzen der Gesellschaft (oder im Fall von Wohlfahrt auch des Individuums) verstanden wird. Das von diesen Autoren verwendete Wohlfahrts-Verständnis entspricht dem des „Wohlergehens“, womit Fragen des guten Lebens aufgeworfen werden. Diese Verwendung des Begriffs deckt sich weitestgehend mit dem vorliegend vertretenen Verständnis von Gemeinwohl, dem Martha Nussbaums Fähigkeitenansatz als Explikation einer Gerechtigkeitstheorie zugrunde gelegt wird (siehe an anderer Stelle). Eine tiefergehende begriffliche Verhältnisbestimmung von Gemeinwohl und Wohlfahrt ist vorliegend nicht zielführend und würde den Rahmen sprengen.

[53] Entsprechend erfolgt eine Akteurs-Analyse im Rahmen dieses Projekts in AP 4.

andere Gruppen zu exkludieren und ihnen dadurch die Vorteile bestimmter Gemeinwohl-Maßnahmen vorzuenthalten.[54]

In Bezug auf den *temporal scope* präzisiert Offe:

> "The common good is thus a moral-political condition of society, to which present efforts are dedicated, which is realized in the future, and which is validated as such by looking back from a second future." ((Offe 2012), S. 676)

Vor der Implementierung einer Maßnahme, die dem Gemeinwohl dienen soll, muss also gefragt werden, wie gut die Aussichten sind, dass diese auch retrospektiv als geeignet bewertet wird und ob die Werte, auf denen sie aufbaut, universell genug sind, um auch zukünftig noch als Wertefundament zu gelten. Als geeignetste *Metrik*, um Gemeinwohl messen zu können, dient die Kategorie der Gerechtigkeit. Die zentrale Bedeutung derselben sowohl für eine Konzeption Nachhaltiger Entwicklung als auch für einen überzeugenden Gemeinwohl-Ansatz ergänzen sich und zeigen eine von mehreren Überschneidungen von NE und Gemeinwohl. Um zu konkretisieren, welche Gerechtigkeitstheorie als überzeugend gelten kann, verweisen wir abermals auf Nussbaums Fähigkeiten-Ansatz (vgl. oben) als eine pragmatische Herangehensweise, um wichtige Aspekte eines guten menschlichen Lebens zu berücksichtigen und keine zu engen inhaltlichen Bestimmungen zu oktroyieren.

Die Frage nach den *Akteur*innen* muss kontextabhängig beantwortet werden und lässt sich nicht verallgemeinernd beantworten. Wichtig sind hier Fragen der Verantwortung sowie danach, wer am stärksten profitiert bzw. geschädigt wird.

Der Zusammenhang von *Planetaren Grenzen*, *Nachhaltiger Entwicklung* und *Gemeinwohl* ist nicht in ein einfaches Schema zu bringen. Am ehestens wäre das Planetare Grenzen-Konzept als physisch-chemisch-ökologische Bedingung der Möglichkeit einer gerechten Gegenwart und Zukunft zu verstehen, wobei auch hier neben dem Problem der empirischen Ungewissheit über Kipppunkte und anderes ggf. sehr spezifische Aspekte durch den rein globalen Blick verloren gehen.[55] NE und Gemeinwohl stellen eine unterschiedliche Herangehensweise an Fragen der Gerechtigkeit dar, wobei Gemeinwohl stärker auf kleinere bis mittlere politische Einheiten (Gruppen, Regionen, „Völker", Nationalstaaten[56]) fokussieren und NE von Umweltproblemen und globalen Unterschieden der Entwicklungspfade ausgeht. Mit etwas gutem Willen lassen sie sich als material-inhaltlich weitgehend gleichbedeutend, gleichsinnig orientiert, aber von unterschiedlichen Perspektiven ausgehend, verstehen.

Der WBGU ((WBGU 2019a, 2019b), S. 35f.) hat in seinem Gutachten zur Digitalisierung einen „normativen Kompass" vorgeschlagen, der letztlich auf ganz Ähnliches abzielt: Er stellt ins normative Zentrum[57] den eher rechtlich als ethisch verstandenen Menschenwürdebegriff und hat drei „Pole" mit „Natürliche(n) Lebensgrundlagen" mit einem Leitplankenansatz, der im Grundsatz den Planetaren Grenzen entspricht, mit „Teilhabe" als Übernahme des Fähigkeiten-Ansatzes sowie mit „Eigenart" als individuellem Schutzrecht der „Entfaltungsfreiheit". Insgesamt

[54] Offe (Offe 2012, S. 675) schlägt den Menschenrechts-Ansatz als einen Ansatz vor, um den *„social referent"* festzulegen, aus dem sich die gesamte Menschheit als solcher ergibt. Da das vorliegende NE-Verständnis eng mit den Menschenrechten verzahnt ist, erscheint Offes Argumentation auch für das vorliegende Vorhaben stimmig.

[55] Ein Versuch, die Planetaren Grenzen für eine Region zu spezifizieren, bietet die Europäische Umweltagentur: https://www.eea.europa.eu/publications/is-europe-living-within-the-planets-limits

[56] Für eine umfassende Studie, die die Zugrundelegung des Bruttoinlandprodukts als Maß für – im Sinne des Gemeinwohls verstandener – Wohlfahrt kritisiert (vgl. (Diefenbacher et al. 2016)).

[57] Die Visualisierung mit Würde im Zentrum und drei Polen einer sternförmigen Struktur macht aus der ‚einnordenden' einsinnigen Richtung eines Kompasses eine Triangulation und Balancierung zwischen den drei Polen – wobei dann aber Würde statt „Ausgangspunkt und Zielbild" (WBGU 2019b, S. 42) eher als resultierender Ort gelungener Triangulation erscheint und nicht als zentrale Orientierung der Richtung. Hier kommen die Kompass-Metaphorik ebenso wie die Idee einer grafischen Visualisierung an ihre Grenzen bzw. eröffnen mögliche Missverständnisse.

sehen wir hier keine Widersprüche zu unserem normativen Rahmen, sondern wiederum variierende perspektivische Zugänge.

Für unsere Analyse sei betont, dass die Unhintergehbarkeit der Menschenwürde und der Menschenrechte auch den Konzepten der NE und des Gemeinwohls, wie wir sie verstehen, zugrunde liegen.[58]

3.2 Überblick zu ethischen und anthropologischen Fragen Künstlicher Intelligenz

In der KI-Debatte bestehen etliche moralisch gehaltvolle Narrative, nach denen KI-Technologien in verschiedenen Bereichen stark negative Konsequenzen mit sich bringen werden. Hierzu gehören das Narrativ, dass KI eine Gefahr für die Demokratie darstellt, oder dass KI-Technologien problematische Konsequenzen mit sich bringen, da sie stark vernetzt (→ Hyperkonnektivität, Abschnitt 2.7), ubiquitär (→ Big Data, Abschnitt 2.6) und dezentral und daher einer aktiven Steuerbarkeit entzogen sind und zugleich das Leben umfassend beherrschen. Prominent ist auch das Narrativ, in dem das Auftreten einer starken bzw. v. a. Super-KI als Gefahr betont wird, die mit den Fähigkeiten erwachsener Menschen ausgestattet ist und nicht auf wenige Anwendungsbereiche beschränkt bleibt, sondern ihre Intelligenz variabel einsetzen kann (starke KI) und sich gegebenenfalls zur ‚omnipotenten' Super-KI entwickeln kann, die menschlichen Fähigkeiten überlegen ist. Dazu äußert sich Hagendorff (2019, S. 4):

> In this context, it is noteworthy that the fear of the emergence of superintelligence is more frequently expressed by people who lack technical experience in the field of AI – one just has to think of people like Stephen Hawking, Elon Musk or Bill Gates – while "real" experts generally regard the idea of a strong AI as rather absurd.

Dem gegenüber stehen die verschiedenen auch moralisch sehr positiv konnotierten Narrative und/oder Zukunftsszenarien, die KI-Technologien, je nach Einsatzort, als optimale oder zumindest richtige Lösung unterschiedlichster Probleme proklamieren. So soll KI dazu beitragen, Expert*innen-Entscheidungen zu vereinfachen, zu beschleunigen und damit zuverlässiger zu machen (Medizin, Rechtswesen, Bankensektor), Menschen gesünder leben zu lassen, unbequeme Arbeitssituationen abzubauen (Fließbandarbeit, unverfügbare Räume), mehr Beteiligung zu ermöglichen (Social Media, Zugang zu Informationen) und die Umwelt zu retten (Müll-Bots, Grow-Bots, Agrartechnik).

Zu berücksichtigen ist – und dies gilt sowohl für dystopische als auch utopische Narrative – dass die meisten Veränderungen doch graduell vonstattengehen: Z. B. ändert sich die evidenzbasierte Medizin seit Jahren durch die Verwendung von KI (zentrale Fragen hierbei sind und bleiben die danach, was eine Evidenz ist und wie Daten genutzt werden). Die sowohl positiven wie negativen Erwartungen des „fundamental Anderen" oder „Neuartigen" werden bislang und vielleicht auch in Zukunft in den meisten Bereichen möglicherweise nicht erfüllt – und dies betrifft vor allem die in den Zukunftsvisionen „„groß" diskutierten.

3.2.1 KI-Leitlinien vor dem Hintergrund von Nachhaltiger Entwicklung und Gemeinwohl – eine Bestandsaufnahme

In der ethischen Forschung gibt es sehr verschiedene Ansätze zu einer „Ethik der KI" (vgl. (Hagendorff 2020)):

[58] Und wir sind uns dabei bewusst, dass zu diesem Thema eine fast unüberschaubare eigene ethische Debatte besteht, vgl. Düwell et al. 2014; in rechtsvergleichender Hinsicht (vgl. Becchi und Mathis 2019).

- Reflexionen darüber, wie man ethische Prinzipien in Entscheidungsroutinen von KI-Systemen integrieren kann,
- empirische Studien darüber, wie die sogenannten „Trolley Cases“ gelöst werden, also moralische Dilemmata, in denen die Schädigung bestimmter Personen statt anderer unvermeidlich ist,
- diverse Meta-Studien über allgemeine Aspekte einer KI-Ethik sowie Reflexionen zu spezifischen Problemen, wie Datensicherheit, Transparenz etc.,
- Zusammenstellung von KI-Richtlinien.

Hagendorff (Hagendorff 2020) präsentiert seine Analyse verschiedenster KI-Richtlinien und stellt die dort genannten ethischen Aspekte zusammen, die in Forschung und Entwicklung von KI-Anwendungen laut Autor*innen der jeweiligen Leitlinien reflektiert und berücksichtigt werden sollten. Genannt werden folgende Forderungen:

- Der Schutz der Privatheit von Nutzer*innen muss gewährleistet sein.
- KI-Anwendungen müssen zuverlässig und stabil sein.
- Bei der Erhebung von Daten sowie deren anschließender Anwendung in KI-Technologen muss gewährleistet sein, dass keine Diskriminierung erfolgt.
- Der Zugang zu KI-Anwendungen muss gerecht sein.
- Die KI-Anwendungen müssen transparent, interpretierbar und erklärbar sein.
- Sicherheit und Cybersicherheit müssen gewährleistet sein.
- Die Systeme müssen weiterhin von Nutzer*innen steuerbar und kontrollierbar sein.

Tatsächlich werden auch die Aspekte „Gemeinwohl“ und „Nachhaltigkeit“ in diversen Richtlinien, sowohl aus der Forschung als auch der Industrie und der Politik explizit angesprochen (z. B. AI4People, Asilomar AI Priciples, AI Now 2016 & 2017 Report, IEEEs Ethically Aligned Design, AI at Google, IBM Everyday Ethics for AI, vgl. (Hagendorff 2020) sowie im WBGU Gutachten „Unsere Gemeinsame Digitale Zukunft“).

In weniger als der Hälfte der von Hagendorff (Hagendorff 2020) untersuchen Richtlinien berücksichtigt werden die Aspekte

- Dual Use Möglichkeiten und militärische Entwicklungen,
- Solidarität, Inklusion und Sozialer Zusammenhalt,
- die Verbindung zwischen Wissenschaft und Politik (z. B. politischer Missbrauch)
- ethnische Diversität und Diversität der Geschlechter.

Sehr selten (in vier und weniger) angesprochen werden die Aspekte

- Auswirkungen auf die Zukunft der Arbeit,
- Öffentliche Aufmerksamkeit, Fragen der digitalen Bildung in Bezug auf KI,

- Menschliche Autonomie.

Fast nie in den Leitlinien adressiert werden folgende Aspekte:

- Schutz von Whistleblowern, die der Allgemeinheit Informationen aus den Entwicklungsorten der KI-Technologien zur Verfügung stellen wollen und damit gegen die Auflagen ihrer Arbeitgeber verstoßen,
- versteckte Kosten und Schäden, wie z. B. Clickwork, Energiebedarf, Ressourcenverbrauch.

Die umfassende Untersuchung von Thilo Hagendorff ist hilfreich, um sich einen ersten Eindruck zu verschaffen, welche ethischen Aspekte in der internationalen Debatte um KI-Ethik vornehmlich adressiert werden, welche nur manchmal Eingang finden und welche Themenbereiche oder Werte praktisch ganz außen vor bleiben.

Was sich beispielsweise in den von Hagendorff untersuchten KI-ethischen Richtlinien nicht findet, ist eine explizite Adressierung und Untersuchung der Manipulierbarkeit von Menschen durch KI-Technologien. Wir gehen davon aus, dass die in den Richtlinien benannten Werte und Problemfelder häufig genau aus dem Grund angesprochen werden, dass die Manipulierbarkeit der Menschen verhindert werden soll. Explizit benannt wird das jedoch nicht, was ein wichtiges Desiderat darstellt und im Projekt in der Untersuchung zu *Affective Computing* intensiv diskutiert wird (vgl. Kapitel 4). Ebenfalls zentral ist die Untersuchung, in wessen Verantwortungsbereich die Tätigkeiten von KI-Technologien fallen. Auch dies wird in den untersuchten Leitlinien indirekt mitverhandelt, sollte dort jedoch auch direkt adressiert und lösungsorientiert analysiert werden. Gleiches gilt für die – vor allem in Europa erhobene – Forderung nach der Ausrichtung der Entwicklungen an der Menschenwürde. In den untersuchten Richtlinien wird dieser Aspekt auf „Autonomie“ verkürzt als zentral angesehen, wie sich z. B. im WBGU-Bericht ((WBGU 2019a, 2019b)) prominent zeigt. Jenseits der von Hagendorff zusammengestellten Leitlinien besteht bereits eine lebhafte ethische Debatte gerade auch über die Desiderate der Leitlinien – wie z. B. der WBGU-Bericht prominent zeigt.

Es bestehen weitere Aspekte, die in den von Hagendorff untersuchten Richtlinien nicht benannt werden, unseres Erachtens jedoch eine wichtige Rolle in der Debatte um KI-Ethik spielen und in der ethischen Rezeption entsprechend diskutiert werden sollten, und die sowohl aus Perspektive des Fähigkeitenansatzes als auch unter Gemeinwohlaspekten einschlägig sind:

- die Notwendigkeit der demokratischen Kontrolle und die Möglichkeiten, mit Governance-Strukturen Entwicklungen zu steuern,
- die Problematik, dass bestimmte Entwicklungen Vorstellungen von Life-Style oder „gutem Leben“ verändern, ohne dass man sich davor schützen kann (z. B. Vorstellungen von ständiger Selbstoptimierung, die durch *Smart-Watches* verstärkt werden).
- *Smart Cities* bieten durch die Vernetzung vieler Endgeräte und die Preisgabe vieler privater Daten in *Smart Homes* Grundlagen für umfassende Datensammlungen. Dasselbe gilt für Datensammlungen in der natürlichen Umwelt, wie sie sich z. B. im Rahmen der Projekte *Digital Earth*[59] oder *Digital Ocean* finden. Daraus ergibt sich die Frage, ob das politisch-

[59] Vgl. https://ec.europa.eu/jrc/en/research-topic/digital-earth, zuletzt geprüft am 23.10.2021.

ethische Problem besteht, verstärkt in Richtung einer Technokratie oder Expert*innen-Herrschaft zu gehen.

- Die Leistungsfähigkeit der digitalen/KI-Technologie nimmt stetig zu. Gegenwärtig werden in Bezug auf Maschinelles Lernen Durchbrüche in den Bereichen *Few* bzw. *One Shot Learning* (nur noch wenige bzw. ein einziges Bild ist notwendig, um die KI zu trainieren) und beim schnelleren Lernen durch *Duelling Networks*[60] erwartet. Hieraus ergibt sich die Herausforderung, als Gesellschaft auf juridischer, ethischer und politischer Ebene der enormen Geschwindigkeit der KI-Entwicklungen „standzuhalten" und in Bereichen, die traditionell viel Zeit in Anspruch nehmen, in einem kürzeren zeitlichen Intervall gleichwohl zumindest darauf zu reagieren, von eigentlich gebotenen antizipierenden ethischen Diskursen ganz zu schweigen.
- Navigationssysteme, Übersetzungssoftware, Lehr-Lernprogramme: Es wird diskutiert, wie ein „Recht auf Vergessen" umgesetzt werden kann, da online gespeicherte Daten über sehr lange Zeiträume verfügbar bleiben.

Im Allgemeinen findet sich in der bestehenden Debatte zur KI-Ethik zumeist eine sehr binäre Rahmung von Chancen auf der einen und Risiken auf der anderen Seite. Diese wird u. a. auch in Abbildung der Kernchancen von KI des WBGU ((WBGU 2019a, 2019b), S. 79) deutlich, denen die Autor*innen Risiken gegenüberstellen (vgl. Abb. 4). Diese „großen Linien" innerhalb der öffentlichen (und wissenschaftlichen) Debatte wurden in der hier getätigten ethischen Analyse mitgedacht, dabei aber differenzierter betrachtet. Durch den starken Bezug zur Nachhaltigen Entwicklung gehen wir davon aus, über eine binäre Rahmung „hinauszugehen", wobei gleichwohl die in Abbildung 4 angesprochenen „Kernchancen" und ihre korrespondierenden „Risiken" auch im Folgenden von Bedeutung sein werden.

Abbildung 4: Chancen und Risiken der (Nicht)Nutzung von KI

Quelle: (WBGU 2019b), S. 79

[60] *Duelling Networks* basieren auf dem Prinzip, dass eines der beiden neuronalen Netzwerke realistische Artefakte wie Fotos generiert (Generator) und dass das andere Netzwerk zwischen Fake und Fakt unterscheidet (Diskriminator). Durch Feedbacks lernen Generator und Diskriminator sehr schnell (projektinterne Fassung von AP 2, S. 21).

3.2.2 Veränderungen der Mensch-Technik-Umweltbeziehung durch KI: Unterschiedliche anthropologische, natur- und technikphilosophische Zugänge und Implikationen

Potenzielle Veränderungen in Hinblick auf die Mensch-Technik-Umwelt-Beziehungen, die sich aus der Entwicklung und Etablierung von KI-Technologien ergeben, stellen einen zentralen Analyseaspekt des Projekts dar. Sie sind eines der Kriterien, nach denen KI-Technologien ethisch evaluiert wurde. Für jede einzelne (oder das Zusammenwirken bestimmter) KI-Technologien lassen sich etliche mögliche Veränderungen in diesen Beziehungen ausfindig machen. Wenn über KI-Technologien „im Allgemeinen" gesprochen wird, sind solche Veränderungspotenziale hingegen schwer zu bestimmen, da die meisten mit Implikationen verbunden sind, die sich aus der jeweils konkreten Technologie und ihrem Anwendungskontext zusammen ergeben. Auch die große Breite dessen, was alles als KI bezeichnet oder darunter subsumiert wird, erschwert den Zugriff.

Technikphilosophisch betrachtet, lassen sich idealtypisch zwei „Basisnarrative" unterscheiden. Das erste ordnet einer (mehr oder weniger bestimmten) Technik eine inhärente (ziemlich) eindeutige Richtung zu. Beispiele wären zwei sich gegenüberstehende, aber jeweils eindeutige, Narrative zur Gentechnik, (a) das des als Schlüssels zum Menschheitsfortschritt bzw. (b) das der vollständigen Selbstentfremdung. Dieses binäre Narrativ, lässt sich analog auf KI übertragen. Das zweite idealtypische Narrativ ist das Ambivalenz-Narrativ, welches eng mit einem instrumentalistischen Technikverständnis einhergeht: Technik, und damit auch sehr viele KI-Anwendungsbereiche, bietet inhärent die Möglichkeit, dass dieselbe Technologie sich positiv oder negativ auf Mensch-Technik-Umwelt-Beziehungen auswirken kann, je nach Art und Kontext der Anwendung. Nicht fremd sind Bezüge der beiden Narrative zu deontologisch bzw. konsequentialistisch geprägten Ethikansätzen.

Doch hier soll es im Folgenden nicht um normative Aspekte der Ethik gehen, sondern um die Frage, wie sich Beziehungen und Konstellationen ändern können, die sowohl die Mensch-Technik- als auch die Mensch-Umwelt- und die Technik-Umwelt-Relationen betreffen. Das macht die Aufgabe nicht eben einfach, und es sei vorausgeschickt, dass wir eher davon ausgehen, dass die genannten Beziehungen sich nicht alle gleichermaßen und gleichsinnig mit Bezug auf „die KI" ermitteln lassen. Dennoch werden wir versuchen, allgemeinere Tendenzen zu formulieren, wo dies sinnvoll erscheint.

Die Veränderungen in Mensch-Technik-Umwelt-Beziehungen, die sich für KI-Technologien „im Allgemeinen" bestimmen lassen, legen die folgenden Unterkapitel, gleichwohl unter Verweis auf illustrierende Beispiele, dar. Basale anthropologische Fragen werden dabei immer wieder neu aufgeworfen. In einer klassischen Unterscheidung wird beispielsweise zwischen dem biologischen Körper (‚den ich habe') und dem phänomenal empfundenen Leib (‚der ich bin') unterschieden (vgl. (Plessner 2003); (Ammicht Quinn 2004) lehnt aus einer feministischen Perspektive diese begriffliche Dualität übrigens ab). Inwiefern Technik das Körper-Leib-Verhältnis modifiziert oder gar zum Verschwinden bringen könnte, wird immer wieder erörtert. Dazu gehört auch die Debatte um Kompetenz- und Kontrollverlust oder auch -gewinn durch Technik. Zugleich ist auf der sozialen Ebene zu fragen, wie Technik die Gesellschaft verändert, aber eben auch wie Gesellschaften agieren, so dass sie bestimmte Technikentwicklungen fordern oder fördern.

Was verbindet Menschen mit ihren Artefakten, einschließlich Technik?

Es kann zu Verlusten hinsichtlich menschlicher Kompetenzen kommen, wenn KI-Systeme Menschen Aufgaben abnehmen, weil bestimmte diese Kompetenzen erfordernde und zugleich fördernde Praktiken nicht mehr eingeübt werden. Ein bereits klassisches Beispiel ist die

Orientierung im topographischen und Straßenraum, wie es seit der Einführung von Navigationsgeräten der Fall ist. Nebenbei kann bei diesem Fall bereits gefragt werden, ob die Nutzung solcher Systeme wirklich den Kern dessen ausmacht, was KI ist. Gleichzeitig können KI-Systeme jedoch auch menschliche Kompetenz(entwicklung) fördern, wo KI-Technologien so eingesetzt werden können, dass bestimmte Fähigkeiten/Kompetenzen trainiert werden, beispielsweise der Einsatz von AC-Technologien, um vulnerable Gruppen auf sozial schwierige Situationen vorzubereiten.
Generell lässt sich vielleicht sagen, dass sich das Kompetenzen-Spektrum von Menschen ändern wird, wenn und weil bestimmte bislang nötige Kompetenzen durch KI übernommen und andere gefördert werden; der Trend würde hier darin liegen, dass die Rolle und die Bedeutung technischer Hilfsmittel zur Kompetenzentwicklung und -förderung weiter betont wird, und es insofern in der Tat zur *weiteren „Technisierung" des menschlichen Lebens führt, die auch Kompetenzen bezüglich der Mensch-Umwelt-Beziehung* umfasst (z. B. Foto-Apps zu Bestimmung von Pflanzen und Tieren).

Würde es mithilfe von KI gelingen, neuen Problemstellungen und ad-hoc Reaktionen erfordernden Situationen auf für Menschen bislang nicht mögliche Weise gerecht zu werden, so würde sich der Handlungsspielraum durch KI qualitativ erweitern ((Nassehi 2019); vgl. projektinterne Endzusammenfassung von AP 2, S. 23). *Ganz grundsätzlich wäre hier KI das Mittel zur besseren Kontingenzbewältigung in allen Aspekten der conditio humana.*

Unüberwachtes Maschinelles Lernen ist dadurch gekennzeichnet, dass sich die Software in Laufe ihrer Tätigkeiten durch Feedback aus der Umwelt ohne menschliches Eingreifen verändert. Sie stellt damit die erste industriell genutzte Technologie der Menschheit dar, die sich durch ihre Nutzung der Umgebung anpasst, ohne dass Menschen dazwischengeschaltet sind, die Daten aufbereiten und dann wieder ins System einspeisen („AI, formerly known as Software"). Das Besondere ist also die Automatisierung („ohne menschliches Eingreifen"), die durch steigende Rechenleistung sowie massenhafter Datenverfügbarkeit zu geringen Kosten technisch-industriell nutzbare Anwendungen ermöglicht. *Hier ist ein Überschreiten der bisherigen Grenze zwischen Menschen als stets mit entscheidendem Gestaltenden der Technik(entwicklung) zu konstatieren, so dass aus der Technik selbst eine ‚kreative Kraft' wird.*[61]

Was macht das Menschsein und das Zusammenleben aus?

Durch das Zusammenspiel von KI und Big Data werden sich menschliche Kommunikationsformen qualitativ stark verändern, und Auswirkungen dieser Änderung auf Gemeinwohlbelange werden erwartet (Altmeppen et al. 2019). Der Wandel wird sich sowohl in Bezug auf Akteure und Strukturen der öffentlichen Kommunikation als auch in Bezug auf Inhalte ergeben (ebd., S. 62). Den großen Plattformunternehmen kommt einerseits eine nicht zu unterschätzende Wirkmacht in Hinblick auf öffentliche Kommunikations-Formen und -Inhalte zu, weshalb sie zum qualitativen Wandel ein großes Stück beitragen. Andererseits sind auch sie von diesen betroffen. Der Prozess dieses Wandels ist entsprechend rekursiv. Es gilt, zu untersuchen, wie das Zusammenspiel von KI und Big Data Gemeinwohlbelange durch die medienvermittelte, öffentliche Kommunikation beeinflussen und welche Rolle den „Verantwortungsrelationen" (ebd.) der einzelnen Akteure zukommt, da Auswirkungen auf Gemeinwohlbelange das menschliche Zusammenleben (z.T. stark) tangieren können.

KI-Technologien haben das Potenzial, zu verstärkter Inklusion vulnerabler Gruppen beizutragen (z. B. durch spezielle Trainingsprogramme oder leichteren Zugang zu Wissen). Gleichzeitig können sie einer solchen auch entgegenwirken, z. B., wenn der Besitz der Technik und digitaler

[61] Hier ließen sich aus Literatur und Film zahlreiche fiktionale Vorwegnahmen dieser Situation ‚kreativer Maschinen' nennen.

Zugang für gesellschaftliche Teilhabe notwendig sind und nicht allen Menschen zur Verfügung stehen.

Wo steht der Mensch in der natürlichen Umwelt?

KI ist nicht ‚körperlos' (dies ist ein nach wie vor verbreitetes Narrativ); die Technologien benötigen sehr große Mengen an Energie und die Hardware, die die KI-Software benötigt, benötigt Ressourcen– auch Ressourcen deren Abbau häufig unter ökologisch problematischen Bedingungen (und der Verletzung von Menschenrechten) geschieht, z. B. seltene Erden und Kobalt. *Diesbezüglich wird die Mensch-Umwelt-Beziehung weiter stark in Richtung Extraktivismus und Instrumentalismus geleitet.* Dieser Aspekt ist zentral für eine KI-Bewertung aus NE-Perspektive und jenseits NE-sensitiver Communities noch kein *Common Ground*. Er gilt für alle KI-Technologien, entsprechend auch in Hinblick auf die Mensch-Umwelt-Beziehung der Anwendungsfelder *Affective Computing* und *Autonome Systeme zur Erschließung bisher unverfügbarer Räume*, bei denen er entsprechend nicht nochmal separat genannt wird.

Durch KI-Technologien erweitert sich der Handlungsspielraum von Menschen in Bezug auf seine natürliche Umwelt sehr stark. Bislang unzugängliche Lebensräume oder Lebensräume, die für Menschen problematische Bedingungen aufweisen, können mithilfe der Technologien erschlossen und genutzt werden, wie z. B. die Tiefsee (für Tiefseebergbau), Wüsten oder Hochgebirge. *Die reale und gedachte Unverfügbarkeit von Natur ohne Zugang für Menschen wird durch AS noch weiter zum Verschwinden gebracht, was sich in entsprechende Narrative zum Anthropozän einfügt.*

Es ergeben sich massive Veränderungen in der Wahrnehmung der natürlichen Umwelt durch die Anwendung von *Augmented* bzw. *Virtual Reality (VR)*-Technologien sowie durch Artbestimmungs-, Naturerfahr- und Erforschungs-Apps. Inwiefern diese das Naturverständnis der digitalisierten Gesellschaften ändern werden, stellt ein wichtiges Forschungsdesiderat dar. Es ergibt sich daraus die Frage, wie eine solche Änderung zu bewerten ist: Dient sie partiell Naturschutzmaßnahmen durch geringere Beeinträchtigung sensitiver Gebiete oder schadet sie vor allem auf lange Sicht, da Menschen die Wertschätzung für die ‚echte' natürliche Umwelt verloren geht?[62] *Die Verhältnisbestimmung von Virtualität und Realität des Mensch-Umwelt-Bezugs wird sich mithin verschieben, aber möglicherweise nicht nur in die Richtung Entfremdung, weil die Medialität ggf. neue auch leiblich erfahrbar relevante Bezüge ermöglicht.* Im Kontext der KI lassen sich grundsätzliche anthropologische Fragestellungen nach möglichen Alleinstellungsmerkmalen des Menschen, die dem Menschen „natürlich" gegeben sind stellen, die nun auch bei Maschinen verortet werden könnten (bestimmte kognitive Prozesse, rudimentäre Formen von Selbstbewusstsein, etc.). Die immer schon schwierige Unterscheidung zwischen Natürlichkeit und Künstlichkeit wird durch KI noch komplexer, daran schließen sich auch Fragen der Bildung für Nachhaltige Entwicklung (vgl. Übersicht bei (Bellina et al. 2020)), und zwar hinsichtlich der inhaltlichen Zugänge ebenso wie der Methoden. Betroffen ist hier eine Dimension der Fähigkeit (nach (Nussbaum 2010)), sich in Interaktion mit der natürlichen Umwelt (oder Mitwelt) begeben zu können.

3.2.3 Wenig untersuchte ethische Aspekte zu KI („blinde Flecken") aus NE- und Gemeinwohl-Perspektiven

Ein Bereich, in dem tatsächlich ein qualitativer Wechsel durch KI stattfindet/stattfinden wird, ist der der öffentlichen Kommunikation. Algorithmische Prozesse ändern tatsächlich Arten und

[62] Aktuell scheint beispielsweise ungeklärt, ob das bereits bestehende massive Angebot an Naturdokumentationen tatsächlich dem Schutz der natürlichen Umwelt durch Förderung der Wertschätzung und „Nicht-Behelligung" genutzt oder insgesamt das Engagement für Natur nicht gefördert sowie die Nachfrage nach entsprechenden Reisen und Belastungen sogar verstärkt hat – und was dies für eine künftige Virtualisierung von Naturzugängen bedeuten kann und soll.

Weisen der Publikumsbeteiligung, so dass neue Formen der Öffentlichkeit entstehen. Sie wirken sich sowohl auf den Inhalt als auch auf die Strukturen der öffentlichen Kommunikation aus und führen zu grundlegenden Verschiebungen, u. a. in Hinblick auf Verantwortungsfragen – sowohl bezüglich Struktur als auch Inhalt (vgl. (Altmeppen et al. 2019)). In den entsprechenden Narrativen zur KI werden zwar einzelne Beispiele, die diesen Wandel mitvorantreiben, wie z. B. Filterblasen diskutiert, der umfassende Wandel in Struktur und Inhalt öffentlicher Kommunikation wird jedoch selten adressiert (vgl. ebd.). Das mag daran liegen, dass KI hierbei nicht die Schlüsseltechnologie darstellt, allerdings spielt sie eine sehr große Rolle in den verschiedenen Prozessen dieses Wandels. Dieser Wandel der Öffentlichkeiten durch KI betrifft die Grundlagen sowohl des Fähigkeitenansatzes als auch des Gemeinwohls, weil beide darauf aufbauen, dass ethische Fragen auf angemessener Basis (faire und sachgerechte Diskursbedingungen) öffentlich ver- bzw. ausgehandelt werden können müssen.

Vor dem Hintergrund der sehr binär angelegten Zukunftsdarstellungen oder Narrativen ist es umso dringlicher und selbst eine große Herausforderung, die jeweiligen Technologien kritisch und ethisch reflektiert zu betrachten. Auf Basis der oben ausgeführten ethischen Rahmenüberlegungen ergeben sich diverse Aspekte und Überlegungen, die es im Blick zu behalten bzw. zu klären gilt:

Mit welchen und wessen Werten die künstlich intelligenten Systeme übereinstimmen, und welche Werte programmiert werden sollten, stellt für Iason Gabriel (Gabriel 2020) eine zentrale Fragestellung dar. Er weist auf die hierzu bisweilen viel zu kurz gekommene Forschung hin (vgl. Überblickspapier KI AP 2, und auch (Spiekermann 2019) für eine detaillierte Auseinandersetzung mit Werten in digitalen Systemen). Über bestimmte Werte besteht weitestgehend Konsens (diese sind zwar nicht ausreichend, es besteht v. a. Konsens bzgl. Werten wie Verantwortung, Privatsphäre und Fairness; nicht bzgl. Werten wie Erhaltung natürlicher Ressourcen und Biodiversität, und es mangelt z.T. noch an der Einsicht, dass Diskriminierungen und Stereotype sowie ein westlich zentriertes Weltbild einprogrammiert werden), in Hinblick auf diese stellt sich die große Frage: wie füllt man sie mit Leben?

In vielen Bereichen, in denen KI-Technologien eingesetzt werden, sind entweder vulnerable Gruppen betroffen (z. B. Bildungssektor, Pflege) oder es handelt sich um intime Bereiche (z. B. Smart Home oder medizinische Daten), so dass auf diese Bereiche immer besonders sorgfältig geschaut werden muss. Hier sind die Fragen nach Aushandlungsarenen und Governance-Strategien mit angesprochen, die oft in den Richtlinien (s.o.) nicht enthalten sind.

Die Perspektive der Nachhaltigen Entwicklung macht besonders deutlich, welche der in vielen Richtlinien eher seltener angesprochenen Punkte auf keinen Fall vernachlässigt werden dürfen, wenn KI-Technologien so entwickelt werden sollen, dass sie dem Gemeinwohl zuträglich und gerecht sind. „Solidarität, Inklusion und sozialer Zusammenhalt“, „Diversität“, „Bildung“ und „Fragen einer guten Governance“ sind für eine umfassend verstandene Nachhaltige Entwicklung zentral. Berücksichtigt man also diese Perspektive, so müssen diese Aspekte deutlich gestärkt werden. Dies gilt umso mehr für die fundamentale Forderung der NE-Perspektive, die ärmsten und unterprivilegiertesten Bevölkerungsteile der Welt zentral in den Blick zu nehmen!

3.3 Vertiefungsstudie A: Affective Computing unter besonderer Berücksichtigung des Bildungsbereiches

3.3.1 Technikfolgenbezogene Analyse

Das Problem einer Bewertung von Technologien, die sich, wie Affective Computing, noch in frühen Stadien der Entwicklung befinden und entsprechend nur eingeschränkt einsatzbereit

sind, wird im sogenannten „Collingridge-Dilemma" formuliert: „When change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult, and time-consuming." ((Collingridge 1980), S. 11). Das Dilemma spricht damit zwei grundsätzliche Dimensionen an (vgl. (Manzeschke und Assadi 2019), S. 168): Die Informationsdimension umfasst das Problem, dass solange eine Technik noch nicht sehr bekannt oder etabliert ist, man noch nichts über Folgen und Nebenwirkungen sagen kann. Die Macht- und Steuerungsdimension spricht das Problem an, dass, sobald eine Technik etabliert ist, sie nur noch schwer zu beeinflussen (z. B. aufgrund von Bedingungen der spezifischen Art und Weise, wie sie eingeführt ist) oder gar zurückzunehmen ist. Eine kritische Begleitung solcher Entwicklungen ist daher notwendig, die frühzeitig und auch vorausgreifend Fragen im Hinblick auf Steuerungsnotwendigkeiten stellt, ohne diese Entwicklungen unnötig zu behindern aber gleichzeitig notwendige Begrenzungen im Blick zu haben. Dieses Dilemma trifft insofern auf die Anwendung von AC im Bildungsbereich zu, als dass sich erste Szenarien in der Entwicklung befinden, also noch keine Abschätzung der Folgen eines breiten Einsatzes angestellt werden kann. Allerdings ist die Forschung breit aufgestellt und die Einsatzmöglichkeiten sind vielfältig und auch für privatwirtschaftliche Zwecke zu nutzen, so dass zu erwarten ist, dass marktreife Systeme schnell einen Absatz und damit eine entsprechende Verbreitung finden. Im Folgenden werden erste Überlegungen angestellt, die jeweils drei Formen des Lernens (formell, informell sowie Trainingsszenarios) darauf hin untersuchen, welche Anwendungspraxis momentan besteht, wie diese Praxis durch den Einsatz emotionssensitiver Technologie verändert wird und was der von Forschenden und Entwickler*innen erwartete Mehrwert ist.

Welche konkrete Anwendungspraxis besteht momentan?

Im Falle des *formellen Lernens* werden vor allem in Deutschland noch klassische Unterrichtsszenarien in Klassenräumen mit der Unterstützung der üblichen Literatur (Schulbücher, Arbeitshefte etc.) sowie einigen Medien (Filme, Dokumentarreihen) eingesetzt. Zusätzlich werden immer mehr digitale Lernunterstützungen (Vokabeltrainer als Apps, Webpages zur Festigung von Lernhinhalten) angeboten. In englischen Sprachraum sind digitale Lehr-Lernplattformen bereits im breiteren Einsatz und stehen entsprechend in der Kritik, aufgrund der nicht geregelten Zugänglichkeit der Performance-Daten der Schüler*innen, z. B. im Falle von Google-Classroom (Persson 2021; Lane 2015). Im Bereich der Hochschulbildung und der Erwachsenenbildung werden zunehmend Konzepte des sogenannten „Blended-Learning" eingesetzt, die selbstständiges Arbeiten anhand einer Internetplattform (meistens Ilias oder Moodle) mit Präsenzphasen verbinden. Außerdem werden komplette Onlineseminare angeboten („Web-based Training"), die allerdings oft von Lehrpersonal begleitet werden. Einige der Programme bzw. Plattformen, die zu Unterstützung von online-basiertem formellen Lernen eingesetzt werden, bieten bereits jetzt Aufmerksamkeitsmessungen anhand von Eye-Tracking an (z. B. Zoom).

Informelles Lernen, also (teilweise) selbstgesteuertes und intrinsisch motiviertes Lernen findet auch z. B. in Museen, Science-Centern oder Zoos statt. Unterstützt werden die Lernenden bzw. die Besucher*innen durch Info-Tafeln, aber auch durch Audioguides, spezielle Apps oder durch Führungen in Gruppen oder Einzelbetreuung. Angebotene Apps werden heute schon teilweise mit Positions-Tracking verbunden, so dass das Programm zielgerichtet über die am Aufenthaltsort der Besucher*innen zu erwerbenden Informationen unterrichtet wird (Lane 2015). Zudem erfolgt informelles Lernen noch stärker selbstgesteuert und ungeregelt über selbstgewählte Lektüre, online wie offline und die Nutzung weiterer Informationsdienste, ebenfalls online wie offline. Systeme zur Erkennung von Emotionen sind hier in der Erprobung. Sie dienen momentan vor allem der Auswertung der Resonanz auf die Ausstellungsstücke und zur Verbesserung der Besucherführung durch die Ausstellungen selbst.

Formen des *Sozialtrainings* finden momentan vor allem als Schulungen oder Beratungen in Einzel- oder Gruppensituation statt. Zudem steht eine Unzahl an Ratgeberliteratur (on- und offline) zu allen denkbaren Themen zur Verfügung. Messungen von Emotionen finden hier bislang noch meist mit den klassischen, nicht-digitalen Methoden der Psychologie statt. In der Erprobung befinden sich Systeme, die anhand von Gesichtserkennungssoftware Frustrationsmomente und Depressionspotenzial erfassen.

Was genau wird durch den Einsatz von AC verändert?

Der Einsatz von KI-Technologien in Kombination mit Affective Computing ermöglicht stark personalisierte Programme, die die Nutzer*innen direkt durch Avatare oder Roboter (je nach Szenario) oder auch durch Expertensysteme ansprechen. Dies gilt sowohl für formelles als auch informelles Lernen. Die Erkennung von Emotionen der Nutzer*innen soll dazu dienen, auf persönliche Präferenzen, vor allem aber auf Frustrationserlebnisse umgehend zu reagieren. Dadurch soll das Lernerlebnis mehr interessensgeleitet gestaltet, entsprechend intensiviert und somit der Lernerfolg deutlich verbessert werden. Gleichzeitig soll sichergestellt werden, dass die Erarbeitung der Lerninhalte, die unabhängig vom Eigeninteresse erworben werden müssen mit weniger Frustrationserlebnissen verbunden sind. Dadurch soll AC, vor allem bei vollständig asynchronen Lehr-Lernformaten einerseits ein auf die emotionale Verfasstheit der Nutzer*innen abgestimmtes Feedback geben können. Das soll zwar eine menschliche Lehrperson nicht ersetzen, die technische Lösung aber dem zwischenmenschlichen Lehrer-Schülerverhältnis annähern. Außerdem wird die Möglichkeit gesehen, die Lehrmethoden und die Art der Präsentation der Lehrinhalte an persönliche Lerneigenschaften immer besser und gezielter anzupassen. Ziel ist eine optimierte direkte Einzelbetreuung von Nutzer*innen, die fokussierter und deutlich enger getaktet erfolgt, als dies in den bislang möglichen Lernsettings möglich ist. Anstatt also – wie bisher – einen Vortrag oder eine Präsentation online zu verfolgen und diesen dann anhand von Leitfragen zu reflektieren würde ein Avatar bereits bei der Präsentation einhaken, sobald ein Nutzer oder eine Nutzerin die Stirn runzelt, um nachzufragen, ob es ein Verständnisproblem gibt.

Eine kulturspezifische Ansprache der Nutzer*innen durch entsprechend trainierte Avatare soll zudem ermöglichen, dass weniger Diskriminierung stattfindet und die Nutzer*innen sich noch persönlicher betreut fühlen. Dies wird unterstützt durch Technologien, die komplexe „natürliche“ Kommunikation in Echtzeit beherrschen (NLP) und übliche in Kommunikationssituationen auftretende emotionale Marker integrieren, sodass sich die Nutzer*innen höflich und unterstützend behandelt fühlen. Ein Beispiel für solche Ansätze stellt z. B. der Chatbot Anna von IKEA dar, der – je nach Kulturkreis – unterschiedlich auf die Kund*innen reagiert, z. B. wenn Anna sich dafür entschuldigt, Informationen nicht finden zu können.[63]

Bei den sozialen Trainingsprogrammen wird der Einsatz von AC hingegen selbst Teil des Trainings. Z. B. können die Nutzer*innen von Programmen, die auf sozial schwierige Situationen vorbereiten sollen, die emotionalen Reaktionen des Agenten jeweils so anpassen, dass es ihrer jetzigen Trainingssituation entspricht. Das Lesen von Emotionen der Nutzer*innen dient bereits jetzt der Kontrolle des Trainingssettings sowie der Überprüfung des Trainingserfolges.

Was ist der angekündigte Mehrwert des Einsatzes von AC?

Sowohl die aktuellen Debatten als auch die spekulativen Zukunftsvisionen im Bereich dieses Feldes zeichnen sich im Bereich der Forschung dadurch aus, dass bei einigen Anwendungen die ethischen Probleme durchaus gesehen werden. Grundsätzlich motiviert sind die Forscher*innen

[63] https://www.chatbots.org/virtual_assistant/anna3/, 29.07.2020

zugleich durch den Gedanken, Kommunikation mit digitalen Agenten zu vereinfachen, angenehmer zu gestalten und im Bildungsbereich dadurch Bildungsinhalte einer größeren und diverseren Gruppe von Personen zugänglich zu machen:

> The link between emotion and inclusion is not conclusive, but there are several reasons for taking it seriously. At the most general level, it seems reasonable a priori to assume that the more natural the interface, the more widely usable it will be; and conversely, the more people have to adjust their communicative style, to use an interface, the fewer people will be able to make the adjustment. ((Cowie 2012), S. 415)

Entwickler*innen sehen im Einsatz von Affective Computing vor allem einen Mehrwert darin, digitale Technologien für die Nutzer*innen allgemein angenehmer zu gestalten, bzw. sie den Interessen und den Bedürfnissen der Nutzer*innen spezifischer anzupassen. Dies gilt für alle Formen des Lernens und für alle Angebotsvarianten. So sollen unangenehme Erfahrungen in der Kommunikation mit künstlichen Systemen verringert oder vermieden werden:

> "[O]ne of the obvious roles of affective computing is remedial. lt is to spare people distress that would otherwise be caused by interactions with affectively incompetent systems." ((Cowie 2015), S. 339)

Durch eine möglichst „natürliche" – also vertraute – Kommunikation soll die Akzeptanz der Nutzer*innen im Allgemeinen und der Lerneffekt im Bildungsbereich im Spezifischen erhöht werden. Durch "sympathische(re)" Avatare bzw. Roboter soll nicht zuletzt der Verkaufserfolg solcher Systeme erheblich gesteigert werden.

Erfahrungen aus der Kommunikation mit kognitiven Assistenten belegen, dass zunehmend einfachere Satzkonstruktionen benutzt werden sowie die Ausbildung eines Befehlstons. Bislang handelt es um eine reduzierte Form der Kommunikation, ein Manko, dass auch durch die verschiedenen Einsatzmöglichkeiten von AC nach und nach behoben werden soll:

> […] Technology is already used in ways that impact on people's emotions, whether the designers intend it to or not. Given that it is used in those ways, it certainly seems fair to say that if emotion-oriented technology can make malinteraction less common, then they are doing good. ((Cowie 2012), S. 415)

Im Bildungsbereich erhofft man sich zusätzlich die Möglichkeit, Trainingssettings, die ohne KI mit AC in dieser Spezifizität nicht möglich wären, zu entwickeln. So könnten z. B. sehr spezifische kulturelle Details oder sehr spezifische Charaktere programmiert werden, anhand derer jederzeit und überall angemessener Umgang geübt werden könnte. Dies wäre im Austausch mit menschlichen Gegenübern nicht möglich. Außerdem verfügen künstliche Systeme über Möglichkeiten, direkte, datengestützte Rückmeldung über die „Performance" der Nutzer*innen zu geben, die detailliert und informativer sein sollen als Rückmeldungen von menschlichen Trainer*innen oder Berater*innen ((Endrass et al. 2013), (Gebhard et al. 2014)).

Dies gilt auch für die bereits erwähnten „Affective Learning Companions", die dazu entwickelt werden, um Kinder das Lernen zu erleichtern und ihnen zu ermöglichen, für sich selbst gute Lernstrategien zu entwickeln:

> "The aim of this project is to build a computerized learning companion that facilitates the child's own efforts at learning. […] The companion is not a tutor that knows all the answers, but a player on the side of the student, there to help him or her learn, and in so doing, learn how to learn better."[64]

[64] https://affect.media.mit.edu/projectpages/lc/, 04.08.20

3.3.2 Veränderung der Mensch-Technik-Umwelt-Beziehung durch AC im Bildungsbereich

Was verbindet Menschen mit ihren Artefakten, einschließlich Technik?

Mensch-Maschine Interaktionen (Human Computer Interaction, HCI) können als sogenannte „parasoziale Interaktionen" bezeichnet werden. Parasoziale Interaktionen zeichnen sich dadurch aus, dass ein Akteur mit einem Gegenüber interagiert, dessen Existenz nicht unbedingt real gegeben sein muss oder sich soziale Beziehungen nicht mehr auf natürliche Lebewesen beschränken, sondern einen technischen Beziehungspartner einschließen (vgl. (Horton und Wohl 1956)).

Manzeschke und Assadi stellen eine Sammlung von Fragen hinsichtlich AC-Systemen vor, die den Einfluss solcher Agenten auf die menschliche Interaktion und die Unterschiede zwischen menschlicher Interaktion und HCI adressieren ((Manzeschke und Assadi 2019), S. 167):

- Wie verändern sich menschliche Interaktionen durch die Anwendung von „emotionalisierter" Technik?
- Welche neuen Interaktionsformen werden möglich?
- Wie können veränderte und/oder neue Formen genutzt werden, um einen Beitrag zum guten menschlichen Leben zu leisten?
- Welche Rückwirkungen könnte es durch die Intensivierung von HCI auf zwischenmenschliche Interaktionen geben?

Dabei unterscheiden die Autor*innen folgende grundsätzlich mögliche Interaktionsverhältnisse (Manzeschke / Assadi, 167):

- Herr – Knecht
- Technischer Lehrer – menschlicher Schüler
- Menschlicher Lehrer – technischer Schüler
- Kollegen
- Freundschaft
- Liebe

Ein Beispiel für ein *Herr-Knecht*-Verhältnis ist der oft von menschlicher Seite bereits jetzt nicht besonders höfliche Umgang mit virtuellen Assistenzsystemen wie Siri oder Alexa (die noch nicht mit AC-Technologien ausgestattet sind), wobei im Befehlston Anweisungen ausgesprochen werden. In diesem Kontext ist das auf Kant beruhende sogenannte Verrohungs-Argument verortet. Wenn Personen Tiere, so die Argumentation von Kant, dauerhaft schlecht behandeln, dann stumpft der Mensch ab und empfindet weniger Mitleid, nicht nur gegenüber Tieren, sondern am Ende auch gegenüber Menschen. Damit, so Kant, ist das Quälen von Tieren den Pflichten der Menschen gegen sich selbst entgegengesetzt ((Kant 1990), MdS 2, § 17, 84). Dieses Argument übertragen einige Autor*innen (vgl. z. B. (Brahnam 2006)) von Tieren auf virtuelle Agenten und Roboter und stellen damit die Frage, wie wir mit solchen Agenten umgehen sollten.

Liebe wird in Verbindung mit den angedachten Sexbots gebracht, aber auch im Kontext unterschiedlicher Formen von Companions, sowohl in Avatar-Form (in rudimentären Formen z. B. in Computerspielen) als auch in Roboterform (vgl. z. B. (Ess 2016), (Loh 2019)) diskutiert. Da noch keine Sexbots mit AC existieren und die Datenlage zur emotionalen Bindung an Avatare mehr als rar ist, bleibt es zunächst eine nicht geklärte empirische Fragestellung, welche Form von Interaktionen sich tatsächlich entwickeln, wenn AC in virtuelle Systeme implementiert wird.[65]

Ähnliches gilt für die Interaktion der *Freundschaft*, welche im Begriff des „Companion" schon mitgedacht ist. Avatare oder Roboter als Companions werden sowohl im Pflegesektor als auch im Bildungsbereich im Rahmen der Entwicklung von AC angedacht. Tatsächlich liegt hierin ein Versprechen der Technologie, welches kritisch hinterfragt werden muss. Denn, wie Manzeschke und Assadi betonen, waren Menschen bislang, sofern eine Form der Abhängigkeit bestand, von der Funktionalität von Maschinen aller Art abhängig. Wenn Menschen sich aber emotional von künstlichen Agenten abhängig machen und sich von ihnen verstanden fühlen, dann können diese Maschinen das Maß für Emotionen vorgeben. Dies würde eine neue Dimension für die *Conditio Humana* eröffnen (vgl. (Manzeschke und Assadi 2019), S. 170). Fraglich ist in diesem Zusammenhang, welche Rolle eine mehr oder weniger stark ausfallende Anthropomorphisierung der Agenten spielt. Die „Uncanny Valley" These scheint eher dafür zu sprechen, dass eine größere Identifikation mit Agenten stattfindet, die gerade nicht vollständig menschenähnlich gestaltet sind. Auch hierbei handelt es sich allerdings zunächst um eine empirische Fragestellung, zu der bislang die entsprechende Datenlage nicht vorhanden ist.

Hinsichtlich des Bildungsbereichs trifft sicherlich die „*Technischer Lehrer – menschlicher Schüler*"-Interaktion vor allem auf die Varianten des formalen Lernens zu. Varianten des informellen Lernens sowie der Trainingsprogramme könnten auch im Rahmen der *kollegialen Interaktion* beschrieben werden. Da diese Systeme vor allem als eine Erweiterung des Lehr-Lernangebotes bzw. als zusätzliche Unterstützung gedacht sind, sind sie nicht so angelegt, die zwischenmenschliche Interaktion zu ersetzen. Eine grundsätzliche Veränderung der Verbindung von Menschen zu ihren Artefakten ist in diesem Rahmen also dann eher nicht zu erwarten, wenn keine „Companion"-Strategien in die Technologien implementiert werden.

Kritisch befragen lässt sich allerdings in jedem Fall folgende These von Manzeschke und Assadi ((Manzeschke und Assadi 2019), S. 170): „Die Konstruktion von emotional kompetenten technischen Artefakten in der Interaktion mit Menschen wird auf eine Anerkennung dieser als sozialer und moralischer Gegenüber hinauslaufen". Die oben beschriebenen technischen Möglichkeiten von AC sowie die Entwicklungsziele von Forscher*innen und Entwickler*innen lassen doch eher Zweifel an dieser sehr weitreichenden Behauptung aufkommen. Ob es zudem notwendig ist, ein neues Vokabular zu entwickeln, um „emotionssensitive Mensch-Technik Interaktion" zu beschreiben ((Manzeschke und Assadi 2019), S. 169) lässt sich zum jetzigen Zeitpunkt ebenfalls schlecht abschätzen. Bislang sind die Techniken so rudimentär, dass unsere Sprache ausreichend scheint, die Entwicklungen zu beschreiben und einzuschätzen, wobei offenbleiben muss, ob damit alle wesentlichen Aspekte der neuen HCI erfasst und umschreibbar sind.

Im Zusammenhang mit der dem Affective Computing zugrundeliegenden Technik der Verarbeitung natürlicher Sprache (NLP) ist wiederholt das Problem angesprochen worden, dass es in Zukunft zunehmend schwieriger wird, zu unterscheiden, ob Sprache oder ein anderes

[65] In Japan werden von der Firma A-Fun, einer der wenigen Reparaturdienstleister für die Roboter-Hunde der Aibo-Reihe buddhistische Trauerzeremonien initiiert. Die irreperablen Aibos werden vorher als „Organspender" benutzt. A-Fun-Chef Nobuyuki Norimatsu erklärt: „Wir möchten die Seelen an die Besitzer zurückgeben und den Roboter zu einer Maschine machen, deren Teile wir verwenden. Wir entnehmen keine Teile, bevor wir eine Bestattung für sie abhalten." https://www.techbook.de/easylife/japan-abschied-roboterhunde, letzter Zugriff 09.03.2020.

Artefakt von einem Menschen oder einer Maschine generiert wurde und mit was für einer Art Gegenüber, Mensch oder Maschine, wir es in einer Kommunikationssituation zu tun haben (*Ununterscheidbarkeit*, vgl. Abschnitt 2.4.1). Damit kann die weiter voranschreitende Vermenschlichung (*Anthropomorphisierung*) durch AC noch verstärkt werden, indem Roboter und Avatare nicht nur in ihrem Erscheinungsbild den Menschen immer ähnlicher werden (sollen), sondern auch in ihrem Verhalten. Das Problem der Unterscheidbarkeit könnte also einerseits verschärft werden, wenn die Kommunikation auf der maschinellen Seite durch die Ergänzung von emotionalen Markern noch natürlicher gestaltet wird. Andererseits könnte gerade der vermehrte Einsatz von Avataren, die die Urheberschaft von Artefakten verdeutlichen, das Problem der Ununterscheidbarkeit beheben. Wenn Schriftstücke, Audiodateien, etc., die mit Hilfe von NLP-Techniken generiert werden, immer durch einen Avatar vorgelesen oder gesprochen werden müssten, dann wäre die Unterscheidbarkeit automatisch gegeben. Da im Bildungsbereich mit Avataren gearbeitet wird, ist die Herausforderung der Ununterscheidbarkeit hier nicht gegeben.

Sollte jedoch eine Entwicklung auszumachen sein, dass Menschen sich durch den Einsatz von AC-Technik häufiger und stärker emotional von künstlichen Agenten abhängig machen, so besteht hier tatsächlich ein Problem. Dann nämlich können Maschinen Menschen nicht nur leichter manipulieren, sie könnten auch das Maß für Emotionen vorgeben, indem die Maschine „vorlebt" welche Emotionen wann und wie angemessen oder unangemessen sind. Dies kann eine neue Dimension für die *Conditio Humana* darstellen, weil der prägende Vorbildcharakter nun von der Maschine vorgegeben wird.

Was macht das Menschsein und das Zusammenleben aus?

Bislang gelten Emotionen bzw. Emotionalität als Alleinstellungsmerkmal des Menschen bzw. der Lebewesen (vgl. (Manzeschke und Assadi 2019)). Es ist unklar, wie sich auf lange Sicht das Verständnis und die Einschätzung der Bedeutung von Emotionen für das Menschsein an sich aber z. B. auch für die Güte von Urteilen entwickeln wird, wenn künstliche Agenten in der Lage wären, Emotionen überzeugend zu simulieren. Ein interessanter Gesichtspunkt wird z. B. von Roddy Cowie ((Cowie 2012), S. 414) angemerkt. Er argumentiert, dass das lange bestehende Primat der Rationalität über die Emotionalität dazu geführt hat, dass Menschen einen bedeutenden Teil ihrer Natur, die Emotionalität, nicht mehr wertschätzen. Affective Computing, so Cowie, könnte helfen, dieses Missverhältnis zu korrigieren:

> "Humans have learned to compare ourselves unfavorably with computers: The are wholly rational, free from emotion, and we are neither. Affective Computing turns that around, and invites people to realize that the deficiency is on the machines' side, and the rational path is to make them, to the very limited extent that we can or want to, more like us." ((Cowie 2012), S. 414)

Die Mensch-Maschine Interaktion ist bislang deshalb unbefriedigend, weil Maschinen keine Emotionen zeigen, so das Argument. Dadurch wird deutlich, wie wichtig Emotionen für Menschen tatsächlich sind, sowohl im alltäglichen Umgang miteinander als auch bei der Urteilsbildung.[66] Diese Erkenntnis führt dann dazu, den emotionalen Teil des Menschseins wieder deutlich mehr zu schätzen. Hinter diesen Überlegungen stecken grundlegende Fragen nach dem Menschenbild, dass Forscher*innen, Entwickler*innen und Nutzer*innen leitet, bzw. welche genaue Werte sie Emotionen zuschreiben. Was ist ein „gutes" Urteil oder gar ein „guter"

[66] Die Bedeutung von Emotionen für die Urteilsbildung betont neben Cowie sowohl die Moralpsychologie (vgl. z. B. Goldie 2000) als auch die normative Ethik in Form der die Fürsorgeethik. Letztere ist gerade als Korrektiv zu rein rational argumentierenden Ethiktheorien entstanden (vgl. Gilligan 1985). Fürsorgeethiker*innen betonen sinnvollerweise die Bedeutung von Emotionen und Beziehungen in der Moral, jedoch können sie keine überzeugenden Antworten auf Gerechtigkeitsfragen geben, weshalb sie vorliegend nicht als normatives Framework für die ethische Analyse dienen kann.

Mensch? Ein emotionaler oder jemand, der seine Emotionen im Griff hat? Stören Emotionen eher oder wann helfen sie und bei was bzw. wofür? ((Cowie 2015), S. 342)

Eher negative Auswirkungen sind mit der Befürchtung verbunden, dass die durch das Affective Computing geförderte emotionale und parasoziale Mensch-Maschine-Beziehung zwischenmenschliche Beziehungen nicht nur ergänzen, sondern gar ersetzen könnten (Gutmann 2011). Inwiefern und wann dies der Fall sein könnte, ist allerdings ebenso wenig abzuschätzen, wie die eher positiven Vermutungen von Cowie.

Eine aus anthropologisch-philosophischer Perspektive bedeutsame Frage ist mit dem Themenfeld der sogenannten „Intersubjektivität" verknüpft. „Intersubjektivität" bezeichnet den Raum geteilter Bedeutungen zwischen Subjekten, der durch zwischenmenschlichen Austausch entsteht (vgl. z. B. (Husserl 1973)). Diese geteilten Bedeutungen können als Träger der sozialen Welt verstanden werden, die durch ihre Mitglieder konstituiert werden. Bislang wird diese soziale Welt von menschlichen Wesen im Dialog (Buber 2009) gestaltet bzw. erschaffen. Wenn nun Avatare oder Roboter in einen (auch) emotional sinnstiftenden Dialog eintreten, dann gilt es zu diskutieren, ob unsere Standardauffassung von Intersubjektivität verändert bzw. erweitert werden muss. Weiterhin stellt sich die Frage, ob und wenn ja welche Folgen der Prozess der intersubjektiven Einbeziehung von Robotern und Avataren in unsere Lebenswelt auf unser Verständnis des Konzeptes einer Person hat bzw. hätte (Brand 2015). Diese Fragenkomplexe gilt es im Rahmen philosophischer Forschung weiter zu beleuchten. Welche Auswirkungen solche Veränderungen im Kontext von Bildungsszenarien und Lehrer*innen-Schüler*innen-Beziehungen haben könnten, stellt eine zusätzliche interessante Forschungsfrage dar.

Diese Unwägbarkeiten gelten auch für die sehr allgemeine Frage nach möglichen Veränderungen des menschlichen Selbstverständnisses. Je nachdem, wie, auf welche Weise und wie breit AC-Technologien in Zukunft eingesetzt werden können oder sollen, könnte sich das Selbstverständnis verändern oder auch nicht. Diese Frage gilt es – auch für ethische Überlegungen – im Sinne einer reflektierten Begleitung der Forschung und Entwicklung im Blick zu behalten.

Wenn Roboter und Avatare in der Lage wären, eine der menschlichen emotionalen Bandbreite sehr nahekommende Simulation zu erzeugen, dann wäre das menschliche Selbstverständnis herausgefordert, sich zu diesen Agenten neu/anders einzustellen. Ebenso herausgefordert wäre die Sichtweise von uns auf uns selbst, wenn wir nicht mehr sicher sein können, dass unsere emotionale Verfasstheit nicht replizierbar, privat und somit nur uns direkt zugänglich wäre. Hier wären wir gezwungen, einen bedeutenden intimen Teil unseres persönlichen Selbstverständnisses als Verfügbar zu denken.

Arne Manzeschke und Galia Assadi weisen darauf hin, dass für die Einschätzung möglicher Veränderungen des Zusammenlebens verschiedene Ebenen des Sozialen berücksichtigt werden müssen: die individuelle, die organisationale und die gesellschaftliche Ebene, wobei alle Ebenen jeweils von kulturellen Einstellungen begleitet werden ((Manzeschke und Assadi 2019), S. 167). Hinsichtlich möglicher Veränderungen im Bildungsbereich zeigt sich, dass tatsächlich alle Ebenen angesprochen werden. Welche Auswirkungen genau zu prognostizieren sind, bleibt allerdings eine offene Frage, solange die Anwendungen nur exemplarisch eingesetzt werden. Absehbar ist, dass sich das individuelle Lernerlebnis insofern ändert, als dass mehr personalisierte Formate angeboten werden können und die Landschaft des digitalen Lernens vielfältiger wird. Auf organisationaler Ebene wird beobachtet werden müssen, welche Institutionen und privatwirtschaftlichen Akteure die Angebote bestimmen werden und mit welchen Qualitätskriterien hier gearbeitet wird. Damit stellt sich direkt im Anschluss die Frage, wer für solche Kriterien zuständig sein sollte. Auf gesellschaftlicher Ebene kann und soll

weiterhin diskutiert werden, wie man gerechte Zugangsmöglichkeiten schaffen und Inklusion und Vielfältigkeit mithilfe von AC-basierten digitalen Lehr-Lern-Methoden fördern kann (ohne dabei aber den entsprechenden Rohstoff-Extraktivismus zu befördern).

Festzustellen bleibt, dass für alle sich in der Entwicklung befindlichen Technologien diese Dimensionen ausgelotet werden sollten. Zusätzlich sollten auch weitreichende Veränderungen in den Blick genommen werden, wie z. B. das „Verständnis von Freundschaft, Pflege, Erziehung oder Sorge“ (Manzeschke und Assadi 2019). Ein weiteres Verständnis dieser Art wäre zudem ein sich veränderndes Konzept von „*Intimität*“, wenn Maschinen in der Lage sind, menschliche Emotionen zu erfassen, zu speichern und unbegrenzt zu verarbeiten und weiterzuleiten. Diese Frage stellt sich jetzt schon im Zusammenhang mit privaten Gesprächen sowie Geräuschen des alltäglichen Lebens, die Systeme wie Alexa und Siri permanent aufzeichnen (Daum 2019b).

Eine positive Transformationsmöglichkeit besteht darin, dass emotional sensible technische Assistenten zum Abbau von Kommunikations- und Zugangsbarrieren beitragen könnten und somit eine breitere Inklusion von Menschen ermöglichen könnten. Emotionssensitive Sprachassistenten können helfen, den Alltag besser zu bewältigen und durch verschiedene Landessprachen oder technische Sprache verursachte Kommunikationsbarrieren abzubauen (Burchardt und Uszkoreit 2018b). Weniger technik-affinen Menschen könnte der Zugang zu verschiedenen Interfaces erleichtert werden, wenn diese Systeme auf negative Reaktionen der Nutzer*innen reagieren könnten und zugleich ein freundliches und sympathisches Gegenüber darstellen (Cowie 2015).

Wo steht der Mensch in der natürlichen Umwelt

Im Kontext dessen, wie sich der Einsatz von AC-Technologien auf die Mensch-Umwelt-Beziehung auswirkt, ist bedenkenswert, „[…] dass wir durch den Einsatz von Emotionen einen Aspekt des Menschlichen adressieren, der für uns selbst nur teilweise verständlich und beherrschbar ist […].“ ((Manzeschke und Assadi 2019), S. 169) Damit könnten sich veränderte Vorstellungen von Natürlichkeit (in Abgrenzung zu Künstlichem bzw. Kulturellem) hinsichtlich unserer emotionalen Verfasstheit ergeben, die allerdings nur schwer absehbar sind.

Weiterhin könnte sich der menschliche Bezug zu einer ‚echten‘ oder ‚natürlichen‘ Umwelt verändern, wenn die „künstlichen“ Umwelten insofern „natürlicher“ werden, als dass die künstlichen Agenten durch die Simulation von Emotionen auf uns immer menschlicher wirken. Die Attraktivität künstlicher Umgebungen könnten dadurch gesteigert werden. So könnten z. B. Schüler*innen, die bereits jetzt gewisse Schwierigkeiten mit sozialen Kontakten haben, durch personalisierte Lehr-Lern-Angebote den Kontakt zu menschlichen Sozialkontakten im Kontrast zu den auf sie abgestimmten Avataren als noch anstrengender und weniger wünschenswert erleben und persönlichen Kontakt immer mehr meiden.

Von entscheidender Bedeutung ist hier erneut die Frage der Bedeutsamkeit einer leiblichen Dimension von Begegnung, Erfahrung und Wahrnehmung im Bildungskontext und die Frage, ob das Verschwinden einer Unterscheidung natürlicher und künstlicher Umwelten letztlich negative Folgen für eine umfassende Orientierungsfähigkeit in der Welt hat.

Konkretisiert können die AC-Technologien, v. a. in Form von Bots und Avataren, auch für die Umweltbildung/Naturpädagogik/Bildung für Nachhaltige Entwicklung (BNE) angewandt werden. Hier bestehen allerdings genau die erwähnten grundlegenden Fragen, weil diese Bildungsansätze davon ausgehen, dass eine leiblich-sinnliche Erfahrung gerade nicht vollständig virtuell ersetzbar ist (dies ist eine empirische Frage) und es, falls denn überhaupt möglich, nicht sein sollte (ethische Frage), weil die Idee der Naturerfahrung genau auf maßgeblich –wenn auch

nicht ausschließlich – nicht-virtuellen Zugängen beruht. In diesem Bereich besteht ein größeres Forschungsdesiderat.

Neben einer Untersuchung der eben benannten empirischen und normativen Fragen, die sich beim Einsatz von AC-Technologie für BNE stellen – und ihre Nutzung dafür gegebenenfalls stark in Frage stellen (können) –, ist ferner wichtig im Blick zu behalten, auf welchen materiellen und energetischen Ressourcen die Technologien beruhen, da diese widersprüchlich zu einer (B)NE-Perspektive sein können.[67]

3.3.3 Ethische Analyse

Stand der Forschung: Ethische Überlegungen zu AC

Das UK Research Council for Engineering and Physical Science (EPSRC), hat eine Formulierung der Asiwovschen Gesetzte entwickelt, die auch auf Avatare anwendbar ist und explizit einen Bezug zum Umgang mit Emotionen enthält (nach (Cowie 2015), S. 338):

1. Robots should not be designed as weapons, except for national security reasons.

2. Robots should be designed and operated to comply with existing law, including privacy.

3. Robots are products: as with other products, they should be designed to be safe and secure.

4. Robots are manufactured artefacts: the illusion of emotions and intent should not be used to exploit vulnerable users.

5. lt should be possible to find out who is responsible for any robot.

Diese Arbeit zeigt zunächst, dass ethische Überlegungen im Kreis der Wissenschaftler*innen, die sich in diesem Feld bewegen, durchaus angestellt und in Richtlinien umgesetzt werden, wie sich ebenfalls dem Vergleich der KI-Richtlinien entnehmen lässt ((Hagendorff 2020); vgl. Kap. 3)

Im Anbetracht des Collingridge-Dilemmas (s.o.) ist die folgende Frage „[…] ob und wenn ja wie und mit welchen Gründen dem Einsatz von Technik gesellschaftliche Grenzen gezogen werden sollen“ ((Manzeschke und Assadi 2019), S. 167) auch für das Feld AC von besonderer Relevanz. Letztendlich dient eine ethische Analyse genau der Beantwortung dieser Fragestellung.

Der Stand der ethischen Forschung zu AC-Technologien, bezieht sich oft nicht auf einzelne Einsatzbereiche, sondern beschäftigt sich mit allgemeineren Fragen bezüglich der technischen Agenten sowie der menschlichen Akteure. Die grundständige ethische Reflektion von AC-Technologien im Bildungsbereich befindet sich (im Gegensatz z. B. zum Einsatz in der Pflege) noch in einem Anfangsstadium. Hier besteht noch deutlicher Forschungsbedarf. Die in der Literatur zu ethischen Überlegungen hinsichtlich AC bereits angesprochenen Punkte werden im Folgenden kurz zusammengestellt.

Hinsichtlich der Gestaltung der technischen Agenten sollte jeweils untersucht werden, welche Konzepte Forschung und Entwicklung leiten und welche Erwartungen, die durchaus unterschiedlich sein können, von Forscher*innen und Nutzer*innen an die Systeme gestellt werden (z. B. die Entwicklung eines Werkzeuges vs. die Entwicklung eines „fühlenden“ Systems, ((Manzeschke und Assadi 2019), S. 166). Bezüglich der Akteure, hier bezogen auf die Forscher*innen und Entwickler*innen, sollte analysiert werden, welche „[…] Welt- und Menschenbilder […] den Produktentwicklungen im Feld der emotionalisierten MTI [Mensch-Technik-Interaktion, CB] eingeschrieben werden“ ((Manzeschke und Assadi 2019), S. 166; vgl. oben). Zu beachten, so Manzeschke und Assadi ((Manzeschke und Assadi 2019), S. 166) ist in

[67] Ähnlich wie bei Smartphones mag das einzelne Gerät diesbezüglich irrelevant sein, die große Menge erzeugt jedoch Probleme (vgl. Meisch et al. 2018); hier stellen sich dann Suffizienzfragen, hier nach materiell und energetisch nachhaltigen Lebens- und Bildungsweisen.

diesem Kontext folgende zweigleisige Fragestellung: „Wirkt die conditio humana normativ auf die technische Realisierung oder wirkt umgekehrt das technisch Mögliche normativ auf den Akteur?“ ((Manzeschke und Assadi 2019), S. 166).

Für die ethische Analyse der Technologie liegt es allein aufgrund des Designs Maschinellen Lernens (als Output-Maximierung) nahe, utilitaristische Kalküle zur Bewertung von AC zu benutzen (Gabriel 2020). So finden sich in der AC-spezifischen Literatur folgerichtig auch grundsätzliche Überlegungen zur Maximierung von Wohlergehen („Benefincience“), die zum Teil als „net happiness“ adressiert werden: “[I]f we believe that affective computing can increase the net happiness of humanity, our ethical duty would include countering misguided fears that might prevent that – and, of course, ensuring that we do not inflame the fears.” ((Cowie 2015), S. 339). Allerdings sind sich die meisten Autor*innen einig, dass eine alleinige Betrachtung der Thematik aus der Perspektive eines normativen ethischen Ansatzes nicht ausreicht, um handlungsleitende ethische Entscheidungen zu diskutieren. Damit sind grundsätzliche methodische Fragen einer angewandten oder anwendungsbezogenen Ethik angesprochen, denn diese Beobachtung gilt nicht nur für ethische Analysen und das Ableiten von Handlungsoptionen für AC. Es sollte allerdings allein aus dem oben zitierten Punkt zur „net happiness“ deutlich werden, dass die Maximierung der Internet-Glücklichkeit der Nutzer*innen nicht die einzige Perspektive zur Beurteilung von AC bleiben sollte.

Als Alternativen werden entweder mittlere Prinzipien (z. B. (Goldie 2000), (Cowie 2015)) oder die Deklaration der Menschenrechte (z. B. (Pasquinelli 2019)) als Bezugsrahmen vorgeschlagen. Die vorliegenden Überlegungen setzen als Bezugsrahmen die Prinzipien einer Nachhaltigen Entwicklung sowie die Gemeinwohlorientierung voraus und beziehen die allgemeinen Menschenrechte mit ein.

Grundsätzlich sind alle ethischen Problemstellungen, die auf KI-Technologien im Allgemeinen zutreffen, auch im Kontext von Affective Computing relevant (vgl. (Hagendorff 2020); Kapitel 3.1). Gleicht man diese Punkte mit der spezifischen ethischen Auseinandersetzung in der Fachliteratur zu AC ab, die dort als besonders relevant diskutiert werden, so zeigen sich folgende Aspekte, die in den „ethical guidelines“ zur KI-Richtlinien angeführt sind, als zentral:

- Schutz der Privatheit von Nutzer*innen und Cybersecurity
- Autonomie
- Fragen des guten Lebens
- Schutz vor Diskriminierung/Diversität
- Zugangsgerechtigkeit
- Öffentliche Aufmerksamkeit und Fragen der digitalen Bildung

Darüber hinaus ergeben sich aus den Besonderheiten der AC-Technologie spezifische Punkte, die sich anhand des Einsatzes im Bildungsbereich gut zeigen lassen, wie z. B. das Problem der Täuschung, die Gefahr von Stigmatisierung und ein Missbrauchspotenzial der Daten. Die allgemeineren Punkte aus KI-Richtlinien werden im Folgenden andiskutiert. Die speziellen Punkte sind Gegenstand der Betrachtung in Kapitel 5. 2 (maßgebliche ethische Herausforderungen).

Schutz der Privatheit von Nutzer*innen und Cybersecurity

Im Falle der Aufzeichnung und Analyse menschlicher Emotionen durch technische Systeme besteht für alle Nutzer*innen eine besondere *Vulnerabilität*, da es sich ganz grundsätzlich um besonders intime und sensible Daten handelt. Dies erfordert einen sehr sorgfältigen Schutz dieser Daten sowie besondere Sorgfältigkeit bei der Herstellung von Transparenz im Umgang mit den erhobenen Informationen (vgl. (Cowie 2015)). Ein Vorbild kann hier der Umgang mit medizinischen Daten sein, für den es bereits entsprechende Vorgaben für den sehr sorgfältigen und transparenten Umgang mit und Verwertung dieser Daten gibt. Allerdings stoßen auch diese Reglements bei zunehmender Vernetzung und automatischer Verarbeitung auf problematische Grenzen, insbesondere, was die Anonymisierung sowie die informierte Zustimmung zur (Nicht-) Weitergabe/-verarbeitung angeht ((Jörg 2018); Kapitel 5.3). Cowie fasst zwei Ebenen der besonderen Vulnerabilität so zusammen:

> First, persons should respect other persons' control of access to information about their emotional states. Second, the fact that persons obtain or are entrusted with knowledge about emotional states of a person imposes special responsibilities upon them: They must not misuse the information and exploit the vulnerabilities of that person. ((Cowie 2015), S. 340)

Autonomie

Für die Autonomie von für Personen wichtige Faktoren sind die Fähigkeiten der Selbstreflexion sowie die Möglichkeiten rationale Entscheidungen zu treffen. Eine solche Form der prozeduralen Unabhängigkeit muss gegeben sein um als autonom zu gelten (Cowie 2015). Wenn Dritte, im Fall von Affective Computing also Maschinen oder die Betreiber dieser Maschinen oder virtuellen Systeme, Informationen über oder gar Zugang zu emotionalen Zuständen von Personen haben, könnte dies Einfluss auf die *prozedurale Unabhängigkeit* haben (Baumann und Döring 2011), indem z. B. die Handlungsoptionen limitiert werden. Wenn jemand meinen emotionalen Zustand kennt, dann werde ich möglicherweise anders handeln, als wenn diese Informationen nicht bekannt sind, und einige Handlungsmöglichkeiten dann gar nicht mehr in Erwägung ziehen. Zudem könnten besonders vulnerable Gruppen (Kinder oder Personen mit psychischen Einschränkungen) unbeabsichtigte und/oder unerwünschte Bindungen zu künstlichen Agenten aufbauen ((Cowie 2015), S. 338) und so in Abhängigkeitsverhältnisse geraten, die die Autonomie einschränken könnten.

Ebenfalls problematisch für autonome Handlungen und Entscheidung ist es, wenn Momente der Täuschung vorliegen. *Täuschungen* können auftreten, wenn virtuelle Systeme entweder Emotionen lesen und dadurch Handlungsoptionen aufweisen, die sich aus den Emotionen nicht ableiten lassen. Dann ist es Nutzer*innen nicht möglich, rational zwischen unterschiedlichen Optionen zu entscheiden ((Cowie 2015), S. 339). Dieses Problem wird noch verstärkt, wenn der Agent für die Unterbreitung oder Untermauerung von Handlungsoptionen auf Sprach- und Textdateien aus dem Internet zurückgreift, deren Genese und Vertrauenswürdigkeit für die Nutzer*innen nicht überprüfbar sind.

Im Zusammenhang mit dem Problem der Täuschung stehen die oben schon angesprochenen Bedenken zur voranschreitenden Anthropomorphisierung von Agenten. Allerdings zeigt die „Uncanny Valley"-These mögliche Grenzen der Vermenschlichung auf, die bei der Gestaltung der Agenten berücksichtigt werden muss. Sollte sich die „Uncanny Valley"-These also weiter bestätigen, dann werden viele Designs einen gewissen Grad der Anthropomorphisierung nicht überschreiten, um von den Nutzer*innen nicht per se abgelehnt zu werden. Im Bildungsbereich besteht allerdings noch die Besonderheit, dass vor allem für bestimmte Trainingsprogramme eine größtmögliche Ähnlichkeit der Agenten zum Gegenüber gegeben sein muss, um den

beabsichtigten Trainingseffekt überhaupt erzielen zu können. Bei Agenten, die im Bereich des formellen und informellen Lernens eingesetzt werden, ist eine große Ähnlichkeit zum Menschen hingegen nicht notwendig. Dennoch sollte im Blick behalten werden, dass für Personengruppen, denen es aus kognitiven oder emotionalen Gründen nicht möglich ist, die Unterschiede zwischen Mensch-Mensch und Mensch-Maschine Interaktionen zu erkennen, ein absichtliches oder zufälliges Verwischen dieser Unterschiede problematisch bis hin zu gefährlich werden kann (vgl. oben).

Eine Herausforderung, die Affective Computing allerdings in der Lage sein soll zu lösen oder zumindest abzuschwächen, betrifft die Kommunikation im Umgang mit Agenten. Kommunikationsformen, die dem Gegenüber als einem autonomen Wesen *Respekt* verweigern, sind problematisch. Eine sehr einfache Form von Respekt ist Höflichkeit. Diese Konventionen sollten Agenten beherrschen und Affective Computing trägt genau hierzu bei ((Cowie 2015), S. 340 f.), bzw. sollte es künftig sozusagen „per default" tun.

Schutz vor Diskriminierung

Grundsätzlich muss bei der Gestaltung von Systemen des Affective Computings immer mitgedacht werden, dass Emotionen nicht wertneutral sind (s.o. die Charakterisierung von Manzeschke und Assadi). Wenn wir Emotionen beschreiben, dann enthält eine solche Beschreibung meist auch eine Bewertung dieser Emotionen: „Hence to use them [emotions, die Autoren] is to pass a kind of moral judgment, and it is not obvious when a machine has the right to pass that kind of judgment" ((Cowie 2015), S. 341). So könnte ein Avatar z. B. feststellen, dass man ärgerlich ist. Diese Äußerung steht aber allein noch in keiner Beziehung zu der wichtigen Information, ob der Ärger angemessen bzw. gerechtfertigt ist. Solche Feststellungen können daher leicht zu *Stigmatisierung* führen. Besonders problematisch ist dies im Kontext des Einsatzes von sogenannten SIFF-Systemen (vgl. (Cowie 2015)). SIFF-Systeme sind „semi-intelligente Informationsfilter", die eigesetzt werden, um eine Menge von Daten auszuwerten, zu interpretieren und Schlussfolgerungen zu ziehen, die sehr problematisch sein können, z. B., wenn Menschen dann „verdächtiges Verhalten" oder „Aggressionsbereitschaft" attestiert wird:

> "The danger is that they [SIFFs, die Autoren] will have limitations that are poorly understood by those who deploy them and, as a result, people will be subjected to actions that they do not deserve or will not receive responses that they ought to." ((Cowie 2015), S. 341)

Auf der anderen Seite können AC-Technologien dabei helfen, Diskriminierungen durch den Einsatz kultursensitiver Systeme abzubauen (André 2015). Wenn Entwickler*innen in der Lage wären, Agenten deutlicher an kulturelle Settings, in denen sie eingesetzt werden sollen, anzupassen, dann würde einerseits das Bewusstsein für diese Unterschiede – und damit vielleicht der Respekt gegenüber diesen Unterschieden – steigen. Andererseits würden so weniger kulturelle Wertvorstellungen aus den Entwicklernationen auf andere Weltregionen und die dort lebenden Nutzer*innen übertragen. Eine etwas vertiefte Betrachtung hinsichtlich der Problematik, eine Werte-Basis für KI zu entwickeln, findet sich in der zweiten Vertiefungsstudie (Kapitel 3.4).

Zugangsgerechtigkeit

Ein Hauptziel der Entwicklung von AC-Technologien im Bildungsbereich gilt der Verbesserung der Nutzbarkeit von digitaler Technik und damit vermehrten Zugangsmöglichkeiten. Die (rudimentäre) emotionale Sensibilität der Technik soll es ermöglichen, Personen die Nutzung leichter und angenehmer zu gestalten, so dass sie Angebote vermehrt wahrnehmen könnten. Hierzu gehört z. B. die durch die AC-Technologie ermöglichte direkte Reaktionsmöglichkeit der

Agenten auf Frustrationserlebnisse seitens der Nutzer*innen. Zudem sollen kulturspezifische Anpassungen durch entsprechend abgestimmte emotionale Settings der Agenten dafür sorgen, dass kulturelle Hürden abgebaut werden. So sollen auch die Zugangsmöglichkeiten zu Bildungsinhalten auf globaler Ebene gesteigert werden. Zu hinterfragen ist allerdings, ob der enorme Entwicklungsaufwand bei bestehenden ethischen Herausforderungen tatsächlich zu rechtfertigen ist, da fraglich ist, wie groß der Personenbereich tatsächlich sein wird, dem hier eine erweiterte Zugangsmöglichkeit geboten werden soll. Zugleich aber legt das Vorsorgeprinzip hier sehr große Sorgfalt nahe.

Öffentliche Aufmerksamkeit, Fragen der digitalen Bildung in Bezug auf KI

Cowie ((Cowie 2015)) betont, dass Entwickler*innen in der Lage sein müssen, ethische Aspekte ihrer Arbeit, vor allem auch im Hinblick auf vulnerable Gruppen, erkennen und reflektieren zu können. Darüber hinaus müssen sie befähigt werden, ihre Motivationen und Ziele in der Öffentlichkeit transparent zu machen, um gerade bei besonders sensiblen Anwendungen – wie dem Umgang mit Emotionen – zu einer informierten gesellschaftlichen Debatte beizutragen. Zur Illustration führt Cowie einige Bedenken aus der öffentlichen und populärwissenschaftlichen Debatte an, die immer wieder auftauchen ((Cowie 2015), S. 345):

- die Vorstellung, dass künstliche „Surrogat-Welten“ entstehen, die so vereinnahmend sind, dass Personen den Willen oder sogar die Fähigkeit verlieren, diese Welten von der Realität zu trennen,
- die Angst, dass künstliche Agenten Menschen irgendwann überlegen sein werden,
- die Befürchtung, dass Agenten, die mit Emotionen ausgestattet sind, letztendlich eine Form von Autonomie erhalten, die dem Menschen vorbehalten ist (oder sein sollte).

Cowie schlussfolgert:

> “If that were a realistic possibility, then people ought to worry about it. The ethical issue here would seem to be helping nonexperts to gauge the probability. If people's fears are unnecessary, then those who know the reality have an ethical obligation to avoid inflaming it and, if possible, to reduce it by exposing the limits of the machines that can actually be built or envisaged.” ((Cowie 2015), S. 345)

Das bedeutet, dass Entwickler*innen zunächst in der Lage sein müssen, die Möglichkeiten und die Begrenzungen ihrer Systeme nicht nur zu kennen und zu reflektieren, sondern auch befähigt werden müssen, diese in die Gesellschaft hinein zu kommunizieren. Darüber hinaus müssen Wissenschaftler*innen die ethischen Implikationen ihrer Entwicklungen kennen:

> “They should understand the premises on which ethical judgments are likely to be based, and the fact that others may rationally hold ethical premises different from their own. […] Scientists should be sensitive to the moral implications attached to terms that they use and models that they propose. Where the fields in which they work raise other ethical issues, they should become familiar with them.” ((Cowie 2015), S. 346)

Diese ethischen und wissenschaftskommunikativen Fähigkeiten sollten entsprechend bereits in der Ausbildung gezielt geschult werden. Wäre dies flächendeckend der Fall, so könnten – zumindest über Umwege – auch im Bereich kommerzieller Anbieter vielleicht mehr ethische Reflektion in die Entwicklungsprozesse einfließen als es bislang der Fall ist.

Darüber hinaus sollten, im Sinne einer Nachhaltigen Entwicklung die wissenschaftlichen Bemühungen immer auch darauf ausgerichtet sein, zur Bildungsgerechtigkeit beizutragen. Im

Falle von AC-Technologien kann dies z. B. durch die Entwicklung von kultursensitiven Agenten (bei Vermeidung von Diskriminierung und Stigmatisierung) geleistet werden.

Cowie fasst den möglichen Mehrwert von AC-Technologien so zusammen:

> "Ethically positive aspirations involve mitigating problems that already exist by supporting humans in emotion-related judgments, by replacing technology that treats people in dehumanized and/or demeaning ways, and by improving access for groups who struggle with existing interfaces. Emotion-oriented computing may also contribute to revaluing human faculties other than pure intellect." ((Cowie 2012), S. 410)

3.3.4 Identifikation der maßgeblichen ethischen Herausforderungen

Vor dem Hintergrund des gegebenen normativen Rahmens sowie möglichen Veränderungen der Mensch-Technik-Umwelt-Beziehungen sind folgende Punkte als maßgebliche ethische Herausforderungen anzusehen: Verletzlichkeit, Täuschung, Diskriminierung sowie besondere Sicherheitsanforderungen. Die Punkte werden im Folgenden nochmals einzeln kurz reflektiert.

Menschen sind konstituiert als *verletzliche* und durch Abhängigkeiten gekennzeichnete Wesen, die ein hohes Bedürfnis nach Anerkennung und Bindung auszeichnet (vgl. (Manzeschke und Assadi 2019), S. 169). Im Bildungsbereich werden sehr häufig vulnerable Gruppen adressiert, so dass sich dies verstärkt und etliche ethisch relevante Gefahren wie Manipulation, Täuschung und starke Abhängigkeit birgt. Diese Gefahren bergen das Potenzial nicht nur die Kategorie der Gefühle in den von Nussbaum aufgeführten Fähigkeiten zu tangieren. Darüber hinaus sind die Dimensionen „Kontrolle über die eigene Umwelt", „Zugehörigkeit", „Sinne, Vorstellungskraft und Denken" sowie „Praktische Vernunft" angesprochen (vgl. (Nussbaum 2010), S. 112 - 114; Kapitel 2). Gerade die Dimension der praktischen Vernunft ist aus ethischer Perspektive besonders interessant – hier besteht in diesem Zusammenhang ein Forschungsdesiderat. Darüber hinaus gilt es zu untersuchen, wie genau sich Bindungen zwischen Menschen und Maschinen aber auch zwischen Mitmenschen verändern könnten und welche Bedeutung diese Veränderungen tatsächlich für die Lebenspraxis haben sollten. Berücksichtig werden sollte der Gedanke, dass es die Interaktion selbst zu gestalten gilt, und damit den maschinellen Aktanten, jedoch nicht die menschlichen Akteur*innen. Zu vermeiden ist eine Assimilation der menschlichen Akteur*innen an die Technik. Letztere sollte sich an den Menschen anpassen und von ihm gestaltet werden, nicht umgekehrt.

Täuschungen[68] können auftreten, wenn virtuelle Systeme Emotionen lesen und dadurch Handlungsoptionen aufweisen, die sich aus den Emotionen nicht ableiten lassen. Dann ist es Nutzer*innen nicht möglich, rational zwischen unterschiedlichen Optionen zu entscheiden: "[People] do object if the [emotional] signs mislead them about the way a system is likely to behave – particularly if the false impression affects their own choice of action." ((Cowie 2015), S. 339; s.o.)

Eine andere Form der Täuschung kann man als „pars pro toto"-Handeln bezeichnen: "lt occurs when a system shows some behaviors associated with an emotion and people infer that it has a complex of other characteristics that would be associated which that emotion in a human." ((Cowie 2015), S. 339). Gerade für den Bildungsbereich und den Umgang mit vulnerablen Gruppen ist dieser Punkt besonders eindrücklich, wie Cowie anhand eines Beispiels schildert:

[68] Cowie (Cowie 2015, S. 339 f.) weist darauf hin, dass unter dem Phänomen der Täuschung zwei Extreme auftreten: ersten eine unausweichliche Täuschung (insofern AC Emotionen grundsätzlich nur simuliert) und zweitens deliberative Täuschung (ein System programmieren, dass Leute in betrügerischer Absicht in die Irre führt). Die erst Art der Täuschung ist dem System inhärent und damit nur zu begegnen, wenn AC grundsätzlich abgelehnt wird. Bei der zweiten Täuschung handelt es sich um Betrug, welcher letztlich rechtlich zu fassen ist. Die hier angesprochenen Punkte liegen in einer tatsächlich eher problematischen Grauzone.

> "The obvious illustration is where an agent acting as a teacher or a companion uses facial and vocal gestures that give an impression of caring. That may help the agent in its intended function, but it is a problem if the user drifts into assuming that it will show other kinds of caring behavior and relies on it for help that it cannot actually provide. Teacher and companion roles are mentioned because it is a problem that we might expect to be worse where users did not have full adult judgment." ((Cowie 2015), S. 340).

Entsprechend muss für den Bildungsbereich (ebenso wie in den Bereichen Gaming und Pflege) sichergestellt werden, dass die Nutzer*innen über klare Handlungsoptionen verfügen und transparent ist, für welche Handlungsbereiche die Agenten entwickelt wurden und wo ihre Grenzen liegen, so dass Nutzer*innen realistische und verbindliche Erwartungen bilden können. Dazu muss auch darüber nachgedacht werden, wie authentisch die von den Agenten simulierten Emotionen im besten Fall sein sollten, denn nicht überall ist es notwendig, auf eine perfekte Kopie (wenn überhaupt jemals möglich) hinzuarbeiten. Es kann durchaus hilfreich sein, wenn die Übereinstimmung nicht vollständig gegeben ist (vgl. auch „Uncanny-Valley"-These):

> "Concerns specifically related to emotion involve creating a lie, by simulate emotions that the systems do not have, or promoting mechanistic conceptions of emotion. Intermediate issues arise where more general problems could be exacerbated-helping systems to sway human choices or encouraging humans to choose virtual worlds rather than reality." ((Cowie 2012), S. 410).

Grundsätzlich gilt es – wie bei allen KI-Technologien – die verschiedenen Möglichkeiten der gewollten oder ungewollten Beeinflussung bis hin zur Manipulation einzuschätzen und einem entsprechenden Missbrauch zuvorzukommen. Gelingt dies nicht, ist ein mit der Manipulierung einhergehender Souveränitätsverlust der menschlichen Akteur*innen zu konstatieren. Dies gilt unbeschadet der Tatsache, dass auch im zwischenmenschlichen Bereich Täuschungen und Manipulationen nicht zuletzt im Bereich des Emotionalen vorkommen (und dort bereits oft ethisch problematisch sind). Außerdem muss dem in den KI-Richtlinien (vgl. Kap. 3.2) stark gemachten Punkt der Beherrschbarkeit und (Selbst)-Steuerung der Technik Rechnung getragen werden, indem durch möglichst große Transparenz der Funktionsweisen ein möglichst informierter Umgang sichergestellt wird. Dies gilt auch und gerade für die Zielgruppe von Kindern und Jugendlichen.

Aspekte von *Diskriminierung* und *Stigmatisierung* sind angesprochen, da bei der Beschreibung von Emotionen meist eine Bewertung dieser Emotionen (auch durch die Maschine) enthalten ist, die Datensammlungen, die das Lernmaterial der Algorithmen darstellen, kulturell spezifisch sind und bei bestimmten Einsatzbereichen, vor allem bei sogenannten SIIF-Systemen (semi-intelligenten Informationsfiltern) zu falschen Beurteilungsergebnissen führen können: "SIIF systems […] are particularly problematic. These uses simplified rules to make judgments about people that are complex, and have potenzially serious consequences." (Cowie 2012). Daraus ergibt sich, dass nicht nur aber auch speziell für den Bildungsbereich besonders über kulturspezifische Praktiken nachgedacht werden muss. In diesem Zusammenhang kann kritisch nach der Fähigkeit zur Zugehörigkeit (vgl. Kapitel 2.3.1 und 2.3.3) gefragt werden, da die AC-Systeme sowohl dazu beitragen könnten, weniger privilegierten Gruppen eine größere gesellschaftliche Teilhabe zu ermöglichen als auch das genaue Gegenteil bewirken könnten.

Die für alle KI-Technologien geltenden *Sicherheitsmaßstäbe* sind im Falle von Affective Computing besonders relevant, da es sich – vergleichbar mit medizinischen Daten – um besonders sensible Informationen handelt. Die massenhafte Erfassung, Erzeugung, Verbreitung, Kommodifizierung und Manipulation von Emotionen durch Affective Computing in den Bereichen der digitalen Bildungsplattformen und der Online-Transaktionsplattformen sind aus

ethischer Sicht besonders kritisch zu hinterfragen und bedürfen besonderer Regulierung. Hier ist es sicherlich sinnvoll, zunächst dem Vorsichtsprinzip zu folgen.

Im Blick zu behalten ist zudem, ob und inwiefern sich unser Verständnis von *Intimität* ändert, wenn die emotionalen und besonders intimen Bereiche des menschlichen Erlebens lesbar, speicherbar, etc. sind. Damit verbundene ethische Fragestellungen sind allerdings eng verknüpft mit einer empirischen Datenlage, die auf absehbare Zeit nicht zur Verfügung stehen wird. Dennoch gilt es, diese Komponente der *Conditio Humana* mitzudenken und kritisch zu reflektieren.

Nachhaltige Entwicklung

AC-Technologien werden unter anderem speziell für die Bereiche Bildung und Pflege entwickelt. Hierbei adressieren sie vor allem die gemeinwohlorientierten Aspekte einer umfassend verstandenen Nachhaltigen Entwicklung. Damit wird deutlich in den Blick gerückt, dass Nachhaltige Entwicklung nicht nur im Denkschema der Planetaren Grenzen gedacht werden darf, sondern sehr viel breiter verstanden werden muss (vgl. Kapitel 3.1). Bildungsgerechtigkeit gilt entsprechend als zentraler Aspekt von NE, wie im SDG Nr. 4 ausgedrückt: „Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all". Deutlich wird anhand dieses Beispiels auch, dass die in den KI-Richtlinien oft nicht berücksichtige Bildungsspektrum gerade aus der NE-Perspektive eine besondere Bedeutung bei der Entwicklung von AC-Technologien haben sollte. Dies gilt ebenso für den Bereich von Diskriminierung und Stigmatisierung und hiermit ebenfalls verbundenen Gerechtigkeitsaspekten, auf die SDG Nr. 5 verweist: „Achieve gender equality and empower all women and girls."

Die Betrachtung von AC-Technologien im Bildungsbereich hinsichtlich ethischer Herausforderungen betont einen wichtigen Punkt, der auch auf andere KI-Technologien zutrifft, hier aber besonders zum Tragen kommt. Es wird deutlich, dass für Entwickler*innen und Nutzer*innen ein quasi *doppelter Bildungsauftrag* besteht. Wie besonders Cowie (s.o.) betont, müssen Entwickler*innen in der Lage sein, die ethischen Implikationen ihrer Systeme zu kennen, zu reflektieren und zu kommunizieren. Nutzer*innen wiederum müssen in die Lage versetzt werden, sich über die Möglichkeiten und Limitationen der von ihnen eingesetzten Systeme zu informieren und diese Informationen auch zu verstehen. Hier setzen die Programme digitaler Bildung an, die entsprechend erweitert werden sollten.

Ein im Kontext der Nachhaltigen Entwicklung naheliegender Gedanke bezieht sich auf den Einsatz von KI-Technologien (mit oder ohne AC) im Bereich „Bildung für Nachhaltige Entwicklung" ((WBGU 2019b), S. 360). Hierzu könnten und sollten diese Technologien einen sinnvollen Beitrag leisten. Geht man z. B. – wie oben bereits angesprochen – davon aus, dass BNE im normativen Sinne zumindest auch einen Fokus auf sinnliche Erfahrungen haben sollte, dann könnten die AC-Technologie insofern einen eigenständigen Beitrag leisten, als dass Emotionen als ein Aspekt einer sinnlichen Erfahrung Formen der digitalen Bildung im BNE-Bereich bereichern könnten, wobei die Frage der notwendigen Leiblichkeit im Kontext von AC zu klären wäre. Immersive Lernansätze sind nicht per se als technikavers oder technikaffin zu beurteilen, hier kommt es stark auf den Anwendungskontext an – und was sozusagen an früheren Zugängen ersetzt oder abgelöst wird, bzw. zusätzlich hinzukommt.

Allgemein lässt sich festhalten, dass die Entwicklung von KI-Technologien grundsätzlich darauf ausgelegt sein sollte, einen globalen Beitrag zur Bildungsgerechtigkeit zu leisten, das ist zunächst eine Forderung, die auf die Möglichkeit der gerechten Entfaltung von Fähigkeiten abstellt. Wie bereits erwähnt, ist hinsichtlich der Bildung für Nachhaltige Entwicklung dann zu klären, inwiefern die damit verbundenen Kompetenzen wirksam durch AC-Ansätze gefördert

werden können oder nicht. Dies hat wiederum einen unmittelbaren Bezug zum Fähigkeiten-Ansatz:

- Sinne, Vorstellungskraft und Denken: grundsätzliche Beeinflussung in unterstützender wie beeinträchtigender Weise möglich
- Gefühle: direkte Beeinflussung
- Praktische Vernunft: interessante Verknüpfung, wenn Emotionen als Grundvoraussetzung für Moralfähigkeit gesehen werden.
- Zugehörigkeit: mögl. Optimierung der Zugangsgerechtigkeit
- Kontrolle über die eigene Umwelt: grundsätzliche Beeinflussung unterstützender wie beeinträchtigender Weise möglich.

Über die Thematik der Bildung (SDG 4) hinaus ist aus der Perspektive einer Nachhaltigen Entwicklung noch ein weiteres der SDGs durch AC-Technologien angesprochen: SDG Nr. 8 „Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all." Einige AC-Entwicklungen für den „Office-Bereich" (z. B. Microsoft VIBE) werden damit beworben, dass sie für mehr Lebensqualität und bessere Arbeitsbedingungen sorgen können. Hierzu gehört z. B. sogenannte „Flow-Erlebnisse" zu erkennen und die Aufrechterhaltung dieser zu unterstützen, indem die entsprechende Anwendung dann z. B. automatisch eingehende Nachrichten ausblendet. Allerdings werfen solche Technologien – ebenso wie die z. B. bereits jetzt bei ZOOM und anderen Videokonferenz-Tools nachverfolgbare Aufmerksamkeitsspanne der Teilnehmer*innen – grundsätzliche Fragen nach der Überwachung am Arbeitsplatz auf. Zudem sollte hier im Blick behalten werden, inwiefern eine stetige Selbstverbesserungslogik nicht nur quantitativ verstärkt wird, sondern eventuell sogar qualitativen Veränderungen Vorschub geleistet wird. Dies wäre z. B. der Fall, wenn nicht nur immer mehr Aufmerksamkeit von uns gefordert wird, sondern zudem eine besondere Qualität solcher Aufmerksamkeit, „im Flow sein", als erstrebenswert und womöglich einforderbar erachtet wird.

Die fließenden Übergänge zwischen Training, Wohlbefinden und Überwachung gelten also sowohl im Bildungsbereich als auch in der Arbeitswelt. Hier ist erneut der doppelte Bildungsauftrag zu betonen: Entwickler*innen müssen wissen bzw. abschätzen, was sie anrichten könnten und Nutzer*innen müssen in der Lage sein zu beurteilen, worauf sie sich einlassen, wenn sie AC-Technologien nutzen. Genauso wie im Konsum-Bereich ist es aber nicht nur Aufgabe der Entwickler*innen, für Transparenz zu sorgen. Es bedarf eine gesellschaftliche Diskussion, wie und wo AC-Technologien eingesetzt werden sollen und wo gerade auch nicht. Wenn solche Technologien problematische Optimierungs-Vorstellungen oder Diskriminierung fördern, dann ist es eine Governance-Aufgabe regulierend einzugreifen.

Ebenso wie im Pflege-Sektor betrifft dies nicht nur die notwendige transparente Gestaltung der Technologie, sondern gerade auch die Besonderheiten, die sich daraus ergeben, wenn die Zielgruppe vulnerable Personen umfasst und dabei zudem sensitive Daten erhoben werden. Sowohl die angesprochenen Intimitäts- und Autonomie-Problematiken also auch das Thema der Zugangsgerechtigkeit sind im Bildungs- wie im Pflegebereich vergleichbar.

Diversitäts- und Diskriminierungsfragen, die im Bildungsbereich angesprochen sind, lassen sich auch für das Gaming-Feld ausmachen; sicherlich ist das Spielen selbst eine unterstützenwerte menschliche Fähigkeit, aber eine weiterhin offene Forschungsfrage besteht darin, welche Spiele

und Spielformen eher hilfreich oder schädlich zur Entfaltung sind. Erneut gilt aber in jedem Fall: Die Entwickler*innen müssen in der Lage sein, die Folgen ihrer Handlungen abschätzen zu können und die implizit und explizit vermittelten Wertvorstellungen kritisch hinterfragen zu können. Dazu bedarf es einer grundständigen Ausbildung ihrer ethischen Urteilskompetenz.

Die bei der Betrachtung von AC im Bildungsbereich in den Fokus rückenden Punkte sind, wie kurz angesprochen wurde, auch im Kontext anderer Entwicklungsbereiche relevant. Genau wie hier analysiert werden sie dort jedoch oft – wenn überhaupt – nur generell betrachtet und nicht im Detail kritisch beleuchtet. Gerade der Fokus auf Nachhaltige Entwicklung und Gemeinwohl verleiht diesen Aspekten ein dringend benötigtes größeres Gewicht und trägt so dazu bei, eine umfassendere ethische Reflexion von Technologien zu ermöglichen.

3.3.5 Points to Consider

Der hier angelegte normative Rahmen der Nachhaltigen Entwicklung und des Gemeinwohls zeigt, dass gerade diese besonderen Perspektiven die üblichen KI-ethischen Überlegungen in wertvoller Hinsicht ergänzen und deren Fokus sinnvoll erweitern bzw. stärken können. Folgende Punkte haben sich dabei als besonders relevant erwiesen:

- Das Problem möglicher Täuschung der Nutzer*innen sollte vor dem Hintergrund der zu gewährleistenden Selbstwirksamkeit und Autonomie bereits bei der weiteren Entwicklung sowie der Implementierung von AC-Technologien unter dem Aspekt der Transparenz mitgedacht werden. Eine Möglichkeit, unbeabsichtigte Täuschungsmöglichkeiten zu vermeiden besteht darin, zukünftige Nutzer*innen bereits an der Entwicklung teilhaben zu lassen: „The obvious prescription is that users and/or their representatives should be involved in identifying possible misinterpretations at the design stage." ((Cowie 2015), S. 340).

- Das Problem der Diskriminierung und Stigmatisierung, dass durch AC entweder verstärkt (z. B. im Falle der SIIF-Systeme) oder abgeschwächt werden kann (zum Beispiel im Bereich der interkulturellen Trainings-Settings oder durch den Einsatz von entsprechend sensiblen „Learning-Companions"). Zu beachten ist dabei, dass hier die Technik nicht wirklich neutral ist, da bei der Erstellung der Trainingsdaten-Sets immer schon Biases entstehen. Diese Biases sind allerdings bekannt und es wird daran gearbeitet, sie so weit wie möglich zu minimieren und/oder wenigstens transparent zu machen.

- Das Problem des Missbrauchs sehr persönlicher und intimer (quasi medizinischer) Daten, die gerade im Bildungsbereich (vor allem bei den sozialen Trainingsprogrammen) gesammelt und gespeichert werden. In diesem Zusammenhang stellt sich die grundsätzliche Frage nach dem Zusammenspiel von Wissenschaft und Politik sowie die Frage nach gesellschaftlich/politischer Regulierung. Skeptisch anfragen kann man hier sicherlich, ob die Weiterentwicklung von SIIF-Systemen gesellschaftlich bzw. politisch gewünscht sein kann und ob AC für den freien unternehmerischen Gebrauch zugelassen werden sollte, bzw. welche Regulierungsmöglichkeiten es hier gibt.

- Grundsätzlich förderungswürdig oder ausbaubar scheinen AC-Technologien dann zu sein, wenn sie für eine größere Zugangsgerechtigkeit zu digitalen bzw. virtuellen Systemen sorgen, die Nutzung digitaler Systeme für den Menschen weniger Stress bzw. unangenehme

Erfahrungen verursacht oder gar kulturelle Barrieren sichtbar gemacht bzw. abgebaut werden können.

- Ferner zeigt die Auseinandersetzung mit den AC-Technologien, die das sensible Thema menschlicher Emotionen ansprechen, wie dringend der im Zusammenhang mit der gesamten KI-Technologie auftretende doppelte Bildungsauftrag wahrgenommen werden muss. Es bedarf sowohl der Implementierung des Erwerbs ethischer Kompetenzen in die Ausbildung von Forscher*innen und Techniker*innen als auch des weiteren Ausbaus der digitalen Bildung für alle Mitglieder unserer Gesellschaft.

Wie in Abschnitt 4 gezeigt, können mit der AC-Technologie einige Veränderungen im Mensch-Technik-Umwelt-Verhältnis einhergehen. Hierzu gehören:

- Mensch-Technik-Beziehung: die Interaktion der Freundschaft könnte sich vor allem durch die Entwicklung von „Companions“ verändern, durch stets ansprechbare und auf die eigenen Bedürfnisse angepasste bzw. optimierte „künstliche“ Gegenüber.

- Mensch-Mensch-Beziehung: mit den oben genannten möglichen Veränderungen in der Mensch-Technik-Beziehung gehen mögliche Veränderungen in der Mensch-Mensch-Beziehung einher. Je ähnlicher die Mensch-Technik-Beziehung einer menschlichen sozialen Interaktion wird, desto eher besteht die Möglichkeit, dass die Mensch-Technik-Beziehungen Mensch-Mensch Beziehungen ersetzen könnten. Damit könnten sich für Menschen neue Fragen nach dem Selbstverständnis stellen, bzw. die alte Frage nach der Grenze zwischen Simulation und Existenz kognitiver und affektiver Zustände neu stellen.

- Zur Mensch-Mensch-Beziehung gehört auch die Frage, wie sich leibliche Dimension von Erfahrung einschließlich Emotion durch Virtualität und AC verändern. Damit zusammen hängen Fragen der Bildung, denn gerade die BNE betont die Notwendigkeit einer auch leiblichen Bildungsdimension, die verloren gehe könnte – mit unklaren Folgen.

- Mensch-Umwelt-Beziehung: Wiederum in Wechselwirkung mit den oben bereits genannten Beziehungen könnten die Verständnisweisen von „natürlich“ und „künstlich“ verschieben und sich damit Wahrnehmungen der Einstellungen zur „natürlichen“ Umwelt verändern. Auch dieser Punkt hat eine kritische BNE-Dimension bezüglich authentischer Erfahrungsräume.

Die in Zukunft tatsächlich eintretenden oder konkreter werdenden Veränderungen der Mensch-Umwelt-Beziehung durch den Einsatz von AC kann in diesem Rahmen nicht geklärt werden. Allerdings sollte eine die Entwicklung begleitende ethische Analyse die benannten potenziell auftretenden Aspekte im Blick behalten. Vor dem Hintergrund des angesprochenen Collinridge-Dilemmas ergibt sich bei AC gerade die Chance, eine Technologie bereits während ihrer Entwicklung hinsichtlich möglicher ethischer wie anthropologischer Aspekte kritisch zu begleiten.

Als bislang ungelöste ethische Probleme bzw. offene Forschungsfragen konnten wir im Laufe der Auseinandersetzung folgende Punkte identifizieren:

- Simulierte Emotionen und die Basis praktischer Vernunft (metaethisch)

- Roboter und Avatare als Mitgestalter*innen der intersubjektiven Lebenswelt
- Die Rolle von „Intimität“ für ethische Analysen
- Frage nach dem Status virtueller Erlebnisse für eine Bildung für Nachhaltige Entwicklung und für Umweltbildung (empirisch und normativ)
- Frage nach möglichen Veränderungen der Mensch-Umwelt Beziehung durch den Einsatz von simulierten Emotionen.

3.4 Vertiefungsstudie B: Erschließung von für Menschen schwer zugänglichen Räumen unter besonderer Berücksichtigung der Ozeane

3.4.1 Technik(folgen)bezogene Analyse

Welche konkrete Anwendungspraxis besteht momentan?

In den Ozeanen und der Tiefsee werden sogenannte Autonome Unterwasser Vehikel (AUV) eingesetzt. Zudem gibt es für Gewässer eine Spezialform des *Internet of Things* (IoT), das sogenannte *Internet of Underwater Things* (IoUT). Dieses funktioniert über ein kabelloses Netzwerk an Unterwasser-Sensoren (*Underwater Wireless Sensory Networks*, UWSNs), wobei die Sensoren verschiedene ökologisch wie ökonomisch relevante Daten messen, wie beispielsweise Wasserqualität und Verschmutzungsgrad, Wasserdruck und -temperatur oder die Wahrscheinlichkeit von Tsunamiwellen (vgl. (Kao et al. 2017)).

AUVs werden für zahlreiche Missionen eingesetzt und dienen der Unterstützung und Entlastung der zuständigen beteiligten Menschen. Sie werden in für Menschen unwirtlichen Umgebungen wie der Tiefsee eingesetzt und reduzieren Risiken für Menschen weitgehend, weil bzw. sofern die Missionen unbemannt durchgeführt werden können. AUVs sind in der Lage ein gutes Unterwasser-Lagebild zu verschaffen, Wartungsarbeiten und Inspektionen an Unterwassereinrichtungen durchzuführen (vgl. (Lernende Systeme- Die Plattform für künstliche Intelligenz o.J.)) sowie Daten zu sammeln für eine weitere Erforschung und Erschließung der Ozeane und vor allem der Tiefsee. Die Daten, die die AUVs sammeln, werden mit KI-Technologien ausgelesen und analysiert und die Systeme lernen aus ihnen, so dass sie nach häufigen Begegnungen mit bestimmten Gegenständen (z. B. Schiffswracks) eigenständig klassifizieren können, dass es sich um diesen Gegenstand handelt.

Um Plastik und anderen Müll aus Gewässern zu sammeln, hat das Startup-Unternehmen Ocean Cleanup den *Ocean Cleanup Interceptor* entwickelt. Dieser Müllsammler ist 24 Meter lang, solarbetrieben, autonom und angeblich in der Lage, rund um die Uhr Kunststoff aus Flüssen aufzusammeln, so dass er nicht in die Ozeane gelangt (vgl. (dw 2019)). Ende 2019 waren zwei Schiffe bereits in zwei Flüssen in Indonesien und Malaysia im Einsatz, ein drittes Schiff wurde für den Einsatz im Mekongdelta in Vietnam und ein viertes für einen Einsatz im Río Ozama in der Dominikanischen Republik vorbereitet (ebd.). In Großbritannien ist ein autonomer mariner Roboter, der sogenannte *WasteShark*, im Einsatz, der in technischer Analogie zu dem große Wassermengen in den Kiemen filtrierenden Walhai (*Rhincodon typus*), Müll aus den küstennahen Gewässern ‚siebt‘. Der vom Unternehmen RanMarine entwickelte *WasteShark* kommt bereits in fünf Ländern zur Anwendung. Gegenwärtig beruht sein autonomer Einsatz lediglich auf einer vorgefertigten Route, von der er nicht abweicht. In naher Zukunft könnte der *WasteShark* jedoch, basierend auf Technologien Maschinellen Lernens (vgl. Abb. 2), mit selbstlernenden autonom agierenden Systemen ausgestattet werden, um z. B. Routen

selbstständig festzulegen und autonom anzupassen und dadurch zu einem noch besseren „Müllsammler“ zu werden.

Neben den zwei genannten arbeiten etliche weitere Unternehmen daran AUVs zu entwickeln, die Müll aus den Ozeanen sammeln können. So arbeiten Forscher*innen der FH Kiel, der Universität Kiel und der Universität Lübeck in Zusammenarbeit mit den Unternehmen SubCtech und Emma Technologies an AUVs, die Müll und Altmunition aus Gewässern sammeln sollen und durch KI-Technologien dazu befähigt werden, aus getätigten Missionen zu lernen (Holbach 2019).

Auch an der Entwicklung der Aufforstungs-Roboter arbeiten gegenwärtig verschiedene Unternehmen wie BioCarbon (Entwicklung von Pflanz-Drohnen), SkyGrow (Entwicklung des sogenannten *GrowBots*) und Universitäten wie die University of Victoria (Entwicklung des sogenannten *TreeRover*). Während sich die Pflanz-Drohnen und die *TreeRover* momentan noch in Testphasen befinden, sind die *GrowBots* bereits im Einsatz.[69] Es ist dabei schwer ersichtlich, inwiefern es sich um selbstlernende, autonome Technologien handelt und ob die genannten Aufforstungs-Roboter bereits der schwachen KI zuzuordnen sind oder nicht. Unabhängig vom gegenwärtigen Stand ist davon auszugehen, dass an selbstlernenden, autonomen Aufforstungs-Robotern geforscht und deren Entwicklung vorangetrieben wird, so dass sie zukünftig aus vorangegangenen Aufforstungstätigkeiten selbstständig Informationen beziehen können.

Was genau wird durch den Einsatz von KI verändert?

Für bisher schwer zugängliche Räume und die Bereiche Müllreduktion in Gewässern, Tiefseebergbau sowie Aufforstung kommt der Veränderung zweier Aspekte eine besonders gewichtige Rolle zu, der *Geschwindigkeit* sowie der *Anpassungsfähigkeit*.

In Hinblick auf die *Geschwindigkeit* sind zwei verschiedene Aspekte von Bedeutung. Zum einen geht die (Weiter)Entwicklung lernender autonomer Systeme mit großer Schnelligkeit vonstatten und die Geschwindigkeit, mit der lernende autonome Systeme ihren Aufgaben nachkommen können, steigt rasant. Zum anderen ermöglichen die lernenden autonomen Systeme eine schnellere, kostengünstigere und effizientere Erschließung und Erforschung der Tiefsee sowie der Ozeanböden, was grundlegend – und vorbereitend – ist für einen möglichen Tiefseebergbau. Diese Erschließung findet in einer für das Ausmaß und die Reichweite potenzieller Konsequenzen und ethischer Herausforderungen bisher nicht dagewesenen Geschwindigkeit statt; diese können politische Entscheidungsinstanzen, wissenschaftliche Evaluationen und ethische oder rechtliche Untersuchungen möglicherweise schwer folgen.

Zum anderen ist die Möglichkeit der *Anpassung* bedeutend. Die lernenden autonomen Systeme können sich an veränderte Situationen anpassen, ohne *zusätzlich* dafür programmiert werden zu müssen (vgl. (Lernende Systeme- Die Plattform für künstliche Intelligenz o.J.)). Neue Missionen lernen autonom von vorherigen Missionen. Die AUVs werden einerseits in die Lage versetzt, selbst aus ihren „Erfahrungen“ zu lernen und das Erlernte bei der nächsten Mission umzusetzen. Andererseits lernen sie nicht lediglich, negative „Erfahrungen“ beim nächsten Mal zu vermeiden und positive zu wiederholen, sie sind auch fähig, ihre Aufgaben in einer veränderten Umgebung auszuführen, ohne dass speziell hierfür eine Programmierung stattgefunden hat. Für Arbeiten in der Tiefsee ist das ein großer Vorteil, da Tiefseeböden und -umgebungen so divers sein können, dass die Technologien bei einer Mission auf sehr unterschiedliche Umgebungen treffen (können). Aber auch für Arbeiten in beispielsweise unterschiedlich kontaminierten Umgebungen oder bei Aufforstungsarbeiten stellt diese Eigenschaft einen enormen Vorteil dar.

[69] Ihre Erforschung und Entwicklung wird durch die EU im Rahmen eines Horizon-2020-Forschungprojekts gefördert, vgl. https://cordis.europa.eu/project/id/824074 (letzter Zugriff 25.10.2021).

Darüber hinaus ändert sich in den Bereichen Müllreduktion in Gewässern und Aufforstung das Ausmaß, in dem beides betrieben werden kann. Müllsammel-Roboter sind in der Lage, Kunststoffe und anderen Müll aus Gewässern zu fischen, wie es ohne den Einsatz der Maschinen und nur von Menschenhand kaum möglich wäre. Roboter sind bei Aufforstungsarbeiten um ein Vielfaches effizienter und schneller als Menschen (vgl. (Grossmann 2017), (Mielczarek 2018), (Simpson 2018)). Entsprechend ermöglichen die autonom agierenden Systeme bestimmte Praktiken wie eine großflächige Müllreduktion in den Ozeanen erst, bzw. sie beschleunigen andere Praktiken wie das Aufforsten bestimmter Gebiete enorm. Campbell Simpson (Simpson 2018) stellt jedoch fest, dass die Aufforstungsroboter menschliche Expert*innen und Arborist*innen nicht vollständig ersetzen (können). Ihr Einsatz ist vor allem nur in den Regionen sinnvoll, wo Bepflanzung physisch anstrengend ist, wo giftige Spinnen oder Schlangen vorkommen, oder im Umgang mit unkomplizierteren Pflanzen, die kein Expert*innenwissen verlangen.

Was ist der angekündigte Mehrwert des Einsatzes von KI?

Lernende autonome Systeme können Tätigkeiten an Unterwasserstrukturen sehr stark verändern. Je nach Anwendung sorgen sie für eine Steigerung der

- Stabilität: Lernende Roboter-Assistenzsysteme können Ausfälle von Teilsystemen vorhersagen und kompensieren;
- Wirtschaftlichkeit: AUVs machen die notwendigen Inspektionen von Unterwasserinfrastrukturen wirtschaftlicher und manche Einsätze überhaupt erst möglich;
- Sicherheit: Der Betrieb von Unterwasserinfrastrukturen wird insgesamt sicherer (…). Gleichzeitig verringert der Einsatz von AUVs das Gesundheitsrisiko für Fachpersonal (v. a. für Taucher).“ ((Lernende Systeme- Die Plattform für künstliche Intelligenz o.J.), S. 15)

Diese Aspekte treffen auf alle Tätigkeiten zu, die unter Wasser durchgeführt werden müssen, sei es, um die Tiefsee und Ozeanböden aus wissenschaftlichen oder wirtschaftlichen Interessen zu erkunden, um Offshore- und Nearshore-Plattformen autonom zu warten oder um Vermisste zu suchen. Der Einsatz unbemannter Fahrzeuge ist zum einen kostengünstiger als der bemannter, zum anderen umgeht man eine Risiko-Exponierung von Pilot*innen und/oder Taucher*innen. Im Fall wissenschaftlicher Expeditionen in die Tiefsee kann eine unbemannte Mission durch ein lernendes autonomes System zusätzlich zu einem verbesserten, gerechteren Zugang zu wissenschaftlichen Daten und Erkenntnissen führen, da die bemannten Unterwasserfahrzeuge in der Regel nur mit wenigen Menschen aus privilegierten Ländern besetzt werden können. Einer Echtzeitübertragung der Kamera eines AUVs können deutlich mehr Personen folgen (vgl. (Böhm 2020)).[70]

Da das Ausführen von Tätigkeiten unter Wasser oder an der Wasseroberfläche mit zahlreichen Risiken verbunden ist, ist in Bezug auf das Themenfeld Ozeane und Tiefsee ebenso wie im Bereich kontaminierter und schwer zugänglicher Umgebungen und lawinengefährdeten Gebieten die *Nicht-Exponierung von Menschenleben* in bedrohlichen Umgebungen ein bedeutender Gewinn des Einsatzes lernender autonomer Systeme. Diese nehmen Menschen jene Arbeiten ab und ermöglichen es durch ihre Lernfähigkeit, tendenziell langfristig keine erheblichen Qualitätsverluste zu verursachen und zugleich erheblich kostengünstiger zu arbeiten. Zudem liegt in der Anwendung dieser Systeme die Hoffnung auf *schnellere und*

[70] Zur Debatte darum, ob die Authentizität von Naturerlebnissen, die durch virtuelle Erfahrungen nicht gegeben sei, vgl. Kap. 3.5 unten. Es müsste sich in Bezug auf die wissenschaftliche Arbeit die Frage anschließen, ob eine solche Authentizität für eine korrekte Interpretation der erhobenen Daten notwendig ist oder nicht. Dies übersteigt den Umfang der vorliegenden Studie.

effektivere Rettungen von Menschenleben, sowohl unter Wasser als auch in verstrahlten, verseuchten oder brennenden Gebieten und aus Lawinenregionen. Das Ziel autonomer Lawinenwarnsysteme liegt darin, Menschen direkt vor dem Lawinenunfalltod zu bewahren und hohe Sachschäden durch Präventionsmaßnahmen vermeiden zu können. Im Bereich der Aufforstung steht weniger die Reduzierung einer menschlichen Risiko-Exponierung im Vordergrund, jedoch wird es auch in diesem Feld als großer Vorteil gesehen, dass die autonom agierenden Roboter Menschen physisch strapaziöse Tätigkeiten abnehmen, wobei sie zugleich effektiver und kostengünstiger arbeiten. Zudem wird auch hier von den Unternehmen die *Effektivität* der Technologie beworben: „The average human being can plant around 1,500 seeds per day whereas a pair of BioCarbon's drones can manage nearly 100,000 in the same period." ((Interface o.J.), S. 1), „The Growbots are to plant trees in a fast, safe and low cost manner and will be able to plant in different terrains and soil types where land clearing has previously occurred".[71] Für die Roboter, die Gewässer von Müll befreien sollen, liegt ein bedeutender Mehrwert darin, dass sie etwas ermöglichen, was von Menschenhand alleine nicht möglich wäre. „RanMarine aims to empower people and organizations across the planet to restore the marine environment to its natural state. Our data-driven autonomous technology creates this opportunity by cleaning and monitoring our waters."[72]

Sowohl in Bezug auf den Tiefseebergbau als auch auf Offshore- oder Nearshore-Windparkanlagen ist zudem das Narrativ verbreitet, dass solche Techniken, künftig maßgeblich unterstützt durch schwache KI, notwendig sind, um eine Transformation hin zu Nachhaltiger Entwicklung (NE) zu ermöglichen. Windkraftanlagen produzieren erneuerbare Energien und stellen in Deutschland derzeit die Form erneuerbarer Energie dar, die den größten Teil zur Stromgewinnung beiträgt (vgl. (BMWi 2021)). Erneuerbare Energien spielen eine gewichtige Rolle, um auf den Verbrauch nicht nachhaltiger fossiler Energien verzichten zu können. Befürworter*innen des Tiefseebergbaus wie das Unternehmen The Metals Company (zuvor: DeepGreen) argumentieren für seine Notwendigkeit damit, dass die Ressourcen, die dabei am Meeresboden abgebaut werden (wie Kobalt, Kupfer, Gold, Seltene Erden, Silber, Nickel), für eine Transformation der gegenwärtigen Mobilitätskonzepte hin zu E-Mobilität sowie für das weitere Voranbringen der Digitalisierung benötigt werden.[73] Kobalt und Seltene Erden werden für die Produktion der Batterien von Elektro-Autos und -Fahrrädern sowie „smarter" Endgeräte der digitalen Vernetzung eingesetzt. Sowohl der verstärkte Einsatz von E-Mobilität, der „herkömmliche" Mobilität ersetzt, als auch eine klug vorangetriebene Digitalisierung können einen wichtigen Beitrag zur NE-Transformation leisten.[74] Die Gewinnung dieser Ressourcen in der Tiefsee wird durch lernende, autonome Systeme in großem Maßstab und vor allem in absehbarer Zeit ermöglicht. Hier besteht das Wechselspiel zwischen KI (in Form lernender autonomer Systeme) für den Tiefseebergbau – Tiefseebergbau für KI (in Form der Endgeräte bzw. der Hardware).

[71] Vgl. https://www.skygrow.com.au/about (Zugriff am 20.10.2021).

[72] Vgl. https://www.ranmarine.io/about-ranmarine-and-our-vision/ (Zugriff 25.10.2021).

[73] Vgl. https://metals.co/nodules/ (Zugriff 25.10.2021). Das Unternehmensziel liest sich wie folgt: „We're building a carefully managed metal commons that will be used, recovered, and reused again and again" (ebd.)

[74] Wobei dabei nicht in Vergessenheit geraten darf, dass in Bezug auf Mobilität eine Auseinandersetzung mit der Suffizienz-Frage unhintergehbar ist, wenn NE tatsächlich erreicht werden soll. Auch in Bezug auf die Digitalisierung gilt es, bestimmte Weichen zu stellen, damit diese zu Gunsten von nachhaltigeren Lebensstilen vonstattengeht und nicht auf Kosten derselben (vgl. WBGU 2019b). Beide Aspekte werden in Kap. 3.5 aufgegriffen.

3.4.2 Veränderung der Mensch-Technik-Umwelt-Beziehung durch die Erschließung bisher unverfügbarer Räume

Was verbindet Menschen mit ihren Artefakten, einschließlich Technik?

Für die Veränderung der Mensch-Technik-Beziehung lassen sich kaum Aspekte finden, die spezifisch sind für das hier untersuchte Anwendungsfeld. Viele der Aspekte, die für KI-Technologien allgemein gelten, lassen sich auch auf AS zur Erschließung bisher unverfügbarer Räume anwenden. Beziehungsänderungen in Hinblick auf die Mensch-Technik-Beziehung, die sich ausschließlich bezogen auf diese AS ergeben, lassen sich keine ausfindig machen.

Dennoch soll an dieser Stelle kurz auf die Ausführungen dazu verwiesen werden, dass Maschinen zwar moralisch relevante Handlungen ausführen können, jedoch nicht zur (moralischen) Verantwortung gezogen werden können (u. a. (Misselhorn 2018), S. 126). Im Fall ethisch relevanter Herausforderungen in Bezug auf AS zur Erschließung bisher unverfügbarer Räume, wie Dilemma-Situationen oder Dual Use-Praktiken, tragen nach wie vor Menschen die Verantwortung für die Entscheidung der AS. Im Zusammenspiel Programmierung, Entwicklung und Nutzung entsteht jedoch eine Art „kollektive Verantwortung“ ((Misselhorn 2018), S. 135), woraus sich die Konsequenz ergibt, dass sich niemand mehr „richtig“ verantwortlich fühlt und dass Maschinen als ‚Sündenböcke‘ herhalten müssen. Catrin Misselhorn sieht hier sogar die Gefahr einer „Erosion“ von Verantwortung für Gesellschaften (ebd., S. 134). Diese ist ebenfalls nicht spezifisch für AS zur Erschließung bisher unverfügbarer Räume, sie ist für dieses Anwendungsfeld jedoch relevant, da hier Fälle adressiert werden, in denen sich problematische Konsequenzen ergeben können (wie z. B. Todesfälle oder militärischer Einsatz der Technik), für die bestimmte Menschen oder Personengruppen Verantwortung tragen müssen.

Ein Aspekt, der in Bezug auf AS zur Erschließung bisher unverfügbarer Räume relevanter erscheint als bei anderen Anwendungsgebieten, ist die Zentralität dessen, dass die Technologien dafür eingesetzt werden sollen, um Menschen zu schonen. Menschen sollen davor bewahrt werden, in für sie lebensbedrohlichen Umgebungen Arbeiten verrichten zu müssen. Anders als in vielen anderen Bereichen steht also die Hilfsfunktion der Technologie (noch) deutlich stärker im Fokus als das – oftmals angenommene – Szenario, dass die eigene Expertise durch eine Technologie ersetzt werden könnte. Hierbei wird jedoch die – sich bezüglich vieler Technologien ergebende – Möglichkeit einer starken Abhängigkeit besonders virulent. Gibt man die Aufgabe, Menschenleben zu retten, an Maschinen ab, so kommt ihnen eine besonders bedeutende Aufgabe zu. Besteht in diesem Anwendungsfeld weniger die „Gefahr“, dass Menschen durch Maschinen ersetzt werden, so ergibt sich dagegen umso mehr ein Potenzial, beim Überleben von Menschen in Situationen der Abhängigkeit von den Technologien zu geraten.

Was macht das Menschsein und das Zusammenleben aus?

In Bezug auf das menschliche Zusammenleben ergibt sich im Bereich des hier untersuchten Anwendungsfeldes vor allem das Szenario der Entstehung einer Art erneutem „Öko-Kolonialismus“. Die „zu erschließenden“ Gebiete sind keine komplett losgelöst von Menschen zu denkenden Wildnis-Gebiete, sondern stellen entweder – wie im Fall von Wüsten – den konkreten Lebensraum bestimmter Menschengruppen dar oder grenzen nahe an diesen an. Letzteres trifft auch auf Tiefsee-Regionen zu. Diese können entweder im Hoheitsgebiet bestimmter Staaten liegen, oder aber sie grenzen näher an das Hoheitsgebiet bestimmter Staaten als an das Hoheitsgebiet anderer Staaten. So liegt die Clarion-Clipperton-Zone, in der Tiefseebergbau-Vorhaben am intensivsten diskutiert werden, zwar in internationalen Gewässern, grenzt aber nahe an einige pazifische Insel-Staaten an. Diese würden von negativen Auswirkungen, wie sie beispielsweise für die Fischerei-Wirtschaft erwartet werden, besonders betroffen. Der Großteil des Nutzens, der durch Tiefseebergbau generiert wird, liegt jedoch im

Globalen Norden. Koloniale Strukturen ergeben sich auch dadurch, dass die Inselstaaten auf Kooperationen mit westlichen Tiefseebergbau-Unternehmen angewiesen sind, da die notwendigen Lizenzen für Tiefsee-Erforschung und späteren -bergbau sehr teuer sind. Dabei bleibt intransparent, in welcher Höhe die Insel-Staaten an Profiten beteiligt werden sollen (vgl. (IntKom et al. 2018)).

Auch lässt sich hinterfragen, ob die – in der Suffizienz-Debatte häufig diskutierte – sogenannte Entschleunigung westlich geprägter Lebensstile, sich nicht auch auf die räumliche Dimension beziehen sollte. Suffizienz fordert einen Lebensstilwandel hin zu nachhaltigeren Lebensführungen. Das bezieht sich nicht lediglich auf den direkten Konsum von Gütern, sondern auch auf Mobilitätskonzepte, Reisen etc. Die räumliche Dimension bezieht sich dabei sowohl auf private Reisen als auch auf gesellschaftliche Vorhaben wie eine Erschließung von Gebieten, die bis dato anderen Lebewesen und Lebensgemeinschaften vorbehalten waren. Ob und wie diese angesichts von Suffizienzforderungen zu rechtfertigen sind, gilt es weiter zu untersuchen und gesellschaftlich zu verhandeln. Damit stehen zahlreiche Wertdimensionen des menschlichen Zusammenlebens zur Debatte. Neben der Frage, was für ein gelingendes Leben tatsächlich essenziell ist, steht auch zur Frage, ob und in welcher Hinsicht es menschliches Leben auf dem Planeten Erde und das gesellschaftliche Zusammenleben tatsächlich besser macht, wenn bislang unerschlossene und weitgehend unerforschte Gebiete wie die Tiefsee, zum einen erforscht, dann erschlossen und letztlich zur Ressourcengewinnung genutzt werden. Oder wenn von Wüstenbewohner*innen beheimatete Gebiete auf eine Weise erschlossen werden, dass sie anderen Gruppen zugänglich werden, was zahlreiche relevante Konsequenzen sowie gegebenenfalls Konkurrenzdruck um Lebensraum mit sich bringt. Fordert es die Suffizienz nicht eher, die bislang unverfügbaren und wenig zugänglichen Gebiete in Ruhe zu lassen, und ergibt sich aus einer stetig andauernden räumlichen Ausdehnung des Menschen ein langfristig nennenswerter Vorteil?

Wo steht der Mensch in der natürlichen Umwelt

Beim hier untersuchten Anwendungsfeld ergeben sich die meisten zu erwartenden Änderungen in Bezug auf die Mensch-Umwelt-Beziehung. Zentral dabei ist die – aufgrund der Erschließung bisher unverfügbarer Räume intensivierte – notwendige Auseinandersetzung mit der Frage nach menschlicher „Begrenzung“. Dies bezieht sich vor allem auf die beim menschlichen Zusammenleben bereits angesprochene räumliche Dimension, so z. B. die Frage, ob ein stetiges Weiter-Vordringen in die letzten unverfügbaren Gebiete des Planeten oder aber in den Lebensraum von Indigenen Völkern in irgendeiner Weise zu einem gelungenen menschlichen Leben beiträgt. Und falls ja, für wen, und gibt es eine zeitgleiche Einschränkung der Lebensqualität anderer?

Das weitere Voranbringen von Ressourcen-Prospektion und -Exploration, welches durch die hier diskutierten AS fortgeführt und durch KI effektiver bzw. in manchen Bereichen erst möglich wird, stützt dabei ein extraktivistisches Mensch-Natur-Verhältnis und reproduziert Handlungsweisen, welche zahlreiche globale Umweltkrisen der Gegenwart mit zu verantworten hat. Dies gilt speziell auch für den Tiefseebergbau. Dieser stellt Ressourcenabbau in einem der letzten Gebiete der Erde dar, wo er bisher nicht betrieben wird und das zudem nur punktuell bekannt ist.

Auf den ersten Blick kann der Einsatz von Müll- oder Aufforstungs-Robotern dem Mensch-Umwelt-Verhältnis sowohl zuträglich als auch abträglich sein. Zuträglich in dem Sinne, dass beide Technologien mit dem Grund beworben werden, Ökosysteme und die Biosphäre zu schützen. Abträglich insofern, als dass sie eine „*end of pipe*“-Technologie darstellen, die als sogenannte Techno-Fix-Lösung die Probleme der Vermüllung von Gewässern und der globalen

Entwaldung nicht wirklich löst, weil sie nur an der Reduzierung der Symptome ansetzt, nicht jedoch an den Ursachen. Auch das bestehende extraktive Verhältnis zur natürlichen Umwelt kann dadurch verfestigt werden, wenn die Ursachen nicht klar adressiert werden – ein Bedarf, die Ursachen zu ändern, wird dann nicht mehr gesehen. Konkret könnte das bedeuten, dass keine Notwendigkeit gesehen wird, Entwaldungsmaßnahmen oder Produktion und Verbrauch von Plastikmüll zu reduzieren. Ebenso könnte das Szenario eintreten, dass verheerende Unfälle als weniger katastrophal gewertet werden, wenn AS in die durch Gifte, Viren oder Strahlung verseuchten Gebiete vordringen müssen, aber keine Menschen mehr. Die ökologischen und anderen Folgeschäden solcher Katastrophen bleiben jedoch in gleichem Maße bestehen.

Ein bedeutender Aspekt ist zudem die anzunehmende Änderung im Mensch-Natur-Verhältnis, die sich dadurch ergibt, wenn es durch die Erschließung bisher unverfügbarer Räume keine „unerforschte Wildnis“[75] mehr auf dem Planeten Erde gibt. Es ist anzunehmen, dass sich sowohl das emotionale als auch das kognitive Verhältnis zur Natur ändern wird, wenn mittels Technologien alle bislang unverfügbaren Gebiete erschlossen werden, selbst für den Menschen so lebensfeindliche Umgebungen wie die Tiefsee. Für die einen stellt dieses Verloren-Gehen von „unerforschter Wildnis“ einen Triumph über Nicht-Wissen und potenzielle „Gefahren aus der Natur“ dar, sowie die Möglichkeit, durch weitere Erschließung von Ressourcen das Wohlstands-Niveau von Menschen zu heben. Auf die Spitze getrieben befördert dies die Idee eines Anthropozän im Sinne einer völligen Dominanz von Menschen auf ‚unserem‘ Planten – eine sich selbst befördernde Prophezeiung des Verschwindens alles Unverfügbaren. Für andere stellt der Verlust bisher für die technische Zivilisation unverfügbarer Wildnis einen großen Verlust dar, da diese entweder um ihrer selbst willen als moralisch direkt zu berücksichtigen und zu schützen angesehen werden (vgl. (Gorke 2010)) oder aus anderen Gründen als zwingend zu erhalten gelten. Die Gründe hierfür sind vielfältig. Sie können beispielsweise spirituellen oder emotionalen Argumentationslinien folgen (vgl. für eine Argumentation für die Erhaltung von Biodiversität, die sich auf Wildnis übertragen lässt, (Ott 2007)) oder ihre Begründung in der Forderung nach Respekt vor der Natur haben, der verloren gehe, wenn man sie bis auf den letzten Zentimeter durchdringen und erforschen möchte. Sie können daher rühren, dass diese Gebiete als das Zuhause anderer Spezies gesehen werden, die man aus Achtung vor diesen ‚unberührt‘ lassen solle (z. B. (Taylor 1989)) oder auf einer ökologischen Argumentation aufbauen, wonach Prozessschutz (vgl. (Potthast 2016), S. 33 - 36) für die Erhaltung der Ökosystemfunktionalität zielführend sei, was durch menschliche Eingriffe und Erschließungen gestört werde. Diese bereits explizit ethischen Aspekte beruhen aber auf einem ihnen vorgelagerten Mensch-Natur-Verhältnis, das die Unterscheidung eben im Begriff des Anthropozäns unterläuft – oder aber die spezifische Gestaltungsweise von Menschen naturalisiert und so einebnet. Beide Extreme befördern die Entdifferenzierung des Mensch-Natur-Verhältnisses an entgegengesetzten Polen.

3.4.3 Ethische Analyse

Für die ethische Bewertung der KI-basierten Technologien, die zur Erschließung neuer Lebensräume eingesetzt werden (können), besteht eine große Herausforderung. Einerseits ist es sinnvoll, bestehende Narrative, die die Autonomen Systeme als potenziell disruptiv und umwälzend darstellen, auf der empirischen und sinnstiftenden Ebene zu hinterfragen und – sofern dies zutrifft – als übertreibend zu dekonstruieren. So stellen die Praktiken, die durch die KI-Technologien ermöglicht werden, oftmals lediglich Erweiterungen oder radikale

[75] Der Wildnis-Begriff ist für die Umwelt- und Naturschutzethik ebenso zentral wie strittig: „Das deutsche ‚Wildnis‘, …, bilde(t) bis heute eines der zentralen und zugleich strittigsten normativ gehaltvollen Konzepte im Heimat- und Naturschutz.“ (Potthast 2016, S. 31) Wichtig ist die kulturelle Prägung dieses Begriffs mitsamt ihrem normativen Gehalt nicht aus dem Blick zu verlieren (dazu ausführlicher in Abschnitt 3.5).

Optimierungen bereits bestehender Praktiken dar. Insofern kommt ihnen möglicherweise doch nicht per se der behauptete disruptive Charakter zu.[76] Andererseits ist für *einzelne Felder* das Umwälzungspotenzial tatsächlich sehr hoch, da die Ermöglichung dessen, was durch die Autonomen Systeme umsetzbar wird, eine in dieser Form bislang undenkbare Perspektive zur Intervention und Gestaltung mit sich bringt, die vielleicht dem berühmten Umschlag von massiver quantitativer Veränderung in qualitative Neuerung entspricht. Mit dieser Spannung gilt es umzugehen und sie im Rahmen einer ethischen Analyse stets mitzudenken.

Das zentrale normative Framework für die hier vorgenommene ethische Analyse sind *Nachhaltige Entwicklung* und die ethischen Grundlagen, auf denen die Forderung nach dieser aufbaut, also *intra- und intergenerationelle Gerechtigkeit.* Eine Transformation hin zu NE sowie eine Orientierung an NE-relevanten Maßnahmen stellen ein geeignetes ‚Mittel' dar, um Antworten auf die oben erwähnten globalen Herausforderungen inklusive aller sozial-ökologischer Aspekte geben zu können. Aus dieser übergreifenden Bedeutung von NE wird deutlich, dass NE nicht von den Fragestellungen um die Veränderungen der Mensch-Technik-Umwelt-Beziehungen sowie von der Identifikation der maßgeblichen ethischen Herausforderungen abzugrenzen ist.

So kontrovers die Debatte darum geführt wird, welche Entwicklungen als nachhaltig gelten können und welche nicht, welche priorisiert werden sollten und wie politische Umsetzungen gestaltet werden können, so einig sind sich die Protagonist*innen dieser Debatte dennoch darüber, dass intra- und intergenerationelle Gerechtigkeit die zentrale ethische Grundlage Nachhaltiger Entwicklung (NE) ist, und dass dabei ein rein instrumentelles Verhältnis zur nichtmenschlichen Natur nicht mehr überzeugen sein kann (vgl. Kapitel 3.1).

Stand der Forschung: Ethische Überlegungen zur Erschließung bisher unverfügbarer Räume

Das Themenfeld der Erschließung bisher unverfügbarer Räume durch Autonome Systeme wird in der ethischen Literatur sehr wenig untersucht und rezipiert. „Klassisch" *umwelt*ethische Wissenschaftsmedien bieten bislang wenig bis keine Auseinandersetzungen mit diesem Thema, obwohl vielfältige umweltethische Fragen tangiert werden, wo es um Tiefseebergbau, Vermüllung der Ozeane, Aufforstung und Wüsten- sowie Hochgebirgs-Ökosysteme geht.

In der Debatte um Nachhaltige Entwicklung findet sich aus ethischer Perspektive der „normative Kompass", den der WBGU in seiner 2019 erschienenen Studie *Unsere gemeinsame digitale Zukunft* ausgearbeitet hat ((WBGU 2019b), S. 35). Diese legt jedoch zum einen den Fokus auf Digitalisierung und nicht auf Künstliche Intelligenz und zum anderen wird das Themenfeld Erschließung bisher unverfügbarer Räume nicht adressiert. In Bezug auf ethische Aspekte Nachhaltiger Entwicklung macht auch der WBGU ((WBGU 2019b), S. 137) deutlich, dass innerhalb der Tech-Community zwar ethisch relevante Aspekte wie ein Mangel an Diversität und Inklusivität, Diskriminierung und schlechte Arbeitsbedingungen reflektiert werden, (sozial-)ökologische Aspekte wie Ressourcen- und Energieverbrauch im umfassenden Sinne oder der Einfluss auf das Klima finden jedoch weniger Beachtung.

Zum gleichen Schluss kommt Hagendorff (Hagendorff 2020), der in seinem Überblicksbeitrag internationale Ethik-Regelwerke für den Umgang mit KI evaluiert. Sogenannte „versteckte Kosten" der KI-Technologien wie Ressourcen- und Energieverbrauch werden lediglich in *einem* Regelwerk benannt (Whittaker 2018). Aspekte, die bezüglich der Erschließung neuer Lebensräume eine Rolle spielen wie Gerechtigkeit und Dual-Use werden zwar in mehreren Regelwerken benannt (Gerechtigkeit in 13 von 14 untersuchten Studien, Dual Use lediglich in 6),

[76] „As Yarden Katz has remarked, AI is just a marketing operation used to rebrand what was known a decade ago as large-scale data analytics and data centre business." (Pasquinelli 2019, S. 3)

allerdings nicht auf den Tiefseebergbau bezogen, innerhalb dessen sie von besonders großer Relevanz sind. Hagendorffs Analyse macht deutlich, dass in der aktuellen Debatte zum ethischen Umgang mit KI-Technologien bestimmte Aspekte unterbestimmt, teilweise sogar unbenannt, bleiben, es aber genau diese Aspekte sind, die aus einer NE-Perspektive sehr bedeutend sind, da es Aspekte sind, die sehr eng mit einer umfassend verstandenen Gerechtigkeit[77] verwoben sind (vgl. auch Kapitel 3.1). Neuere Beiträge nehmen dieses Desiderat nun ansatzweise in den Blick, wie die Interviewsammlung „KI und Nachhaltigkeit" der Plattform Lernen Systeme anzeigt.[78]

Zum Tiefseebergbau gibt es eine Fülle an Literatur, wobei die meisten Beiträge naturwissenschaftliche, politikwissenschaftliche oder technische Forschungsergebnisse präsentieren. Explizit philosophisch-ethische Bewertungen finden sich bislang selten und eher in der sogenannten „grauen Literatur". So zum Beispiel die beiden Studien *Solwara 1 – Bergbau am Meeresboden vor Papua-Neuguinea* des Vereins für Internationalismus und Kommunikation e.V. in Zusammenarbeit mit Brot für die Welt auf (IntKom et al. 2018) und (Greenpeace 2019)), die die durch den Tiefseebergbau hervorgerufenen Verletzungen von Gerechtigkeitsgrundsätzen benennen und aufzeigen, inwiefern dieser mit Prinzipien globaler Gerechtigkeit kollidiert.

3.4.4 Identifikation der maßgeblichen ethischen Herausforderungen

Die Erforschung und Entwicklung autonomer Systeme für die Erschließung bisher unverfügbarer Räume bringt zentrale ethische Herausforderungen mit sich, die im Folgenden herausgearbeitet werden. Vorab sei darauf hingewiesen, dass bei umweltethischen Analysen unterschiedliche Grundposition zum Tragen kommen. Je nach vertretener Position kommt Argumenten gegebenenfalls eine unterschiedliche Gewichtung zu. Anthropozentrische Positionen stellen Menschen und ihre Interessen in den Mittelpunkt von Handlungsbewertungen, da ausschließlich diesen ein direkter moralischer Selbstwert zukommt. Physiozentrische Positionen schreiben auch anderen Lebewesen (Sentientismus und Biozentrik) und/oder Ganzheiten wie Ökosystemen, Arten und Prozessen (Ökozentrik und Holismus) einen moralischen Selbstwert zu.[79] Je nachdem, welcher Ansatz eingenommen wird, sind bestimmte Aspekte sehr bedeutend, die bei einem anderen Ansatz in den Hintergrund rücken können. Im Folgenden steht auf Basis des Bezugs auf Nachhaltige Entwicklung die anthropozentrische Perspektive im Vordergrund, andere Positionen fließen jedoch mit ein, wo ansonsten wichtige Aspekte wie das Wohl leidesfähiger nichtmenschlicher Wesen oder die moralische Relevanz aller Lebensformen aus dem Blick zu geraten drohen.

Moralische Dilemma-Situationen

Die Autonomen Systeme (AS) für Unterwasserarbeiten können auch dafür eingesetzt werden, Menschenleben zu retten, sollte es Unterwasser zu einem Unfall gekommen sein. Für AS für den Einsatz in kontaminierten Umgebungen stellt die Rettung von Menschenleben – neben dem Umgang mit Gefahrenstoffen – eine „typische" Aufgabe dar, ebenso wie für Bergungsroboter in den Gebirgen (die bislang jedoch nicht autonom agieren können, es handelt sich hierbei also nicht um AS). In diesen Fällen können sich sogenannte *Rettungsboot-Szenarien*, also *moralische Dilemmata-Situationen*, ergeben, in denen von mehreren gefährdeten Menschen nur eine:r gerettet werden kann oder zumindest eine:r zuerst gerettet wird, und somit in vielen Fällen die

[77] Damit ist gemeint, dass wir aus einer NE-Perspektive ökologische, soziale und ökonomische Aspekte zusammendenken und sie nicht als separierbare Bereiche von Gerechtigkeit und/oder Nachhaltiger Entwicklung ansehen (vgl. Kapitel 3.1).

[78] https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/PLS_KI_und_Nachhaltigkeit_2021.pdf (Zugriff 25.10.2021). Allerdings arbeitet dieses Dokument überwiegend nicht mit dem anspruchsvollen Begriff Nachhaltiger Entwicklung, sondern eher in einem Verständnis von Nachhaltigkeit als Umwelt- und Ressourcenschutz.

[79] Für Überblickbeiträge zu diesen Positionen(vgl. Ott et al. 2016). Zum Begriff des Selbstwerts in Abgrenzung zum Eigenwert (vgl. Eser und Potthast 1999, S. 54-55).

Überlebenschancen gesteigert wird. Moralische Dilemmata sind so definiert, dass sie „auch durch eine noch so sorgsame Abwägung nicht befriedigend gelöst werden können“ ((Ott 1996), S. 111) bzw. es liegt ein Dilemma vor, „wenn es keine befriedigende Auflösung eines moralischen Konflikts gibt“ ((Nida-Rümelin und Weidenfeld 2018), S. 103). Auch wenn die ausgeführte Handlung in solch einer Dilemma-Situation nicht moralisch befriedigend sein kann, muss evaluiert werden, welche Handlung die befriedigend*ere* sein wird. Daher sollten für solche Situationen handlungsleitende Orientierungsmaßnahmen erarbeitet werden. Philosophische Theorien arbeiten hierfür häufig Prinzipien aus, mittels derer abgewogen und entscheiden wird, wer (zuerst bzw. überhaupt) gerettet wird (vgl. z. B. (Taylor 1989), (Gorke 2010)). Diese Prinzipien müssen per Programmierung in die AS Eingang finden, die in solchen Dilemma-Situationen darüber „entscheiden“ müssen, wen sie retten. Dabei stellt sich die Frage, ob die lernenden AS in solche Situationen eigene „Erfahrungen“ miteinbeziehen können – was wäre die moralische Erfahrung eines AS? – und sollten, oder ob in solch schwierigen über Lebenschancen entscheidenden Situationen die Entscheidungsmacht ausschließlich in menschlicher Kontrolle verbleiben sollte. Beziehen die AS eigene Erfahrungen mit ein, stellt sich im Falle von misslungenen Rettungsversuchen die Frage nach Verantwortung auf eine andere Art und Weise, wie wenn sie dies nicht tun. Es muss dabei sichergestellt sein, dass es nicht zu Verantwortungslücken kommt. Zudem ergibt sich die Notwendigkeit einer Gefahrenbewertung bezüglich eines potenziellen Kontrollverlusts in solchen Situationen, wenn lernende Systeme eingesetzt werden, deren Aneignungs-Eigenschaft in den meisten Fällen durchaus gewollt und positiv bewertet wird. In Hinblick auf die Wahl von Prinzipen folgen einige sehr kontroverse Fragen.

1. Welche Prinzipien werden ausgewählt, um in einer Dilemma-Situation entscheiden zu können?
2. Wer bestimmt diese Prinzipien und wessen Wertvorstellungen entsprechen sie (nicht)?
3. Sollten die Prinzipien global und universell gelten oder kulturelle Prägungen miteinbeziehen? Sind die AS entsprechend universell einsetzbar oder sollten sie sich von Kultur zu Kultur unterscheiden? Wie kann mit dieser Frage in Bezug auf Situationen in internationalen, kulturell losgelösten Tiefseeregionen umgegangen werden?

Programmierte Werte-Basis

Damit zusammenhängend ergibt sich die – für alle KI-Technologien – wichtige Frage, mit wessen Werten die künstlich intelligenten Systeme übereinstimmen, und welche Werte programmiert werden sollten. Iason Gabriel (Gabriel 2020) benennt das als zentrale Fragestellung und weist auf die hierzu bisweilen viel zu kurz gekommene Forschung hin (vgl. Kapitel 3.2; (Spiekermann 2019) für eine detaillierte Auseinandersetzung mit Werten in digitalen Systemen). Der vorliegend stark gemachte Bezug zu ethischen Grundlagen Nachhaltiger Entwicklung sowie der Menschenrechtskonvention (vgl. Kapitel 3.2; (Gabriel 2020), S. 11), der bei der Entwicklung von und Forschung an AS stark gemacht werden sollte, kann als Orientierungshilfe dienen, wenn grundsätzlich darüber verhandelt wird, welche Prinzipien in die AS-Programmierung Eingang finden und welche nicht. Für den konkreten Fall von den eben angesprochenen moralischen Dilemmata ist diese normative Rahmung jedoch nicht weiterführend. Moralische Dilemma-Situationen stellen für die Erforschung, Entwicklung und Anwendungserprobung autonom fahrender Fahrzeuge ebenfalls eine zentrale ethische Herausforderung dar. Auf die für dieses Feld erarbeiteten und diskutierten Ergebnisse kann folglich für eine Anwendung im Bereich der für Menschen lebensbedrohlichen Räume zurückgegriffen werden; allerdings zeigt diese auch an, wie schwierig und voraussetzungsreich so etwas ist.

Dual Use

Eine weitere ethische Herausforderung, die für die AS zur Erschließung bisher unverfügbarer Räume ebenso wie für die AS in zahlreichen anderen Feldern analysiert werden sollte, ist der sogenannte *Dual Use*. Hiermit ist die Einsetzbarkeit von Gütern (im vorliegenden Fall von Technologien) für sowohl die zivile als auch die militärische Nutzung gemeint. Wieso dies ein Problem darstellen kann, welches ethischer Reflexion bedarf, bringt Alard von Kittlitz (Kittlitz 2012) etwas überspitzt formuliert auf den Punkt: „‚Dual Use' heißt der zentrale Begriff. Forschung, die zum Wohle der Menschheit betrieben wird, die in den falschen Händen aber zur Katastrophe führen kann."

Technikphilosophisch betrachtet sind hier zwei Aspekte zu unterscheiden: (a) Mit einem rein instrumentellen Verständnis der Neutralität von Technik, deren Wertdimension sich erst in der konkreten Anwendung zeige, wird nur zwischen gutem und bösem Einsatz unterschieden; dabei kann dann auch eine militärische Präzisionswaffe „gut" sein, wenn sie die „Richtigen" trifft. (b) „Dual use" im engeren Sinne trennt dagegen zwischen ziviler und militärischer Nutzung, wobei zunächst offenbleiben kann, ob jede zivile Nutzung „gut" und jede militärische „böse" ist. Sehr wohl wird aber darauf hingewiesen, dass die Kontrollmechanismen, Transparenz und Entscheidungsstrukturen im Militär oft einer sehr anderen Logik folgen als im zivilen Bereich, und das ist ethisch relevant.

In Bezug auf den Tiefseebergbau muss auf Grund der unternehmerischen Verwertungsziele desselben die Zwecksetzung „zum Wohle der Menschheit" kritisch geprüft werden. Die teilweise enge Zusammenarbeit mit dem Militär, aus der sich die Dual-Use-Herausforderung ergibt, zeigt sich beispielsweise deutlich im Fall von Großbritannien. Die britische Regierung hat die Durchführung des Tiefseebergbaus als politisches Ziel etabliert. Partner der Regierung ist dabei das größte Rüstungsunternehmen der Welt Lockheed Martin (vgl. (Malter und Steeger 2019)). Diese Dual-Use-Praktik muss der als normative Richtschnur zugrunde gelegten Gerechtigkeit nicht zwangsläufig widersprechen, sie ist jedoch sehr anfällig dafür. Zahlreiche militärischen Verwendungszwecke von Gütern und Technologien lassen sich mit Gerechtigkeitstheorien wie dem Fähigkeitenansatz von Nussbaum als widersprüchlich zu globaler Gerechtigkeit einstufen. Zudem stellt sich auch hier die Frage danach, in wessen Händen die Kontrolle der Einsätze liegt, welche Machtasymmetrien zwischen „Kontrolleuren" und Betroffenen existieren und ob die Gefahr besteht, dass die *Kontrollierbarkeit* der Technologien verloren geht, speziell im Fall selbstlernender Systeme.

Irreversibilität und Unwissenheit

In Bezug auf die Erschließung der Tiefsee mit Hilfe von Autonomen Systemen und dem Einsatz der Systeme für den Tiefseebergbau ist ein wichtiger ethisch zu bewertender Aspekt die *Irreversibilität* der geplanten Bergbau-Einsätze, da das Ökosystem entweder irreversibel zerstört wird (z. B. beim Abbau von Manganknollen) oder aus menschlicher Perspektive sehr lange benötigt, um sich zu regenerieren. Mögliche Schäden sind entsprechend nicht lediglich mit dem Problem behaftet, dass trotz aktueller Forschungstauchgänge die Tiefsee ein weitgehend *unbekanntes* Ökosystem ist, so dass der Schweregrad entstehender Schäden schwer bewertbar ist, sondern zusätzlich mit einer Langfristigkeit dieser Schäden. Auch hier spielt der Aspekt der *Kontrollierbarkeit* eine bedeutende Rolle, da auf Grund der Unkenntnis über das Ökosystem und aufgrund der lediglich punktuellen Erforschung desselben die Auswirkungen des Bergbaus leicht außer Kontrolle geraten können. Kritische Stimmen sehen in einem zeitnah gestarteten Tiefseebergbau daher eine Art Experiment: „Heute den Tiefseebergbau zu etablieren wäre nicht nur übereilt, es wäre offensichtlich ein unverantwortlicher Testversuch." ((IntKom et al. 2018), S. 55). Die Beurteilung als unverantwortlich ergibt sich aus den katastrophalen Umweltfolgen in Kombination mit der kritischen Suffizienzfrage, ob denn die scheinbar so nötig zu fördernden

Ressourcen wirklich zu einem guten Leben beitragen oder eher nicht-nachhaltige Lebensweisen weiter verstärken (vgl. (Meisch et al. 2018)).

Die Irreversibilität der Eingriffe ist auch für Einsätze AS in Hochgebirgsregionen ein bedeutender Aspekt, der in die Bewertung solcher Systeme und deren Anwendung einfließen muss. Auch hier handelt es sich um hochsensible Ökosysteme, die durch ein unbedachtes Erschließen mittels AS mittel- bis langfristig zerstört werden können.

Unwissenheit, die in die ethische Bewertung der Erschließung bisher unverfügbarer Räume einfließen muss, besteht nicht lediglich in Hinblick auf die Tiefsee, sondern ebenfalls in Hinblick auf Hochgebirgsregionen und anderer sensibler Ökosysteme, die zwar besser erforscht sind als die Tiefsee, aber für die die Konsequenzen, welche einer Erschließung und nachfolgenden (anderen) Eingriffen, folgen, nicht vorhersehbar sind. Über Handlungen in Fällen von Unwissenheit besteht eine intensiv geführte Debatte und solche Szenarien sind für die ethische Bewertung nahezu aller KI-Technologien sehr bedeutsam. Wie für KI „allgemein" ist es auch für die den Einsatz von AS zur Erschließung bisher unverfügbarer Räume sinnvoll, dem Vorsichtsprinzip (*precautionary principle*) zu folgen.

Vernachlässigung der Fehleranfälligkeit

Ein Problem, das für *alle* Anwendungsfelder autonomer intelligenter Systeme gilt, kommt auch im Bereich der Erschließung bisher unverfügbarer Räume zum Tragen. So kann ein unreflektiertes Vertrauen in die Technologie und die Ignoranz ihrer Fehleranfälligkeit ((Pasquinelli 2019); (Misselhorn 2018), S. 134) negative Konsequenzen haben. Gerade in Hinblick auf ein weitgehend unbekanntes, fragiles Ökosystem wie der Tiefsee sowie ethisch brisanter Fälle wie moralische Dilemmata-Situationen oder auch Rettungsmaßnahmen ohne Dilemma ist besonders zu prüfen, ob das Vertrauen in die AS gerechtfertigt ist und wo potenzielle Fehlerquellen drastische Auswirkungen hätten. Im Fall von Lebensrettungs-Situationen beispielsweise muss gewährleistet sein, dass das AS nach den zuvor austarierten Prinzipien vorgeht und nicht auf Grund diskriminierender Annahmen bestimmte Personengruppen anstelle anderer Personengruppen rettet. Die Fehleranfälligkeit betrifft ferner die Frage, was passiert, wenn Abbausysteme außer Kontrolle geraten und autonom Zerstörungen vorantreiben, sowie die Frage, ob bei einem Defekt der Abbausysteme kontaminierende Substanzen und Bauteile o.ä. ins Ökosystem gelangen können.

Techno-Fix-Perspektive

Der Einsatz Autonomer Systeme, um Menschenleben zu retten und um Menschen davor zu bewahren, sich in für sie lebensbedrohliche Gebiete begeben zu müssen, ist ausgesprochen positiv zu bewerten. Ebenso ist jedes Kilogramm Plastikmüll, dass durch ein AS aus Gewässern gefischt wird, begrüßenswert. Trotz dieser positiven Einschätzung geht mit Einsätzen wie den genannten eine bekannte Gefahr einher – die verharmlosende Perspektive: „Die Technologie wird es richten, es besteht daher kein wirklich ernstzunehmendes Problem". Es besteht die Gefahr einer möglichen Überkompensation (Rebound-Effekt) der Plastikmüll-Reduktion in Gewässern durch geringere Bestrebungen, Plastikmüll zu vermeiden, sofern davon ausgegangen wird, dass derselbe mithilfe von AS wieder aus den Ökosystemen entfernt werden kann. Ebenso besteht die Gefahr, dass weniger Anreiz geboten ist, die Kontaminierung von Gebieten mit toxischen Stoffen zu vermeiden, wenn dadurch keine Menschen mehr direkt betroffen wären, sondern lediglich die AS in diese Gebiete vordringen müssen. Das Einnehmen eines sogenannten Techno-Fix-Verständnisses ist daher stets mit Problemen behaftet, die es zu reflektieren gilt.[80] Für solch eine Reflexion und um das Auftreten einhegender Umkehr-Effekte zu vermeiden, ist

[80] Für eine kritische Auseinandersetzung mit Techno-Fix (vgl. Huesemann und Huesemann 2011).

die ausreichende Information der davon betroffenen bzw. davon profitierenden Bevölkerung notwendig – ein Bildungsthema.

Argumente physiozentrischer Positionen

Für sentientistische, biozentrische und auch holistische Positionen, wonach alle empfindungsfähigen bzw. alle Lebewesen direkt moralisch zu berücksichtigen sind, ergibt sich als ein zentrales Argument bezüglich der Erschließung unerschlossener Räume, dass diese Umgebungen das Habitat anderer Spezies darstellen. Es stellt sich die Frage, ob es ethisch zulässig ist, dieses zu erschließen oder ob es nicht vielmehr ethisch geboten wäre, es für Menschen unerschlossen zu lassen, um das Wohlergehen und Überleben der anderen Wesen zu sichern. Einig sind sich diese Positionen darin, dass eine Erforschung und mögliche Erschließung ausschließlich in größtmöglichen Einklang mit den dort lebenden Spezies erfolgen sollte und den einzelnen Individuen dabei möglichst kein oder ein möglichst geringer Schaden zugefügt werden darf – eine Forderung, die auch kompatibel in einer nicht-reduktionistischen anthropozentrischen Position ist und bzw. weil sie im Einklang mit den Forderungen der UN-Biodiversitätskonvention (cdb.int) steht.

Räumliche Begrenzung menschlicher Aktivitäten

Die grundsätzliche Frage nach einer räumlichen Eingrenzung menschlicher Aktivitäten wird je nach Perspektive anders bewertet. Die Frage, ob Wildnis-Gebiete oder Gebiete sekundärer Wildnis etwas an sich Wertvolles darstellen und um ihrer selbst willen in ihrer Unerschlossenheit und/oder Unverfügbarkeit erhalten bleiben sollen, ist so alt wie die Umweltethik selbst. Positionen, die Ökosystemen einen direkten moralischen Wert zuschreiben (Holismus und Ökozentrik) werden dies klar bejahen. Positionen, die (allen oder nur bestimmten) Lebewesen einen direkten moralischen Wert zusprechen (Sentientismus und Biozentrik), werden dies ebenfalls bejahen, allerdings aus Gründen der Lebensgrundlage für Organismen. Ebenso bestehen innerhalb anthropozentrischer Positionen viele Argumente, wonach diese Gebiete speziell in ihrer Unerschlossenheit und/oder Unverfügbarkeit wertvoll für Menschen sind bzw. sein können.[81] Für die hier untersuchte Fallstudie sind die Fragen relevant,

- ob (umwelt)ethische Gründe gegen eine Erschließung und Nutzung, eventuell auch gegen eine Erforschung, dieser Räume sprechen,
- ob eine Erschließung durch AS anders zu bewerten ist als direkt „durch Menschenhand“ und
- wie ein sich änderndes Mensch-Natur-Verhältnis zu bewerten ist, das daher resultiert, dass (sekundären) Wildnis-Gebieten ihr – oftmals positiv bewerteter – „geheimnisvoller“ Charakter genommen wird, wenn sie erschlossen und nutzbar gemacht werden und damit ein extraktivistisches Verhältnis zur Natur reproduziert wird.

Die Begründungslast, die sich ergibt, um für die Erschließung von und Ressourcen-Nutzung in bisher unzugänglichen Gebieten zu argumentieren, wird dadurch erschwert, dass in zahlreichen Fällen empirisch belegt ist, dass die Ausweitung des Radius erschlossener Gebiete und besonders die Explorations-Praktiken (zum Teil stark) negative Auswirkungen auf die betroffenen Gebiete haben. Gerade auch in Zeiten einer globalen Pandemie sind diese Fragen besonders virulent. Wie Sandra Junglen vom Institut für Virologie der Charité betont, kann „die

[81] Argumente lassen sich mit der Biophilie-Hypothese verknüpfen, wonach der Mensch evolutiv eine Zuneigung zu den Ökosystemen und Habitaten ausgebildet habe, die das vielfältige Leben auf dem Planeten Erde ermöglichen (Wilson 1984). Doch auch unabhängig davon gibt es vielfältige ethische Begründungen wie die sinnstiftender Erfahrungsräume, Orte für Abenteuer, transformative direkte Naturerfahrungen etc. (vgl. Eser und Potthast 1999, Potthast 2014).

Entstehung zahlreicher Krankheiten […] mit dem Vordringen des Menschen in vormals unberührte Natur erklärt werden."[82]

Ob dabei die Unerschlossenheit oder die Unverfügbarkeit dieser Räume verhandelt wird, macht für die Bewertung einen Unterschied dahingehend aus, dass *unverfügbare* Räume für ihre bloße Existenz als ebensolche Räume wertgeschätzt werden, *unerschlossene* Räume dagegen durchaus als verfügbar wertgeschätzt werden können. Die Erschließung solcher Räume direkt durch Menschen würde es ermöglichen, dass sie ein Gespür für den Raum bekommen, der dann aber der Gefahr einer starken Umgestaltung durch zu viele Besuchende oder gar weitergehende Nutzung unterworfen wird. Hier steht man vor einem bekannten Phänomen und Dilemma. AS könnten gegebenenfalls nun unerschlossene Räume störungsfreier erkunden, aber auch eine solche Erkundung könnte Erschließungsbegehrlichkeiten wecken. Zudem schließt sich daran die Frage der Authentizität und Bedeutung von direkten und realen (als abgegrenzt von virtuellen) Naturerfahrungen an.

Nachhaltige Entwicklung

Sowohl aus NE-, als auch aus Gemeinwohl-Perspektive *steht intra- und intergenerationelle Gerechtigkeit* im Fokus (auch implementiert durch SDG Nr. 16 „Frieden, Gerechtigkeit und starke Institutionen"). Wie etliche Wissenschaftler*innen gegenwärtig warnen, läuft der (oder mindestens ein vorschnell etablierter) *Tiefseebergbau* Gefahr, in Konflikt mit intra- und intergenerationeller Gerechtigkeit zu gelangen. Die hierfür benannten Gründe werden im Folgenden kurz umrissen. Sie beziehen sich dabei nicht direkt auf die Autonomen Systeme, jedoch auf indirekte Weise, da der Tiefseebergbau ohne AS nicht auf die gleiche Weise umsetzbar wäre.

Die Ressourcen der internationalen Tiefsee werden von der *International Seabed Authority* (ISA) verwaltet, die in Folge der *United Nations Convention on the Law of the Sea* (UNCLOS) gegründet wurde. Kritiker*innen werfen der ISA und vor allem dem Leiter Michael Lodge jedoch eine sehr enge Verbindung zu den zentralsten Tiefseebergbau-Unternehmen vor (beispielsweise zu *DeepGreen/metals.co*, vgl. (Malter und Steeger 2019)) und damit zusammenhängend ein Mangel an Objektivität, wenn es darum geht, das sogenannte gemeinsame Erbe der Menschheit (*common heritage of humankind*) auf tatsächlich gerechte Art und Weise zu verteilen. Ferner vermissen Kritiker*innen die Möglichkeit der Teilhabe der Öffentlichkeit, wenn es um eben dieses gemeinsame Erbe geht, die laut Lodges eigenen Worten offensichtlich nicht vorgesehen ist (vgl. ebd.).

Die polymetallischen Knollen, die die wertvollen Rohstoffe enthalten, sind als Bestandteil der Tiefsee ebenso ein gemeinsames Erbe der Menschheit. Von ihrem Abbau profitieren jedoch vor allem die Unternehmen, die sie abbauen, von denen sich mindestens die größten und einflussreichsten wie Nautilus Minerals und DeepGreen im Globalen Norden befinden.[83] Von den Endprodukten, für die die Rohstoffe genutzt werden, profitieren ebenfalls die Menschen im Globalen Norden sowie wohlhabende Eliten des Globalen Südens, da die Ressourcen vor allem für luxusorientierte Güter wie die Batterien elektrischer Mobilitätsfahrzeuge und digitaler Endgeräte benötigt werden (vgl.) (IntKom et al. 2018).[84] Der Abbau der Tiefseerohstoffe und

[82] Vgl. https://www.bmu.de/pressemitteilung/schulze-weltweiter-naturschutz-kann-risiko-kuenftiger-seuchen-verringern/ (Zugriff 22.10.2021; vgl. Settele 2020). Die Fragen, ob es diese „unberührte Natur" noch gibt und wenn nicht, seit wann nicht mehr bzw. auch, in welchem Fall tatsächlich von „unberührter Natur" gesprochen werden kann, ohne dass dies ein eurozentrisches/westliches Weltbild offenbart, welches indigene Bevölkerungsgruppen ignoriert, seien an dieser Stelle nur am Rande erwähnt; es kann nicht weiter diskutiert werden. Für die Tiefsee und den Weltraum ist es ohnehin nicht einschlägig.

[83] Wobei einer der „Big Player" des Tiefseebergbaus Nautilus Minerals inzwischen Insolvenz angemeldet hat. Einer der Investoren von Nautilus Minerals, Gerard Barron, ist der Gründer des „Big Players" DeepGreen, weswegen beide Unternehmen ihren Hauptsitz in Vancouver haben. Auch britische und deutsche Unternehmen forcieren den Tiefseebergbau.

[84] Im Zusammenhang mit der Forderung nach einem Menschenrecht auf digitale Bildung und digitalen Zugang (vgl. Kapitel 3.3 „Affective Computing") kann argumentiert werden, dass digitale Endgeräte im Zeitalter der Digitalisierung nicht mehr länger ein

entsprechend potenzielle Auswirkungen desselben (auf z. B. lokale Fischbestände, die Teilen der Bevölkerung weniger-wohlhabenden Insel-Staaten das Überleben sichern) finden dagegen im Globalen Süden statt, da das meiste Vorkommen der polymetallischen Knollen sich im Pazifischen Ozean befindet. Es stellt sich die Frage nach einer gerechten Verteilung von Vor- und Nachteilen, von Chancen und Risiken des Tiefseebergbaus, wobei die gegenwärtige Lage darauf verweist, dass der Globale Norden davon deutlich stärker profitiert als der Globale Süden. Letzterer wird hingegen deutlich stärker negative Auswirkungen zu spüren bekommen.

Von einer gerechten Verteilung kann folglich nicht ausgegangen werden. In diesem Zusammenhang liegt der Vorwurf eines *Neokolonialismus* nahe, den man eng mit Nussbaums zehnter Fähigkeit „Kontrolle über die eigene Umwelt" in Zusammenhang bringen kann, an der es den Menschen vor Ort unter solchen Gegebenheiten mangelt. Die Lizenzen für die Erforschung der Tiefsee, welche Voraussetzung für spätere Abbaulizenzen ist, sowie mögliche spätere Abbaulizenzen sind sehr teuer. Aus diesem Grund gehen viele Pazifik-Staaten Partnerschaften mit Unternehmen der G20-Staaten ein.[85] Die Pazifik-Staaten liegen zum einen nahe an den Bereichen der internationalen Tiefsee, die momentan als mögliche Abbauregion erforscht wird (die sogenannte Clarion-Clipperton-Zone), zum anderen finden sich wertvolle Ressourcen in der Tiefsee ihrer nationalen wirtschaftlichen Zone. Kritiker*innen bemängeln, dass die Verträge intransparent sind, so dass unklar ist, mit wieviel Prozent die Pazifikstaaten am Gewinn beteiligt sein werden. Die Länder des Globalen Südens können durch den Tiefseebergbau in Zusammenarbeit mit Unternehmen des Globalen Nordens zwar kurzfristig gewinnen, es besteht jedoch die ernstzunehmende Gefahr, dass dadurch langfristig das funktionierende Ozean-Ökosystem geschädigt wird (vgl. auch SDG Nr. 14 „Leben unter Wasser" und Nussbaums achte Fähigkeit „Andere Spezies"), von dem diese Länder profitieren. Aus eben diesem Grund argumentiert die Fischereiwirtschaft gegen den Tiefseebergbau (vgl. (IntKom et al. 2018)). All diese Aspekte widersprechen dem Konzept Nachhaltiger Entwicklung aus dem Brundtland Report (WCED 1987), gemäß der NE den Fokus auf den Grundbedürfnissen der weltweit am schlechtesten gestellten Menschen haben sollte (vgl. oben).

Diese (Macht-)Asymmetrien zwischen dem Globalen Norden und dem Globalen Süden stellen in Bezug auf Fragen globaler *Gerechtigkeit* keine „neuen" Aspekte dar, sondern verhärten lediglich bestehende Asymmetrien in einem neuen Feld. Allerdings zeigt das Beispiel deutlich, dass Autonome Systeme und KI-Technologien, die den Tiefseebergbau nur schneller und effizienter ermöglichen, ohne ausreichende Einbettung in entsprechende gerechte Regelwerke, die bestehenden Probleme forttreiben und teilweise vertiefen, anstatt mehr Inklusion, Chancengleichheit und gerecht verteilte Teilhabe am Wohlstand zu fördern. Dem WBGU (WBGU 2019b) ist daher beizupflichten, dass die Erforschung, Entwicklung und Nutzung dieser neuen Technologien in Bahnen gelenkt werden muss, die die NE-Transformation unterstützen, anstatt sie zu hemmen.

Das bestehende – von der Tiefseebergbau-Lobby und Unternehmen stark vorangetriebene – Narrativ, dass Tiefseebergbau notwendig sei, um nachhaltigere Gesellschaften aufzubauen (wie durch SDG Nr. 11 „Nachhaltige Städte und Gemeinden" gefordert),[86] da er für eine breitflächigere Etablierung der E-Mobilität und einer Abkehr von kohlenstoffgetriebener Mobilität notwendig

Luxusprodukt sein sollten, sondern dass alle Menschen dazu Zugang bekommen sollen. Diese Forderung ist wichtig, gegenwärtig lassen sich digitale Endgeräte dennoch als etwas klassifizieren, das sich vor allem Menschen in privilegierteren, wohlhabenderen Staaten leisten können.

[85] Neben Papua-Neugineas Partnerschaft mit dem inzwischen insolventen Unternehmen Nautilus Minerals und Naurus Partnerschaft mit DeepGreen(metals.co kooperieren Tonga, Kiribati und die Cookinseln mit Firmen der G20-Staaten.

[86] So formuliert das Unternehmen auf seiner Homepage Sätze wie: "Society has a growing need for battery metals to enable a full transition to clean energy and electric transportation. And that means we need to find new, more responsible sources for those metals. We believe that polymetallic nodules are the source with by far the least environmental and social impacts." https://metals.co/nodules/ (Zugriff 25.10.2021)

sei, suggeriert einen Zielkonflikt zwischen Klimaschutz (SDG Nr. 13) und Ozeanschutz (SDG Nr. 14): zwischen der Erhaltung fragiler, weitgehend unbekannter Ökosysteme, die als Kohlenstoffspeicher fungieren, und der Gewinnung von Rohstoffen für E-Mobilität, die dazu dienen soll, CO_2-Emissionen zu reduzieren. Was bei dieser Perspektive gänzlich außen vorgelassen wird, ist die *Suffizienz*-Frage, die sich mit nachhaltigeren Lebensstilen auseinandersetzt. Für eine Transformation hin zu nachhaltigeren Gesellschaften ist es nicht lediglich notwendig, bestehenden Lebensstilen einen „grüneren Anstrich zu verpassen" und wie bisher, nur effizienter, weiterzumachen. Notwendig ist eine Abkehr von bestimmten, als nicht-nachhaltig klassifizierten, Praktiken bei gleichzeitiger Etablierung nachhaltigerer Alternativen. Hierbei geht es nicht um eine dualistische Gewinn-Verzicht-Debatte. Suffizienz ist nicht gleichzusetzen mit Verzicht (vgl. (Meisch et al. 2020), S. 27 - 28), sondern dreht sich um die Frage, was Menschen für ein gutes Leben tatsächlich benötigen und wie nachhaltigere Praktiken zur Gewohnheit werden können (vgl. (Kopatz 2016)). In Bezug auf den angeblichen Zielkonflikt bedeutet das, dass es nicht darum gehen kann, alle „klassischen" PKWs durch E-PKWs zu ersetzen, sondern dass neue Mobilitätskonzepte erdacht und angewandt werden müssen – mit Implikationen für den Tiefseebergbau.

Die *Plastikmüll-Reduzierung* in Flüssen und Ozeanen durch AS ist aus NE-Perspektive grundsätzlich positiv zu bewerten, da der Plastikmüll in den Gewässern verheerende Auswirkungen auf die Gewässerökosysteme und die in ihnen lebenden Tiere hat. Zu einer negativen Bewertung der Müllsammel-Roboter wie dem WasteShark kann es dann kommen, wenn ein zu hoher Rebound-Effekt (vgl. oben) eintritt, so dass die Mülleinbringung in Gewässer bzw. das ohnehin zu hohe Plastikaufkommen nicht mehr zu reduzieren versucht wird, da sich die Ansicht etabliert, es sei ohnehin reversibel. Um das zu verhindern, sollten die Technologien nicht als Techno-Fix zur Lösung des Problems betrachtet werden, sondern als geeignetes Hilfsmittel, um die Symptome eines Problems zu bekämpfen, dessen Ursachen aber ebenso mit geeigneten Mitteln adressiert werden müssen.

Breitflächige und effektivere *Aufforstungen* mit Hilfe von AS sind aus NE-Perspektive ebenfalls positiv zu bewerten, u. a. da Bäume effektive CO_2-Speicher darstellen, aber auch aus zahlreichen anderen Gründen. So können Bäume beispielsweise Habitat für viele Tiere darstellen (SDG Nr. 15 „Leben an Land") und sie sind auch für menschliches Wohlergehen wichtig. In Bezug auf autonome Aufforstungs-Technologien muss hinsichtlich ihrer Bewertung jedoch von Fall zu Fall differenziert und genau hingeschaut werden: Welche Gebiete werden aufgeforstet, was war dort vorher? Wem gehören diese Gebiete bzw. gibt es Machtfragen, die in diesem Zusammenhang adressiert werden sollten? Ist die Aufforstung ökologisch sinnvoll? Wird mit endemischen oder ‚exotischen' Arten aufgeforstet und wenn letzteres, wie ist das zu bewerten? Unterschiedliche Antworten auf diese Fragen führen zu verschiedenen ethischen Bewertungen und dazu, dass die Aufforstungsmaßnahmen aus NE-Perspektive mehr oder weniger sinnvoll sind.

Ähnliches lässt sich für den Einsatz autonomer Systeme in *Wüstenregionen* festhalten. Werden AS in Wüstengebieten zur Aufforstung eingesetzt, muss kontextabhängig geprüft werden, welche Flächen aufgeforstet werden sollen, woher die Wasserversorgung für die Bepflanzung kommen kann, mit welchen Arten aufgeforstet wird und wie sich eine solche Aufforstung auf globale Kreisläufe auswirkt, da beispielsweise Sahara-Sand – der durch Aufforstung gebunden werden könnte – durch Winde über den Planeten getragen wird und nährstoffärmere Gebiete mit Nährstoffen versorgt. Legt man das Habitatschutz-Argument der physiozentrischen Positionen zugrunde, gilt es zudem zu prüfen, wessen Habitate durch eine Aufforstung zerstört werden und welche Arten bzw. Individuen durch den Einzug anderer Arten bzw. Individuen in den aufgeforsteten Regionen verdrängt werden würden. Aus Gerechtigkeitsperspektive besonders zentral ist die Frage, welche Menschengruppen in den Regionen leben und ob diese in

Entscheidungsprozesse eingebunden werden. Ist letzteres nicht der Fall besteht auch hier die Gefahr einer Form von *Neokolonialismus*. Da auch Wüsten Ökosysteme mit komplexen Interaktionen der in ihnen lebenden Arten darstellen, spielt auch hier der Aspekt der *Unwissenheit* eine nicht zu unterschätzende Rolle für die ethische Bewertung. Durch AS durchgeführte Eingriffe sowie die Erschließung solcher Ökosysteme können Auswirkungen haben, die trotz wissenschaftlicher Untersuchungen im Vorfeld nicht bekannt waren. Dual-Use-Aspekte können hier relevant werden, wenn die Wüstenregionen militär-strategisch vorteilhafte Standorte darstellen. Ob Dual-Use-Praktiken vorliegen, muss konkret geprüft und bewertet werden. Wie für die Anwendung aller KI-Technologien, ist auch für einen Einsatz autonomer Systeme in Wüsten die – häufig ignorierte – Fehleranfälligkeit der Systeme eine Herausforderung. Die Konsequenzen eines Technologie-Versagens sollten im Vorfeld geprüft und evaluiert werden. Auch die, je nach Perspektive gerechtfertigte oder ungerechtfertigte, Forderung nach einer räumlichen Begrenzung menschlicher Aktivitäten ist hier relevant, jedoch in geringerem Maßstab als für andere Gebiete, da Wüstenregionen bereits von Menschengruppen bewohnt werden. Hier ist die bereits genannte Gefahr von Neokolonialismus-Tendenzen von größerer Bedeutung.

Für AS zum Einsatz in *kontaminierten bzw. lebensbedrohlichen Gebieten* ist besonders die Frage zentral, wessen Werte in die Systeme einprogrammiert werden, da die AS in Dilemma-Situationen auf dieser Grundlage entscheiden, wer zuerst (bzw. überhaupt) gerettet wird. Für Katastrophenfälle ohne menschliche Verunglückung besteht das Risiko, dass diese als weniger dramatisch angesehen werden, wenn AS für Aufräumarbeiten in die verseuchten/brennenden Gebiete vordingen müssen, anstatt menschliche Einsatzkräfte. Die negativen Umweltauswirkungen solcher Geschehnisse bleiben jedoch bestehen, so dass die Nicht-Exponierung von Menschen in Gefahrengebieten begrüßt werden muss, die Katastrophen aber dennoch weiterhin als ebensolche gewertet werden müssen. Auch für diesen Anwendungsbereich gilt, dass die Fehleranfälligkeit der AS einer eingehenden Evaluierung bedarf. Fehltritte bei Rettungsaktionen können schwerwiegende Folgen haben, ebenso wie beim Hantieren mit Gift- und Gefahrenstoffen.

Für die Erschließung des *Weltraums* stellt sich die Frage nach einer ethisch gebotenen räumlichen Begrenzung menschlicher Aktivitäten am stärksten. Dabei fallen bestimmte Argumente für eine Nicht-Erschließung weg, wie z. B. das Habitatschutzargument oder auch ein Verweis auf Wilsons Biophilie-Hypothese (vgl. oben). Virulent werden Argumente wie der Verlust des ‚Geheimnisvollen' oder ‚Andersartigen' sowie Argumente einer grundlegenden Art von Respekt gegenüber den Menschen ohne aufwendige Technologien nicht zugänglichen Gebieten, die auf Grund einer – unterschiedlich begründeten – Achtung unerschlossen bleiben sollten. Daneben sind Gerechtigkeitsfragen von großer Brisanz, sofern es tatsächlich eine greifbare und umsetzbare Option werden sollte, menschliche Siedlungsaktivitäten auf den Weltraum zu erweitern, unter anderem um dem Planeten Erde ‚entfliehen' zu können, sollten die Lebensbedingungen auf Grund der Klimakrise und/oder Kriegsszenarien zu schlecht werden. „Flucht"-Optionen wie diese wären vermutlich wohlhabenden globalen Eliten vorbehalten, die durch extravagante Lebensstile größere Mitverursacher*innen des Problems darstellen als die Menschengruppen, die Klima- und andere Krisen am wenigsten zu verantworten haben und denen die Option einer „Weltraum-Flucht" nicht offensteht. Hier werden zahlreiche Fragen globaler, aber auch intergenerationeller, Gerechtigkeit berührt, die für die NE-Debatte zentral sind. Zudem gilt auch für dieses Anwendungsfeld, dass potenzielle Weltraum-Besiedlungsaktivitäten und der – kurz- und mittelfristig relevantere – Ressourcenabbau im Weltraum nicht von den relevanten Problemen auf der Erde ablenken dürfen, wie beispielsweise, dass etliche Ressourcen endlich sind und nicht ausschließlich die Erschließung anderer Ressourcen und eine effizientere Nutzung notwendig sind, sondern ebenso die

Etablierung von Suffizienzmaßnahmen und damit einhergehend eine Abkehr von sehr ressourcenintensiven Lebensstilen. Dies ist notwendig, möchte man eine tatsächliche Transformation zu nachhaltigeren Gesellschaften erreichen.

3.4.5 Points to consider

Vorab lässt sich für das KI-Anwendungsgebiet der AS zur Erschließung bisher unverfügbarer Räume festhalten, dass es auch hier keiner „ganz neuen Ethik" bedarf, um dieses Feld ethisch zu analysieren, sondern die Ethik vor der Aufgabe steht, die bestehenden, grundlegenden ethischen Maßstäbe auf ein neues Handlungsfeld anzuwenden und bei der Abwägung ethischer Argumente keine Zeit zu verlieren, um mit der Geschwindigkeit der Änderungen in diesem neuen Handlungsfeld mithalten zu können. Unsere Annährung an diese Aufgabe hat folgendes ergeben:

- Es hat sich herausgestellt, dass in der Nicht-Exponierung von Menschen in gefährliche Situationen und Umgebungen sowie in der Kostenersparnis, die sich durch den Einsatz von AS ergeben, enorme Vorteile liegen. Menschenleben Unterwasser oder in kontaminierten bzw. lebensbedrohlichen Umgebungen können durch den Einsatz von AS (anstatt anderer Menschen, die je nach Situation die Umgebungen gar nicht betreten können) schneller gerettet werden.

- Ebenso sehen wir einen großen Vorteil in der Möglichkeit der Technologien, sich selbst durch Lernen an veränderte Situationen und andere Umgebungen anzupassen, wodurch eine kosteneffiziente und qualitativ hochwertigere Arbeit möglich wird.

- Es ergeben sich dabei jedoch kritische Fragen nach der Notwendigkeit einer Ressourcenprospektion oder anderweitiger räumlicher Ausdehnung des Menschen vor dem Hintergrund einer, unterschiedlich begründbaren, Achtung vor „Wildnis"-Gebieten bzw. vor unverfügbaren Gebieten sowie vor dem Hintergrund der – für die Nachhaltige Entwicklung sehr wichtigen – Suffizienz. Die (u. a. lebensstilbezogenen) Fragen, welche KI-Anwendungen und AS-Einsätze wirklich notwendig sind und welche lediglich nichtnachhaltige Lebensweisen weiter stützen und ein extraktivistisches Mensch-Natur-Verhältnis reproduzieren, dürfen nicht vernachlässigt werden. Diese Fragen sind nach wie vor offen und bedürfen weiterer (umwelt-)ethischer Evaluierung.

- Erhebliche Herausforderungen bestehen durch die potenzielle verstärkte Etablierung eines „technologischen Öko-Kolonialismus", der aus Gerechtigkeitsgründen abgelehnt werden muss und einer NE-Transformation im Weg stehen würde. Auch vor diesem Hintergrund stellt sich die Frage, welche KI-Anwendungen und AS-Einsätze stattfinden und welche Möglichkeiten besser unausgeschöpft bleiben sollten vor dem Hintergrund, dass Menschen die Wahl haben, nicht lediglich Dinge durchzuführen oder anzuwenden, sondern dies auch zu unterlassen.

- Der Einsatz von AS ermöglicht ein besseres Verständnis für Ökosysteme wie die Tiefsee und die AUVs ermöglichen neue Möglichkeiten der Erforschung bisher weitestgehend unbekannter Ökosysteme. Dies kann einerseits Naturschutzmaßnahmen – im Fall der Tiefsee Meeresschutzmaßnahmen – zugutekommen, andererseits macht es die Etablierung von Nutzungspraktiken wie dem Tiefseebergbau wahrscheinlicher. Die Technologie

ermöglicht somit sowohl ein besseres und schnelleres Kennenlernen bisher unbekannter Ökosysteme und damit theoretisch auch deren besseren Schutz als auch deren bessere und schnellere Exploration. Vor diesem Hintergrund können die Technologien für eine NE-Transformation sowohl Beschleuniger als auch Hindernis sein.

- Mithilfe von AS Plastik- und anderen Müll aus den Flüssen und Meeren zu sammeln, ist grundsätzlich sehr positiv zu bewerten, wie auch effektivere Aufforstungen, sofern keine ökologischen Gründe gegen die Aufforstung sprechen. Die Hilfsdienstleistungen der AS bieten Optionen, die in dieser Weise bisher nicht geboten waren. Gleichzeitig kann es durch diese Praktiken zur Reproduktion sogenannter Techno-Fixe kommen, durch die die Gefahr entsteht, von anderen (z. B. Suffizienz- und Effizienz-) Maßnahmen abzulenken und lediglich Symptome statt Ursachen zu beheben.

Wie oben gezeigt, gehen mit vielen dieser Aspekte (mögliche) Änderungen im Mensch-Technik-Umwelt-Verhältnis einher. Die wichtigsten Potenziale dieses Anwendungsfelds, modifizierend auf dieses Verhältnis einzuwirken, sehen wir in folgenden Aspekten:

1. Die Technologie soll eingesetzt werden, um Menschenleben zu schonen, nicht um Menschen mittel- oder langfristig zu ersetzen und kann sich dadurch positiv auf eine Mensch-Technik-Beziehung auswirken.
2. Die räumliche Dimension der Entschleunigungs-Forderung westlicher Lebensstile bietet bei konsequenter Durchdringung und Umsetzung zahlreiche Optionen, das menschliche Zusammenleben zu verändern. So kann beispielsweise die unmittelbare Umgebung wieder in den Fokus rücken und eine Ressourcen-Prospektion in weit entfernten bzw. unzugänglichen Räumen als unerwünscht betrachtet werden.
3. Damit zusammenhängend hat die Frage nach einer ethisch gebotenen Begrenzung menschlicher Aktivitäten das Potenzial Mensch-Natur-Verhältnisse zu modifizieren. Entweder dahingehend, dass durch das Verschwinden unverfügbarer Räume Natur das (letztlich gebliebene bisschen an) Geheimnisvolle(m) genommen wird oder aber, dass durch eine tatsächliche Etablierung menschlicher Begrenzung vom bestehenden extraktivistischen Mensch-Natur-Verhältnis Abstand genommen wird.

3.5 Schlussfolgerungen: Synthese ethischer und anthropologische Herausforderungen der KI im Kontext von Nachhaltiger Entwicklung und Gemeinwohl

3.5.1 Unterschiede und Gemeinsamkeiten im Vergleich der beiden Vertiefungsstudien (AC und ASEuR)

Die beiden Vertiefungsstudien „Affective Computing“ (AC) und „Autonome Systeme zur Erschließung von (bislang) unverfügbaren Räumen“ (ASEuR) wurden unter anderem deshalb zur Vertiefung ausgewählt (vgl. Kapitel 3.1), weil sie Schwerpunkte auf unterschiedliche gesellschaftliche Aspekte legen, die maßgeblich von den KI-gestützten Technologien beeinflusst werden (können). Dabei bestehen allerdings auch Gemeinsamkeiten. Hier sollen über die grundsätzlichen, für alle Anwendungsfelder der KI-Technologien zutreffenden, ethisch bedeutsamen Aspekte hinaus diese Gemeinsamkeiten vor allem mit Blick auf die Verknüpfung Mensch-Technik-Umwelt (MTU – oder auch: Mensch-Maschine-Mitwelt) ausgewiesen werden. Im Folgenden fassen wir zunächst die Unterschiede zusammen und verdeutlichen anschließend die Gemeinsamkeiten, die sich in der Zusammenschau ergeben.

Die *zentrale Differenz* der beiden in den Studien behandelten Zugänge besteht darin, auf wen bzw. auf *welcher Ebene* sich die maßgeblichen Konsequenzen aus der Etablierung der Technologien auswirken. AC wirkt insbesondere auf individueller menschlicher und zwischenmenschlicher Ebene, ASEuR auf ökosystemarer Ebene sowie auf einer internationalen politischen Ebene – beides aber stets in einem MTU-Kontext. Viele der aufgeworfenen Fragen im Rahmen der AC-Studie fallen daher unter die Schnittstelle von KI-technikethischen und sozialethischen *gesellschaftspolitischen* Fragen, wohingegen bei der ASEuR-Studie technikbezogene umwelt- und naturschutzethische Fragen sowie Fragen *globaler* Gerechtigkeit aufgeworfen sind. Für eine Transformation hin zu Nachhaltiger Entwicklung sind all diese Aspekte relevant und werden als verwoben anstatt separiert angesehen. Nimmt man sie trotz der Verwobenheit in ihren charakteristischen Eigenheiten ernst, so ergibt sich als größter Unterschied eben der Hauptfokus – einmal aufs menschliche Individuum und seine zwischenmenschlichen gesellschaftlichen Beziehungen und einmal auf umfassendere Gesamtheiten wie Ökosysteme, Nationalstaaten und transnationale Institutionen. Daraus ergeben sich für *Governance*-Strukturen unterschiedliche Anforderungsprofile und zu bedenkende mögliche Maßnahmen. AC kann, wenn es für *Umweltbildung* eingesetzt wird, dabei ansetzen, gesellschaftliche *Naturverhältnisse* auf der individuellen Ebene mitzugestalten, wobei in diesem Fall Natur*verständnisse* vermittelt und verhandelt werden, die sich auf praktische Naturverhältnisse auswirken (können). Der Einsatz von KI-gestützten Technologien für die Erschließung bisher für Menschen schwer zugänglichen Räumen kann erwartbar – je nach Raum – dazu führen, bestehende extraktivistische Naturverhältnisse zu verstärken, anstatt sie, im Sinne einer NE, zu reduzieren und auf geänderte gesellschaftliche Naturverhältnisse hinzuarbeiten. Entsprechend kommt auch hier der Umweltbildung im Sinne der Förderung kritischer Reflexionskompetenzen eine erhebliche Rolle zu, die sich neben der individuellen Ebene auch stark an politische Institutionen auf national- und übernationalstaatlicher Ebene richten sollte.

Eine weitere wichtige Differenz im Ergebnis der beiden Studien, die sich aus der Adressierung zentraler sozialethischer Aspekte bei AC und umweltethischer Aspekte bei ASEuR ergibt, stellen die durch die Technologien modifizierten *Beziehungen* dar. Durch die Anwendung von AC-Technologien können sich Mensch-Technik-Beziehungen dahingehend ändern, dass das „künstliche Gegenüber“ durch das Suggerieren von Emotionen mehr und mehr den Stellenwert einer Freundin/eines Freundes einnimmt. Dies kann in einem zweiten Schritt Auswirkungen auf Mensch-Mensch-Beziehungen haben dergestalt, dass manche Mensch-Maschine-Beziehung eine zwischenmenschliche Beziehung ersetzen könnte. Bei der Erschließung von für Menschen bisher schwer zugänglichen Räumen ist das Potenzial gegeben, die Mensch-Umwelt-Beziehung zu ändern. Dies kann in einer Weise geschehen, dass durch das Verschwinden unverfügbarer Räume der Natur genau diese Eigenschaft noch weiter entzogen wird und beispielsweise dabei die Anmutung des Geheimnisvollen sowie des „ganz Anderen“ genommen wird. Eine gegensätzliche Option wäre, die Unverfügbarkeit von (manchen) Räumen als moralisch wünschenswert auszuweisen durch eine tatsächliche Etablierung letztlich auch vom bestehenden rein instrumentellen und v. a. extraktivistischen Mensch-Natur-Verhältnis Abstand zu nehmen.

Zwar fokussieren die beiden Studien unterschiedliche Ebenen, sie adressieren aber auch zentrale ethische Fragestellungen, die ihnen gemeinsam sind. So ergibt sich die Frage nach normativer Unverfügbarkeit bei ASEuR offensichtlich in Hinblick auf die (Un)Verfügbarkeit der betreffenden Räume; im Rahmen eines Einsatzes von AC bezieht sie sich auf menschliche Emotionen. Wenn diese in digitalisierter Form vorliegen, können sie gespeichert und ausgetauscht werden. Somit werden sie jenseits der jeweiligen Person verfügbar und nutzbar. Hier gilt es die Frage zu verhandeln, ob solche Daten nicht schlicht unverfügbar bleiben sollten,

sowohl für den Staat als auch für Unternehmen. Wichtig zu betonen ist, dass solche Gefahren auch ohne KI bestehen, sie aber maßgeblich und vielleicht bei AC sogar in einer neuen Qualität vorliegen. Darüber hinaus muss analysiert werden, in welchem Rahmen und in welchem Maße Räume bzw. Emotionen unverfügbar sind. Es gilt hier zu vermeiden, Unverfügbarkeit in einer simplifizierenden binären Art als Entweder- Oder zu denken, also als absolute Un- oder totale Verfügbarkeit (und tertium non datur). Innerhalb des Anwendungsfeldes AC kann man z. B. nicht von einer totalen Verfügbarkeit von Emotionen sprechen, da sowohl Erkennung als auch Simulation von Emotionen auf sehr basalen Mechanismen beruhen, die die Vielfältigkeit unseres emotionalen Erlebens nicht abbilden können. Außerdem muss man anders über Verfügbarkeit nachdenken, wenn solche Daten mit einem medizinischen Hintergrund erhoben bzw. verwendet werden, als wenn sie im Konsumbereich oder im Sicherheitsbereich genutzt werden. In Bezug auf für Menschen schwer zugängliche Räume sollten unterschiedliche normative Schlussfolgerungen gezogen werden, je nachdem, ob eine Verfügbarmachung des Raums zu dessen erhöhtem Schutz oder verstärkter Degradation führt, wie man es als gegenübergestellte Optionen in Hinblick auf die Tiefsee findet. Prinzipiell gilt für alle bisher für Menschen schwer zugänglichen Räume, dass sie durch das Zugänglich-Werden nicht in ihrer Gänze verfügbar werden, sondern der Übergang vom unzugänglichen Raum zum zugänglichen Raum graduell verläuft und in den meisten Fällen nach wie vor einer Spezialausrüstung bedarf, so dass diese Räume nicht allen Menschen zugänglich werden.

Die Gefahr der *Täuschung und Manipulation* ist ebenfalls in beiden Studien aufgeworfen: Durch AC können einzelne Individuen emotional manipuliert werden, und sie können sich hinsichtlich der tatsächlichen emotionalen Fähigkeiten der Avatare oder Roboter täuschen bzw. getäuscht werden. Beide Formen des Missbrauchs führen zur Einschränkung der selbstbestimmen Handlungsfähigkeit. Im Fall der Erschließung bisher schwer zugänglicher Räume geht es um die Frage, welche Narrative der Erschließung, umweltfreundlichen Machbarkeit und Verfügbarkeit sich durchsetzen und welchen Grad an Realistik sie enthalten bzw. leugnen. Über die „normative Kraft des Fiktionalen“ kann an den an ASEuR beteiligten Unternehmen und/oder staatlichen Konsortien eine Täuschung der globalen Öffentlichkeit gelingen, sowohl was die Machbarkeit als auch den Zerstörungsgrad der Eingriffe vorgeworfen werden. Dies ist wiederum kein Spezifikum der KI-gestützten Technologien, aber hinsichtlich der Debatten um empirische und normative Unverfügbarkeit von größter Bedeutung, weil alte Träume der Erschließung „völlig neuer Welten“ bedient werden.

Zu beachten sind übereinstimmend auch Fragen nach *Militarisierung bzw. dual use*: Bei ASEuR ist der Aspekt wichtig, dass die KI-Techniken maßgeblich (mit) vom Militär entwickelt und genutzt werden, so dass die Frage besteht, ob hier schlicht territoriale Herrschaftsansprüche befördert werden, in denen bislang die Staatengemeinschaft eine nationale Unverfügbarkeit forderte. Bei AC finden sich diverse Aspekte, die für AI und dual use allgemein zutreffen und die vor allem die Kontrolle von Personen und den Sicherheitssektor im weiteren Sinne betreffen.

Die Diskriminierung und Stigmatisierung, die sich aus der Anwendung von AC-Technologien ergeben können, verstoßen gegen die für NE zentrale ethische Forderung intragenerationeller Gerechtigkeit. Beim Tiefseebergbau treten zwangsweise Verstöße gegen globale Gerechtigkeitsforderung durch ungleiche Verteilung von Vorteilen und Nachteilen auf, aber auch gegen die intergenerationelle Gerechtigkeit aufgrund irreversibler Schäden am Ökosystem Tiefsee. Auch hier spiegeln sich die unterschiedlichen Ebenen der Betroffenheit wider. Bei AC besteht das Diskriminierungspotenzial bzw. die auf individueller Ebene bzw. der gesellschaftspolitischen Ebene marginalisierter Personengruppen; bei Anwendung von KI-gestützten Technologien zur Erschließung von bisher schwer zugänglichen Räumen sind es internationale Nord-Süd- und planetare Umweltgerechtigkeit. Man kann argumentieren, dass

Gesamtheiten wie Ökosysteme und Nationen nicht ungerecht behandelt werden können, sondern dass auch in diesem Fall letztlich Individuen als Teil von Kollektiven Unrecht geschieht. Dennoch sind es in diesem Fall „abstraktere" Einheiten von Individuen, da es im Fall der gegenwärtig Betroffenen größere Gruppen (an Individuen) trifft und im Fall der zukünftig Betroffenen die Individuen noch nicht existieren.

Beide Studien legen entsprechend den für Nachhaltige Entwicklung – und immer auch im Sinne des Gemeinwohls – zentralen Punkt der Gerechtigkeit offen. Beide verweisen auf die Fragen der gerechten Verteilung und Zugänglichkeit von Technologien und ihren Folgen in intra- und intragenerationeller Perspektive. Zudem ist Frage der Verpflichtung gegenüber der Mitwelt in sentientistischer, biozentrischer oder holistischer Perspektive zumindest ethisch bedenkenswert, wenn beispielsweise der Eigenwert der biologischen Vielfalt im Sinne der CBD berücksichtigt wird.

Dabei ist gleichzeitig stets die Frage, welches Potenzial zur Verstärkung von Abhängigkeiten jeweils in der Etablierung Technologien enthalten ist und ob der techno-fix insgesamt die zu bevorzugende Lösung des jeweiligen Problems ist. Eine weitere Gemeinsamkeit, die sich für alle KI-Technologien verallgemeinern lässt, ist die Einsicht, dass die Technologien für eine NE-Transformation sowohl Beschleuniger als auch Hindernis sein können. Wie Kapitel 4 und 5 gezeigt haben, können sie je nach Anwendung der Etablierung eines „mehr" an Gerechtigkeit im Sinne von NE dienlich sein oder dem entgegenwirken. Um dies nicht als unverbindlichen Allgemeinplatz zu belassen, bedarf es im Rahmen der Planung, der Entwicklung und – nicht erst (!) – des Einsatzes dieser Technologien der steten ethischen und politischen Debatte und gegebenenfalls stringent durchgesetzter Kontrollmaßnahmen. Hier kommt abermals der – für die Forschung und Entwicklung von AC ebenso wie ASEuR bedeutende – doppelte Bildungsauftrag zum Tragen, welcher darauf hinweist, dass Entwickler*innen in der Lage sein müssen, die ethischen Implikationen der Zielsetzung und der Folgen ihrer Systeme zu reflektieren und dass Nutzer*innen in die Lage versetzt werden müssen, die Reichweite und Limitationen der angewandten Systeme zu verstehen.

3.5.2 Übertragbarkeit der Untersuchungsergebnisse aus den Vertiefungspapieren auf andere KI-Felder

Affective Computing

Beim *Natural Language Processing* handelt es sich um einen Grundbaustein der AC-Technologie, ohne den weder die Emotionserkennung noch die Simulation funktionierten würde, da beide auch auf die Sprachverarbeitung zurückgreifen. Entsprechend treffen die für das Natural Language Processing diskutierten ethischen relevanten Aspekte ebenso auf AC-Technologien zu. Verstärkt in den Fokus rückt dabei das mögliche Problem der Manipulation und Täuschung, das im Rahmen von Natural Language Processing bereits diskutiert wird. Entsprechend problematischer wird die AC-Technologie hinsichtlich ihrer Konsequenzen sowohl im Überwachungssektor als auch im Konsumbereich. Mögliche Potenziale hinsichtlich der Beteiligung und Zugangsgerechtigkeit können in der Kombination von NLP und AC hingegen verstärkt genutzt werden.

Innerhalb des Bereichs der *Extended/Augmented Reality* lassen sich die Überlegungen zur Entfremdung von einer Realität, zum möglichen Verlust zwischenmenschlicher Beziehungen sowie erneut das Thema Manipulation und Täuschung übertragen. Diese Befürchtungen könnten auch nochmal verstärkt vorliegen, wenn beide Technologien miteinander verknüpft werden. Auf der anderen Seite könnte in der Kombination von Extended/Augmented Reality und AC auch ein größeres Potenzial hinsichtlich der Nutzung im Bildungsbereich bestehen.

Daraus ergibt sich dann aber auch noch deutlicher die Frage der Bedeutung rein digitaler Erfahrungen z. B. für die Umweltbildung. Hier wird ein bislang technisch-empirisch unverfügbarer „natürlicher“ Weltzugang technisch verfügbar und verschiebt damit anthropologische Grenzen. Zudem wäre hier ebenfalls die ethische Frage der Grenze technisch vermittelter Verfügbarkeit zu erörtern.

Die der *Enhancement*-Debatte zugrundeliegenden Natürlichkeitsverständnisse und Künstlichkeitsvorstellungen sowie die angedeuteten Fragen zur Verfügbarkeit werden nochmal stärker in Frage gestellt, wenn Emotionen technisch zugänglich und produzierbar werden. Letztlich geht es um die Technisierbarkeit der Emotionen, die zuvor eben zur Natur des Menschen gehörten und dies nun nicht mehr nur sind. Ebenfalls gemeinsam sind beiden Technologiefeldern auf der einen Seite die Frage nach therapeutischem Nutzen, z. B. bei der Diagnose von Depressionen oder der Unterstützung emotionaler Kontrollprozesse und auf der anderen Seite das Spektrum der Optimierungsfragen. Verstärkt unter dem Aspekt des Enhancements zudem die Frage des Missbrauchs persönlicher, intimer Daten und die damit verbundenen Sicherheits- und Privatheitsaspekte.

Erschließung von für Menschen schwer zugänglichen Räumen

Wie im Fall der Erschließung von für Menschen schwer zugänglichen Räumen, dient KI zur Unterstützung *autonomer Systeme* auch bei anderen Anwendungsfeldern einer Entlastung der Gefährdung von Menschen in entsprechenden Umgebungen. Wenn diese Systeme der *Reparatur oder Wiederherstellung* wünschenswerter Zustände dienen, kann dies einerseits erstrebenswert im Sinne der Zielsetzung sein, andererseits aber auch entsprechende *Vorsorgemaßnahmen* schwächen, die dazu führen, dass Schäden gar nicht erst entstehen (techno-fix vs. Vorsorge).

Wie im Fall der ASEuR mit Fokus für die Tiefsee gezeigt, werden KI-Technologien „allgemein“ angesehen als Mittel zur *Überschreitung aller Grenzen* der Räume, die bisher dem Zugriff von Menschen entzogen schienen. Damit gehen für alle diese Technologien gewisse „Science-Fiction-Phantasien“ einher in Bezug auf die Erschließung „unendlicher Weiten und Tiefen“ in den Weltmeeren (bis hin zum tiefsten Punkt des Mariannen-Grabens) und darüber hinaus einer immer weiteren Erschließung des Weltraums. Damit verbunden ist die immer stärkere Etablierung (zumindest in bestimmten Thinktanks von „Visionären“und bestimmten Forschungs- und Entwicklungs-Gemeinschaften) einer Anthropologie der Grenzüberüberschreitung als Charakteristikum des Menschen. Mit solch grenzüberschreitenden Entwicklungen wird die Trennung von Natur als „dem Anderen“ des Menschen durch das Mittel der Technik obsolet gemacht: Natur wird zwar nicht selbst technisiert, aber durch die technische Verfügbarmachung letztlich aufgelöst, weil sie den Charakter des Anderen und Unverfügbaren verliert. Die ethische Debatte der Unverfügbarkeit fragt – analog zur Wildnisdebatte im Naturschutz – danach, welche Räume der Welt und des Weltraums aus welchen Gründen unverfügbar gelassen werden sollen, selbst wenn sie verfügbar sein könnten – in Zukunft auch aufgrund der immer breitflächigeren Etablierung von KI-basierten Technologien.

3.5.3 Übergreifende Aspekte und Forschungsdesiderate

Zentrale normative Aspekte, die wir in Hinblick auf AC und ASEuR aufgeworfen haben, wie Herausforderungen für die Sicherheit, die Privatheit, Fragen nach Transparenz, Kontrollierbarkeit und insgesamt die – aus Gerechtigkeitsperspektive besonders relevanten – Aspekte der Zugänglichkeit und Ermöglichung von Teilhabe (möglichst) aller, stellen sich in ähnlichem Maß auch beispielsweise für Big Data-Anwendungen und Hyperkonnektivitäts-Praktiken.

Allgemein lassen sich nochmals fünf übergreifend wichtige Aspekte zugleich als ethische Forschungsdesiderate formulieren:

1. **Manipulation und Täuschung** in kritischer Verbindung mit **Beteiligung und Zugangsgerechtigkeit** inkl. der Rolle privater Unternehmen
2. Extended/Augmented Reality: besseres Lernen vs. Entfremdung als **Entpersonalisierung und Entkörperlichung (des Lernens)**
3. Digitales Enhancement von Emotionen: siehe 1) sowie **Privatheit persönlichster Daten**, **Natürlichkeit vs. Künstlichkeit/Authentizität**
4. Autonome Systeme zur Reparatur/Wiederherstellung vs. Vorsorge: **Symptom- oder Ursachenorientierung**?
5. Frontier-Ideologie der nützlichen Technisierung der Natur vs. Forderungen nach (partieller) **Unverfügbarkeit von Räumen**
6. **Dual use bzw. starke militärische Beteiligung/Förderung KI**

Im Ergebnis scheint die **(Un)Verfügbarkeit** mit Bezug auf Emotionen, Umwelt-Räume, Daten, Vernetzung und vieles mehr in unterschiedlichsten Facetten einer *der* zentralen ethisch weiter zu diskutierenden Punkte: Anscheinend bisher Unverfügbares wird empirisch verfügbar – soll es aber vielleicht nicht besser unverfügbar bleiben und von welchen Graden von Unverfügbarkeit/ Verfügbarkeit und verfügbar für wen ist die Rede? Dies im Detail zu analysieren, ist ein Forschungsdesiderat.

3.5.4 Kritische Reflexion der Implikationen des gewählten normativen Rahmens

Der normative Rahmen mit Nachhaltiger Entwicklung nach Brundtland (WCED 1987) und die umweltethisch relevante Erweiterung durch die Biodiversitätskonvention haben den Vorzug, dass sie eine völkerrechtlich breit anerkannte normative Basis liefern und zugleich internationale und nationale rechtliche Regelwerke prägen; dies gilt unbeschadet der Tatsache, dass die konkrete Interpretation der Gerechtigkeitsprinzipien im Einzelfall durchaus strittig sein kann. Politisch noch strittiger ist sicherlich die Frage, wie Gemeinwohl im Detail zu fassen ist, weil hier Grundfragen der politisch-ökonomischen Ordnung verhandelt werden. Gleichwohl ist die Orientierung am Gemeinwohl in Deutschland, und nicht nur dort, Basis der Rechtsordnung. Nicht im Fokus standen sicherlich kulturelle Spezifika der Frage nach gerechten Gesellschaften (Fokus aufs Individuum oder eher auf Gemeinschaft/Gruppen) sowie nicht-westlich geprägten Mensch-Natur-Verhältnissen.

Nicht aus Gründen der Wahl einer ethischen Theorie oder des politisch-normativen Rahmens, sondern aus pragmatisch-arbeitsökonomischen Gründen konnten wichtige Aspekte nicht genug in die Betrachtung einbezogen werden. Dies sind vor allem:

- Eine umfassende Analyse der Politischen Ökonomie der KI hinsichtlich der maßgeblichen Treiber und der damit verbundenen Interessengruppen (siehe dazu ausführlich (Nemitz und Pfeffer 2020)), wobei diese Dimension mit Bezug auf Gemeinwohlfragen gut bearbeitbar wäre.

- Eine ethisch-politische Tiefenanalyse der Institutionen und der Governance der KI, die genauer nach Machtstrukturen und Fragen gerechter politischer Verfahren und Teilhabe fragt.

- Eine umfassende Analyse der unmittelbaren Implikationen eines breiten KI-Einsatzes hinsichtlich des ökologischen Fußabdrucks, der THG-Bilanz und weiterer externer

(Umwelt)Kosten der Technologie und ihren Folgen (vgl. u. a. (Lange und Santarius 2020), (WBGU 2019b), sowie die thembezogene Webseite reset.org)

- Eine detaillierte Kritik an der Rhetorik und Begrifflichkeit der KI mit Blick auf wirtschaftliche und wissenschaftliche Vermarktungs- und Lobby-Interessen „Technically speaking, it would be more accurate to call Artificial Intelligence machine learning or computational statistics but these terms would have zero marketing appeal for companies, universities and the art market." ((Pasquinelli 2019), S. 4).

3.5.5 Fazit und Ausblick: Zentrale Entwicklungslinien der KI im Kontext NE und GW zwischen ethischer Analyse und Narrativen

Die Frage nach der Ethik, nach einer KI-Ethik oder Ähnlichem, ist selbst eingewoben in sinnstiftende Narrative der wissenschaftlich-technisch geprägten Welt. Unter einem Narrativ kann eine erzählende und dabei auch moralisch orientierende, vor allem aber sinnstiftende Struktur verstanden werden, die die Welt und darin bestimmte Entwicklungen zu verstehen und einordnen hilft. In den letzten Jahrzehnten wurde mit Bezug auf Fortschritte in Wissenschaft und Technik oft davon gesprochen, dass sich „ganz neue" ethische Fragen stellen, es mithin einer ganz neuen Ethik bedürfe. Auch die Entwicklungen der Digitalisierung und KI legen solche Narrativ-Momente nahe, denn eine radikal veränderte – gar: posthumane – *conditio humana* scheint eben auch eine neue Ethik zu verlangen.

Wir vertreten allerdings eine andere Position, die auf Erfahrungen einer Ethik in den Wissenschaften (Ammicht Quinn et al. 2015) aufbaut. Diese geht davon aus, dass sich in neuen gesellschaftlichen und technischen Konstellationen zwar grundlegende ethische Herausforderungen stellen und es dabei keine einfachen und auch keine bewährten Umgangsweisen für diese gibt. Es sind jedoch die Kontexte, die neu sind, nicht die Ethik mit ihrem Bestand an Prinzipien, Normen, Werten, Tugenden. Es braucht also keine ‚ganz neue' Ethik, um KI-Technologien ethisch zu analysieren, weil die anwendungsbezogene Ethik Methoden und Prinzipien bereithält, die auch zur Bewertung der Digitalisierung bzw. von KI-Technologien herangezogen werden können.

Wissenschaft und Gesellschaft stehen allerdings vor der Aufgabe, die bestehenden, grundlegenden ethischen Maßstäbe in einem neuen Handlungsfeld zu verorten und gleichzeitig bei der Erwägung ethischer Argumente der Geschwindigkeit der Änderungen in diesem neuen Handlungsfeld gerecht zu werden – zumal Durchbrüche im *Few* bzw. *One Shot Learning* und bei *Duelling Networks* die Entwicklungen im Bereich der KI-Technologien nochmals beschleunigen werden. Dies gilt umso mehr, als dass eine Ethik prospektiv und umfassender problembezogen statt reaktiv und technologieinduziert vorgehen sollte. Die Frage der Ethik an der und für die Grenze des technisch-gesellschaftlichen Fortschritts („*Frontier*") gehört ebenso hierher, wenn mit Hans Jonas eine „Ethik der technischen Zivilisation" (Jonas 1979) aufgerufen wird. Nicht die Ethik ist bei Hans Jonas jedoch neu, denn sie ist stark an Kants Kategorischem Imperativ und zugleich an Klugheitserwägungen (im Sinne eines Vorsorgeprinzips) sowie an Fragen des guten, „wahrhaft menschlichen" Lebens orientiert. Vielmehr ist es die technisch vermittelte Macht und das damit verbundene Gefahrenpotenzial planetaren Ausmaßes bzw. der biotechnischen Eingriffstiefe in Lebensprozesse und -strukturen selbst, die den neuen Kontext bilden. Damit sind die hier vorgestellten Überlegungen auch kompatibel mit aktuellen Narrationen zur Verantwortung im Anthropozän.

Für die ethische Analyse müssen in jedem Fall grundsätzliche gerechtigkeitstheoretische Erwägungen angestellt werden. Zugleich gilt die Berücksichtigung der in vielen ethischen

Debatten angesprochenen Besonderheiten der Abwägung bei Handlungen unter Unsicherheit. Für diese müssen Prinzipien wie das *Vorsorgeprinzip* herangezogen werden und im Speziellen auf KI-Technologien angepasst werden. Weiterhin gilt es, zum dritten, in verschiedenen Bereichen der anwendungsbezogenen Ethik diskutierte Themenfelder, wie z. B. Fragen des Konsums, medienethische Herausforderungen, intensiver Ressourcenabbau und Bildungsgerechtigkeit zusammenzuführen. Viertens gilt es schließlich, das Problem des Sich-Einlassens auf hypothetische oder gar kontrafaktische Szenarien und Narrative zu beachten. Eine Ethik, die sich zu sehr auf jedes spekulative „if – then" (Nordmann 2007) bezieht, vernachlässigt zweierlei: Sie übersieht die „normative Kraft des Fiktionalen" (Mieth 2002), also die Veränderung der Wirklichkeitswahrnehmung und -bewertung mittels „machtvoller" Narrative – oft erzählt von charismatischen „großen" Männern wie Elon Musk oder Stephen Hawking. Bei aller Antizipation (Mieth 2002) möglicher heute noch unrealistisch erscheinender Zukünfte muss eine anwendungsbezogene Ethik sich dagegen auf einen seriösen Stand des Wissens und der Technik und der Ungewissheit(en) beziehen, um angemessene Beurteilungen entwickeln zu können.

Die größten Widersprüche

Grundsätzlich finden sich, wie im gesamten Bereich von KI-Technologien, auch zum Thema **Affective Computing** zwei Standard-Narrative, ein utopisches und ein dystopisches. Auf der *utopischen* Seite (Entwicklung, Industrie) wird – wenig überraschend – der Mehrwert für die Nutzer*innen betont. Argumentiert wird hier vor allem so, dass es Menschen einfach glücklicher macht, wenn künstliche Agenten freundliche und respektvolle Gegenüber sind (Cowie 2015). Kommunikation mit einem Bot, so die Argumentation, wird einfacher, wenn sie uns mehr an zwischenmenschliche Kommunikation erinnert. Weniger anstrengende und enervierende Auseinandersetzungen mit virtuellen Agenten tragen dazu bei, dass wir unsere Ziele schneller und entspannter erreichen, was uns letztendlich im Alltag zufriedener macht. Geworben wird außerdem damit, dass AC Diskriminierungen aufgrund unterschiedlicher kultureller Praxen, die Emotionalität beeinflussen, minimieren kann, indem sie spezifisch auf diese Unterschiede reagiert. Hier stellt sich allerdings die Frage, inwiefern eine solche Spezifizierung nicht gerade dazu beiträgt, bestimmte Gruppenzugehörigkeiten noch weiter zu manifestieren.

Die *dystopische* Seite (populärwissenschaftliche Autor*innen ohne genaue Fachkenntnisse, z. B. aber nur teilweise ((Bostrom 2014), (Hawking 2018)) proklamiert den weiteren Verlust eines vermeintlichen Alleinstellungsmerkmals der Menschheit. Die emotionale Verfasstheit des Menschen und ihre besondere Verknüpfung mit unseren rationalen Fähigkeiten, die zu einer einmaligen Urteilsbildungs- und Reflexionsfähigkeit führt, steht, so die Befürchtung, dann nicht mehr nur Menschen, sondern auch nichtmenschlichen Wesen zur Verfügung. Wäre dies der Fall, so die weitere Argumentation, wären Maschinen Empathie- und am Ende leidensfähig, was als eine Voraussetzung dafür gilt in die Gruppe der moralisch zu berücksichtigenden Wesen aufgenommen zu werden und entsprechend Rechte in Anspruch nehmen zu können. Zwar verfügen auch andere Tiere über Emotionen, im Falle der zudem rational agierenden Maschinen wird die größere Ähnlichkeit allerdings zur größeren Bedrohung. Weiterhin stark betont wird die Gefahr, von „den Maschinen" nicht nur (emotional) abhängig zu werden – sondern ihnen letztendlich vollständig unterlegen zu sein (vgl. (Kehl und Coenen 2016), (Pasquinelli 2019)).

Die Studie zu **Autonomen Systemen zur Erschließung von für Menschen (bisher) unverfügbaren Räumen** spiegelt ebenfalls die widersprüchlichen Erzählungen wider, die sich in Hinblick auf KI-Technologien für quasi jedes Anwendungsfeld finden. Auch hier steht der Annahme, dass KI-Technologien den Status Quo massiv verbessern werden, die Annahme gegenüber, dass sie den Status Quo massiv verschlechtern werden. In dem konkreten Fall der Erschließung bisher unverfügbaren Räume zeigt sich dies in der Annahme einerseits, dass die

Aktivitäten, die mithilfe von KI-Technologien in der Tiefsee durchgeführt werden, zu einem besseren Verständnis der Tiefsee führen werden. Daraus leiten Befürworter*innen als Folge einen besseren Schutz der Ozeane und speziell der Tiefsee mitsamt ihren Bewohnern ab. Konträr dazu steht die Annahme, dass diese Aktivitäten zu einem intensiven Raubbau an der Tiefsee führen werden und die Tiefsee-Ökosysteme (mindestens für einen menschlichen Maßstab) irreversibel zerstören werden. Es zeigt sich entsprechend auch hier, dass die Technologien auf sehr unterschiedliche Art und Weise eingesetzt werden können. Ethisch evaluiert und verhandelt werden müssen folglich die menschlichen Zielsetzungen und Handlungen sowie politische Maßnahmen, weniger die Technologie als solche. Ebenso wichtig ist aber die bereits erwähnte Frage, wie mit dem Auseinanderklaffen von zunehmend verschwindender empirischer Unverfügbarkeit und der normativen Forderung, eben diese Unverfügbarkeit solcher Räume aus naturphilosophischen und umweltethischen Gründen erhalten zu sollen, umzugehen ist.

Offene Diskussionspunkte hinsichtlich Mensch-Technik-Umwelt sowie Ethik

Hinsichtlich möglicher Veränderungen der Mensch-Technik-Umwelt-Beziehungen lässt sich festhalten, dass diese eher allgemein bleiben müssen, wenn über KI-Technologien im Allgemeinen gesprochen wird. Aus diesem Grund sind die hier genannten Aspekte recht offen formuliert. Die beiden vertiefenden Betrachtungen des *Affective Computing* und der *Erschließung bisher unverfügbarer Räume* haben aber exemplarisch gezeigt, dass verschiedene Technologien manche Aspekte (z. B. Mensch-Technik im Bereich AC und Mensch-Umwelt im Bereich der bisher unverfügbaren Räume) besonders beeinflussen können, während andere kaum direkt angesprochen sind. Das gilt ungeachtet der ebenfalls oben skizzierten allgemeineren übergreifenden Aspekte.

Für „KI im Allgemeinen“ lässt sich in Bezug auf die Veränderung der *Mensch-Technik-Beziehung* ausmachen, dass sich das Kompetenzen-Spektrum von Menschen ändern wird, wenn und weil bestimmte bislang nötige Kompetenzen durch KI übernommen und andere gefördert werden. Ferner lässt sich ein Überschreiten der bisherigen Grenze zwischen Menschen als einzigen Gestaltenden der gestalteten Technik konstatieren, wenn aus der Technik selbst eine „kreative Kraft“ wird. In Bezug auf das *menschliche Zusammenleben* ermöglichen KI-Technologien eine stärkere Annährung an globale Gerechtigkeit, wenn sie so gestaltet und zur Verfügung gestellt werden, dass sie die Inklusion bisher benachteiligter Gruppen gewährleisten. *Mensch-Umwelt-Beziehungen* ändern sich hinsichtlich der realen und gedachten Unverfügbarkeit von Natur für Menschen, die noch weiter zum Verschwinden gebracht wird, was sich in entsprechende Anthropozän-Narrative einfügt. Es folgt außerdem eine Verschiebung der Verhältnisbestimmung von Virtualität und Realität des Mensch-Umwelt-Bezugs. Diese verschiebt sich jedoch nicht zwingend nur in die Richtung Entfremdung, sondern es könnten auch neue, zwar virtuell vermittelte, aber letztlich durchaus leiblich erfahrbare Bezüge ermöglicht werden. Gleichwohl bleibt klar, dass die zugrunde gelegte Trias selbst – also Mensch, Technik, Umwelt – in ihrer wechselseitigen Abgrenzung bedroht wird, wenn und weil das Menschliche und das Technische ebenso wie das Technische und die Umwelt ihre Abgrenzungsklarheit verlieren. In diesem Sinne würde die vertraute Einheit in der Verschiedenheit, also eine Welt, in der es identifizierbar und auch ontologisch unterschiedliche und unterscheidbare abgegrenzte Bereiche Mensch-Technik-Umwelt nicht mehr überall so klar geben. Ihre jeweiligen Konstellationen werden sich (weiter – das zeigen die Technikgeschichte und die Geschichte der Naturphilosophie –) verschieben. Dies ist zunächst ein anthropologischer sowie technik- und naturphilosophischer Befund, der dann ethisch zu bewerten ist.

All die eben genannten Aspekte müssen in Wissenschaft, Politik und der öffentlichen Debatte intensiv diskutiert werden. Für jeden einzelnen steht es aus, auszuloten, welche Mittel- und

Langzeitfolgen sie für verschiedene Gesellschaften mit sich bringen, welche Änderungen gewünscht und gewinnbringend sein werden/können und in welchen Zusammenhängen auf die Anwendung bestimmter KI-Technologien besser aus guten Gründen verzichtet werden sollte.

Points to Consider für eine nachhaltigkeitsethisch informierte KI-bezogene Ethik

Aus der ethischen und philosophischen Analyse von KI-Technologien und ihrer Auswirkungen auf Mensch-Technik-Umwelt-Beziehungen ergeben sich folgende *Points to Consider* für eine nachhaltigkeitsethisch informierte KI-bezogene Ethik:

- Große Möglichkeiten bietet die „Fähigkeit" der Technologien, sich selbst durch Lernen an veränderte Situationen und andere Umgebungen anzupassen. Dies ermöglicht u. a. kosteneffizientere und qualitativ hochwertigere Arbeit in vielen Bereichen. Für welche konkreten Arbeitsbereiche man davon ausgehen kann, dass KI-gestützte Technologien diese qualitativ hochwertiger leisten können als Menschen, ohne, dass dabei potenziell sinnstiftende Arbeit für den Menschen verloren geht, stellt eine offene Forschungsfrage dar.

- Kritische Fragen nach der Notwendigkeit der Ressourcenprospektion oder anderweitiger räumlicher Ausdehnung müssen vor dem Hintergrund der Suffizienz gestellt werden. Was ist materiell wirklich nötig für ein gelingendes Leben, was würde lediglich nichtnachhaltige Lebensweisen weiter stützen (Rebound-Effekte, aber auch weit darüber hinaus)? Die Rede von „körperloser KI" sollte nicht nur hier als unhaltbar verworfen werden, weil die Hardware Ressourcen benötigt, deren Abbau z.T. mit ökologischen Schäden und Menschenrechtsverletzungen einhergehen (z. B. Seltene Erden, Kobalt) und die Software sehr große Mengen an Energie benötigt. Es stellt sich hier abermals die Frage, wie Suffizienz-Maßnahmen politisch umgesetzt werden können, so dass nachhaltigere Lebensstile in westlichen Gesellschaften vermehrt Eingang finden – anstatt mithilfe von KI-Technologien weniger nachhaltige Lebensstile zu „reproduzieren" oder zu „perfektionieren" – und welche Rolle KI-Technologien dabei spielen können.

- KI-Technologien ergeben neue Möglichkeiten für Problemlösungen, gleichzeitig besteht in einigen dieser Fälle die Gefahr einen Techno-Fix zu reproduzieren, durch den wiederum die Gefahr entsteht, von anderen (z. B. Suffizienz- und Effizienz-) Maßnahmen abzulenken. Z. B. ist es ein großer Gewinn, wenn in gefährlichen Situationen und kontaminierten Umgebungen nicht mehr Menschen den Gefahren exponiert werden müssen, sondern Autonome Systeme dies übernehmen können. Gleichzeitig besteht die Gefahr, die aus z. B. ökologischer Perspektive dringend notwendige Vermeidung solcher Situationen weniger ernst zu nehmen. Ebenso ist die Möglichkeit, mit Hilfe Autonomer Systeme die Mülleinbringung in den Weltmeeren und anderen Gewässern stark zu reduzieren aus NE-Perspektive positiv zu bewerten, es besteht allerdings die Gefahr, dadurch lediglich die Symptome und nicht die Ursachen des Problems zu beseitigen.

- Bei aller Fragwürdigkeit der Parole: Eine „*KI made in Europe*" zu entwickeln wäre im Sinne Nachhaltiger Entwicklung, wenn die unterstellten „europäischen Werte" (keine reine Marktsteuerung à la USA, keine autoritäre staatliche Steuerung à la China), die implementiert werden sollen, stärker an intra- und intergenerationeller Gerechtigkeit ausgerichtet werden. (Es sei betont, dass auch in Europa Trends dahingehen, die Dominanz

marktliberalistischer oder/und populistisch-autoritär demokratiefeindlicher oder/und fremdenfeindlich-rassistischer Werte gegen Menschenrechte und Demokratie auszuspielen.) In diesem Kontext ist ferner zu fragen, inwiefern die Art und Weise der Entwicklung und Regulierung von KI solche Trends verstärkt oder ob sie auch dagegen einsetzbar wäre. Wie bereits jetzt anhand der Social Media zu beobachten, ist die Stellung von Umwelt-NGOs, Umweltbehörden, ökologischen Start-Ups, die Wahrnehmung und Rezeption von Umweltkatastrophen ausgesprochen ambivalent: Auch in demokratischen Staaten ist „anti-Umweltschutz"-Desinformation verstärkt zu beobachten, in autoritären Staaten spielen – vielleicht mit Ausnahme von China – die neuen Kommunikationstechnologien zum Teil eine subversiv-kritische Rolle. Doch jenseits der konkreten Wirkungen ist abermals die Frage aufgeworfen, welche *inhärente Struktur neue KI-Technologien* haben müssten, damit sie der Förderung von Fähigkeiten und Gemeinwohl eher dienen können. Hier sind sicher die weiter oben von (Hagendorff 2020) aufgeführten Fragen der Transparenz der Technikentwicklung mitentscheidend.

- Um intra- und intergenerationelle Gerechtigkeit tatsächlich zu erreichen, muss ein gleicher Zugang zu digitalen und KI-getriebenen Technologien für alle ermöglicht werden. Geschieht dies, und werden diskriminierende Biases in der Programmierung der KI-Technologien vermieden bzw. minimiert – und in jedem Falle transparent gemacht –, würde eine verstärkte Inklusion bislang benachteiligter sowie vulnerabler Gruppen ermöglicht. Dies wiederum fördert die Bildungsgerechtigkeit und gilt entsprechend als zentraler Aspekt von NE (s. auch SDG Nr. 4). Dabei bezieht sich der Aspekt des gerechten Zugangs aber auch auf die Entwicklung der Technologien selbst. Denn umfassende Teilhabe schließt die Teilhabe an Forschung, Entwicklung und ökonomischen Nutzen mit ein. Wie ein solch gleichberechtigter Zugang global geschaffen werden kann, sollte verstärkt inter- und transdisziplinär untersucht werden.

- Damit zusammenhängend muss der doppelte Bildungsauftrag ernst genommen werden: Es besteht sowohl die Notwendigkeit, die Weltbevölkerung (vgl. Kapitel 3.2, *social referent*) in die Lage zu versetzten, die Technologie nutzen, aber genauso muss die Fähigkeit und das Wissen vermittelt werden, um die Technologien hinterfragen zu können (vgl. Kapitel 3.2, Fähigkeitenansatz). Zudem besteht die Notwendigkeit, Entwickler*innen zu befähigen, gesellschaftliche und moralische Fragen der Technikentwicklung selbstkritisch reflektieren zu können sowie kommunikationsfähig zu werden, um in einen Austausch mit der Gesellschaft treten zu können.

- Hinsichtlich einer Bildung für Nachhaltige Entwicklung, insbesondere für die Natur- und Umweltbildung ist die Frage nach der Möglichkeit oder dem Verlust leiblich-sinnlicher Naturerfahrung und Praxis, die von Mensch zu Mensch vermittelt wird, von entscheidender kritischer Bedeutung. Die Gefahr der Zunahme weiter entkörperlichter Wissensformen und der Verlust der Montessorischen Einheit im Lernen mit *Kopf*, *Herz und Hand* bleibt aktuell. Es gilt weiterhin auszuloten, welche Verluste leiblich-sinnlicher Naturerfahrung von besonders großer Bedeutung für Psyche, aber auch „Geisteshaltung" in Bezug auf Natur der Menschen sind.

- Hinsichtlich des Inhalts und der Strukturen der öffentlichen Kommunikation werden durch den vermehrten Einsatz von KI-gestützten Technologien qualitative Änderungen erwartet bzw. z.T. bereits sichtbar (vgl. Kapitel 3.3). Diese führen zu Modifikationen von Verantwortungsrelationen, welchen bislang nicht die auf Grund ihrer gesellschaftlichen „Durchdringungstiefe“ notwendige Aufmerksamkeit in der wissenschaftlichen Debatte zukommt. Welche konkreten Änderungen in der öffentlichen Kommunikation entstehen, welche Verantwortungswahrnehmungen dadurch erodieren und wo eingelenkt werden sollte, sind diesbezüglich zentrale, offene Forschungsfragen.

Die vielleicht wichtigsten Themen einer KI-bezogenen Ethik, die auf Prinzipien der NE und des Gemeinwohls und damit Gerechtigkeitsprinzipien und dem guten Leben beruht, betreffen Bereiche, in denen ethische Fragen eng verbunden sind mit philosophisch-anthropologischen Fragen der conditio humana, der Mensch-Technik-Umweltbeziehung:

Unverfügbarkeit: Wie stark wird die lebensweltlich vertraute – aber nicht absolut zu denkende – Trennung von Mensch-Technik-Umwelt mittels KI unterlaufen/unterminiert, so dass sich neue anthropologische Orientierungsfragen nach dem spezifisch Menschlichen und seiner Mitwelt in der technischen Zivilisation stellen, wenn die Bereiche mehr und mehr zusammenfallen? Und – als davon zumindest analytisch zu trennende Frage – wo ist dies aus welchen Gründen (nicht) wünschenswert? Diese Frage lässt sich nicht beantworten, wenn allein die Folgen von KI betrachtet werden: Es muss stets der Kontext von Zielen (Problemstellung und Lösungsoptionen), von Mitteln und von möglichen Folgen und Nebenfolgen beachtet werden.

Bildung als/und kritisches Denken: KI-Anwendungen können den Lebenswelt- und Umweltbezug des Lernens radikal modifizieren und virtualisieren. Auch hier stellt sich die Frage danach, was tatsächlich verloren geht im rein technisch vermittelten Weltbezug und was sich quasi vollständig substituieren lässt? Die zweite Frage besteht dann darin, welche Konsequenzen dies für die Frage nach der Bestimmung von Mensch-Technik-Umwelt hat, siehe auch Punkt (a), vor allem aber in der Konsequenz sich Formen des kritischen Denkens und seiner Förderung durch Bildung verändern. Und drittens ist die ethische Frage zu stellen, welche Bildungsprozesse KI-vermittelt wünschenswert sind und welche nicht. Auch hier sind wiederum Ziele, Mittel und Folgen gesamtheitlich zu betrachten.

Es sei abschließend nochmals ausdrücklich betont, dass diese beiden Aspekte stets bezogen sind auf die normative Basis Nachhaltiger Entwicklung – und immer auch im Sinne des Gemeinwohls – also der *Gerechtigkeit* in *globaler intra- und intragenerationeller* Perspektive. Nicht zuletzt für das Umweltressort der Politik sind die beiden Punkte der (Un)Verfügbarkeit und der Bildung für Nachhaltige Entwicklung unter der Gerechtigkeitsperspektive zentrale Kriterien für die weitere Gestaltung von KI. Sie kommen im Punkt der Frage nach Grenzen der Virtualisierung und Grenzüberschreitung zwischen Menschen, Technik und Umwelt zusammen und schließen so auch an Konzepte einer Starken Nachhaltigkeit an, die die Substitution von Natur über Ressourcenfragen hinaus (Ott und Döring 2011) vielmehr auch als Frage der leiblichen Orientierung und Präsenz von Menschen in ihrer Mitwelt versteht, die sich nicht leicht durch KI ersetzen und virtualisieren lässt – und es nicht einfach überall lassen sollte, wo es denn funktioniert.

4 Anknüpfung in Innovationssystem: Akteure und Arenen

4.1 Ziele und konzeptioneller Ansatz

Ziel dieses Arbeitspaketes war die Identifizierung möglicher Anknüpfungspunkte für die Denklinien der nachhaltigkeitsethisch informierten KI-bezogenen Ethik innerhalb des vorhandenen Regime- und Institutionengeflechts. Indem die Operationalisierung von Beginn an mitgedacht wird, sollte erreicht werden, dass die Erkenntnisse in Agenda Setting-Prozesse Eingang finden können. Dazu wurden aktuelle Zeitungsartikel zu den beiden Vertiefungsthemen *Affective Computing und Autonome Systeme zur Erschließung von für den Menschen bislang unverfügbaren Räumen unter besonderer Berücksichtigung der Ozeane* zusammengetragen und auf Themen und Akteure hin analysiert. Ergänzend wurden wissenschaftliche Konferenzen und Reports herangezogen. Die Zwischenergebnisse wurden im Projektteam von Fraunhofer ISI, IZEW und Stefanie Saghri diskutiert und ein Entwurf dieses Dokuments wurde vom IZEW kommentiert. Auf dieser Basis werden schließlich Vorschläge für mögliche Verknüpfungsarenen für die Befunde aus diesem Projekt gemacht.

4.2 Affective Computing

4.2.1 Vorgehen

Ein wesentliches Element der Verknüpfungsanalyse war die Auswertung von Veröffentlichungen in den Medien aus dem vergangenen Jahr (07/2019-07/2020). Dabei wurden neben Affective Computing auch die Begriffe „*Emotional/Emotion AI"* sowie *„empathic media*" einbezogen, da sich herausstellte, dass sich diese Begriffe mittlerweile für die Anwendungen von Affective Computing etabliert haben. Auf diese Weise wurden 465 „News" identifiziert. Zusätzlich wurden 400 Veröffentlichungen zu einzelnen Personen (Biographien) identifiziert und die 30 mit der höchsten Trefferquote zu den Stichwörtern ausgewertet.

Das zweite Element war eine Analyse der Beiträge zu der 2019 abgehaltenen einschlägigen Konferenz „*International Conference on Affective Computing and Intelligent Interaction (ACII)*" mit 109 Beiträgen zu den aktuellsten Forschungsrichtungen.[87] Die Artikel und Konferenzbeiträge wurden auf Anwendungsbereiche und Ausrichtungen der Forschungsfrage sowie auf positive oder negative Bewertungen hin analysiert. Dabei wurde insbesondere die explizite Nennung von Nachhaltigkeitsaspekten untersucht, um mögliche Anknüpfungspunkte zum Diskurs um nachhaltige Entwicklung zu identifizieren. Zweitens wurde die Nennung von aktiven oder betroffenen Personengruppen erfasst und die Herkunft der aktiven Akteure aufgezeichnet, um beurteilen zu können, ob eine Anknüpfung auf nationaler Ebene in Deutschland sinnvoll ist.

4.2.2 Thematische Schwerpunkte

Grundsätzlich lassen sich im Affective Computing zwei Ausrichtungen unterscheiden (vgl. AP2 Screening). Die eine Gruppe zielt primär auf die Erkennung von menschlichen Gefühlen (Beispiel (Huang et al. 2019). Die zweite Gruppe baut auf diesen Erkenntnissen auf, um Maschinen zu programmieren, die emotionales Verhalten simulieren und mit Menschen auf emotionaler Basis interagieren (Beispiel (Ghandeharioun et al. 2019)). Die Mehrzahl der auf der ACCI vorgestellten Forschungen ist im ersteren Bereich angesiedelt, jedoch gut ein Drittel beschäftigt sich mit der

[87] https://ieeexplore.ieee.org/xpl/conhome/8911251/proceeding

Simulation und Beeinflussung von Emotionen, wofür die Erkenntnisse aus der Emotionserkennung in der Regel die Grundlage bilden.

Bei der Emotionserkennung geht es vor allem um die Verbesserung der Algorithmen für maschinelle Lernverfahren. Quellen sind primär Gesichtsaufnahmen und Sprache (gesprochen und Stimme) sowie Messungen der Hautleitfähigkeit (Beispiel (Thammasan et al. 2019)). Daneben werden aber andere Quellen erschlossen vor allem Gestik, Bewegungsmuster, Kopfhaltung, Schreibverhalten (Beispiel (Lopez-Carral et al. 2019)) und Display Wischmuster (Hashemian et al. 2019). Neuste Forschungen bemühen sich bei der Emotionserkennung den Kontext einzubeziehen z. B. die Bedeutung des zeitlichen Ablaufs von emotionalen Regungen (Dudzik et al. 2019; Spencer et al. 2019). Eine wichtige Richtung ist auch die Vervollständigung von Emotionsdatensätzen durch unüberwachtes Lernen sowie die maschinelle Unterstützung von Annotierungsprozessen durch menschliche Bewerter*innen (Heimerl et al. 2019).

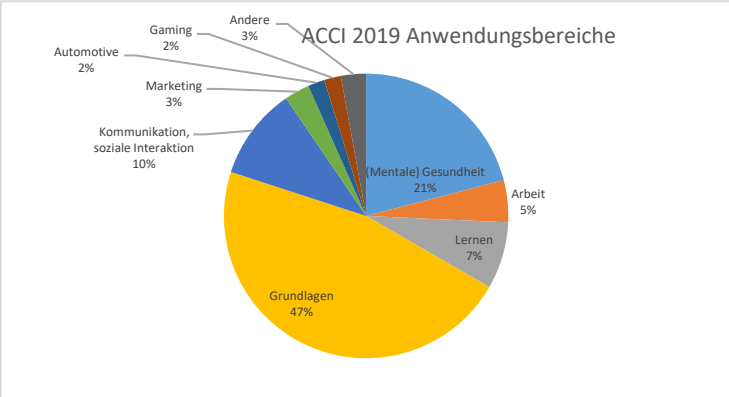
Die mit LexisNexis erfasste Berichterstattung in den Medien fokussiert sich auf Forschungs- und Entwicklungserfolge in den genannten Feldern sowie auf Beispiele für Anwendungen des Affective Computing aus den im folgenden Abschnitt genannten Bereichen. In nur drei Artikeln (z. B. (Lewis 2019)) findet eine kritische Auseinandersetzung mit gesellschaftlichen Folgen des Affective Computing statt.

4.2.3 Anwendungsfelder

Auf der ACCII als wissenschaftlicher Konferenz dominieren naturgemäß grundlagenorientierte Beiträge. Dabei werden jedoch des Öfteren Anwendungsbereiche genannt (vgl. Abbildung 5). Hier dominiert bei weitem der Gesundheitsbereich, gefolgt von Kommunikation und Interaktion (hier sind auch viele der Mensch-Robotik Interaktionsbeiträge verortet).

Abbildung 5: In den wissenschaftlichen Beiträgen der ACCI 2019 adressierte Anwendungsbereiche von Affective Computing

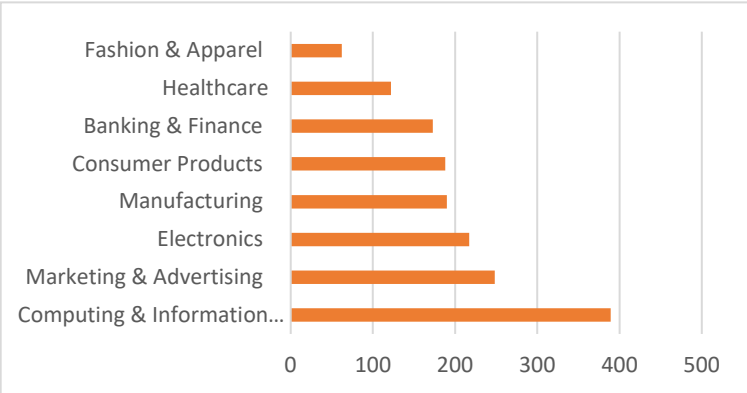
Stichprobe: n=109



Ein anderes Bild ergibt ein Blick auf die Branchenzuordnung der eher anwendungsorientierten Fachartikel aus LexisNexis.

Abbildung 6: Verteilung der LexisNexis Artikel (alle Typen) zum Thema Affective Computing nach Branchen

Stichprobe: n=900; Mehrfachzuordnungen von LexisNexis Artikeln möglich

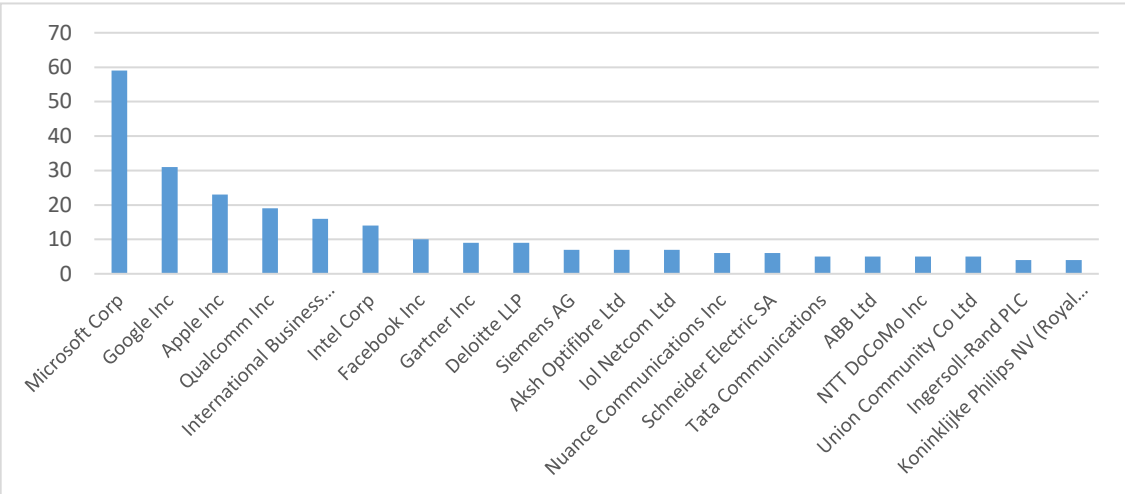


Hier wird die hohe Bedeutung für die Werbebranche deutlich. Dies spiegelt sich auch in den immer wieder genannten Hauptanwendern aus dem Bereich der Konsumprodukte wie Coca-Cola, Mars und Kellogs. Daneben ist auch der Finanzbereich stark präsent.

4.2.4 Akteure

In den Medienbeiträgen sind vor allem die Firmen präsent, die Emotion AI Software maßgeblich entwickeln und einsetzen. Dies ist beides der Fall bei den wichtigen Key Playern der Internetwirtschaft Google, Microsoft, Amazon, IBM, Apple und Facebook, die insbesondere seit ca. fünf Jahren allesamt entsprechende Firmen aufgekauft haben und deren Software in ihre Produkte integriert haben (z. B. in Alexa von Amazon). So ist etwa Microsoft an mehreren der Beiträge der ACCI 2019 beteiligt. Ein weiterer Großanwender, der auch selbst in die Entwicklung eingestiegen ist, ist die online-Lernfirma Pearson.

Abbildung 7: Anzahl von Nennungen von Firmen im Bereich Affective Computing in den LexisNexis Artikeln



Daneben ist eine Reihe kleinerer Firmen in der Entwicklung aktiv.

Die LexisNexis Artikel nennen folgende Akteure auf der **Anwenderseite von Affective Computing**: Banken, Versicherungen, Einzelhandel/Supermärkte, Werbebranche, Militär, Regierungen, Recruiter, Arbeitgeber, globale Bildungsindustrie, Medien, Unterhaltungsbranche, Therapiebranche, Coaching und Telekom-Firmen.

Betroffene Stakeholder kommen in den ausgewerteten Artikeln mit wenigen Ausnahmen nur als Nutznießer*innen der Anwendungen vor (siehe Tabelle 4). Ausnahmen sind eine Warnung vor negativen Auswirkungen auf Schüler*innen durch Überwachung und geringere Lernqualität und auf Bewerber*innen, deren Leistung automatisiert eingeschätzt wird sowie eine generelle Warnung vor den Auswirkungen auf die Gesellschaft.[88]

Tabelle 4: In den LexisNexis Artikeln genannte Nutzungsgruppen von Affective Computing

Menschen mit psychischen Krankheiten
Arbeitnehmer*innen
Menschen die sich nicht verbal äußern können
Schüler*innen
Bewerber*innen
Kinder
Kinder mit ADHS
Schulen
Eltern
Frauen
Charities (disability)
Universitäten
Kund*innen
Gamer
SocialMedia-Nutzer*innen

Die überwiegende Zahl der in den Quellen genannten Akteure stammt aus den USA, insbesondere das MIT Media Lab mit seiner Affective-Computing Gruppe unter Leitung von Rosalind Picard ist prominent aktiv. Daneben sind aber auch Akteure aus UK, den Niederlanden, Japan und Deutschland vertreten. Insbesondere Prof. André ist auch auf internationalem Parkett stark präsent.

[88] https://www.theguardian.com/technology/2019/aug/17/emotion-ai-artificial-intelligence-mood-realeyes-amazon-facebook-emotient

https://www.nzz.ch/meinung/digitale-ueberwachung-wem-gehoert-mein-gesicht-ld.1520534

https://www.faz.net/aktuell/karriere-hochschule/hoersaal/digitalisierung-der-bildungsferne-campus-16411188.html?printPagedArticle=true#pageIndex_2

Abbildung 8: Nationalität der Autor:innen zu Affective Computing auf der ACCI 2019

Stichprobe: n=109

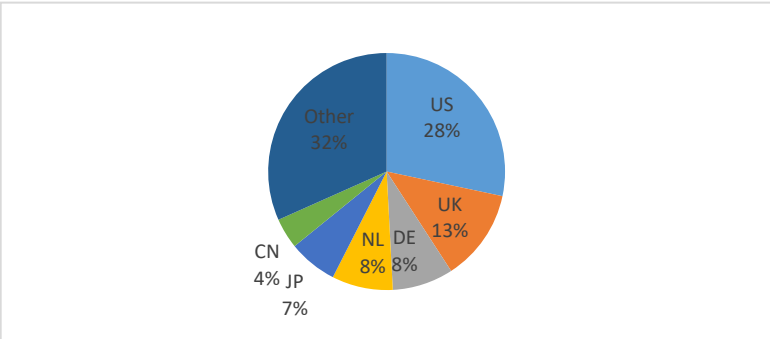
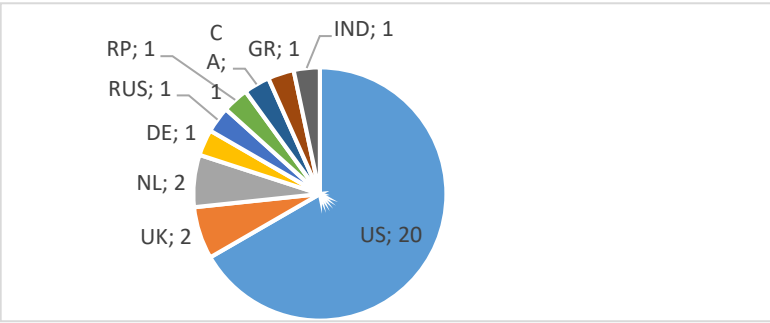


Abbildung 9: Nationalität der portraitierten Personen zu Affective Computing in LexisNexis Biographien

Stichprobe: n=30



Akteure in Deutschland

Kernakteure Forschung

- Universität Augsburg Institut für Informatik, Multi-Modale Mensch Technik Interaktion Prof André
  https://www.informatik.uni-augsburg.de/lehrstuehle/hcm/staff/andre/
- DFKI Saarbrücken Affective Computing Group Schneeberger, Gebhardt
  http://affective.dfki.de/?page_id=257
- Universität Hamburg Informatik Knowledge Technology, Barros
  https://www.inf.uni-hamburg.de/en/inst/ab/wtm/people/barros.html
- Fraunhofer IIS Erlangen Seuß
  https://www.iis.fraunhofer.de/en/ff/sse/imaging-and-analysis/ils/tech/shore-facedetection.html
- Nürnberg Institut für Marktentscheidungen Dieckmann
  https://www.nim.org/ueber-uns/team/dr-anja-dieckmann
- Universität Hamburg Psychologie Forschungsgruppe Veränderungsmechanismen in sozialen Interaktionen Prof. Lehmann-Willenbrock
  https://www.psy.uni-hamburg.de/forschung/interaktion.html

- Uni Saarbrücken Arbeits- und Organisationspsychologie Langer https://www.uni-saarland.de/lehrstuhl/koenig/personen/dr-markus-langer.html

Weitere Forschungsakteure

- Universität Augsburg Chair of Embedded Intelligence for Health Care Shuller, Zong https://www.uni-augsburg.de/en/fakultaet/fai/informatik/prof/eihw/team/
- Universität Magdeburg INSTITUT FÜR INFORMATIONS- UND KOMMUNIKATIONSTECHNIK Kognitive Systeme http://iikt.ovgu.de/KS.html

Entwicklungsfirmen

- Cognitec Systems GmbH https://www.cognitec.com/unser-unternehmen.html
- Charamel https://www.charamel.com/company
- TriCat https://www.tricat.net/unternehmen/
- SemVox https://www.semvox.de/

Anwender

- dSpace (Automotive) https://www.dspace.com/en/pub/home.cfm
- iav (Automotive) https://www.iav.com/en/services/core-expertise/ai-big-data/

Förderer/Vermittler

- BMBF (Mensch-Technik Kooperation) (PT VDI/VDE)

4.2.5 Anknüpfungspunkte

Von Seiten der Affective Computing Community gibt es innerhalb der analysierten Quellen praktisch keine unmittelbaren Anknüpfungspunkte zu Diskursen um Nachhaltige Entwicklung. In keinem der ausgewerteten Texte wurden Nachhaltigkeitsaspekte erwähnt, sieht man einmal ab von den postulierten Beiträgen zur Gesundheit, die als Bestandteil sozialer Nachhaltigkeit interpretiert werden können. Auch ethische Aspekte werden nur sehr vereinzelt erwähnt und beziehen sich dann fast immer auf den Schutz der sensiblen Daten, die im Zuge von Affective Computing Anwendungen erfasst werden. Eine Ausnahme ist die Forschungsgruppe „the emotional AI Lab“ von Prof. Andrew McStay von der Wales University, die sich explizit mit gesellschaftlichen und kulturellen Rahmenbedingungen von Emotional AI befasst[89] und etwa auf den Zusammenhang zwischen Affective Computing und der Wirksamkeit von gezielt personalisierten Fake News hinweist (Bakir und McStay 2018) sowie die grundsätzliche Problematik der auch anonymen Erfassung von Emotionen insbesondere mit der zunehmenden Erfassung des Kontextes thematisiert (McStay 2020; McStay und Urquhart 2019).

Obwohl also von der Affective Computing Community keine unmittelbare Anknüpfung vorliegt, können aus den Publikationen einige inhaltliche Anknüpfungsmöglichkeiten zu NE-Diskursen abgeleitet werden. Die Anknüpfung ist in beide Richtungen zu denken, also sowohl Beiträge von Affective Computing zu NE-Diskursen als auch umgekehrt.

[89] https://emotionalai.org/

Arena 1 Verhaltensbeeinflussung

Affective Computing zielt in hohem Maße auf die Beeinflussung von Verhaltensweisen über den Weg der Gefühle. Hier bestehen in mehrerer Hinsicht Anknüpfungspunkte zu aktuellen Nachhaltigkeitsstrategien:

Wie oben ausgeführt wird Affektive Computing derzeit stark im Werbebereich und im Marketing eingesetzt. Ziel ist es, über Verstehen und Ansprechen von Emotionen eine tiefere wertbasierte Bindung der Käufer*innen an die Produkte zu erreichen. Dies steht in klarer Wechselwirkung mit Bestrebungen nachhaltiges Konsumverhalten zu fördern (Kahlenborn et al. 2018). Dort heißt es auf S. 37:

> „Perspektivisch können mit der Verbreitung und Weiterentwicklung digitaler Assistenten, welche auf Kamera und Mikrophon des Nutzers zugreifen, Emotionen durch die Analyse von Mimik und Stimme situationsbezogen erfasst und für die Anpassung angezeigter Inhalte herangezogen werden. Die Anpassung von Werbebotschaften in Inhalt und Gestaltung an das individuelle Persönlichkeitsprofil ermöglicht es, stark emotionale Reaktionen und Bedürfnisse wesentlich gezielter als bislang hervorzurufen. Weil die Beeinflussung nicht beim Durchschnittskonsumenten, sondern beim Individuum ansetzt, diese Individualisierung wie auch deren Hintergründe aber nicht kenntlich gemacht werden, bekommt diese Werbung einen im Vergleich zur Werbung im analogen Raum noch stärker manipulativen Charakter. Dabei kann die Produktpräsentation nicht nur an positiv besetzte Wünsche einer jeden Person gekoppelt werden, etwa den Wunsch nach Freiheit oder familiärer Geborgenheit, sondern auch an persönliche Ängste oder Schwächen, welche für die Erzeugung neuer Konsumbedürfnisse, das heißt für die Problemerkennung, ausgenutzt werden."

Die wachsende Möglichkeit Kund*innen zu manipulieren führt nach Erkenntnissen dieser Studie tendenziell „zu einer Verstetigung nicht-nachhaltigen Kaufverhaltens" (ebd. S.42) etwa durch Zunahme von Fehlkäufen und Zunahme von Konsum insgesamt. Hier besteht also ein Anknüpfungspunkt zu Debatten über hochtechnisierte manipulative Werbestrategien. Gleichzeitig ist jedoch auch die Nutzung solcher Ansätze zur Beeinflussung in Richtung nachhaltigerer Konsum denkbar. So könnten etwa emotionssensitive Bots und Avatare genutzt werden, um nachhaltigere Verhaltensentscheidungen zu fördern. Hier besteht eine Anknüpfungsmöglichkeit zu Überlegungen aus dem Kontext des „Sustainability Nudging" (Mont et al. 2014), wobei sich die dort vorgenommen Abwägungen bei der Anwendung von Affective-Computing Ansätzen noch einmal verschärft darstellen dürften. Zumindestens jedoch können die Techniken zur Emotions*erkennung* möglicherweise dazu beitragen, auch die wichtigen emotionalen Aspekte nachhaltigeren Konsums besser zu verstehen. So hat etwa das Nürnberg Institut für Marktentscheidungen, das eine Abteilung Behavioural Science unterhält und im Bereich Affective Computing aktiv ist (Seuss et al. 2019), auch im Bereich ethischer Konsum geforscht (Frank et al. 2016). Hier könnte ein Anknüpfungspunkt für Debatten liegen.

Ein Teil der Emotion AI/Affective Computing Community widmet sich dem Design empathischer Technologien. Ziel ist es die „kalte, unmenschliche Technik" durch AI Komponenten dem Menschen zugänglicher und freundlicher zu machen (vgl. Kapitel 3.3 und (Pavliscak 2017)). Auch hier sind Anknüpfungspunkte zu Strategien nachhaltigeren Designs denkbar, etwa der Ansatz des „emotionally durable design", der auf längere Nutzungsdauer durch emotional Bindung an Produkte setzt (Chapman 2012).

Affective Computing wird auch zunehmend zur Gestaltung privater und öffentlicher Räume eingesetzt. So werden etwa großflächige Werbeflächen in Abhängigkeit von den Auswertungen dahinter verborgener Kameras mit Emotionsauswertung der Passant*innen geschaltet (Lewis

2019) und das Smart Home soll sich nicht nur nach den expliziten Befehlen, sondern auch nach den Gefühlen der Bewohner*innen richten. Die Gestaltung von Räumen, die nachhaltigere Verhaltensweisen unterstützen und fördern, ist ein zentrales Element der Nachhaltigkeitsforschung (zur entsprechenden Nudging-Problematik: vgl. Meisch et al. 2018). Entsprechend gibt es auch hier Spannungen und mögliche Befruchtungen mit den Affective Computing Ansätzen.

Arena 2: Bildung

Anwendungen in Lernen und Bildung sind ein Schwerpunkt des Affective Computing. Durch gezieltes Beobachten der Gefühle der Lernenden sollen Inhalte besser angepasst und so das Lernen verbessert werden (vgl. Kapitel 3.3). Hier besteht eine Anknüpfungsmöglichkeit zu Bildung allgemein sowie zur Bildung für Nachhaltige Entwicklung (BNE), denn auch innerhalb des Bildungssektors wird großer Wert auf emotionale Zugänge gelegt (Jung et al. 2015). Einerseits könnten diese Ansätze der emotional ausgerichteten Bildung und BNE geschwächt werden, wenn konsumorientierte Akteure Emotionen monopolisieren, anderseits können Bildung allgemein wie auch BNE im Speziellen vielleicht Ansätze des Lernens über Affective Computing nutzen. Diese Ansätze werden allerdings von einigen Bildungsakteuren sehr kritisch gesehen (Lankau 2019).

Arena 3: Menschliches Selbstverständnis

Die Überlegungen des WBGU zu Potenzialen und Risiken der Digitalisierung für die Stärkung der Eigenart des Menschen als Basis für die Nachhaltigkeitstransformation (WBGU 2019b) gelten für das Affective Computing in besonderem Maße. So verweisen etwa (McStay und Urquhart 2019) explizit auf Gefahren von Affective Computing für die menschliche Identitätsbildung. Damit besteht ein Anknüpfungspunkt an diese Debatten um KI und menschliches Selbstverständnis auf einer sehr grundsätzlichen Ebene. Schließlich gilt für Affective Computing ebenso wie für KI insgesamt die Notwendigkeit an Debatten zur globalen Gerechtigkeit anzuknüpfen (vgl. Kapitel 3.3), um einem umfassenden Verständnis von Menschenwürde gerecht zu werden.

Ein konkreter Ansatzpunkt für solche Überlegungen könnten aktuelle Aktivitäten im Bereich der Normung darstellen. Im Bereich der künstlichen Intelligenz bestehen eine Reihe von Aktivitäten zur Integration ethischer Aspekte in bestehende Standards so etwa die IEEE Initiative „*Open Community for Ethics in Autonomous and Intelligent Systems (OCEANIS)*“ an der auch der deutsche Verband VDE beteiligt ist und speziell das „*Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)*“.[90]

In dem IEEE Handbuch „*Ethically Aligned Design*“ (IEEE 2019) ist dem Affective Computing ein eigenes Kapitel gewidmet, an das die Überlegungen aus AP 3 sinnvoll anknüpfen könnten.

In zwei der in Entwicklung befindlichen Standards könnten die entwickelten Aspekte in ethische Überlegungen zum Umgang mit KI Eingang finden:

- IEEE P7014™ - IEEE Draft Standards Project for Ethical considerations in Emulated Empathy in Autonomous and Intelligent Systems (Entwurf für einen Standard zur Einbeziehung ethischer Aspekte in den Feldern „Emulated Empathy” und „Autonome Intelligente Systeme”)
  Diese Norm definiert ein Modell für ethische Überlegungen und Praktiken beim Entwurf, der Erstellung und dem Einsatz von empathischer Technologie. Dies umfasst Systeme mit der

[90] ethicsstandards.org

Fähigkeit affektive Zustände, wie Emotionen und kognitive Zustände, zu identifizieren, zu quantifizieren, darauf zu reagieren oder zu simulieren. Dies beinhaltet „affective computing", „emotion Artificial Intelligence" und verwandte Gebiete.

- IEEE P7008™ - IEEE Draft Standards Project for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems (Entwurf für einen Standard zur ethisch informierten Nudging für Robotik und Autonome Intelligente Systeme)
  Dieser Standard behandelt typische „Nudges" die derzeit verwendet werden oder erstellt werden könnten. Er umfasst Konzepte, Funktionen und Vorteile, die notwendig sind, um ethisch begründete Methoden für das Design robotergestützter, intelligenter und autonomer Systeme, die solche Nudges beinhalten, zu etablieren und sicherzustellen. „Nudges" durch robotische, intelligente oder autonome Systeme, sind definiert als offene oder versteckte Vorschläge oder Manipulationen, die das Verhalten oder die Emotionen eines Benutzers beeinflussen sollen.

Eine weitere Anknüpfungsmöglichkeit stellen die zahlreichen laufenden Prozesse zur Governance von KI etwa bei EU OECD und UN dar.[91]

4.3 Erschließung von für Menschen bislang unverfügbaren Räumen unter besonderer Berücksichtigung der Ozeane

4.3.1 Vorgehen

Zentraler Bestandteil des Vorgehens war eine Medienanalyse in der Datenbank LexisNexis mit den Stichworten „Deep Sea", „underwater" und „Artificial Intelligence". Einbezogen wurden Artikel von Oktober 2019 bis Oktober 2020. Von den gefundenen 1562 News und 581 Artikeln zu spezifischen Personen (Biographien) wurden 30 Artikel und 22 Biographien als besonders relevant angesehen und näher auf Themen und Stakeholder untersucht. Ausgehend von den einschlägigen Funden wurden weitere in den Artikeln referenzierte Quellen einbezogen.

4.3.2 Thematische Schwerpunkte

In den identifizierten Medien dominieren folgende Themenschwerpunkte:

- Bedrohung durch autonome Unterwasserwaffen und Wege damit umzugehen,
- Wandel der Kriegsführung zur See hin zu autonomen Unterwasser-Waffensystemen und die Rolle von Russland und China in diesem Wandel,
- Wirtschaftswachstum durch Erschließung der Tiefsee,
- Chancen autonomer Systeme in der Tiefsee für Naturschutz.

Folgende Forschungs- und Entwicklungsthemen wurden als aktuell prominent genannt:

- Verbesserung der Qualität der Unterwasser-Bildaufnahmen,

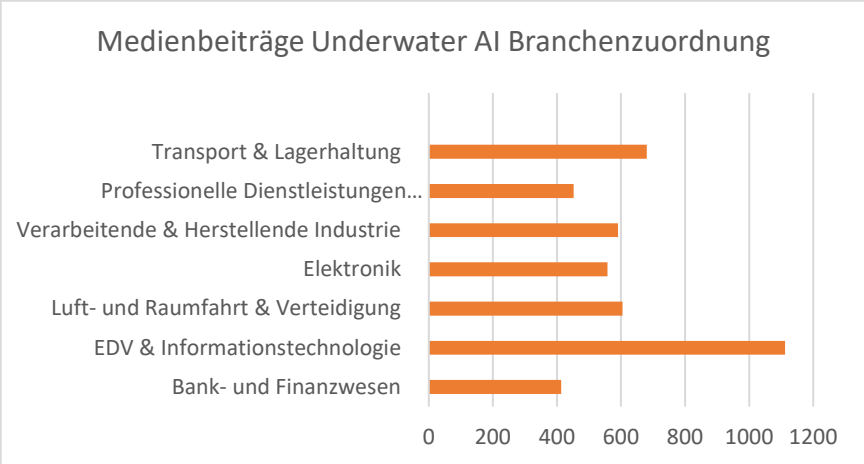
[91] z. B. EU General Data Protection Regulation 2016 (GDPR), EU ePrivacy Directive EPD/EPR, EU AI Strategy Process, OECD Global Partnership AI https://gpai.ai/, UN ITU AI for Global Good Process.

- Autonomie (Wahrnehmung der Umgebung, Erkennen von Objekten (Zielklassifizierung) und selbstständiges Überwinden von Hindernissen und Greifen von Objekten),
- Steuerung von Schwärmen unbemannter autonomer (Unterwasser)-Fahrzeuge (AUV),
- Modulares Design für verschiedene Einsätze (z. B. mit und ohne Waffe),
- Greifverfahren für maritime Kreaturen (z. B. durch Soft Robotics),
- Sammeln von Daten am Meeresboden,
- Sounderkennung im Rahmen von Naturschutzanwendungen etwa zum Tracking von Walen.

4.3.3 Anwendungsfelder

Abbildung 10: Zuordnung der LexisNexis Artikel zum Thema KI unter Wasser zu Branchen

Stichprobe: n=2055; Mehrfachzuordnungen der LexisNexis Artikel zu Branchen möglich



Fehler! Verweisquelle konnte nicht gefunden werden. zeigt die von LexisNexis vorgenommenen Zuordnung der Artikel zu Branchen, diese gibt aber ein nur unvollständiges Bild, da viele der Dienstleistungen und Informationstechnologie-Anwendungen weiteren Anwendungsbereichen zugeordnet werden können. Im Zentrum der Artikel stehen autonome Unterwasser Fahrzeuge (AUV), die KI dazu nutzen, in der Tiefsee zu navigieren und Entscheidungen zu treffen.

Eine manuelle Auswertung der einschlägigsten Artikel ergab folgende am häufigsten in den Medien genannten Anwendungsbereiche dieser Technologie:

Sicherheit und Verteidigung

Ausbau der Seeflotten in Russland, China, US und anderen Ländern mit KI gestützten autonom agierenden Unterwasser-Drohnen, die insbesondere Tiefsee-Gebiete abdecken können.

Infrastruktur

Überwachung und Wartung von Infrastrukturen in den Meeren wie Anlagen der Ölförderung und Offshore Wind Produktion. Insbesondere im Bereich der Offshore Windkraft wird starkes Wachstum erwartet, das auch die Anwendung von AUVs etwa bei der Verlegung der Kabel, aber auch der Überwachung der Anlagen antreiben wird.

Meeresboden-Monitoring

Monitoring des Meeresbodens für die Erschließung von Gebieten für Tiefseebergbau und zur Durchführung seismischer Studien zur Vorsorge vor Erdbeben und Tsunamis: KI unterstützt das autonome Aufsuchen von Standorten für Sensoren.

Ökosystem-Monitoring

Monitoring von maritimen Ökosystemen insbesondere Tracking von bedrohten Arten, aber auch Erkennung von Verschmutzungen wie Plastikmüll und illegalen Eingriffen sowie Risikoabschätzungen für ganze Bereiche: KI unterstützt die Erkennung von Arten und Interpretation von Bewegungen.

Freizeit- & Tiefsee-Archäologie

Aktivitäten aus dem Bereich des Tiefsee-Tourismus wie Erleben von unbekannten Tiefseewelten und Besichtigung von Schiffswracks: In letzterem Bereich kommen KI gestützte Überwachungstechnologien zum Einsatz, um Gefährdungen der unterirdischen Kulturschätze zu verhindern. Auch archäologische Operationen in der Tiefsee werden mit KI-gestützten Robotern durchgeführt.

Gefährliche Tätigkeiten

Übernahme von für Menschen gefährlichen Tätigkeiten z. B. Missionen in vermintem oder radioaktiv verstrahltem Gelände:

4.3.4 Akteure

Die in den Quellen genannten Akteure lassen sich entlang der Anwendungsbereiche aufführen. An erster Stelle stehen militärische Akteure, insbesondere der Marine. Da diese auch die Forschung und Entwicklung in hohem Maße finanzieren, dürften sie hohes Interesse mit hohem Einfluss verbinden und damit das Feld stark dominieren. Dahinter stehen in der Regel die jeweiligen Länder mit ihren Ministerien und Forschungseinrichtungen, die oft eng mit dem Militär verbunden sind. Auch viele der genannten Entwicklungsfirmen sind im Rüstungsbereich angesiedelt.

An erster Stelle bei den Nennungen stehen jedoch Google, Microsoft und IBM, die Software-Plattformen und IT Infrastrukturen für eine Vielzahl von Tiefsee-Projekten bereitstellen.

Nahezu ebenso präsent sind jedoch große Rüstungsfirmen wie BAE Systems (UK), Thales (FR), Lockheed Martin Corporation (US) und Boeing (Liquid Robotics) (US) in diesen Bereichen. Weitere Spezialfirmen sind Modus (UK), M Subs Ltd/ Submergence Group (UK), Rovco (UK), ThayerMahan (US), L3Harris Technologies (US), Huntington Ingalls (Hydroid Inc.) (US), TeledyneMarine (US), Unmanned Surface Vehicle (US) und Kraken Robotik GmbH (KRG) (DE/CA).

Unter den Forschungsakteuren sind ebenfalls das Office of Naval Research (US) sowie die NASA prominent vertreten. Bei den Forschungseinrichtungen ist vor allem das kalifornische Monterey Bay Aquarium Research Institute (MBARI) mit seinen Beiträgen zur Tiefsee-Erforschung in den Medien präsent. Weitere Erwähnungen finden das Ocean One Projekt des Stanford Robotic Labs sowie die mit dem MIT verbundene Woods Hole Oceanographic Institution.

Wie deutlich wird, dominieren in den Medien Akteure aus den US und UK in Forschung und Industrie. Im Zusammenhang mit Tiefseeaufrüstung werden auch Russland und China als Hauptakteure genannt, oft aus der Perspektive anderer weniger einflussreicher Küsten und Inselstaaten.

Tabelle 5: Akteure mit hohem Einfluss und hohem Interesse in der Unterwasserforschung und -nutzung

Militär/Marine (insbesondere US Navy und Russisches Militär)
Nationen (US, Russland, China, Singapur, Pakistan, Indien, UK ...)
Nutzende Industrien (Ölindustrie, Offshore Wind, Tiefsee Bergbau)
Entwicklungsfirmen von Tiefsee-KI
Forschungsakteure (auch Forschungsförderung)
Akteure Rechtssystem (z. B. International Seabed Authority ISA)
Umweltschutz Akteure/NGOs
Küstenstädte, Tiefseehäfen, Nutzer/Anrainer von Handelsrouten
Fischerei Akteure (inklusive Behörden)

Tabelle 6: Genannte Betroffene Akteure von der Unterwasserforschung und -nutzung

Maritime Ökosysteme (Leben im Meer, Korallenriffe, aussterbende Arten, Wale)
Kleinere Küstenstaaten insbesondere Inselstaaten
Photograph*innen, Filmemacher*innen
Meeresarchäologie-Akteure
Freizeit/Tourismus-Akteure
Akteure maritimer Sportarten
Arbeitslose durch Automatisierung

Akteure in Deutschland

In Deutschland beschäftigen sich die großen Forschungseinrichtungen aus verschiedenen Perspektiven mit der Erschließung von für Menschen bislang schwer zugänglichen Räumen und insbesondere der Tiefsee. Die Helmholtz Allianz „Robotische Exploration unter Extrembedingungen“ ROBEX zielt auf „die Nutzung von Synergien von zwei bislang unverbundenen Forschungsfeldern, die beide mit der Nutzung von Technologie unter extremsten Umweltbedingungen zu tun haben: die Tiefsee und der Mond.“[92] Zwei weitere Kernakteure sind das Helmholtz Zentrum für Ozeanforschung GEOMAR in Kiel und das Deutsche Zentrum für Künstliche Intelligenz DFKI in Bremen (beide Mitglied in ROBEX). Am DFKI ist die Abteilung Marine Sensing auf Unterwasser Sensorsysteme fokussiert, das GEOMAR betreibt mehrere Unterwasser Roboter, darunter ein autonomes Fahrzeug (ABYSS) sowie verschiedene autonome Messeinrichtungen.

Am Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB in Ilmenau und Karlsruhe widmet sich die Gruppe „Maritime Systeme/Oberflächenwasser“ unter anderem der Tiefseeerkundung. Das Forschungsteam hat ein autonomes Unterwasserfahrzeug entwickelt, das in großer Stückzahl gebaut werden soll (Deep Diving AUV for Exploration (DEDAVE)).

[92] https://www.dfki.de/web/forschung/projekte-publikationen/projekte-uebersicht/projekt/robex/

Schließlich beschäftigt sich am Institut für transformative Nachhaltigkeitsforschung IASS in Potsdam die Forschungsgruppe „Ocean Governance“ „mit der Frage, wie geeignete Steuerungsregeln und gesellschaftliche Normbildungsprozesse zu einem nachhaltigen Umgang mit den Ozeanen beitragen können“.[93] Hier ist auch das vom Umweltbundesamt finanzierte Projekt zu ökologischen Leitplanken für den Tiefseebergbau angesiedelt.

4.3.5 Anknüpfungspunkte

Ähnlich wie bei dem affektiven Computing zeigt sich auch hier, dass von Seiten der Forschungs- und Entwicklungscommunity selbst kein spezifischer Bedarf für Denklinien einer Ethik, die sich mit diesen KI-Ansätzen auseinandersetzt, artikuliert wird. Auch im Kontext der Debatte um die Erschließung der Tiefsee ist die Rolle von KI nicht thematisiert.[94] Andererseits finden in diesem Vertiefungsfeld im weiteren Umfeld bereits Debatten darüber statt, welche Werte und wessen Werte Eingang finden. Dies gilt für die Frage nach autonomen Waffensystemen ebenso wie für die wirtschaftliche Erschließung der Tiefsee. In beiden Bereichen existieren spezifische Aspekte, die mit dem Einsatz von künstlicher Intelligenz zusammenhängen und die somit Anknüpfungspunkte für die Denklinien einer Ethik, die sich mit diesen Technologien auseinandersetzt, bieten, welche innerhalb des vorhandenen Regime- und Institutionengeflechts ansetzt.

Aus der Verknüpfung mit den Befunden der Stakeholder Analyse und den Überlegungen aus AP3 ergeben sich damit zunächst folgende „Verknüpfungsarenen“:

- **Tiefsee Aufrüstung in der Abrüstungsdebatte:** In die Debatte um die Gefährlichkeit autonomer Waffensysteme[95] sollte das Augenmerk verstärkt auch auf autonome Waffensysteme in den Meeren gelenkt werden. Hier könnten die Denklinien einen Anstoß geben.

- **Rolle von KI in der Debatte um Tiefseeerschließung thematisieren:** Der spezifische Beitrag von künstlicher Intelligenz für die Erforschung und wirtschaftliche Erschließung des Ozeans könnte mithilfe der Denklinien beleuchtet werden. Diese Thematik ist in den aktuellen Debatten zu dem Thema praktisch nicht präsent. Hier bietet es sich an, die „Denklinien“ etwa in die aktuelle Debatte zur Errichtung von Schutzzonen in den Ozeanen einzuspeisen.[96] Dies ist insbesondere deshalb relevant, weil die Notwendigkeit dieser Erschließung mit Nachhaltigkeitsargumenten (Elektromobilität, erneuerbar Energie, etc.) begründet wird. Die in AP 3 betonten Aspekte der Suffizienz und der Rebound-Effekte könnten hier in den entsprechenden Debatten vertieft werden. Auch der Aspekt der globalen Gerechtigkeit, der in dieser Debatte eine zentrale Rolle spielt, ist damit angesprochen. Einen möglichen Rahmen für eine solche Einspeisung bietet die von der UN ausgerufenen Dekade der Ozeanerforschung. [97] Das Motto „The Science we need for the ocean we want“ lädt zu einer Debatte um die Verfügbarkeit und normative Orientierung geradezu ein. In der aktuellen „Preparatory Phase“ sind Beiträge besonders gefragt.

[93] https://www.iass-potsdam.de/de/forschung/governance-der-ozeane

[94] Beispiel Greenpeace Bericht zu Deep Sea Mining (Casson 2019.)

[95] vgl. etwa den TAB Bericht von 2020 (Grünwald & Kehl, 2020) und das zugehörige öffentliche Fachgespräch im Bundestag 2020 https://www.bundestag.de/dokumente/textarchiv/2020/kw45-pa-bildung-waffensysteme-798846.

[96] https://www.dosi-project.org/

[97] https://www.oceandecade.org/

- **Auf KI bezogene ethische Denklinien in die Umwelt- und Naturschutz- und NE-Debatten einspeisen:** Der obige Anknüpfungspunkt kann auf die gesamte Debatte um Nachhaltige Entwicklung hin erweitert werden, in der aktuell vor dem Hintergrund des Anthropozäns verschiedene Strömungen über neue Orientierungen diskutieren. Die Denklinien über den Beitrag von KI könnten hier eingespeist werden.
- **Rolle von Tiefsee und anderen unverfügbaren Räumen in ethische Fragen der KI einbringen:** Umgekehrt kann der Aspekt der Rolle bislang weitgehend unverfügbarer Räume für die menschliche Identität jedoch auch in die Debatten um KI-ethische Herausforderungen eingebracht werden. Die Frage nach dem Beitrag von KI zur Erschließung bislang weitgehend unverfügbarer Räume spielt dort bislang nur eine geringe Rolle und wenn dann eher mit Blick auf den Weltraum.
- **Erkenntnisse zur Tiefseeerschließung in Debatte um Umweltbildung einspeisen (und umgekehrt):** Die in der Umweltbildung geführte Debatte über die Rolle der (normativ verstandenen) Unverfügbarkeit natürlicher Räume ist in besonderem Maße auch für die Tiefsee relevant. Bei den Akteuren der Tiefsee-Erschließung ist das Narrativ verbreitet, dass eine bessere Kenntnis der Tiefsee auch eine größere Achtung mit sich bringt. Dagegen steht die Position, dass gerade die Unverfügbarkeit von Natur für menschliches Selbstverständnis kognitiv und emotional bedeutsam ist (vgl. Kapitel 3.4). Entsprechend gilt auch umgekehrt, dass die Erkenntnisse und kritischen Aspekte aus der Umweltbildung wertvolle Aspekte zu der Debatte um Tiefseeerschließung beitragen können.

4.4 Schlussfolgerungen: Anknüpfungs-Arenen für die Denklinien der KI-bezogenen ethischen Herausforderungen

Die beiden Bereiche *Affective Computing* und *Autonome Systeme für die Erschließung von für Menschen bislang unverfügbaren Räumen unter besonderer Berücksichtigung der Ozeane* weisen jeweils sehr spezifische Anknüpfungs-Arenen auf. Zugleich lassen sich einige übergreifende Aspekte mit konkreteren Unterthemen herausstellen:

1. Arena Bildung für Nachhaltige Entwicklung und nachhaltigeres Verhalten

- Künstliche Intelligenz als Mittel und Gegenstand von Bildung
- Manipulation durch Künstliche Intelligenz als Herausforderung für Bildung
- Erkenntnisse aus der Bildungsforschung für eine auf Künstliche Intelligenz bezogene Ethik

2. Arena Unverfügbare Räume im Anthropozän

- Rolle unverfügbarer Räume für die menschliche Identitätsbildung
- Rolle von Künstlicher Intelligenz bei der Erschließung unverfügbarer Räume (in und um den Menschen)
- Rolle von Künstlicher Intelligenz für den Naturschutz im Anthropozän

3. Arena Governance der Künstlichen Intelligenz (inclusive Ethik und Standardentwicklung)

- Spezifische ethische Herausforderungen von Affective Computing/Emotional Artificial Intelligence
- Rolle von Künstlicher Intelligenz bei der Erschließung von für Menschen bislang unverfügbaren Räumen und ethische Implikationen hinsichtlich normativ verstandener Unverfügbarkeit

5 Storyboards zum Einstieg in die Aushandlung neuer Transformationsnarrative

Arbeitspaket 5 hat im Gesamtprojekt die Aufgabe, die diskursive Auseinandersetzung mit KI-Entwicklungen und den dadurch ausgelösten grundlegenden Veränderungen der Mensch-Technik-Umwelt-Beziehungen, ethischer Aspekte und nachhaltiger Entwicklung über ein neues Erzählformat zu fördern. Das Erzählformat ist als Impuls für unterschiedliche Diskurse gedacht, ohne dass es den erforderlichen partizipativen Aushandlungsprozess vorwegnimmt. Inhaltlich wurden entsprechend der Schwerpunkte in der ethischen Analyse die beiden Themenkomplexe *Affective Computing in der Bildung* und *Autonome Systeme in bislang unverfügbaren Räumen am Beispiel der Tiefsee* adressiert.

5.1 Ziele und konzeptioneller Ansatz

Für die Kommunikation der Befunde aus dem Projekt wurde ein neuer Weg beschritten. Ziel war es, mit einer grafischen Repräsentation einen Einstieg in neue Transformationsnarrative zu schaffen. Hierzu sollen aktuelle Entwicklungen aufgegriffen, Dilemmata und Interessenkonflikte beleuchtet werden. Der Einstieg in neue Narrative lässt aber offen, wie die Erzählung weitergeht. Dafür wurden neue Erzählungsstränge formuliert und als Storyboards im Sinne von Abläufen bzw. Drehbüchern verfasst, visualisiert und animiert. Die Storyboards sollten Alltagsbezug haben und emotional packend sein, sie sollten allgemeinverständlich und für unterschiedliche Zielgruppen, insbesondere das Umweltbundesamt, umweltinteressierte Bürger*innen und Medienschaffende, einsetzbar sein. Einzelpersonen, die auf die UBA-Webseite gelangen, sollen weitergeleitet werden und sich die Geschichten in Ruhe ansehen können, um sich ein eigenes Urteil zu bilden. Es geht in den Erzählungen nicht um „richtig“ oder „falsch“ bzw. „positive“ oder „negative“ Zukünfte, sondern darum, sich zu vergegenwärtigen, welche Fragen auf die Menschen zukommen und wie diese Fragen verhandelt werden können.

Eine weitere Anforderung lautete, dass die Storyboards zu den ausgearbeiteten Vertiefungsfeldern textlich ansprechend dargestellt werden, um einen gewissen Tiefgang der Debatten und ihre Zusammenhänge über die ausgewählten Felder hinaus zu transportieren. Eine ästhetische, eventuell auch emotional ansprechende grafische Darstellung sollte die Texte ergänzen bzw. als grafische Repräsentation des jeweiligen Textteils dienen. Die Storyboards skizzieren die KI-ethischen Dilemmata und Polylemmata bzw. Probleme und Herausforderungen in verständlicher Form, betten diese in Lebenswelten ein und ermöglichen damit den Lesenden eine Grundorientierung bzw. Auseinandersetzung mit Themen einer gemeinwohlorientierten Digitalisierung. Damit gehen die Storyboards thematisch weit über die bislang dominierenden technikzentrierten Narrative zu KI hinaus.

Hintergrund des Einstiegs in „neue Erzählungen“

Von zunehmender Bedeutung auch in der KI-Debatte ist der Begriff „**Narrativ**“. Dabei handelt es sich nicht um eine plausible Argumentation, sondern um eine sinnstiftende Erzählfigur (mehr als eine Metapher, weniger als eine komplexe Geschichte), die dadurch Überzeugungskraft entfaltet. Es gibt vielfältige Detailbedeutungen und Funktionen von Narrativen. Nach (Espinosa et al. 2017) können Narrative a) Kommunikation ermöglichen, b) Bezugspunkte für soziale Akteure bieten, c) Aufzeigen, was getan werden soll, d) Werte verändern oder erhalten, e) Politische Allianzen und kollektives Handeln konfigurieren, f) Politische Positionen und strategische Legitimation produzieren. Nicht alles ist in einem Narrativ zugleich erfüllbar, und strittig ist auch, inwiefern sich Narrative – die immer auch normativ sind oder sogar

wünschbare Zukünfte enthalten – bewusst in ihrer Wirksamkeit planen lassen bzw. erst im Nachhinein klar wird, welche Narrative erfolgreich sind und warum.

Wir verstehen unter Narrativen hier einfach zu verstehende „Geschichten" aus der Alltagswelt, die existierende - unterschiedliche - Erzählstränge aufgreifen und diese auf unterschiedliche Art und Weise in die Zukunft weitererzählen. Wir verstehen darunter nicht (normative) Leitbild-gebende Erzählungen oder Narrative im wissenschaftlichen Sinne[98], sondern Geschichten, die zum Nachdenken anregen und unterschiedliche Perspektiven transportieren.

Die **Einstiege** in neue Erzählungen laden dazu ein, weitere mögliche Verläufe von Zukünften zu antizipieren, auch wenn normative „Ladungen" durch die gewählten Einstiege nicht zu vermeiden sind. Mit der Beschränkung auf den Einstieg in neue Transformationsnarrative wird den von (Espinosa et al. 2017) differenzierten Funktionen nur reduziert Rechnung getragen, was folgende Tabelle veranschaulicht:

Tabelle 7: Funktionen von Narrativen und Anspruch im Projekt „Digitalisierung und Gemeinwohl"

Funktion	Anspruch im Projekt
Kommunikation ermöglichen	Ja
Bezugspunkte für soziale Akteure bieten	Ja
Aufzeigen, was getan werden soll	Sehr eingeschränkt: vgl. auch Erfolgsfaktor „Offenheit und Mehrdeutigkeit" (s.u.)
Werte verändern oder erhalten	als Funktion: Nein, eingeschränkter Transport von Werten und Wertedebatten kann aber nicht ausgeschlossen werden
Politische Allianzen und kollektives Handeln konfigurieren	Nein
Politische Positionen und strategische Legitimation produzieren	Nein

Die Erzählungen in diesem Projekt sollen folglich nicht allen Funktionen von Narrativen gerecht werden, sondern einen Beitrag zu „Kommunikation ermöglichen" und „Bezugspunkte für soziale Akteure bieten" leisten. Das Projektteam verfügt nicht über die Legitimität zu sagen, was getan werden soll, weshalb der Einstieg in die Narrative nicht beabsichtigt, „Werte zu verändern oder zu erhalten". „Aufzeigen, was getan werden soll" können diese Erzählungen nur insofern, als dass die möglichen Konsequenzen verschiedener Handlungsoptionen plausibel erzählt werden. Dementsprechend liegen auch „Politische Allianzen und kollektives Handeln konfigurieren" und „Politische Positionen und strategische Legitimation produzieren" weit entfernt von dem hier verfolgten Einstieg in das Erzählen. Den Sinn der Formulierung eines erzählenden Einstiegs in neue Transformationsnarrative sehen wir darin, begründete Ausschnitte als Rohmaterial für Kontingenz und Gestaltbarkeit aus möglichst neutraler Sicht zu beleuchten, ohne dass dafür durch Stakeholder-Einbindung Legitimität geschaffen werden muss.

[98] Narrative werden in wissenschaftlichen Kontexten viel diskutiert, u. a. wurden die Debatten auf folgenden Konferenzen verfolgt: International Hybrid Conference "Narratives in times of radical transformations", TU Berlin, 19.-20.11.2020 und FEST Heidelberg, „Framing KI. Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz", 4.-5.12.2020.

Vorgehen bei der Erstellung

Um die Voraussetzungen für die Kommunikation mit einer breiteren Öffentlichkeit zu schaffen, und damit zukunftsorientierte Debatten über Digitalisierung/ Künstliche Intelligenz, nachhaltige Entwicklung und ethische Konsequenzen anzustoßen, ist in diesem Projekt mit neuen Erzählformen experimentiert worden. Am Anfang stand das Ausprobieren unterschiedlicher Erzählweisen auf der Basis des generierten Wissens und der normativen Debatten zu den beiden Themenkomplexen. Für die Ausarbeitung dieser beiden Einstiege in Erzählungen wurden zentrale Entwicklungen der Künstlichen Intelligenz und damit verbundene ethisch relevante Themen gesichtet und einige davon für die Storyboards ausgewählt. Zur Entwicklung des Einstiegs in Narrative und neue Erzählungen haben wir insbesondere Methoden der Entwicklung qualitativer Szenario-Storylines mit herangezogen.

Die Erstellung der Storyboards erfolgte in folgenden Schritten:

1. Bestandsaufnahme

Anhand der Leitfragen *Welche Erzählstränge um KI und Gemeinwohlorientierung gibt es?* und *Welche Erzählstränge fehlen?* wurde das bis dato im Projekt generierte Material aus AP2 (Screening), AP3 (Ethik) und AP4 (Anknüpfung) gesichtet, geordnet und reflektiert. Als Strukturierungshilfe diente eine Matrix mit den beiden Dimensionen inhaltliche Themen[99] (horizontal) und den Kategorien der Causal Layered Analysis[100] (vertikal) (Inayatullah und Milojevic 2015), die auch in der narrativen Szenariotechnik verwendet wird.

Tabelle 8: Beispielhafte Illustration der Strukturierungshilfe für die Befunde aus AP2 (Screening), AP3 (Ethik) und AP4 (Anknüpfung)

	Affective Computing	Autonome Systeme in bislang unverfügbaren Räumen	Big Data und Hyperkonnektivität
Fakten (Litanei)	KI kann x Gesichter in y Minuten erkennen	Für Menschen normalerweise unzugängliche Räume können erreicht werden	KI kann Daten in einer Geschwindigkeit identifizieren und kombinieren, wie es Menschen nicht können
Weltanschauungen	Müssen Menschen und Maschinen unterscheidbar sein?	Dürfen wir überall hin vordringen?	Privatheit von Daten
Metaphern	„E-lexa zieht ein“	Tief schürfen	Hyperkonnektivität und Big Data sind selbst Metaphern

[99] hier die drei Stränge (1) Affective Computing, inkl. Mensch-Maschine-Interaktion, (2) Autonome Systeme, insb. zur Erschließung von für Menschen bislang unverfügbaren Räumen sowie (3) Künstliche Intelligenz mit Schwerpunkt auf Machine Learning, Big Data Gesellschaft und Hyperkonnektivität

[100] Causal Layers sind, übersetzt nach (Inayatullah und Milojevic 2015):

Litanei - der Klang und die Wut, mediale Klangexzerpte, Klischees, Bilder, das Empirische, das Sichtbare und Aufscheinende.

Soziale und systemische Ursachen - in der Regel durch akademische Politikforschung moderiert, schafft ein rationales Verständnis von Themen.

Weltsicht/Episteme - das sind die auf der Zivilisation basierenden Annahmen, die von den Menschen selten in Frage gestellt werden, bis wir in andere Gemeinschaften/Kulturen reisen, seien es andere Länder, Forschungszentren, Dörfer, Unternehmen usw.

Mythos/ Metapher - das ist die Basis von Strukturen, die letztlich die intersubjektive Bedeutungsgebung und Identität des Selbst/ des Anderen, des Unbewussten des Universums, vermitteln.

In einem projektinternen Workshop wurden die Themenkandidaten gesammelt, offen diskutiert und auf einem Board dargestellt.

2. Schreiben der Erzählungen

Für die ausgewählten Themenstränge dieses Projektes wurden neue Erzählungen entworfen und in unterschiedlicher Form geschrieben. Wir berücksichtigten in der Ausarbeitung zwar die gesamten Themenkomplexe *Affective Computing in der Bildung* und *Autonome Systeme in bislang unverfügbaren Räumen am Beispiel der Tiefsee*, für die Storyboards wurden jedoch nur bestimmte Aspekte herausgegriffen, die gut kombinierbar oder einfach transportier- und darstellbar sind. Kriterium hierfür war, dass der Aspekt für eine breite Zielgruppe, interessierte Bürger*innen und nicht nur Fachleute, interessant sein dürfte und dass er darstellbar ist. Wie bei alternativen Szenarien[101] ist es möglich, Geschichten um eine Person herum zu erzählen, einen Dialog zwischen Personen zu erfinden, der die Inhalte transportiert, einen kurzen nüchternen, vielleicht sogar mit Fakten geladenen Report zu schreiben, eine Reise in die Zukunft zu unternehmen oder fiktive Alltagssituationen zu erzählen.[102]

Die Entwicklung der Erzählungsdynamik hatte verschiedene aufeinander aufbauende Elemente, von der fachlichen Einführung des Themas, über die Beleuchtung der Arena mit ihrem räumlichen Setting und Personen, der Entwicklung von Alltagsdialogen, die verschiedene Sichtweisen und Perspektiven in Bezug zueinander bringen, das Ausleuchten von Spannungsfeldern, Blockaden und Lösungswegen bis hin zu offenbleibenden Fragen. Wir nutzten dabei ein Framing[103], das die Notwendigkeit betonte, auch bei umwelt- und nachhaltigkeitsrelevanten Fragen ethische Aspekte/ Fragestellungen der KI prominent einzubeziehen. Wir haben jeweils die Perspektiven der Gesellschaft, des Alltags und der Umwelt berücksichtigt. Die Prototypen-Erzählungen wurden in Rohform einer Projekt-internen kritischen Betrachtung unterzogen und mehrfach angepasst. Danach wurde in Abstimmung mit dem Umweltbundesamt als Erzählform die nüchterne Schilderung von gewöhnlichen Situationen in der Zukunft ausgewählt. Trotz der Nüchternheit des Textes sollten die Lesenden durch die Entfaltung der Erzählung emotional angesprochen werden. Abweichend von diesem Bericht werden Genderaspekte aus Gründen der Einfachheit der Sprache in den beiden Geschichten wie von den Massenmedien Tagesschau und Süddeutsche Zeitung praktiziert gehandhabt.

3. Visualisierung

Aus den reinen Texten wurden danach Storyboards im Sinne von Abläufen und Sequenzen von Handlungen und Ereignissen entwickelt. Diese sollten den gleichen Charakter haben, wie die zuvor erstellten schriftlichen Erzählungen. Entsprechend der Erzählform fiel die Auswahl der Darstellung auf Bilder und die Überleitung des Textes in Form eines sogenannten „Scrollytelling“: Hierbei handelt es sich um Bilder mit Text, teilweise wie ein Comic mit Sprechblasen, bei dem die Lesenden selbst scrollen können und damit die Geschwindigkeit des Lesens und Ansehens selbst bestimmen. Die Illustratorin Stefanie Saghri hat die zwei Storyboards im Scrollytelling umgesetzt und programmieren lassen.

[101] Ein Beispiel zu KI-Szenarien sind die „Vier Zukunftsszenarien für Künstliche Intelligenz in der öffentlichen Verwaltung“ (Opiela et al. 2018).

[102] siehe z. B. die „Geschichten aus der Zukunft“ des zweiten BMBF Foresights (Zweck et al. 2015).

[103] Das Framing selbst erfordert möglicherweise einen kompletten Perspektivwechsel - die Perspektive, aus der formuliert werden soll, ist die Gesellschaft, die einen Blick auf Technik und ihre Einbettung wirft.

Die Erzählungen selbst spielen in der Zukunft, daher werden am Ende der Storyboards wenige Bilder bereits existierender KI-Anwendungen gezeigt, um zu verdeutlichen, dass Ansätze dieser Entwicklung durchaus schon zu beobachten sind. Abbinder der neuen Erzählungen sind Fragen, die Lesende zum weiteren Denken anregen sollen.

Zum Testen wurden die fertigen Geschichten Personen aus dem Konsortium, dem wissenschaftlichen Beirat des UBA-KI-Projektes, einer Gruppe Wissenschaftler*innen aus dem Fraunhofer ISI, Schüler*innen sowie dem Publikum einer Veranstaltung zu KI in Rottweil vorgelegt. Die dabei erhaltenen Hinweise befinden sich in Kapitel 5.3.

5.2 Die neuen Erzählungen aus der Zukunft

Die beiden Storyboards *Affective Computing in der Bildung* und *Autonome Systeme in bislang unverfügbaren Räumen am Beispiel der Tiefsee* führen zunächst das Fachthema ein und entwickeln dann eine alltagsnahe Geschichte. Am Ende stehen heute schon existierende reale Beispiele für die thematisierten Anwendungen Künstlicher Intelligenz sowie offene Fragen, die zur Reflexion und Diskussion einladen, ohne dabei bereits eine normative Position zu beziehen.

Die Storyboards liegen als Scrollytelling umgesetzt vor. Das Scrollytelling erlaubt es, die Storyboards in der eigenen Geschwindigkeit am Bildschirm zu rezipieren.

Die beiden Geschichten können als Scrollytelling im Internet ab voraussichtlich Januar 2022 im eigenen Tempo angesehen werden. Sie liegen auf einem Webserver der Fraunhofer Gesellschaft bereit. Zum Anschauen, klicken Sie bitte hier:

www.uba-ki-storyboard.de

Im Folgenden werden Ausschnitte der beiden Geschichten gezeigt, um einen Eindruck von ihren Inhalten und den Darstellungsformen zu vermitteln.

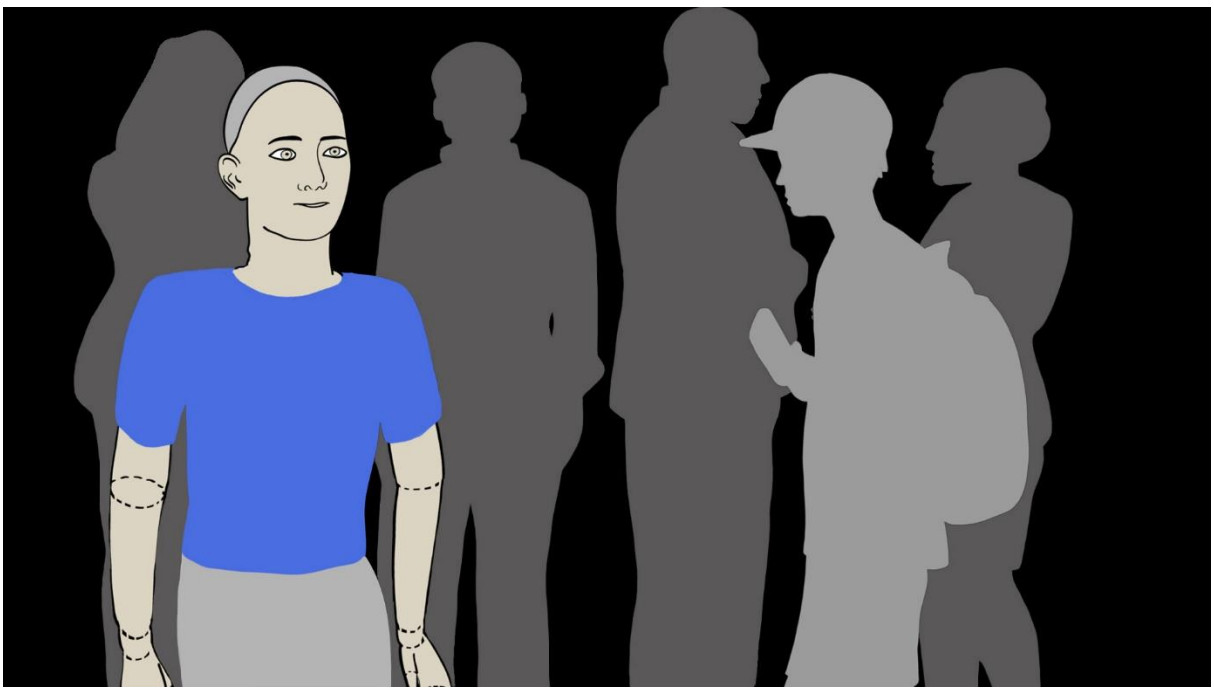
5.2.1 Kim will lehren

In der ersten Erzählung zum Affective Computing geht es um die Humanoidisierung von Robotern im Schulunterricht. Für die Unterstützung von Lehrenden im Schulalltag an einer Modellschule, wird der Lehroid „Kim“ (Abbildung 11) angeschafft. Kim sieht wie ein Mensch aus, ist aber als humanoider Roboter erkennbar. Er ist so groß wie die meisten Schüler*innen, um auf Augenhöhe zu kommunizieren (Abbildung 12). Als akuter Lehrermangel eintritt, bietet Kim an, den Unterricht zu übernehmen. In der Erzählung sehen wir zunächst die Reaktionen von Rektor und Lehrkräften. Danach folgen die ersten Versuche allein zu unterrichten, wobei sich der Lehroid anthropomorpher Interaktionsmuster bedient und in einigen Situationen wie ein Mensch Emotionen hervorruft. Kim lernt jeden Tag dazu und kann Emotionen erkennen (Abbildung 13). Die Schüler*innen mögen ihn, weil er nicht unfair ist (Abbildung 14) und auch nie die Geduld verliert. Kim ist mit dem weltweiten Wissen verbunden (Abbildung 15). Als noch mehr Lehrende ausfallen, wird ein „Kollege“ von Kim hinzugezogen, der die Basisdaten zwar übernehmen kann, danach aber selbständig lernt.

Abbildung 11: Kim, ein Lehroid

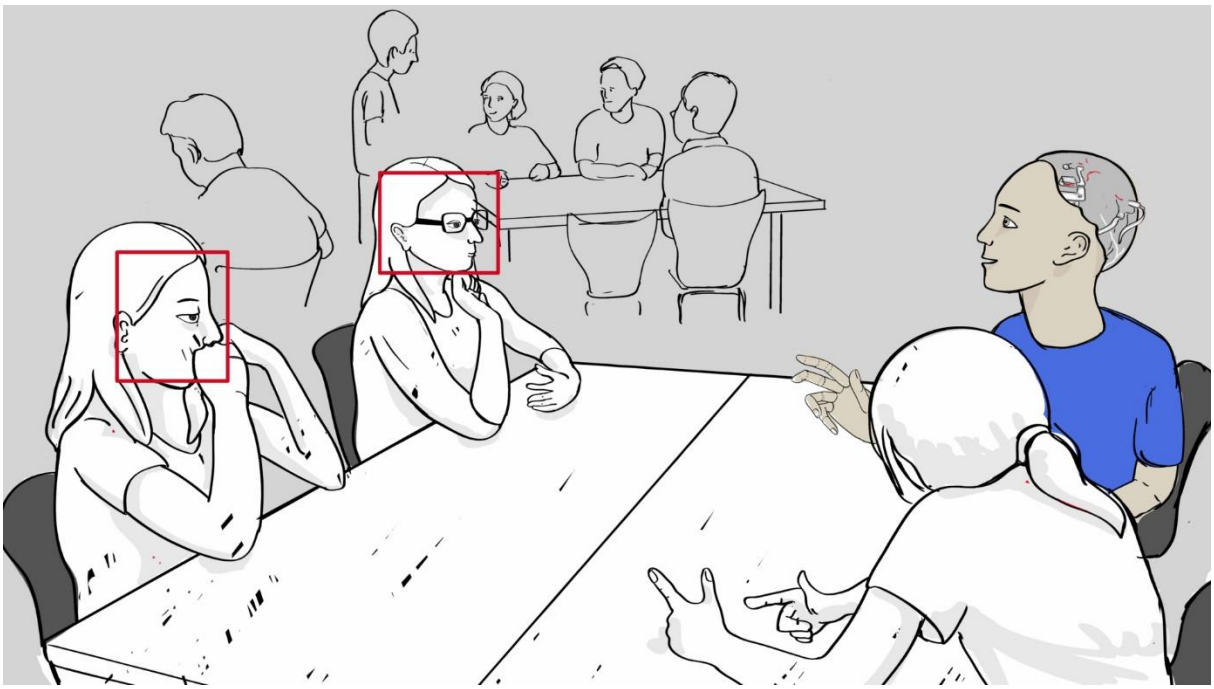
Quelle: Stefanie Saghri und Fraunhofer ISI

Abbildung 12: Kim auf dem Schulhof



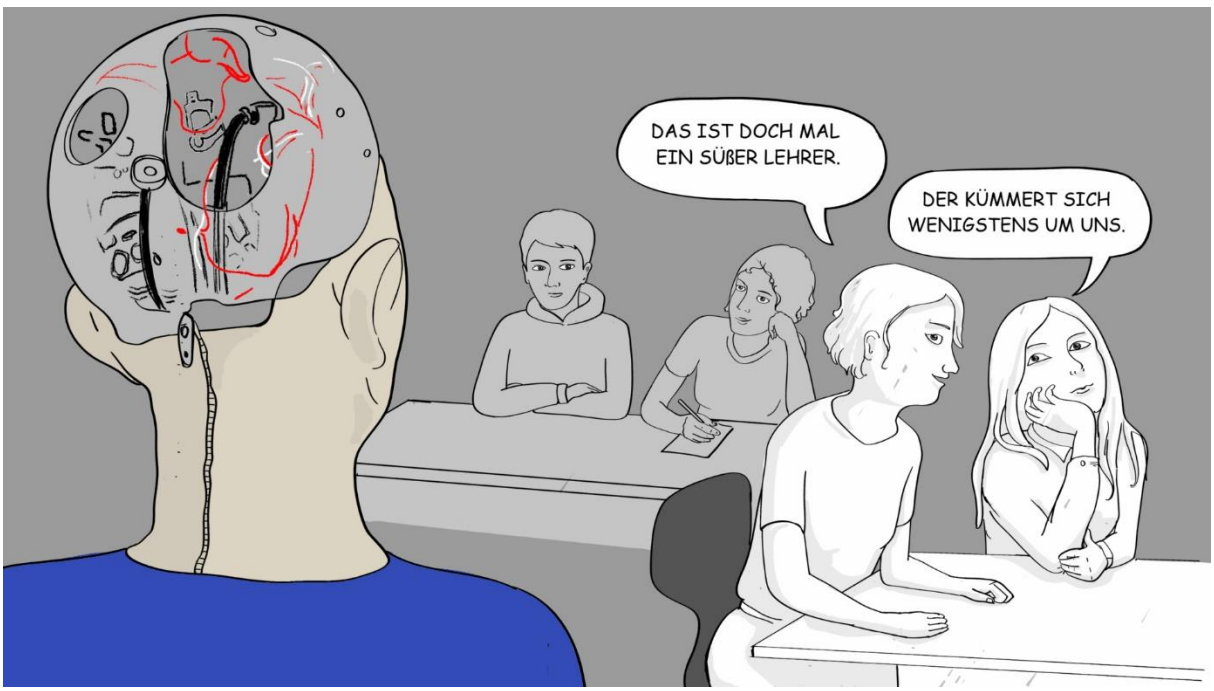
Quelle: Stefanie Saghri und Fraunhofer ISI

Abbildung 13: Kim erkennt Emotionen



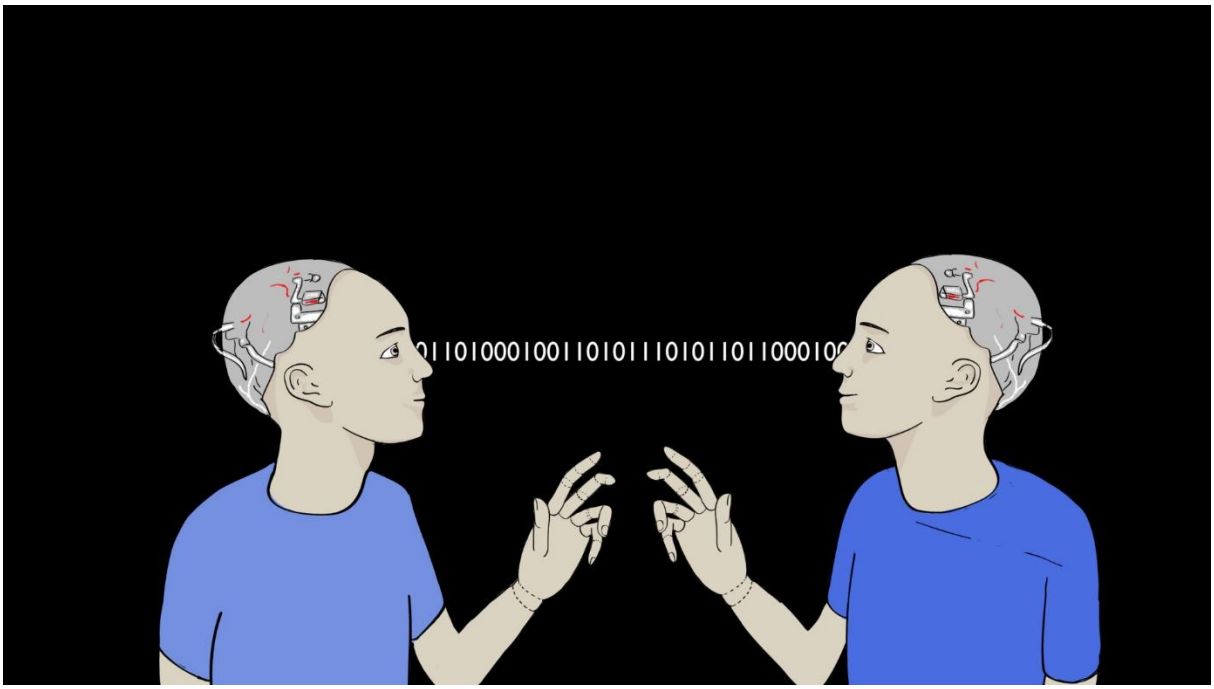
Quelle: Stefanie Saghri und Fraunhofer ISI

Abbildung 14: Kim wird angehimmelt



Quelle: Stefanie Saghri und Fraunhofer ISI

Abbildung 15: Kim überträgt seine Daten auf einen Lehroid-Kollegen



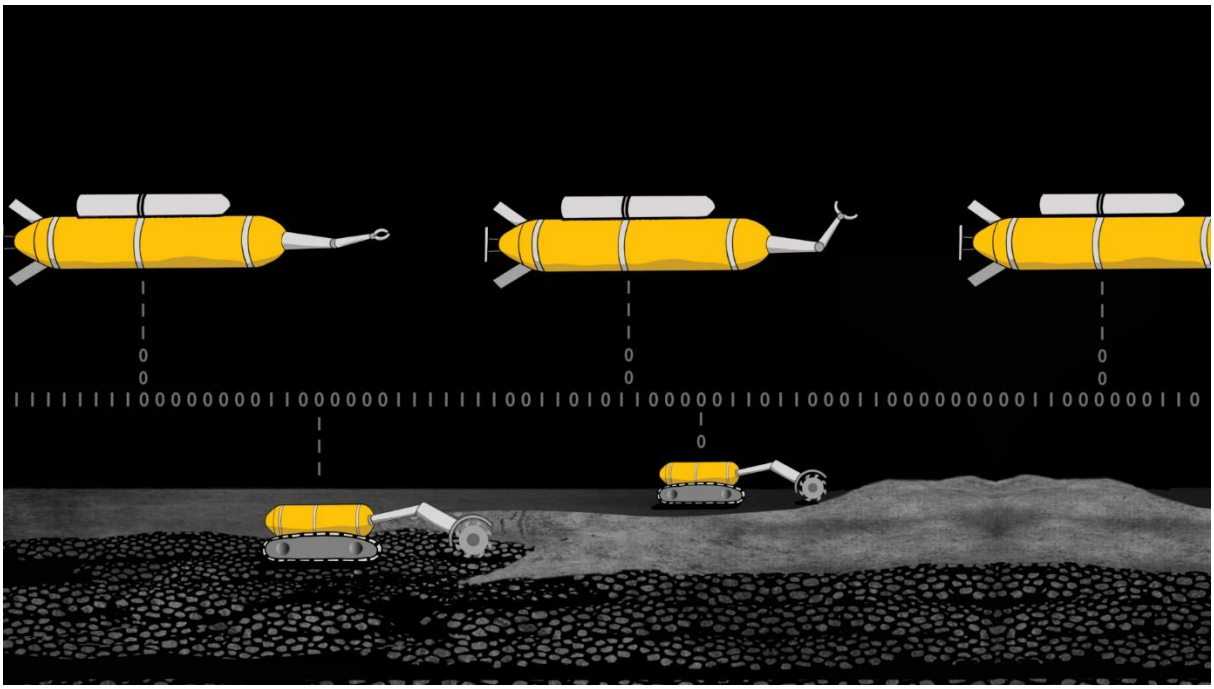
Quelle: Stefanie Saghri und Fraunhofer ISI

Den weiteren Verlauf der Erzählung und die damit verbunden ethischen Aspekte sehen Sie bitte online an.

5.2.2 Autonome Systeme in bislang unverfügbaren Räumen – in der Tiefsee

Die zweite Erzählung hat „Autonome Systeme in bislang unverfügbaren Räumen“ zum Thema. Diese Geschichte spielt in einem Kulturzentrum, wo eine Vortrags- und Diskussionsveranstaltung stattfindet. Die Veranstaltung beginnt mit schönen Bildern aus der Tiefsee (Abbildung 20) und von neuen Unterwasserfahrzeugen (Abbildung 17). Schrittweise entwickeln sich heiße Debatten beispielsweise über die Rolle der Forschung, KI-Unternehmen und Militär oder die Finanzierung von Projekten beim längst im Gang befindlichen Einsatz autonomer Unterwassersysteme sowie Debatten über Gerechtigkeitsfragen (Abbildung 18). Die Geschichte thematisiert die Bedeutung unverfügbarer Räume und weist über die Erschließung der Tiefsee durch Autonome Systeme hinaus auf weitere Ausweitungen der menschlichen Grenzen (*frontier)* und ihre Bedeutung für Nachhaltige Entwicklung hin (Abbildung 19, Abbildung 20).

Abbildung 16: Autonome Unterwasservehikel erkunden die Tiefsee und bauen dort Rohstoffe ab

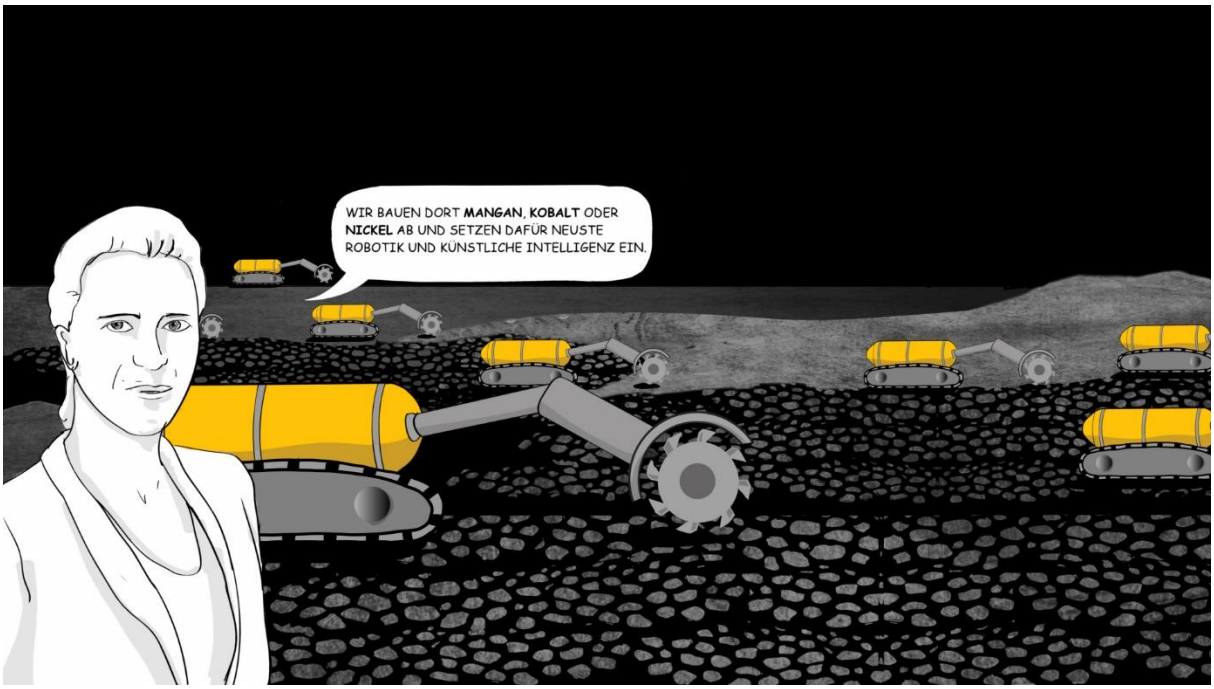


Quelle: Stefanie Saghri und Fraunhofer ISI

Abbildung 17: Was ist uns die Erkundung der Tiefsee Wert?

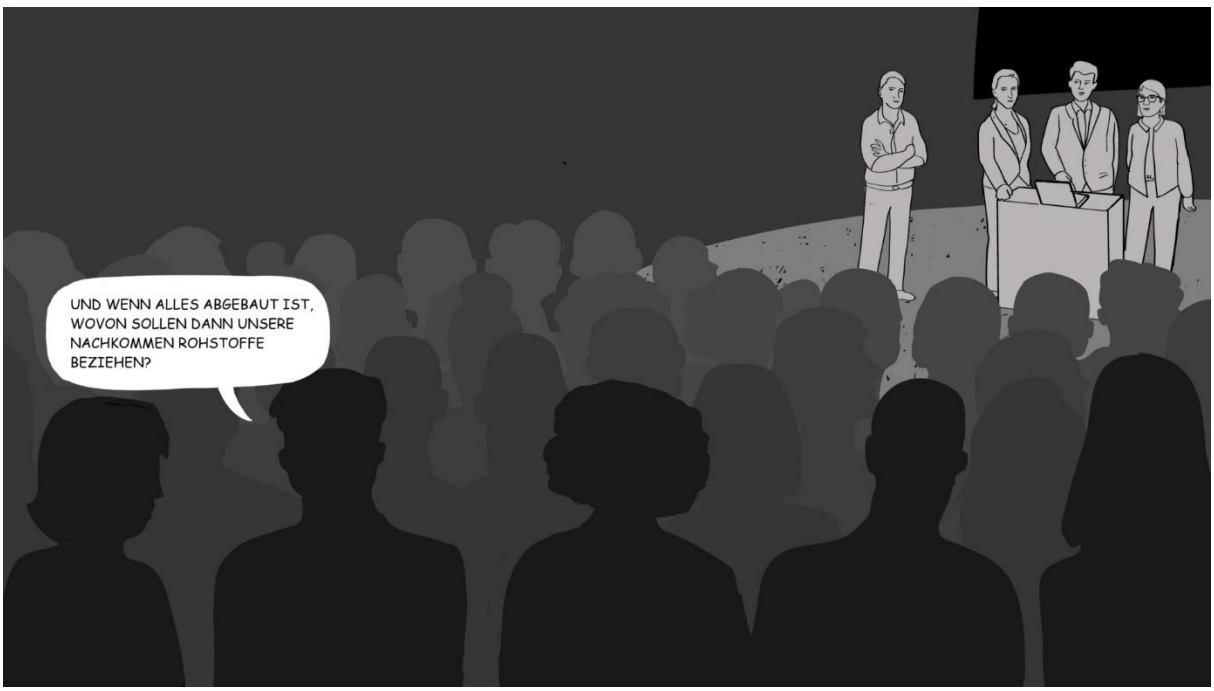
Quelle: Stefanie Saghri und Fraunhofer ISI

Abbildung 18: Robotik und KI im Tiefseebergbau



Quelle: Stefanie Saghri und Fraunhofer ISI

Abbildung 19: Generationengerechtigkeit beim Rohstoffabbau



Quelle: Stefanie Saghri und Fraunhofer ISI

Abbildung 20: Zuerst die Tiefsee - dann das Weltall?

Quelle: Stefanie Saghri und Fraunhofer ISI

Den weiteren Verlauf der Erzählung und die damit verbunden ethischen Aspekte sehen Sie bitte online an.

5.3 Schlussfolgerungen: Erste Rezeptionserfahrungen

Mit der Entwicklung von Storyboards wurde ein medialer Einstieg in die Schaffung neuer Erzählungen für gesellschaftliche Transformationen zur Verfügung gestellt, der neben der Fachwelt auch interessierte Laien dazu anregen soll, sich mit normativen Fragen der KI zu befassen. Die entstandenen Scrollytelling-Erzählungen sind für die Rezeption der Storyboards am eigenen Bildschirm gedacht. Im Projektverlauf wurde das Scrollytelling versuchsweise bereits (teilweise in abgefilmter Form) verschiedenen Zielgruppen präsentiert, darunter Jugendliche, Mitarbeiter*innen am Competence Center Foresight des Fraunhofer ISI, den Beiratsmitgliedern des Projektes und einzelne Ausschnitte auch anlässlich einer gemeinsamen Veranstaltung des Fritz-Erler-Forums Baden-Württemberg im Kooperation mit dem Forum Soziale Technikgestaltung und der Evangelischen Kirchengemeinde Rottweil auch dezidiert einem Publikum, in dem sich überwiegend interessierte Laien befanden. Jede Rezeption der Storyboards erfolgte bislang in einem bestimmten Kontext, variierend hinsichtlich des gesellschaftlichen Klimas, der örtlichen Umstände, anwesender Personen und so weiter.

Folgende Einsichten können aus den verschiedenen Rezeptionskontexten extrahiert werden:

- Das Fachpublikum (Beirat, Fraunhofer ISI) konnte die Storyboards gut verstehen und kommentieren. Die anvisierten Wirkungen hinsichtlich Inspiration und Öffnung von Debatten wurde in dieser Zielgruppe gut erfüllt.
- In medialer Hinsicht kann vermerkt werden, dass es den meisten Menschen Spaß macht, die Storyboards anzuschauen (Beirat, Fraunhofer ISI, Jugendliche). Manche wünschten sich zusätzlich eine Tonspur (was beim Scrollytelling nicht sinnvoll ist, aber in abgefilmter Form

möglich wäre), andere betonten, dass die Storyboards mit Social Media Teasern hohe Aufmerksamkeit erzielen könnten.

- Interessierte Laien haben teilweise einen sehr unterschiedlichen Kenntnisstand hinsichtlich der Digitalisierung (von Excel über Smart Phone bis hin zum KI-Bibliothek-Nutzenden). Für Menschen mit geringerem Kenntnisstand wird empfohlen, der Präsentation der Storyboards eine ausführliche Einführung zu Digitalisierung und KI sowie zu den verschiedenen Akteursgruppen mit ihren jeweiligen Interessen voranzustellen.
- Interessierte Laien sind den populärwissenschaftlich vertretenen Entwicklungen hin zu einer „Superintelligenz“ so medial ausgesetzt, dass sie leicht Ängste entwickeln. Das Voranschreiten Chinas bei der Entwicklung von KI und die Kolportation immer gleicher Stereotype darüber in den Medien erzeugt bei vielen Menschen Ohnmachtsgefühle und es ist möglich, dass in bestimmten Veranstaltungskontexten eine allgemeine Wissenschaftsskepsis vorherrschen kann. Diesbezüglich wird empfohlen, den Standpunkt des Projektes und sein aufklärerisches, gemeinwohlorientiertes Interesse sehr deutlich herauszustellen.

Abschließend lässt sich festhalten, dass die Präsentation der Storyboards nicht in jedem Fall für sich steht. Sie ist bei Veranstaltungen kein Selbstläufer ist, sondern erfordert eine zielgruppenspezifische Einführung sowie eine Moderation der Rezeption und der aufkommenden Fragen.

6 Forschungsbedarf

Der Forschungsbedarf wurde durch einen kritischen Review von Projektprodukten und -prozessen, einschließlich der Beiratssitzungen identifiziert.

6.1 Kritische Würdigung des Forschungsansatzes

Das vom Umweltbundesamt geförderte Vorlaufforschungsprojekt „Digitalisierung und Gemeinwohl: Transformationsnarrative zwischen Planetaren Grenzen und Künstlicher Intelligenz“ nahm einen für das Umweltressort ungewöhnlichen Ausgangspunkt ein, nämlich durch Künstliche Intelligenz getriebene anthropologische Entwicklungen in dem Sinne, dass sich die vertrauten Vorstellungen zu Mensch-Technik-Umwelt-Beziehungen grundlegend ändern können. Der Anspruch des Umweltressorts, Transformationen in Richtung Nachhaltigkeit zu fördern bringt es mit sich, dass weit über den Umwelt- und Naturschutz hinaus Entwicklungen in den Blick zu nehmen sind.

Wesentliche Fachgebiete, die zur Bearbeitung der Ziele und Forschungsinhalte dieses Projektes herangezogen wurden (darunter Innovationsforschung, philosophische Anthropologie, Ethik und narrative Szenariotechnik) liegen in Wissensbereichen, denen sich das Umweltressort normalerweise wenig(er) bedient. In diesem Projekt wurden jene eher untypischen Wissensbestände sondiert und so aufbereitet, dass sie für das Umweltressort fruchtbar gemacht werden konnten.

Hinsichtlich des fachlichen Hintergrundes waren im Projektteam zahlreiche Disziplinen vertreten, darunter Technischer Umweltschutz, Maschinenbau, Kommunikationswissenschaften, Soziologie, Philosophie, Ethik, Biologie, Landschaftsökologie, Kulturwissenschaften und Foresight. Einige Disziplinen wie die Technikanthropologie und die Narratologie waren mit den Qualifikationsprofilen der Projektbearbeiter*innen nur am Rande vertreten.

Der ausgesprochen breite und vielschichtige Untersuchungsgegenstand erforderte ein pragmatisches Sondieren und Strukturieren von Wissensbeständen. Die zentralen Auswahlprozesse basierten auf Kriterien, die mit dem Auftraggeber abgestimmt wurden und in diesem Bericht transparent und nachvollziehbar dokumentiert sind. Dennoch lassen sich unter den gegebenen Projektbedingungen gewisse intuitiv basierte Schritte nicht vollständig vermeiden. Ein kritisches Momentum stellte die Konstruktion von Anwendungsfeldern dar, die durch KI maßgeblich getrieben werden. KI kann faktisch überall dort verwendet werden, wo Daten verarbeitet werden. Selbstverständlich haben wir mithin bei den genannten Anwendungsfeldern der KI keinen Vollständigkeitsanspruch, das gewählte Verfahren war jedoch gut dazu geeignet, Unterschiede zwischen Anwendungsfeldern anhand grundlegender Veränderungen der Mensch-Technik-Umwelt-Beziehungen zu begründen und damit die Anwendungsfelder zu konstituieren. Beispielsweise wurden im Zuschnitt des Feldes „Simulation natürlicher Sprache“ Sprachassistenten und Computational Creativity zusammen behandelt, weil es um die gemeinsame grundlegende Veränderung der Nichtunterscheidbarkeit der Herkunft eines Textes oder des Gegenübers in einem Sprechakt zwischen Menschen und KI-basierter Technik geht. Die Fokussierung auf KI brachte es mit sich, dass Digitalisierungsfelder wie Blockchain oder Quantencomputing nicht als vordringlich eingestuft wurden, auch wenn sie ebenfalls mit grundlegenden Veränderungen der Mensch-Technik-Umwelt-Beziehungen einhergehen können.

Für die ethische Analyse wurden existierende anerkannte normative Frameworks herangezogen und die darin enthaltenen Kategorien, Prinzipien und Methoden auf den Untersuchungsgegenstand –entsprechend des „Tübinger Ansatzes“ – bezogen. Es gibt auch

andere Ansätze, die anhand anthropologischer Veränderungen durch Künstliche Intelligenz die Kategorien, Prinzipien und Methoden der etablierten normativen Frameworks in Frage stellen. Beispielsweise stellt sich anhand der Durchdringung des menschlichen Körpers mit Technik die Frage, inwiefern medizinethische und oder technikethische Konzepte Anwendung finden und ob sie in ihren jeweiligen begrifflichen Instrumentarien dazu geeignet sind, der hybriden Wirklichkeit bzw. erwünschten hybriden Menschenbildern zutreffend Rechnung zu tragen. Auch stellt sich hierbei die Frage nach der Notwendigkeit eines erweiterten Gerechtigkeitsverständnisses (Diefenbacher et al. 2014), , wenn sich Mensch-Technik-Verständnisse grundlegend verschieben.

Nicht aus Gründen der Wahl einer ethischen Theorie oder des politisch-normativen Rahmens, sondern aus pragmatisch-arbeitsökonomischen Gründen konnten wichtige Aspekte nicht genug in die Betrachtung einbezogen werden. Dies sind vor allem:

- Eine umfassende Analyse der Politischen Ökonomie der KI hinsichtlich der maßgeblichen Treiber und der damit verbundenen Interessengruppen (siehe dazu ausführlich (Nemitz und Pfeffer 2020), wobei diese Dimension mit Bezug auf Gemeinwohlfragen gut bearbeitbar wäre.

- Eine ethisch-politische Tiefenanalyse der Institutionen und der Governance der KI, die genauer nach Machtstrukturen und Fragen gerechter politischer Verfahren und Teilhabe fragt.

- Eine umfassende Analyse der unmittelbaren Implikationen eines breiten KI-Einsatzes hinsichtlich des ökologischen Fußabdrucks, der Treibhausgas-Bilanz und weiterer externer (Umwelt)Kosten der Technologie und ihren Folgen (vgl. u. a. (Lange und Santarius 2020), (WBGU 2019b) sowie die thematische Webseite reset.org)

- Eine detaillierte Kritik an der Rhetorik und Begrifflichkeit der KI mit Blick auf wirtschaftliche und wissenschaftliche Vermarktungs- und Lobby-Interessen, wie folgendes Zitat illustriert: „Technically speaking, it would be more accurate to call Artificial Intelligence machine learning or computational statistics but these terms would have zero marketing appeal for companies, universities and the art market.“ ((Pasquinelli 2019), S. 4).

Die breite lexikalische Analyse der Diskurse zu Affective Computing und Autonomen Systemen in der Tiefsee ergab, dass Technikentwicklung auf der einen Seite und Gemeinwohlorientierung/Nachhaltige Entwicklung auf der anderen Seite weitgehend disparate Sphären sind. Diese unvermittelten Diskurse sind ein wichtiger Befund. In Zukunft könnte auch noch gezielter nach wenigen hochwertigen Schlüsselquellen gesucht oder eben neue erstellt werden, die beiden Sphären gerecht werden.

Der ursprüngliche Ansatz, prospektiv weiter in neue Transformationsnarrative einzudringen wurde fallengelassen, weil der Projektauftrag und -umfang kein ausreichendes Fundament für eine erweiterte Beteiligung von Stakeholdern am Aushandlungsprozess ermöglichte. Deshalb wurde ein Erzählformat in Form eines Einstiegs, der eine Grundlage für die Aushandlung neuer Transformationsnarrative sein kann, gewählt. Mit der Ausarbeitung dieses Einstiegs in Form des Storyboards und der medialen Umsetzung als Scrollytelling wurde damit etwas Neues geschaffen, dessen Wirkung nur punktuell getestet, nicht aber über den Verlauf eines substanziellen Entstehungsprozesses von Transformationsnarrativen im Hinblick auf seine Wirkungen verfolgt werden konnte. Das bewusste ‚in die Welt setzen‘ von neuen Narrativen

durch öffentliche Einrichtungen erfordert in einer Demokratie aber eine Legitimierung durch Partizipation verschiedener gesellschaftlicher Gruppen sowie letztlich entsprechende Beschlüsse (vgl. die analogen Ausführungen zu Nudging bei (Meisch et al. 2018)). Hier besteht eine klare (Macht-)Asymmetrie gegenüber Interessengruppen, für ihre Agenda strategisch bestimmte Narrative zu konstruieren und einzusetzen.

Der Beirat gab wertvolle Sichtweisen und Anregungen für das Projekt. Dessen multidisziplinäre Zusammensetzung bewirkte jedoch auch, dass zahlreiche und teilweise wenig miteinander vereinbare Anregungen für die Ausgestaltung formuliert wurden. Diese Konstruktion verbesserte sicherlich die inhaltliche Qualität des Projektes, spielte die anwachsenden Ausrichtungsmöglichkeiten der Arbeit jedoch auch immer wieder in die Hände des Forschungsteams zurück.

Der pragmatische Forschungsansatz mit seiner Analyse auf einem mittleren Abstraktionsniveau hat – wie gezeigt – einige Grenzen, aber insbesondere den Vorteil, angesichts der großen Breite und Vielschichtigkeit des Untersuchungsgegenstandes vergleichsweise rasch und effizient die Wissensbestände zu strukturieren und handlungsorientiert aufbereiten zu können. Das mittlere Abstraktionsniveau bewahrt vor zu allgemeinen und oberflächlichen Aussagen, aber auch – angesichts der mächtigen gesellschaftlichen Dynamiken – vor der Befassung mit weniger relevanten Einzelfällen.

6.2 Relevante Forschungslinien

Das Themenfeld wurde erschlossen, sondiert und zahlreiche offene Fragen wurden identifiziert. Hieraus lassen sich wiederum Forschungsbedarfe ableiten, die im Falle transformativer Forschung vom Handlungsbedarf nicht sinnvoll zu trennen sind. Drei übergreifende Forschungsstränge lassen sich unterscheiden:

Die Notwendigkeit der Konzeption einer **umfassenden Forschungsprogrammatik** im Zeitalter der Digitalisierung und des Anthropozäns unter ausdrücklichem Einbezug anthropologischer und ethischer Aspekte wurde ersichtlich; im Einzelnen bedeutet dies:

- Verortung von hybriden Digitalisierungs- und Nachhaltigkeitstransformationen im historischen Geschehen unter Einbezug von steuerungstheoretischen und zeitlichen Bezugspunkten;
- Weitere Ausgestaltung des anthropologischen Rahmens als reflexive und bidirektionale Technik- und Gesellschaftsentwicklung und Analyse ihrer Bedeutung für die zukünftige Technikentwicklung und für Nachhaltigkeitsinnovationen;
- Das im Sinne eines Horizon Scanning strukturierte Suchverfahren und die Aufbereitung der Wissensbestände in diesem Projekt legen folgende Aktivitäten auch für andere digitale Entwicklungen nahe:
    - Erkennung von digitalen Entwicklungen jenseits der Künstlichen Intelligenz, die mit grundlegenden Veränderungen der Mensch-Technik-Umwelt-Beziehungen einhergehen können (Horizon Scanning);
    - Aufgreifen von im Horizon Scanning identifizierten anthropologisch relevanten Schlüsselentwicklungen für Vertiefungsanalysen im Hinblick auf die Bedeutung für das Umweltressort (Vertiefungsanalysen);

- Fruchtbarmachung von grundlegenden Veränderungen von Mensch-Technik-Umwelt-Beziehungen durch Digitalisierung für die operative Arbeit des Umweltressorts, z. B. Nutzung der Möglichkeiten und Minimierung der Risiken von kognitiven Assistenten (einschließlich Chatbots) in der Nachhaltigkeitskommunikation und von Extended Reality beim Umwelt-Nudging;

- Mehrdimensionale Analyse von vorhanden normativen Zukünften (Visionen, Leitbilder, normative Szenarien, Trendextrapolationen, Narrativen, etc.) zur Digitalisierung, ihren Entstehungsbedingungen und Wirkungen, als Ausgangspunkt für einen eigenständigen Entwicklungsprozess für Transformationsnarrative (Vision Assessment).

Eine detaillierte inhaltliche Ausgestaltung einer **umfassenderen Ethik** für das Zeitalter der Digitalisierung und des Anthropozäns unter Einbezug der Umwelt-/Nachhaltigkeitsperspektive sowie des Gemeinwohls ist erforderlich.

Allgemein lassen sich folgende übergreifend wichtige Aspekte zugleich als ethische Forschungsdesiderate formulieren:

1. Manipulation und Täuschung in kritischer Verbindung mit Beteiligung und Zugangsgerechtigkeit inkl. der Rolle privater Unternehmen
2. Extended Reality: besseres Lernen vs. Entfremdung als Entpersonalisierung und Entkörperlichung (des Lernens)
3. Digitales Enhancement von Emotionen: siehe a) sowie Privatheit persönlichster Daten, Natürlichkeit vs. Künstlichkeit/Authentizität
4. Autonome Systeme zur Reparatur/Wiederherstellung vs. Vorsorge: Symptom- oder Ursachenorientierung?
5. Frontier-Ideologie der nützlichen Technisierung der Natur vs. Forderungen nach (partieller) Unverfügbarkeit von Räumen
6. Dual use bzw. starke militärische Beteiligung/Förderung KI

Im Ergebnis scheint die (Un)Verfügbarkeit mit Bezug auf Emotionen, Umwelt-Räume, Daten, Vernetzung und vieles mehr in unterschiedlichsten Facetten einer *der* zentralen ethisch weiter auszuarbeitenden und zu diskutierenden Punkte: Anscheinend bislang Unverfügbares wird empirisch verfügbar – soll es aber vielleicht nicht besser unverfügbar bleiben und von welchen Graden von Unverfügbarkeit/Verfügbarkeit und verfügbar für wen ist die Rede? Dies im Detail zu analysieren, ist ein Forschungsdesiderat.

Es geht um die weitere **Zusammenführung** von Digitalisierung, Ethik und Umwelt-/Nachhaltigkeitsperspektiven, Gemeinwohl **im Sinne einer Zukunftsgestaltung** in Form von partizipativen Prozessen, Diskursen und Narrativen; im Einzelnen:

- Erfassung und Strukturierung des normativen Dissenses zu Künstlicher Intelligenz (vgl. u. a. AI Summit und Social Summit) und Zusammenführung der KI-Positionen (z. B. als analytisches Instrument), um Aushandlungsprozesse zu unterstützen

- Konkretisierung des KI-Diskurses vor dem Hintergrund ethischer Desiderata für Nachhaltigkeitstransformationen (Was soll für Nachhaltigkeitstransformationen geforscht und entwickelt werden? Was nicht? Wer entscheidet darüber?)

- Partizipative Konzeption, Durchführung und Kommunikation eines vollständigen Narrativentwicklungsprozesses unter Einbezug begleitender narratologischer Analysen und

Exploration der Wirkung der in diesem Projekt entwickelten Storyboards auf verschiedene Stakeholdergruppen

- Einspeisung der ethischen Aspekte in die in diesem Projekt identifizierten Arenen, Ausleuchtung dieser Arenen, Aktivierung und Interaktion relevanter Stakeholder zu den konkreten identifizierten Themen und Adressierung der aufgezeigten Chancen und Risiken für Nachhaltigkeit mit konkreten Maßnahmen.

7 Quellenverzeichnis

Ackermann, Nils (2018): Artificial Intelligence Framework: A Visual Introduction to Machine Learning and AI. Hg. v. Towards data Science, zuletzt aktualisiert am 13.12.2018, zuletzt geprüft am 12.08.2019.

AG Lebensfeindliche Umgebungen (2019): Lernende Systeme in lebensfeindlichen Umgebungen. Potenziale, Herausforderungen und Gestaltungsoptionen. Hg. v. Lernende Systeme- Die Plattform für künstliche Intelligenz. München. Online verfügbar unter file:///C:/Users/SCHARM~1/AppData/Local/Temp/AG-7_Bericht_web_final.pdf, zuletzt geprüft am 24.11.2021.

Albrecht, Urs-Vito; Amelung, Volker E.; Aumann, Ines; Breil, Bernhard; Brönner, Matthias; Dierks, Marie-Luise et al. (2016): Chancen und Risiken von Gesundheits-Apps (CHARISMHA). Medizinische Hochschule Hannover. Hannover. Online verfügbar unter urn:nbn:de:gbv:084-16040811153. http://www.digibib.tu-bs.de/?docid=00060000.

Altmeppen, Klaus-Dieter; Bieber, Christoph; Filipovi´c, Alexander; Heesen, Jessica; Neuberger, Christoph; Röttger, Ulrike et al. (2019): Öffentlichkeit, Verantwortung und Gemeinwohl im digitalen Zeitalter. Zur Erforschung ethischer Aspekte des Medien- und Öffentlichkeitswandels. In: *Publizistik* (64), S. 59–77.

Ammicht Quinn, Regina (2004): Körper – Religion – Sexualität. Theologische Reflexionen zur Ethik der Geschlechter. 3. Aufl. Mainz: Grünewald.

Ammicht Quinn, Regina; Potthast, Thomas; Kröber, Birgit; Dietrich, Julia; Heesen, Jessica; Meisch, Simon (2015): Ethik in den Wissenschaften. 1 Konzept, 25 Jahre, 50 Perspektiven. Tübingen: Eberhard Karls Universität Tübingen Internationales Zentrum für Ethik in den Wissenschaften (IZEW) (Materialien zur Ethik in den Wissenschaften, Band 10).

André, Elisabeth (2015): Empathische Reaktionen und ihre Modellierung im Computer. In: Helmut Fink und Rainer Rosenzweig (Hg.): Das soziale Gehirn, S. 55–70.

Arbeitskreis Systemaspekte (2018): Kommunikation im Industrie-4.0-Umfeld. Welchen Herausforderungen hat sich die Welchen Herausforderungen hat sich die industrielle Kommunikation im Kontext von Digitalisierung und Industrie 4.0 zu stellen? Hg. v. Fachverband Automation ZVEI (Whitepaper - Teil 4). Online verfügbar unter https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2018/April/Kommunikation_im_Industrie-4.0-Umfeld/Kommunikation_im_Industrie-4.0-Umfeld_Download-Neu.pdf, zuletzt geprüft am 25.11.2021.

Association for Computational Creativity (Hg.) (2016): Computational creativity. Online verfügbar unter http://computationalcreativity.net/home/about/computational-creativity/, zuletzt geprüft am 12.08.2019.

Avin, Shahar; Wintle, Bonnie C.; Weitzdörfer, Julius; Ó hÉigeartaigh, Seán S.; Sutherland, William J.; Rees, Martin J. (2018): Classifying global catastrophic risks. In: *Futures* (102), S. 20–26. Online verfügbar unter https://doi.org/10.1016/j.futures.2018.02.001.

Bakir, Vian; McStay, Andrew (2018): Fake News and The Economy of Emotions. In: *Digital Journalism* 6 (2), S. 154–175. DOI: 10.1080/21670811.2017.1345645.

Baron, Oliver (2018): Diese Aktien hat der Supercomputer. In: *GodMode Trader*, 26.01.2018. Online verfügbar unter https://www.godmode-trader.de/artikel/diese-aktien-hat-der-supercomputer,5719248.

Bastin, J. F.; Finegold, Y.; Garcia, C.; Mollicone, D.; Rezende, M.; Routh, D. et al. (2019): The global tree restoration potential 365, S. 76–79. DOI: 10.1126/science.aax0848.

Bauer, Thomas (2018a): Die Vereindeutigung der Welt. Über den Verlust an Mehrdeutigkeit und Vielfalt. 7. Aufl. Ditzingen: Reclam ([Was bedeutet das alles?]).

Bauer, Vera (2018b): Überwachung: AR-Brillen mit Gesichtserkennung für chinesische Polizei. Hg. v. Mobile Geeks. Online verfügbar unter https://www.mobilegeeks.de/news/ueberwachung-ar-brillen-mit-gesichtserkennung-fuer-chinesische-polizei/, zuletzt geprüft am 12.08.2018.

Baum, Seth D.; Armstrong, Stuart; Ekenstedt, Timoteus; Häggström, Olle; Hanson, Robin; Kuhlemann, Karin et al. (2019): Long-Term Trajectories of Human Civilization. In: *Foresight* 21 (1), S. 53–83. Online verfügbar unter DOI 10.1108/FS-04-2018-0037.

Baumann, Holger; Döring, Sabine (2011): Emotion-oriented systems and the autonomy of persons. In: Paolo Petta, Catherine Pelachaud und Roddy Cowie (Hg.): Emotion-Oriented Systems. The Humaine Handbook (Cognitive Technologies). Heidelberg: Springer, S. 735–752.

Baumgartner, Christof (2018): Studie: Augmented und Virtual Reality sind in Unternehmen in drei Jahren Standard. Hg. v. Computer Welt. Online verfügbar unter https://computerwelt.at/news/topmeldung/studie-augmented-und-virtual-reality-sind-in-unternehmen-in-drei-jahren-standard/, zuletzt geprüft am 12.08.2019.

Becchi, Paolo; Mathis, Klaus (2019): Handbook of Human Dignity in Europe. Berlin: Springer.

Bellina, Leonie; Tegeler, Merle Katrin; Müller-Christ, Georg; Potthast, Thomas (2020): Bildung für Nachhaltige Entwicklung (BNE) in der Hochschullehre. BMBF-Projekt „Nachhaltigkeit an Hochschulen: entwickeln – vernetzen – berichten (HOCHN). 2. Auflage (erweitert und überarbeitet). Online verfügbar unter https://www.hochn.uni-hamburg.de/-downloads/handlungsfelder/lehre/hochn-leitfaden-lehre-2020-neu.pdf, zuletzt geprüft am 10.11.21.

Beni, Gerardo (2005): From Swarm Intelligence to Swarm Robotics. In: Erol Şahin und William M. Spears (Hg.): Swarm robotics. Swarm Robotics Workshop held after the] SAB 2004 International Workshop, Santa Monica, CA, USA, July 17, 2004 ; revised selected papers, Bd. 3342. Berlin, Heidelberg, 2005. Swarm Robotics Workshop; SAB International Workshop. Berlin: Springer (Lecture notes in computer science, 3342), S. 1–9.

Bitkom (2016): Digitale Prozesse. Begriffsabgrenzung und thematische Einordnung. Hg. v. Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. (Bitkom). Online verfügbar unter https://www.bitkom.org/sites/default/files/file/import/160803-Whitepaper-Digitale-Prozesse.pdf, zuletzt geprüft am 15.11.2019.

Bitkom (Hg.) (2018): Digitalisierung gestalten mit dem Periodensystem der Künstlichen Intelligenz. Ein Navigationssystem für Entscheider. Online verfügbar unter https://www.bitkom.org/sites/default/files/2018-12/181204_LF_Periodensystem_online_0.pdf, zuletzt geprüft am 25.11.2021.

Bitkom (23.04.2018): Industrie 4.0: Jede vierte Maschine ist smart. Online verfügbar unter https://www.bitkom.org/Presse/Presseinformation/Industrie-40-Jede-vierte-Maschine-ist-smart.html, zuletzt geprüft am 12.08.2019.

BMWi (2021): Entwicklung der Erneuerbaren Energien in Deutschland im Jahr 2020. Stand: Februar 2021. Bundesministerium für Wirtschaft und Energie. Online verfügbar unter https://www.erneuerbare-energien.de/EE/Navigation/DE/Service/Erneuerbare_Energien_in_Zahlen/Entwicklung/entwicklung-der-erneuerbaren-energien-in-deutschland.html, zuletzt geprüft am 24.11.2021.

Boden, Margaret A. (2004): The creative mind. Myths and mechanisms. 2. Aufl. London: Routledge. Online verfügbar unter http://www.loc.gov/catdir/enhancements/fy0651/2003046533-d.html.

Böhm, Andrea (2020): Unter uns das All. In: *Die Zeit* (3), S. 15–17. Online verfügbar unter https://www.zeit.de/2020/05/tiefsee-meer-lebewesen-entdeckungen-umweltverschmutzung?utm_referrer=https%3A%2F%2Fwww.google.com%2F, zuletzt geprüft am 10.11.2021.

Bonabeau, Eric; Dorigo, Marco; Theraulaz, Guy (2010): Swarm intelligence. From natural to artificial systems. New York, Oxford: Oxford University Press (Santa Fe Institute Studies in the Sciences of Complexity).

Borchers, Detlef (2019): Bundeswehr: Drohnenschwarm soll gläsernes Gefechtsfeld aufklären. Online verfügbar unter https://www.heise.de/newsticker/meldung/Bundeswehr-Drohnenschwarm-soll-glaesernes-Gefechtsfeld-aufklaeren-4401975.html, zuletzt geprüft am 14.08.2019.

Bosch (Hg.) (o.J.): KI-fähige und voll autonome Systeme. Neue innovative Produkte für Hunderte von Millionen Menschen. Online verfügbar unter https://www.bosch.com/de/forschung/innovationsfelder/vollstaendig-autonome-systeme/, zuletzt geprüft am 12.08.2019.

Bostrom, Nick (2014): Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press.

Bostrom, Nick (2017): Superintelligence. Paths, dangers, strategies ; [New afterword]. Oxford: Oxford University Press.

Boyle, T. Coraghessan (2017): Die Terranauten. Roman. Unter Mitarbeit von Dirk van Gunsteren. München: Carl Hanser Verlag. Online verfügbar unter http://www.hanser-literaturverlage.de/9783446253865.

Brahnam, Sheryl (2006): Gendered bots and bot abuse. Montreal (Proceedings of the CHI 2006 workshop Misuse and abuse of interactive technologies). Online verfügbar unter https://sherylbrahnam.com/papers/EN1962.pdf, zuletzt geprüft am 16.11.2021.

Braidotti, Rosi (2019): Zoe/Geo/Techno-Materialismus. In: Katrin Klingan und Christoph Rosol (Hg.): Technosphäre. Berlin: Matthes & Seitz (Bibliothek 100 Jahre Gegenwart).

Brand, Cordula (2015): „Wie Du mir so ich Dir“ - Moralische Anerkennung als intersubjektiver Prozess. In: Robert Ranisch, Marcus Rockoff und Sebastian Schuol (Hg.): Selbstgestaltung des Menschen durch Biotechniken. Tübingen: Narr Francke Attempto, S. 21–33.

Brohmann, Bettina; David, Martin (2015): Transformationsstrategien und Models of Change für nachhaltigen gesellschaftlichen Wandel. Tipping Point Konzeptionen im Kontext eines nachhaltigen gesellschaftlichen Wandels. Hg. v. Umweltbundesamt. Öko-Institut e.V. Institut für angewandte Ökologie; KWI – Kulturwissenschaftliches Institut. Dessau-Roßlau (Texte, 67). Online verfügbar unter https://www.umweltbundesamt.de/sites/default/files/medien/378/publikationen/texte_67_2015_tipping_point_konzeptionen_im_kontext_eines_nachhaltigen_gesellschaftlichen_wandels_1.pdf, zuletzt geprüft am 24.11.2021.

Buber, Martin (2009): Das dialogische Prinzip. Gütersloh: Gütersloher Verlagshaus.

Bundesministerium für Bildung und Forschung (2018): Rahmenprogramm Gesundheitsforschung der Bundesregierung. Hg. v. Bundesministerium für Bildung und Forschung, Referat Grundsatzfragen, Digitalisierung und Transfer. Berlin. Online verfügbar unter https://www.gesundheitsforschung-bmbf.de/files/Rahmenprogramm_Gesundheitsforschung_barrierefrei.pdf, zuletzt geprüft am 03.04.2019.

Bundesregierung (Hg.) (2018a): Forschung und Innovation für die Menschen. Die Hightech-Strategie 2025. Online verfügbar unter https://www.hightech-strategie.de/hightech/shareddocs/downloads/files/hts2025.pdf;jsessionid=5932EC0B5A355CA8F8991681CF6E2505.live381?__blob=publicationFile&v=1, zuletzt geprüft am 24.11.2021.

Bundesregierung (Hg.) (2018b): Strategie Künstliche Intelligenz der Bundesregierung. Berlin. Online verfügbar unter https://www.bundesregierung.de/resource/blob/975226/1550276/3f7d3c41c6e05695741273e78b8039f2/2018-11-15-ki-strategie-data.pdf?download=1, zuletzt geprüft am 26.11.2021.

Bundesregierung (2018c): Strategie Künstliche Intelligenz der Bundesregierung. Stand: November 2018.

Burchardt, Aljoscha; Uszkoreit, Hans (2018a): IT für soziale Inklusion. Digitalisierung - Künstliche Intelligenz - Zukunft für alle. Berlin, Boston: De Gruyter Oldenbourg.

Burchardt, Aljoscha; Uszkoreit, Hans (Hg.) (2018b): IT für soziale Inklusion. Digitalisierung – Künstliche Intelligenz – Zukunft für alle. München, Wien: De Gruyter Oldenbourg. Online verfügbar unter https://www.degruyter.com/isbn/9783110561371.

Carman, Ashley (2019): North Focals Glasses review: A $600 smartwatch for your face. Hg. v. The Verge. Online verfügbar unter https://www.theverge.com/2019/2/14/18223593/focals-smart-glasses-north-review-specs-features-price, zuletzt geprüft am 12.08.2019.

Casson, Louisa (2019): In Deep Water. The emerging threat of deep sea mining. Unter Mitarbeit von Sebastian Losada, Sofia Tsenikli, Dr David Santillo, Duncan Currie, Gargi Sharma, Will McCallum. Hg. v. Greenpeace International. Online verfügbar unter https://www.greenpeace.org/international/publication/22578/deep-sea-mining-in-deep-water/, zuletzt geprüft am 02.02.2021.

Ceruzzi, Paul E. (2012): Computing. A concise history. Cambridge, Mass: MIT Press (MIT Press essential knowledge series). Online verfügbar unter http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=463391.

Chapman, Jonathan (2012): Emotionally Durable Design: Routledge.

Cherney, Mike (2021): BUZZ OFF, BEES. POLLINATION ROBOTS ARE HERE. Advances in artificial intelligence are helping some startups develop another way to pollinate plants, which could increase yield compared with insects and human workers. Hg. v. The Wall Street Journal. Online verfügbar unter https://www.wsj.com/articles/buzz-off-bees-pollination-robots-are-here-11625673660, zuletzt geprüft am 16.11.2021.

Cockburn, Harry (2019): Ocean cleaning contraption successfully removes waste from Great Pacific garbage patch for first time. Technology designed by 25-year-old Dutch inventor provides hope for large-scale clean-up operation. In: *The Independent* 2019, 03.10.2019. Online verfügbar unter https://www.independent.co.uk/environment/great-pacific-garbage-patch-ocean-cleaning-successful-inventor-boyan-slat-a9138286.html, zuletzt geprüft am 06.11.2019.

Coleman, Jeff (2018): Brain Computer Interface with Artificial Intelligence and Reinforcement Learning. Online verfügbar unter https://medium.com/@askwhy/brain-computer-interface-with-artificial-intelligence-and-reinforcement-learning-9c94b0454209, zuletzt geprüft am 12.08.2019.

Collingridge, David (1980): The social control of technology. London: Pinter.

Colombetti, Giovanna; Stephan, Achim (2013): Affektwissenschaft (*affective science*). In: Achim Stephan und Sven Walter (Hg.): Handbuch Kognitionswissenschaft. Stuttgart, Weimar: J.B. Metzler, S. 501–509.

Colton, Simon; Wiggins, Geraint (2012): Computational Creativity: The Final Frontier. In: L. de Raedt, C. Bessiere, D. Dubois, P. Doherty, P. Frasconi, F. Heintz und P. Lucas (Hg.): Ecai 2012. 20th European Conference on Artificial Intelligence. Fairfax: IOS Press Incorporated (Frontiers in Artificial Intelligence and Applications Ser, v.242), S. 21–26.

ConsorsBank (2018): Wie künstliche Intelligenz den Börsenhandel verändern könnte. Online verfügbar unter https://wissen.consorsbank.de/t5/Blog/Wie-k%C3%BCnstliche-Intelligenz-den-B%C3%B6rsenhandel-ver%C3%A4ndern-k%C3%B6nnte/ba-p/75654, zuletzt aktualisiert am 30.10.2018, zuletzt geprüft am 20.11.2021.

Continental (Hg.) (2019): Augmented Reality Head-up Display. Online verfügbar unter https://www.continental-automotive.com/en-gl/Passenger-Cars/Interior/Display-Systems/Head-Up-Displays/Augmented-Reality-HUD, zuletzt geprüft am 18.11.2019.

Cosgrave, Ellie (2017): The future of floating Cities and the realities. Hg. v. BBC. Online verfügbar unter http://www.bbc.com/future/story/20171128-the-future-of-floating-cities-and-the-realities, zuletzt geprüft am 14.08.2019.

Costanza, Robert; d'Arge, Ralph; Groot, Rudolf de; Farber, Stephen; Grasso, Monica; Hannon, Bruce et al. (1997): The value of the world's ecosystem services and natural capital. In: *Nature* (387), S. 253–260. Online verfügbar unter https://www.nature.com/articles/387253a0.

Cowie, Roddy (2012): The Good Our Field Can Hope to Do, the Harm It Should Avoid. In: *IEEE Transactions on Affective Computing* (Vol. 3, No. 4), S. 410–423.

Cowie, Roddy (2015): Ethical Issues in Affective Computing. In: Rafael Calvo, Sidney D'Mello, Jonathan Gratch und Arvid Kappas (Hg.): The Oxford Handbook of Affective Computing. Oxford, New York: Oxford University Press, S. 334–348.

Daum, Timo (2019a): Die Künstliche Intelligenz des Kapitals. Originalveröffentlichung, Erstausgabe, 1. Auflage. Hamburg: Edition Nautilus (Nautilus Flugschrift).

Daum, Timo (2019b): Die Künstliche Intelligenz des Kapitals. 1. Aufl. Hamburg: Edition Nautilus (Nautilus Flugschrift).

Day, Matt (2019): Amazon Is Working on a Device That Can Read Human Emotions. Hg. v. Bloomberg. Online verfügbar unter https://www.bloomberg.com/news/articles/2019-05-23/amazon-is-working-on-a-wearable-device-that-reads-human-emotions, zuletzt geprüft am 12.08.2019.

DeCleene, John (2019): A Quick Guide to Investment Algorithms. In: *Data Driven Investor*, 31.01.2019. Online verfügbar unter https://www.datadriveninvestor.com/2019/01/31/a-quick-guide-to-investment-algorithms/, zuletzt geprüft am 15.11.2019.

Delgado, Rick (2017): Why IoT should have an artificial intelligence layer. Hg. v. Jaxenter. Online verfügbar unter https://jaxenter.com/iot-artificial-intelligence-131992.html, zuletzt geprüft am 12.08.2019.

Dennett, D. C. (2018): From bacteria to Bach and back. The evolution of minds. Fist published as a Norton paperback. New York, London: W. W. Norton & Company.

Deutsche Stiftung Weltbevölkerung (Hg.) (2018): Soziale und demogafische Daten weltweit DSW-Datenreport 2018. Unter Mitarbeit von Mareike Döring, Greta Theilen und Ute Stallmeister. Hannover. Online verfügbar unter https://www.presseportal.de/pm/24571/4040855, zuletzt geprüft am 25.11.2021.

Deutscher Ethikrat (Hg.) (2016): Jahresbericht 2015. Berlin. Online verfügbar unter https://www.ethikrat.org/fileadmin/Publikationen/Jahresberichte/deutsch/jahresbericht-2015.pdf, zuletzt geprüft am 26.11.2021.

Deutscher Ethikrat (Hg.) (2017): Autonome Systeme. Wie intelligente Maschinen uns verändern. Jahrestagung. Berlin, 21.06.2017.

DFKI (18.04.2018): Wie Mensch und Maschine zusammenarbeiten: Einsatz von KI und AR. Kaiserslautern. Reinhard Karger. Online verfügbar unter https://www.dfki.de/web/news/detail/News/wie-mensch-und-maschine-zusammenarbeiten-einsatz-von-ki-und-ar/, zuletzt geprüft am 18.09.2019.

DHL (08.02.2017): DHL Supply Chain makes Smart Glasses new standard in logistics. Bonn. Online verfügbar unter https://www.logistics.dhl/global-en/home/press/press-archive/2017/dhl-supply-chain-makes-smart-glasses-new-standard-in-logistics.html, zuletzt geprüft am 12.08.2019.

Diefenbacher, Hans; Düwell, Marcus; Philips, Jos; Leggewie, Claus; Sommer, Berd; Petschow, Ullrich et al. (2014): Konzepte gesellschaftlichen Wohlstands und ökologische Gerechtigkeit. Hg. v. Umweltbundesamt. Dessau-Roßlau (UBA-Texte, 45/2014). Online verfügbar unter https://www.umweltbundesamt.de/publikationen/konzepte-gesellschaftlichen-wohlstands-oekologische, zuletzt geprüft am 08.11.2021.

Dudzik, Bernd; Jansen, Michel-Pierre; Burger, Franziska; Kaptein, Frank; Broekens, Joost; Heylen, Dirk K.J. et al. (2019): Context in Human Emotion Perception for Automatic Affect Detection: A Survey of Audiovisual

Databases. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). Cambridge, United Kingdom, 03.09.2019 - 06.09.2019: IEEE, S. 206–212.

Dueck, Gunter (2018): Schwarmdumm. So blöd sind wir nur gemeinsam. 1. Aufl. München: Wilhelm Goldmann Verlag.

Dumitrescu, Roman; Gausemeier, Jürgen; Slusallek, Philipp; Cieslik, Sarah; Demme, Georg; Falkowski, Tommy et al. (2018): Studie „Autonome Systeme". Hg. v. Expertenkommission Forschung und Innovation. Deutsches Forschungszentrum für Künstliche Intelligenz; Acatech; Fraunhofer-Institut für Entwurfstechnik Mechatronik IEM (Studien zum deutschen Innovationssystem, 13-2018). Online verfügbar unter https://www.econstor.eu/bitstream/10419/175555/1/1015315232.pdf, zuletzt geprüft am 24.11.2021.

Dunseath, Sarah; Weibel, Nadir; Bloss, Cinnamon S.; Nebeker, Camille (2018): NIH support of mobile, imaging, pervasive sensing, social media and location tracking (MISST) research. Laying the foundation to examine research ethics in the digital age. In: *npj Digital Med* 1 (1), S. 1679. DOI: 10.1038/s41746-017-0001-5.

dw (2019): Müllfänger soll Plastik aus Flüssen holen. In: *DW Made for Minds* 2019, 27.10.2019. Online verfügbar unter https://www.dw.com/de/m%C3%BCllf%C3%A4nger-soll-plastik-aus-fl%C3%BCssen-holen/a-51004137.

Endrass, Birgit; André, Elisabeth; Rehm, Matthias; Nakano, Yukiko (2013): Investigating culture-related aspects of behavior for virtual characters. In: *Auton Agent Multi-Agent Syst* 27 (2), S. 277–304. DOI: 10.1007/s10458-012-9218-5.

Enquete-Kommission Künstliche Intelligenz (2021): Unterrichtung der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale. Hg. v. Deutscher Bundestag. Berlin (Drucksache, 19/23700). Online verfügbar unter https://dserver.bundestag.de/btd/19/237/1923700.pdf, zuletzt geprüft am 24.11.2021.

Eraslan, Gökcen; Avsec, Žiga; Gagneur, Julien; Theis, Fabian J. (2019): Deep learning. New computational modelling techniques for genomics. In: *Nature reviews. Genetics*. DOI: 10.1038/s41576-019-0122-6.

Erdmann, Lorenz; Röß, Andreas (2020): Warum die Innoationsforschung ein explizites Innovationsverständnis braucht. In: Fraunhofer ISI (Hg.): Beiträge zur Analyse der Digitalisierung aus Innovationsperspektive, Bd. 68. Unter Mitarbeit von Bernd Beckert, Lorenz Erdmann, Alexander Feidenheimer, Matthias Gotsch, Henning Kroll, Andreas Röß und Torben Schubert. Karlsruhe (Fraunhofer ISI Discussion Papers Innovation Systems and Policy Analysis, Nr. 68), S. 63–72.

Ergin, Yasemin; Ammer, Andreas (2021): Alterslose Avatare. Wie ABBA sich unsterblich macht. Das Erste, 07.11.2021. Online verfügbar unter https://www.daserste.de/information/wissen-kultur/ttt/sendung/abba100.html.

Eser, Uta; Potthast, Thomas (1999): Naturschutzethik. Baden-Baden: Nomos.

Espinosa, Cristina; Pregernig, Michael; Fischer, Corinna (2017): Narrative und Diskurse in der Umweltpolitik: Möglichkeiten udn Grenzen ihrer strategischen Nutzung. Zwischenbericht. Hg. v. Umweltbundesamt (UBA-Texte, 86/2017). Online verfügbar unter https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2017-09-27_texte_86-2017_narrative_0.pdf, zuletzt geprüft am 26.11.2021.

Ess, Charles (2016): What's Love Got to Do with It? Robots, sexuality, and the arts of being human. In: Marco Nørskov (Hg.): Social Robots: Boundaries, Potentials, Challenges. Ashgate, Surrey (England): Farnham, S. 57–79.

European Commission (Hg.) (2018): Artificial Intelligence A European Perspective. Brussels. Online verfügbar unter https://publications.jrc.ec.europa.eu/repository/handle/JRC113826, zuletzt geprüft am 25.11.2021.

Evers-Wölk, Michaela; Oertel, Britta; Sonk, Matthias (2018): Gesundheits-Apps. Innovationsanalyse. Arbeitsbericht Nr. 179. Unter Mitarbeit von Mattis Jacobs. Hg. v. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag. Berlin. Online verfügbar unter https://www.tab-beim-bundestag.de/de/pdf/publikationen/berichte/TAB-Arbeitsbericht-ab179.pdf, zuletzt geprüft am 24.11.2021.

Feireiss, Lukas; Najjar, Michael (Hg.) (2018): Planetary echoes. Exploring the implications of human settlement in outer space. Unter Mitarbeit von Michael Najjar. 1. Aufl. Germany: Spector Books.

Fischer, Roland (2018): Beziehungs-Kiste. In: *Technology review* (6), S. 80.

Fong, Terry (2018): Autonomous Systems. NASA Capability Overview, 24.08.2018. Online verfügbar unter https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=2ahUKEwi48c3Sq47hAhUQZFAKHXMsAasQFjACegQICRAC&url=https%3A%2F%2Fwww.nasa.gov%2Fsites%2Fdefault%2Ffiles%2Fatoms%2Ffiles%2Fnac_tie_aug2018_tfong_tagged.pdf&usg=AOvVaw1zOTLO6j1I5VHoWV_B7nhO, zuletzt geprüft am 14.08.2019.

Frank, Ronald; Unfried, Matthias; Schreder, Regina; Dieckmann, Anja (2016): Ethical Textile Consumption: Only a Question of Selflessness? In: *GfK Marketing Intelligence Review* 8 (1), S. 52–58. DOI: 10.1515/gfkmir-2016-0009.

Frankfurter Allgemeine Zeitung (Hg.) (2019): 12.000 Satelliten für Internet auf der ganzen Welt. Online verfügbar unter https://www.faz.net/aktuell/wirtschaft/unternehmen/tausende-satelliten-von-spacex-sollen-internet-in-alle-welt-bringen-16203802.html, zuletzt aktualisiert am 24.05.2019, zuletzt geprüft am 12.08.2019.

Fraunhofer IIS (Hg.) (o.J.): Autonome Systeme. Online verfügbar unter https://www.iis.fraunhofer.de/de/ff/lv/dataanalytics/tech/auto.html, zuletzt geprüft am 12.08.2019.

Friedrich, Jörg Phil (2012): Kritik der vernetzten Vernunft. Philosophie für Netzbewohner: Telepolis.

Frühauf, Markus (2019): Mit Algorithmen die Vorstandschefs durchleuchten. In: *FAZ*, 06.04.2019. Online verfügbar unter https://www.faz.net/aktuell/wirtschaft/diginomics/vermoegensverwalter-invesco-setzt-ki-ein-16126402.html, zuletzt geprüft am 15.11.2019.

Fuchs, Michael; Lanzerath, Dirk; Hillebrand, Ingo; Runkel, Thomas; Balcerak, Magdalena; Schmitz, Barbara (2002): Enhancement. Die ethische Diskussion über biomedizinische Verbesserungen des Menschen. Bonn: Deutsches Referenzzentrum für Ethik in den Biowissenschaften (drze-Sachstandsbericht, 1).

Gabbatiss, Josh (2019): Robot 'shark' that eats plastic waste launched to tackle pollution. In: *The Independent* 2019, 04.03.2019. Online verfügbar unter https://www.independent.co.uk/environment/robot-shark-plastic-pollution-wasteshark-wwf-devon-ilfracombe-a8805681.html, zuletzt geprüft am 05.11.2019.

Gabriel, Iason (2020): Artificial Intelligence, Values, and Alignment. In: *Minds and Machines* (30), S. 411–437. Online verfügbar unter https://link.springer.com/content/pdf/10.1007/s11023-020-09539-2.pdf, zuletzt geprüft am 10.11.21.

Garfield, Leanna (2018): A pilot project for a new libertarian floating city will have 300 homes, its own government, and its own cryptocurrency. Hg. v. Business Insider. Online verfügbar unter https://www.businessinsider.co.za/floating-city-plans-seasteading-institute-peter-thiel-blue-frontiers-2017-12, zuletzt geprüft am 14.08.2019.

Gartner (2018): Hype Cycle for Emerging Technologies, 2018, August 2018.

Gebhard, Patrick; Baur, Tobias; Ionut, Damian; Mehlmann, Gregor (2014): Exploring Interaction Strategies for Virtual Characters to Induce Stress in Simulated Job Interviews. In: Aamas 14 Conference Committee (Hg.): Proceedings of the 13th International Conference on Automous Agents and Multiagent Systems.

Ghandeharioun, Asma; McDuff, Daniel; Czerwinski, Mary; Rowan, Kael (2019): EMMA: An Emotion-Aware Wellbeing Chatbot. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction

(ACII). 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). Cambridge, United Kingdom, 03.09.2019 - 06.09.2019: IEEE, S. 1–7.

Gheorghiu, Radu; Dragomir, Bianca; Andreescu, Liviu; Cuhls, Kerstin; Rosa, Aaron; Curaj, Adrian; Weber, Matthias (2017): New horizons. Data from a Delphi survey in support of future European Union policies in research and innovation. Luxembourg: Publications Office.

Gilligan, Carol (1985): Die andere Stimme. Lebenskonflikte und Moral der Frau. München: Pieper.

Giménez, Blanca; Nolasco-Rózsás; Pichler, Franz; Weibel, Peter (2018): Open Codes. The World as a Field of Data. Unter Mitarbeit von Jens Lutz und Miriam Stürner.

Goffner, Deborah; Sinare, Hanna; Gordon, Line J. (2019): The Great Green Wall for the Sahara and the Sahel Initiativeas an opportunity to enhance resilience in Sahelian landscapes and livelihoods 19, S. 1417–1428. Online verfügbar unter https://link.springer.com/content/pdf/10.1007%2Fs10113-019-01481-z.pdf, zuletzt geprüft am 06.11.2019.

Gohd, Chelsea (2018): Walmart has patented autonomous robot bees. Hg. v. World Economic Forum. Online verfügbar unter https://www.weforum.org/agenda/2018/03/autonomous-robot-bees-are-being-patented-by-walmart, zuletzt geprüft am 14.08.2019.

Goldie, Peter (2000): The Emotions: A Philosophical Explanation. Oxford: Clarendon Press.

Goodwin, Sara; McPherson, John D.; McCombie, W. Richard (2016): Coming of age. Ten years of next-generation sequencing technologies. In: *Nature reviews. Genetics* 17 (6), S. 333–351. DOI: 10.1038/nrg.2016.49.

Goram, Mandy (o. J.): Künstliche Intelligenz. Online verfügbar unter http://www.datenbanken-verstehen.de/lexikon/kuenstliche-intelligenz/, zuletzt geprüft am 12.08.2019.

Gorke, Martin (2010): Eigenwert der Natur. Ethische Begründungen und Konsequenzen. Stuttgart: Hirzel.

Gotsch, Matthias; Erdmann, Lorenz (2018): Digitalisierung ökologisch nachhaltig nutzbar machen. Zwischenbericht zu AP2: Identifizierung und Analyse von Trendthemen mit potenziell hoher Umweltrelevanz. unveröffentlicht. Fraunhofer ISI im Auftrag von Umweltbundesamt. Karlsruhe.

Greenfield, Adam (2018): Radical Technologies. The Design of Everyday Life. London, New York: Verso.

Greenpeace (2019): In Deep Water. The Emerging Threat Of Deep Sea Mining. Online verfügbar unter https://storage.googleapis.com/planet4-international-stateless/2019/06/f223a588-in-deep-water-greenpeace-deep-sea-mining-2019.pdf, zuletzt geprüft am 10.11.2021.

Grossmann, David (2017): How Drones Could Be Able to Plant a Billion Trees Per Year. They'll be shooting seeds into the ground starting this September. In: *Popular Mechanics*, 14.08.2017. Online verfügbar unter https://www.popularmechanics.com/technology/robots/a27743/company-wants-drones-to-plant-a-billion-trees/, zuletzt geprüft am 06.11.2019.

Grunwald, Armin (2019): Der unterlegene Mensch. Die Zukunft der Menschheit im Angesicht von Algorithmen, künstlicher Intelligenz und Robotern. 1. Aufl. München: premium riva.

Gutmann, Mathias (2011): Sozialität durch technische Systeme? (Technikfolgenabschätzung - Theorie und Praxis, 20. Jg. Heft 1). Online verfügbar unter https://tatup.de/index.php/tatup/article/view/790/1446, zuletzt geprüft am 24.11.2021.

Hagendorff, Thilo (2020): The Ethics of AI Ethics: An Evaluation of Guidelines. In: *Minds & Machines* 30 (1), S. 99–120. DOI: 10.1007/s11023-020-09517-8.

Hao, Karen (2019): We analyzed 16,625 papers to figure out where AI is headed next. Hg. v. Technology Review. Online verfügbar unter https://www.technologyreview.com/s/612768/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/, zuletzt geprüft am 12.08.2019.

Harari, Yuval Noah (2016): Homo deus. A brief history of tomorrow. London: Harvill Secker.

Hartung, Gerald (2018): Philosophische Anthropologie. Reclams Universal-Bibliothek. Ditzingen: Reclam Verlag (Reclams Universal-Bibliothek).

Hasenöhrl, Ute (2005): Zivilgesellschaft, Gemeinwohl und Kollektivgüter. Diskussionspapiere. Wissenschaftszentrum Berlin für Sozialforschung. Berlin. Online verfügbar unter https://bibliothek.wzb.eu/pdf/2005/iv05-401.pdf, zuletzt geprüft am 23.11.2021.

Hashemian, Mojgan; Prada, Rui; Santos, Pedro A.; Dias, Joao; Mascarenhas, Samuel (2019): Inferring Emotions from Touching Patterns. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). Cambridge, United Kingdom, 03.09.2019 - 06.09.2019: IEEE, S. 1–7.

Hawking, Stephen (2018): Brief Answers to the Big Questions. New York: Bentam.

Hecker, Dirk; Döbel, Inga; Petersen, Ulrike; Rauschert, Ulrike; Schmitz, Velina; Voss, Angelika (2017): Zukunftsmarkt Künstliche Intelligenz. Potentiale und Anwendungen. Hg. v. Fraunhofer-Allianz BIG DATA. Online verfügbar unter https://www.bigdata-ai.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/KI-Potenzialanalyse_2017.pdf, zuletzt geprüft am 25.11.2021.

Heimerl, Alexander; Baur, Tobias; Lingenfelser, Florian; Wagner, Johannes; Andre, Elisabeth (2019): NOVA - A tool for eXplainable Cooperative Machine Learning. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). Cambridge, United Kingdom, 03.09.2019 - 06.09.2019: IEEE, S. 109–115.

Heßler, Martina; Liggieri, Kevin (Hg.) (2020): Technikanthropologie. Handbuch für Wissenschaft und Studium. Baden-Baden: Nomos (edition sigma).

Hilty, Lorenz; Behrendt, Siegfried; Binswanger; Bruinink, arend; Erdmann, Lorenz; Kuster, Nils et al. (2003): Das Vorsorgeprinzip in der Informationsgesellschaft. Auswirkungen des Pervasive Computing auf Gesundheit und Umwelt. Hg. v. Zentrums für Technologiefolgen-Abschätzung (46/2003). Online verfügbar unter https://www.izt.de/pdfs/pervasive/Vorsorgeprinzip_Informationsgesellschaft_Pervasive_Computing_Langfassung.pdf, zuletzt geprüft am 24.11.2021.

Hoffmann, Verena; Scheffler, Jonathan (2018): Künstliche Intelligenz – Chance und Herausforderung. Hg. v. Führungsakademie der Bundeswehr. Hamburg. Online verfügbar unter https://www.fueakbw.de/index.php/de/aktuelles/410-kuenstliche-intelligenz-chance-und-herausforderung-fuer-politik-gesellschaft-und-streitkraefte, zuletzt aktualisiert am 28.08.2018.

Holbach, Anne (2019): Unterwasserroboter soll autonom agieren. In: *Kieler Nachrichten* 2019, 07.06.2019. Online verfügbar unter https://www.kn-online.de/Nachrichten/Wirtschaft/Kieler-Forschungsprojekt-Unterwasserroboter-soll-autonom-agieren, zuletzt geprüft am 05.11.2019.

Horn, Eva; Bergthaller, Hannes (2019): Anthropozän. zur Einführung. Hamburg: Junius.

Horton, Donald; Wohl, Richard R. (1956): Mass Communication and Para-Social Interaction. Observations on Intimacy at a Distance. In: *Psychiatry* (19), S. 215–229.

Huang, Eddie; Valdiviejas, Hannah; Bosch, Nigel (2019): I'm Sure! Automatic Detection of Metacognition in Online Course Discussion Forums. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). Cambridge, United Kingdom, 03.09.2019 - 06.09.2019: IEEE, S. 1–7.

Huesemann, Michael; Huesemann, Joyce (2011): Techno-Fix: Why Technology Won't Save Us Or the Environment. Gabriola: New Society Publishers.

Huppertz, Guido (2016): Roboter-Schwärme. In: *Europäische Sicherheit & Technik* (8/2016), S. 88.

Husserl, Edmund (1973): Zur Phänomenologie der Intersubjektivität. In: Texte aus dem Nachlass Dritter Teil: 1929-1935. Den Haag: Nijhoff (Husserliana: Edmund Husserl - Gesammelte Werke, Bd. 15).

IEEE (2019): The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition. IEEE, 2019. Online verfügbar unter https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.htm.

Ienca, Marcello; Haselager, Pim; Emanuel, Ezekiel J. (2018): Brain leaks and consumer neurotechnology. In: *Nature Biotechnology* 36, 805 EP -. DOI: 10.1038/nbt.4240.

Inayatullah, S.; Milojevic, I. (2015): CLA 2.0. Transformative Research in Theory and Practice. Taipeh: Tamkang University Press.

Interface (o.J.): Climate Take Back Case Study. BioCarbon Engineering. Online verfügbar unter http://interfaceinc.scene7.com/is/content/InterfaceInc/Interface/Americas/WebsiteContentAssets/Documents/CaseStudies/CTB/BioCarbon%20Engineering/wc_am-biocarbonengineeringctb.pdf, zuletzt aktualisiert am o.J., zuletzt geprüft am 10.11.2021.

Interview 1. anonymisiert.

Interview 3. anonymisiert.

Interview 7. anonymisiert.

Interview 9. anonymisiert.

IntKom; Brot für die Welt; Fair Oceans (2018): Solwara 1. Bergbau am Meeresboden vor Papua-Neuguinea. Hintergründe, Folgen, Widerstand. IntKom (Verein für Internationalismus und Kommunikation e.V.); Brot für die Welt; Fair Oceans. Online verfügbar unter https://www.brot-fuer-die-welt.de/fileadmin/mediapool/blogs/Mari_Francisco/studie_solwara1_final_e-book_oht.pdf, zuletzt geprüft am 21.11.2021.

ITWissen.info (Hg.) (2017): Virtuelle Realität. Online verfügbar unter https://www.itwissen.info/Virtuelle-Realitaet-virtual-reality-VR.html, zuletzt geprüft am 12.08.2019.

Jaramillo, Fernando; Destouni, Georgia (2015): Comment on "Planetary Boundaries: Guiding Human Development on a Chaning Planet". In: *Science* (348).

Jonas, Hans (1979): Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation. Frankfurt am Main: Suhrkamp.

Jörg, Johannes (2018): Digitalisierung in der Medizin. Wie Gesundheits-Apps, Telemedizin, künstliche Intelligenz und Robotik das Gesundheitswesen revolutionieren. Berlin: Springer.

Jung, Norbert; Molitor, Heike; Schilling, Astrid (Hg.) (2015): Natur, Emotion, Bildung - vergessene Leidenschaft? Zum Spannungsfeld von Naturschutz und Umweltbildung. Opladen, Berlin, Toronto: Budrich UniPress Ltd (Eberswalder Beiträge zu Bildung und Nachhaltigkeit, Band 4). Online verfügbar unter http://www.content-select.com/index.php?id=bib_view&ean=9783863882488.

Kadner, Susanne; Winter, Johannes; Blocher, Anselm; Dengler, Dietmar; Fehling, Marcus; Haustein, Berthold et al. (2017): Fachforum Autonome Systeme Chancen und Risiken für Wirtschaft, Wissenschaft und Gesellschaft. Hg. v. Bundesministerium für Bildung und Forschung. Online verfügbar unter https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjCweK787X0AhXFDewK

HZWTBgcQFnoECA4QAQ&url=https%3A%2F%2Fwww.stifterverband.org%2Fdownload%2Ffile%2Ffid%2F7103&usg=AOvVaw3Bj9IRwDhUzx4hGde84AHg, zuletzt geprüft am 25.11.2021.

Kahlenborn, Walter; Keppner, Benno; Uhle, Christian; Richter, Stephan; Jetzke, Tobias (2018): Die Zukunft im Blick: Konsum 4.0: Wie Digitalisierung den Konsum verändert Trendbericht zur Abschätzung der Umweltwirkungen. Hg. v. Umweltbundesamt. Online verfügbar unter www.umweltbundesamt.de/publikationen, zuletzt geprüft am 26.09.2020.

Kahlenborn, Walter; Keppner, Benno; Uhle, Christian; Richter, Stephan; Jetzke, Tobias (2019): Konsum 4.0: Wie Digitalisierung den Konsum verändert. Trendbericht zur Abschätzung der Umweltwirkungen. Hg. v. Umweltbundesamt. Dessau-Roßlau. Online verfügbar unter https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/fachbroschuere_konsum_4.0_barrierefrei_190322.pdf, zuletzt geprüft am 24.11.2021.

Kahn, Peter H. (2011): Technological Nature. Adaptation and the Future of Human Life. Cambridge, MA: MIT Press.

Kahn, Severson, Ruckert (2009): The Human Relation With Nature and Technological Nature. In: *Current Directions in Psychological Science*.

Kant, Immanuel (1990): Die Metaphysik der Sitten. Ditzingen: Reclam.

Kao, Chien-Chi; Lin, Yi-Shan; Wu, Geng-De; Huang, Chun-Ju (2017): A Comprehensive Study on the Internet of Underwater Things: Applications, Challenges, and Channel Models. In: *Sensors (Basel, Switzerland)* 17 (7). DOI: 10.3390/s17071477.

Kaufmann, Franz-Xaver (2002): Sozialpolitik zwischen Gemeinwohl und Solidarität. In: Karsten Fischer und Herfried Münkler (Hg.): Gemeinwohl und Gemeinsinn. Rhetoriken und Perspektiven sozial-moralischer Orientierung. Berlin: Akademie Verlag.

Kehl, Christoph; Coenen, Christoph (2016): Technologien und Visionen der Mensch-Maschine-Entgrenzung. TAB (TAB Arbeitsbericht, Nr. 167). Online verfügbar unter https://www.tab-beim-bundestag.de/de/pdf/publikationen/berichte/TAB-Arbeitsbericht-ab167.pdf, zuletzt geprüft am 24.11.2021.

Kelleher, John D.; Tierney, Brendan (2018): Data Science. Cambridge (MA), London: MIT Press.

Kind, Sonja; Ferdinand, Jan-Peter (2018): Big Social Data – die gesellschaftspolitische Dimension von Prognose und Ratingalgorithmen. Hg. v. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (Themenkurzprofil, 20). Online verfügbar unter https://www.tab-beim-bundestag.de/de/pdf/publikationen/themenprofile/Themenkurzprofil-020.pdf, zuletzt geprüft am 25.11.2021.

Kind, Sonja; Weide, Sebastian (2017): Microtargeting: psychometrische Analyse mittels Big Data. Hg. v. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (Themenkurzprofil, 18). Online verfügbar unter https://www.tab-beim-bundestag.de/de/pdf/publikationen/themenprofile/Themenkurzprofil-018.pdf, zuletzt geprüft am 25.11.2021.

Kittlitz, Alard von (2012): Jenseits von Eden. Die Geschichte eines Virus. Online verfügbar unter https://www.faz.net/aktuell/gesellschaft/die-geschichte-eines-virus-jenseits-von-eden-11749185.html, zuletzt geprüft am 10.11.2021.

Knight, Will (2019): Alexa needs a robot body to escape the confines of today's AI. Online verfügbar unter https://www.technologyreview.com/s/613199/alexa-needs-a-robot-body-to-escape-the-confines-of-todays-ai/, zuletzt geprüft am 12.08.2019.

Kompa, Nikola; Moll, Henrike; Eckardt, Regine; Grassmann, Susanne (2013): Sprache, sprachliche Bedeutung, Sprachverstehen und Kontext. In: Achim Stephan und Sven Walter (Hg.): Handbuch Kognitionswissenschaft. Stuttgart, Weimar: J.B. Metzler.

Kopatz, Michael (2016): Ökoroutine: Damit wir tun, was wir für richtig halten. München: Oekom.

Krail, Michael (2018): Automatisiertes und vernetztes Fahren. Digitalisierung ökologisch nachhaltig nutzbar machen. Fraunhofer ISI. Berlin, 29.06.2018.

Krauss, Christopher; Do, Xuan Anh; Huck, Nicolas (2016): Deep neural networks, gradient-boosted trees, random forests. Statistical arbitrage on the S & P 500. Erlangen-Nürnberg: Friedrich-Alexander-Universität Erlangen-Nürnberg, Institute for Economics (FAU discussion papers in economics, no. 2016/03).

Kühnreich, Katika (2018): Soziale Kontrolle 4.0? Chinas Social Credit Systems. In: *Blätter für deutsche und internationale Politik* (07), S. 49–60.

Kurzweil, Ray (2013): How to create a mind. The secret of human thought revealed. New York, NY: Penguin Books.

Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J. et al. (2001): Initial sequencing and analysis of the human genome. In: *Nature* 409, S. 860–921. DOI: 10.1038/35057062.

Lane, H. Chad (2015): Enhancing Informal Learning Experiences with Affect-Aware Technologies. In: Rafael Calvo, Sidney D'Mello, Jonathan Gratch und Arvid Kappas (Hg.): The Oxford Handbook of Affective Computing. Oxford, New York: Oxford University Press, S. 435–445.

Lange, Steffen; Santarius, Tilman (2020): Smart green world? Making digitalization work for sustainability. Milton Park, Abingdon, Oxon, New York, NY: Routledge (Routledge studies in sustainability). Online verfügbar unter https://www.taylorfrancis.com/books/9781003030881.

Lankau, Ralf (2019): Der bildungsferne Campus. In: *Frankfurter Allgemeine Zeitung*, 05.10.2019. Online verfügbar unter https://www.faz.net/aktuell/karriere-hochschule/hoersaal/digitalisierung-der-bildungsferne-campus-16411188.html?printPagedArticle=true#pageIndex_2, zuletzt geprüft am 26.09.2020.

Latour, Bruno (2005): Reassembling the Social – An Introduction to Actor-Network-Theory. Oxford: Oxford University Press.

Lauchenauer, David (2018): Hyperkonnektivität - warum dieser Trend wichtig ist. Myfactory. Online verfügbar unter https://www.myfactory.com/blogbeitrag/Hyperkonnektivit%C3%A4t_warum_dieser_Trend_wichtig_ist.aspx, zuletzt aktualisiert am 02.03.2018, zuletzt geprüft am 11.09.2ß19.

Lebreton, L.; Slat, B.; Ferrari, F.; Sainte-Rose, B.; Aitken, J.; Marthouse, R. et al. (2018): Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic. In: *Scientific Reports* 8 (1), S. 4666. DOI: 10.1038/s41598-018-22939-w.

Lenzen, Manuela (2002a): Natürliche und künstliche Intelligenz. Einführung in die Kognitionswissenschaft. Frankfurt/Main: Campus-Verl. (Campus-Einführungen). Online verfügbar unter http://www.sub.uni-hamburg.de/ebook/ebook.php?act=b&cid=955.

Lenzen, Manuela (2002b): Natürliche und künstliche Intelligenz. Einführung in die Kognitionswissenschaft. Frankfurt: Campus Verlag (Campus Einführungen).

Lernende Systeme- Die Plattform für künstliche Intelligenz (Hg.) (o.J.): KI-Anwendungsszenarien. Online verfügbar unter https://www.plattform-lernende-systeme.de/anwendungsszenarien.html, zuletzt geprüft am 14.08.2019.

Leung, Michael K. K.; Delong, Andrew; Alipanahi, Babak; Frey, Brendan J. (2016): Machine Learning in Genomic Medicine. A Review of Computational Problems and Data Sets. In: *Proc. IEEE* 104 (1), S. 176–197. DOI: 10.1109/JPROC.2015.2494198.

Lewis, Tim (2019): AI can read your emotions. Should it? Advertisers, tech giants and border forces are using face tracking software to monitor our moods - whether we like it or not. In: *The Guardian*, 17.08.2019. Online

verfügbar unter https://www.theguardian.com/technology/2019/aug/17/emotion-ai-artificial-intelligence-mood-realeyes-amazon-facebook-emotient, zuletzt geprüft am 02.02.2021.

Li, Kaifu (2018): AI superpowers. China, Silicon Valley, and the new world order. Boston: Houghton Mifflin Harcourt. Online verfügbar unter http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=1873289.

Lindner, Roland (2015): Wie geht es mit Googles Datenbrille weiter? Hg. v. Frankfurter Allgemeine Zeitung. Online verfügbar unter https://www.faz.net/aktuell/wirtschaft/netzwirtschaft/google/verkauf-von-google-glass-wird-eingestellt-13374767.html, zuletzt geprüft am 12.08.2019.

Lippert, Jan (2018): Dynamische Preisoptimierung im Handel. Die Potentiale automatisierter Preisfindungsverfahren. prudsys. Online verfügbar unter https://prudsys.de/wissen/whitepaper-dynamic-pricing/, zuletzt geprüft am 26.11.2021.

Loh, Janina (2018): Trans- und Posthumanismus zur Einführung. Hamburg: Junius (Zur Einführung).

Loh, Janina (2019): Über Liebe, Freundschaft und Sex mit nicht-menschlichen Wesen. In: Lukas Rehm, Lisa Charlotte Friedrich und Jim Igor Kallenberg (Hg.): Castor&&Pollux. Hofheim am Taunus: Wolke, S. 30–33.

Lopez-Carral, Hector; Santos-Pata, Diogo; Zucca, Riccardo; Verschure, Paul F.M.J. (2019): How you type is what you type: Keystroke dynamics correlate with affective content. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). Cambridge, United Kingdom, 03.09.2019 - 06.09.2019: IEEE, S. 1–5.

Macovei, Diana (2017): APIS, the pollinator drone - Presented by Anthony Van der Pluijm & Aleksandar Petrov, Delft University of Technology. Online verfügbar unter https://smartfarmingconference.com/speaker/apis-pollinator-drone-presented-anthony-van-der-pluijm-aleksandar-petrov-delft-university-technology/, zuletzt geprüft am 14.08.2019.

Magrabi, Amadeus; Bach, Joscha (2013): Entscheidungsfindung. In: Achim Stephan und Sven Walter (Hg.): Handbuch Kognitionswissenschaft. Stuttgart, Weimar: J.B. Metzler, S. 274–289.

Mainzer, Klaus (2016a): Künstliche Intelligenz - Wann übernehmen die Maschinen? Berlin, Heidelberg: Springer (Technik im Fokus). Online verfügbar unter http://ebooks.ciando.com/book/index.cfm/bok_id/2112686.

Mainzer, Klaus (2016b): Künstliche Intelligenz – Wann übernehmen die Maschinen? Berlin, Heidelberg: Springer Berlin Heidelberg.

Malter, Bettina; Steeger, Gesa (2019): Der letzte Schatz. In: *Die Zeit* (48), S. 49–50.

Manzeschke, Arne; Assadi, Galia (2019): Emotionen in der Mensch-Maschine Interaktion. In: Kevin Liggieri und Oliver Müller (Hg.): Mensch-Maschine-Interaktion. Handbuch zu Geschichte – Kultur – Ethik. Heidelberg: Metzler, S. 165–171.

Marscheider-Weidemann, Frank; Langkau, Sabine; Hummen, Torsten; Erdmann, Lorenz; Espinoza, Luis Tercero (2016): Rohstoffe für Zukunftstechnologien 2016. Hg. v. Deutsche Rohstoffagentur. Online verfügbar unter https://www.isi.fraunhofer.de/content/dam/isi/dokumente/ccn/2016/Studie_Zukunftstechnologien-2016.pdf, zuletzt geprüft am 25.11.2021.

McMullan, Thomas (2019): How swarming drones will change warfare. Online verfügbar unter https://www.bbc.com/news/technology-47555588, zuletzt geprüft am 14.08.2019.

McStay, Andrew (2020): Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. In: *Big Data & Society* 7 (1), 205395172090438. DOI: 10.1177/2053951720904386.

McStay, Andrew; Urquhart, Lachlan (2019): 'This time with feeling?' Assessing EU data governance implications of out of home appraisal based emotional AI. In: *FM*. DOI: 10.5210/fm.v24i10.9457.

Meidert, Ursula; Scheermesser, Mandy; Prieur, Yvonne; Hegyi, Stefan; Stockinger, Kurt; Eyyi, Gabriel et al. (2018): Quantified Self - Schnittstelle zwischen Lifestyle und Medizin. -. 1. Aufl. Zürich: vdf Hochschulverlag (TA-Swiss).

Meisch, Simon; Bossert, Leonie; Voget-Kleschin, Lieske (2020): Landkarte für mehr Fahrvergnügen auf dem Weg zur Suffizienz. In: Leonie Bossert, Lieske Voget-Kleschin und Simon Meisch (Hg.): Damit gutes Leben mit der Natur einfacher wird – Suffizienzpolitik für Naturbewahrung. Marburg: Metropolis, S. 9–31.

Meisch, Simon; Brohmann, Bettina; Kerr, Matthias; Potthast, Thomas (2018): Mengenproblematik: Wenn individuelle Entscheidungsfreiheit (scheinbar) mit der Nachhaltigkeit in Konflikt gerät. Umweltbundesamt Texte 113/2018. Dessau-Roßlau. Online verfügbar unter https://www.umweltbundesamt.de/publikationen/mengenproblematik-wenn-individuelle, zuletzt geprüft am 10.11.2021.

Meyer, Bernd; Ahlert, Gerd; Diefenbacher, Hans; Zieschank, Roland; Nutzinger, Hans (2013): Eckpunkte eines ökologisch tragfähigen Wohlfahrtskonzepts. GWS Research Report 2013/1. Hg. v. Gesellschaft für Wirtschaftliche Strukturforschung mbH. Online verfügbar unter http://www.gws-os.com/discussionpapers/gws-researchreport13-1, zuletzt geprüft am 10.11.2021.

Mielczarek, Michael (2018): Autonomous robot for planting trees to assist in environmental protection. Hg. v. Autonomic Vehicles. Online verfügbar unter https://autonomicvehicles.eu/2018/11/01/autonomous-robot-planting-trees-assist-environmental-protection/, zuletzt aktualisiert am 01.11.2018.

Mieth, Dietmar (2002): Was wollen wir können? Ethik im Zeitalter der Biotechnik. Freiburg i.Br: Herder.

Misselhorn, Catrin (2018): Grundfragen der Maschinenethik. Reclams Universal-Bibliothek. Reclam Verlag.

Monchalin, Eric; Purswani, Purshottam; van Rijn, Do (2019): Swarm Intelligence. Concept, vision and application. Hg. v. Atos. Atos. Online verfügbar unter https://atos.net/wp-content/uploads/2019/07/atos-swarm-intelligence-white-paper.pdf, zuletzt geprüft am 05.11.2019.

Mont, Oksana; Lehner, Matthias; Heiskanen, Eva (2014): Nudging. A tool for sustainable behaviour? Stockholm: Swedish Environmental Protection Agency (Rapport / Naturvårdsverket, 6643).

Mück, Julia E.; Ünal, Barış; Butt, Haider; Yetisen, Ali K. (2019): Market and Patent Analyses of Wearables in Medicine. In: *Trends in biotechnology* 37 (6), S. 563–566. DOI: 10.1016/j.tibtech.2019.02.001.

Mueller, John Paul; Massaron, Luca (2018): Artificial intelligence for dummies. Hoboken, NJ: Wiley (Learning made easy).

Müller-Mall, Sabine (2020): Freiheit und Kalkül. Die Politik der Algorithmen. Ditzingen: Reclam ([Was bedeutet das alles?]).

Muraca, Barbara (2020): Für eine Dekolonialisierung des Anthopozändiskurses: Diagnosen, Protagonisten und Transformationsszenarien. In: Frank Adloff und Sighard Neckel (Hg.): Gesellschaftstheorie im Anthropozän. Frankfurt am Main/New York: Campus Verlag (Zukünfte der Nachhaltigkeit, 1), S. 169–189.

Nadin, Mihai (2019): Machine intelligence. A chimera. In: *AI & SOCIETY* 34 (2), S. 215–242. DOI: 10.1007/s00146-018-0842-8.

NASA (Hg.) (2019): Explore Budget Estimates. FY 2020. Briefing book. Online verfügbar unter https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjQvfvjs47hAhVSmbQKHaRXBcUQFjAAegQIABAC&url=https%3A%2F%2Fwww.nasa.gov%2Fsites%2Fdefault%2Ffiles%2Fatoms%2Ffiles%2Ffy2020_summary_budget_brief.pdf&usg=AOvVaw2wBboUXLq1Y4rvXfnFUQnE, zuletzt geprüft am 26.11.2021.

Nassehi, Armin (2019): Muster. Eine Theorie der digitalen Gesellschaft: CH Beck.

Nemitz, Paul; Pfeffer, Matthias (2020): Prinzip Mensch. Macht, Freiheit und Demokratie im Zeitalter der Künstlichen Intelligenz. Berlin: Dietz.

Nicholson, Chris (2019): Strong AI, Weak AI & Superintelligence. Hg. v. Skymind. Online verfügbar unter https://skymind.com/wiki/strong-ai-general-ai, zuletzt aktualisiert am 09.08.2019, zuletzt geprüft am 12.08.2019.

Nida-Rümelin, Julian; Weidenfeld, Nathalie (2018): Digitaler Humanismus. Eine Ethik für das Zeitalter der künstlichen Intelligenz. 2. Aufl. München: Piper.

Nordmann, Alfred (2007): If and Then: A Critique of Speculative NanoEthics. In: *Nanoethics* 1 (1), S. 31–46. DOI: 10.1007/s11569-007-0007-6.

Nussbaum, Martha Craven (2000): Women and human development. The capabilities approach. Cambridge, New York: Cambridge University Press.

Nussbaum, Martha Craven (2010): Die Grenzen der Gerechtigkeit. Behinderung, Nationalität und Spezieszugehörigkeit. Berlin: Suhrkamp.

OECD (Hg.) (2018): AI: Intelligent Machines, Smart Policies. Conference Summaries (OECD Digital Economy Papers, 270). Online verfügbar unter https://www.oecd-ilibrary.org/docserver/f1a650d9-en.pdf?expires=1637927138&id=id&accname=guest&checksum=144F6A72844773BF55E2C1B882660F30, zuletzt geprüft am 25.11.2021.

Offe, Claus (2012): Whose good is the common good? In: *Philosophy & Social Criticism* 38 (7), S. 665–684. DOI: 10.1177/0191453712447770.

Opiela, Nicole; Kar, Rhesa Mahabbat; Thapa, Basanta; Weber, Mike (2018): Exekutive KI 2030. VIER ZUKUNFTSSZENARIEN FÜR KÜNSTLICHE INTELLIGENZ IN DER ÖFFENTLICHEN VERWALTUNG. Hg. v. Kompetenzzentrum Öffentliche ITFraunhofer-Institut für Offene Kommunikationssysteme FOKUS. Online verfügbar unter https://cdn0.scrvt.com/fokus/08ca2f8c31e340ab/545e16e3df50/Exekutive-KI-2030---Vier-Zukunftsszenarien-f-r-K-nstliche-Intelligenz-in-der--ffentlichen-Verwaltung.pdf, zuletzt geprüft am 24.11.2021.

Ott, Konrad (1996): Zum Verhältnis naturethischer Argumente zu praktischen Naturschutzmaßnahmen unter besonderer Berücksichtigung der Abwägungsproblematik. In: Hans G. Nutzinger (Hg.): Naturschutz – Ethik – Ökonomie. Theoretische Begründungen und praktische Konsequenzen. Marburg: Metropolis, S. 93–134.

Ott, Konrad (2007): Zur ethischen Begründung des Schutzes von Biodiversität. In: Thomas Potthast (Hg.): Biodiversität – Schlüsselbegriff des Naturschutzes im 21. Jahrhundert? Bonn: BfN, S. 89–124.

Ott, Konrad; Dierks, Jan; Voget-Kleschin, Lieske (Hg.) (2016): Handbuch Umweltethik. Stuttgart: Metzler.

Ott, Konrad; Döring, Ralf (2011): Theorie und Praxis Starker Nachhaltigkeit. Marburg: Metropolis.

Panetta, Kasey (2019): 5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, 2018. Online verfügbar unter https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018, zuletzt geprüft am 26.11.2021.

Pasquinelli, Matteo (2019): How a Machine Learns and Fails – A Grammar of Error for Artificial Intelligence. In: *Spheres* (5), S. 1–18.

Pavliscak, Pamela (2017): Designing for Happiness. 1st edition: O'Reilly Media, Inc.

Persson, Jen (2021): A Day-in-the-Life of a Datafied Child – Observations and Theses. In: Ingrid Stapf, Regina Ammicht Quinn, Michael Friedewald, Jessica Heesen und Nicole Krämer (Hg.): Aufwachsen in überwachten Umgebungen. Interdisziplinäre Positionen zu Privatheit und Datenschutz in Kindheit und Jugend. Baden-Baden: Nomos, S. 295–311.

Petermann, Thomas; Grünwald, Reinhard (2011): Stand und Perspektiven der militärischen Nutzung unbemannter Systeme. Endbericht zum TA-Projekt. Berlin: TAB (Arbeitsbericht / Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag, 144).

Plessner, Helmuth (2003): Conditio Humana. Gesammelte Schriften VIII. 1. Auflage. Frankfurt a.M.: Suhrkamp.

Potthast, Thomas (2014): The Values of Biodiversity. In: Dirk Lanzerath und Minou Bernadette Friele (Hg.): Concepts and Values in Biodiversity. London: Routledge, S. 131–146.

Potthast, Thomas (2015): Ethics and Sustainability Science beyond Hume, Moore and Weber – Taking Epistemic-Moral Hybrids Seriously. In: Simon Meisch, Johannes Lundershausen, Leonie Bossert und Marcus Rockoff (Hg.): Ethics of Science in the Research for Sustainable Development. Baden-Baden: Nomos, S. 129–152.

Potthast, Thomas (2016): Wildnis, Evolution, Prozessschutz. In: Konrad Ott, Jan Dierks und Lieske Voget-Kleschin (Hg.): Handbuch Umweltethik. Stuttgart: Metzler, S. 31–36.

Potts, S. G.; Neumann, P.; Vaissière, B.; Vereecken, N. J. (2018): Robotic bees for crop pollination. Why drones cannot replace biodiversity. In: *Science of the Total Environment* 642, S. 665–667.

Prüfer, Tillmann (2019): Sie kommen. Hg. v. Zeit (21). Online verfügbar unter https://www.zeit.de/zeit-magazin/2019/21/roboter-mensch-technologie-robotik-privatleben/komplettansicht, zuletzt aktualisiert am 18.05.2019, zuletzt geprüft am 12.08.2019.

Pschera, Alexander (2016): Das Internet der Tiere: Natur 4.0 und die conditio humana. Medien der Natur. In: *ZMK* 7 (2), S. 111–124. DOI: 10.28937/1000107556.

Ramge, Thomas (2020): Augmented Intelligence. Wie wir mit Daten und KI besser entscheiden. Ditzingen: Reclam (Was bedeutet das alles?).

Raworth, Kate (2018): Die Donut-Ökonomie. Endlich ein Wirtschaftsmodell, das den Planeten nicht zerstört. Unter Mitarbeit von Hans Freundl und Sigrid Schmid. München: Hanser, Carl.

Rehm, Heidi L. (2017): Evolving health care through personal genomics. In: *Nature reviews. Genetics* 18 (4), S. 259–267. DOI: 10.1038/nrg.2016.162.

Remotti, Luca Alessandro; Damvakeraki, Tonia; Nioras, Alexandros; Britzolakis, Stella; Erdmann, Lorenz; Cuhls, Kerstin et al. (2016): European value changes. Signals, drivers, and impact on EU research and innovation policies : final report. Luxembourg: Publications Office.

Reply AG (Hg.) (o.J.): Die Evolution des Consumer IOT. Online verfügbar unter https://www.reply.com/de/topics/internet-of-things/the-evolution-of-the-consumer-internet-of-things, zuletzt geprüft am 12.08.2019.

Röcke, Anja (2017): (Selbst)Optimierung. Eine soziologische Bestandsaufnahme. In: *Berlin J Soziol* 27 (2), S. 319–335. DOI: 10.1007/s11609-017-0338-2.

Rockström, J.; Steffen, W.; Noone, K.; Persson, Å.; Chapin, F. S.; Lambin, E. et al. (2009): Planetary Boundaries: Exploring a Safe Operating Space for Humanity. In: *Ecology and Society* 14 (2). Online verfügbar unter https://www.ecologyandsociety.org/vol14/iss2/art32/.

Roelfsema, Pieter R.; Denys, Damiaan; Klink, P. Christiaan (2018): Mind Reading and Writing. The Future of Neurotechnology. In: *Trends in cognitive sciences* 22 (7), S. 598–610. DOI: 10.1016/j.tics.2018.04.001.

Rosol, Christoph (2019): 1948. In: Katrin Klingan und Christoph Rosol (Hg.): Technosphäre. Berlin: Matthes & Seitz (Bibliothek 100 Jahre Gegenwart).

Salter, Brian; Salter, Charlotte (2017): Controlling new knowledge. Genomic science, governance and the politics of bioinformatics. In: *Social studies of science* 47 (2), S. 263–287. DOI: 10.1177/0306312716681210.

Salzig, Christoph (2016): Was ist Big Data? – Eine Definition mit fünf V. Hg. v. The unbelievable Machine Company. Online verfügbar unter https://blog.unbelievable-machine.com/was-ist-big-data-definition-f%C3%BCnf-v, zuletzt geprüft am 14.08.2019.

Sandoval, Lysette Maurice N. (2019): UNHQ Calls For Floating Cities Due To Climate Change. Hg. v. Science Times. Online verfügbar unter https://www.sciencetimes.com/articles/19777/20190408/unhq-calls-for-floating-cities-due-to-climate-change.htm, zuletzt geprüft am 14.08.2019.

Scharre, Paul (2018): How swarming will change warfare. In: *Bulletin of the Atomic Scientists* 74 (6), S. 385–389. DOI: 10.1080/00963402.2018.1533209.

Scheermesser, Mandy; Meidert, Ursula; Evers-Wölk, Michaela; Prieur, Yvonne; Hegyi, Stefan; Becker, Heidrun (2018): Die digitale Selbstvermessung in Lifestyle und Medizin. Eine Studie zur Technikfolgenabschätzung. In: *TATuP* 27 (3), S. 57–62. DOI: 10.14512/tatup.27.3.57.

Schmid, Ute; Funke, Joachim (2013): Kreativität und Problemlösen. In: Achim Stephan und Sven Walter (Hg.): Handbuch Kognitionswissenschaft. Stuttgart, Weimar: J.B. Metzler, S. 335–343.

Schnabel, Ulrich (2016): Die Vermessung der Gefühle. In: *Die Zeit* 43, 2016.

Schneider, Werner (2020): Sterben und Tod. In: Martina Heßler und Kevin Liggieri (Hg.): Technikanthropologie. Handbuch für Wissenschaft und Studium. Baden-Baden: Nomos (edition sigma), S. 421–429.

Schnurr, Maria; Glockner, Holger; Berg, Holger; Schipperges, Michael (2018): Erfolgsbedingungen für Systemsprünge und Leitbilder einer ressourcenleichten Gesellschaft. Band 3: Leitbilder einer ressourcenleichten Gesellschaft Abschlussbericht. Hg. v. Umweltbundesamt (Texte, 86). Online verfügbar unter https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2018-10-23_texte_85-2018_ressourcenleichte-gesellschaft_band3.pdf, zuletzt geprüft am 25.11.2021.

Schröter, Welf (2019): Auf dem Weg zum "mitbestimmten Algorithmus". Warum der Begriff "KI" nicht als "künstliche" sondern nur als "kleine" oder "keine Intelligenz" ausgeschrieben werden sollte. In: latenz. Journal für Philosophie und Gesellschaft, Arbeit und Technik, Kunst und Kultur (Hg.): Der Künstliche Mensch. Menschenbilder im 21. Jahrhundert, Bd. 04. Unter Mitarbeit von Irene Scherer und Welf Schröter. Mössingen-Talheim: talheimer (04), S. 109–116.

Senanayake, S. G. J. N. (2006): INDIGENOUS KNOWLEDGE AS A KEY TO SUSTAINABLE DEVELOPMENT. In: *The Journal of Agricultural Sciences* 2 (1), S. 87–94.

Settele, Joseph (2020): Die Triple-Krise: Artensterben, Klimawandel, Pandemien. Hamburg: Edel.

Seuss, Dominik; Dieckmann, Anja; Hassan, Teena; Garbas, Jens-Uwe; Ellgring, Johann Heinrich; Mortillaro, Marcello; Scherer, Klaus (2019): Emotion Expression from Different Angles: A Video Database for Facial Expressions of Actors Shot by a Camera Array. In: 2019 8th International Conference on Affective Computing, S. 35–41.

Simpson, Campbell (2018): Skygrow's blue-sky vision: using tree-planting robots to fight climate change. Online verfügbar unter https://exchange.telstra.com.au/skygrow-muru-d-climate-change/, zuletzt aktualisiert am 18.05.2018, zuletzt geprüft am 06.11.2019.

Specia, Lucia; Harris, Kim; Blain, Frédéric; Burchardt, Aljoscha; Macketanz, Vivien; Skadiņa, Inguna et al. (2017): Translation Quality and Productivity. A Study on Rich Morphology Languages, S. 55–71.

Spencer, Christine; Moore, Daniel; McKeown, Gary; Rutherford, Lucy; Morrison, Gawain (2019): Context matters: protocol ordering effects on physiological arousal and experienced stress during a simulated driving task. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). Cambridge, United Kingdom, 03.09.2019 - 06.09.2019: IEEE, S. 1–7.

Spiekermann, Sarah (2019): Digitale Ethik. Ein Wertesystem für das 21. Jahrhundert. München: Droemer.

Steffen, W.; Richardson, K.; Rockström, J.; Cornell, S. E.; Fetzer, I.; Bennett, E. et al. (2015): Planetary boundaries. Guiding human development on a changing planet. In: *Science* 347 (6223).

Steffen, Will; Broadgate, Wendy; Deutsch, Lisa; Gaffney, Owen; Ludwig, Cornelia (2018): The trajectory of the Anthropocene. The Great Acceleration. In: *Anthropocene review*.

Stephan, Achim; Walter, Sven (Hg.) (2013): Handbuch Kognitionswissenschaft. Stuttgart, Weimar: J.B. Metzler.

Stieler, Wolfgang (2018): Die Geister, die wir rufen. In: *Technology review* (6), S. 69.

Stone, Peter; Brooks, Rodney; Bynjolfsson, Erik; Calo, ryan; Etzioni, Oren; Hager, Greg et al. (2016): Artificial Intelligence and Life in 2030." One Hundred Year Study on Artificial Intelligence. Report of the 2015-2016 Study Panel. Hg. v. Stanford Unviversity. Stanford. Online verfügbar unter https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/AI100Report_MT_10.pdf, zuletzt geprüft am 25.11.2021.

Taylor, Paul W. (1989): Respect for Nature. A Theory of Environmental Ethics. New Jersey: Princeton University Press.

Tenzer, F. (2019): Anzahl der Mobilfunkanschlüsse weltweit von 1993 bis 2018 (in Millionen). Hg. v. Deutsches Statistisches Bundesamt. Online verfügbar unter https://de.statista.com/statistik/daten/studie/2995/umfrage/entwicklung-der-weltweiten-mobilfunkteilnehmer-seit-1993/, zuletzt aktualisiert am 09.08.2019, zuletzt geprüft am 12.08.2019.

Thammasan, Nattapong; Stuldreher, Ivo; Wismeijer, Dagmar; Poel, Mannes; van Erp, Jan; Brouwer, Anne-Marie (2019): A novel, simple and objective method to detect movement artefacts in electrodermal activity. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). Cambridge, United Kingdom, 03.09.2019 - 06.09.2019: IEEE, S. 1–7.

The Economist (Hg.) (2019): Drones need to be encouraged, and people protected. Online verfügbar unter https://www.economist.com/leaders/2019/01/26/drones-need-to-be-encouraged-and-people-protected, zuletzt geprüft am 14.08.2019.

Topol, Eric J. (2019): High-performance medicine. The convergence of human and artificial intelligence. In: *Nature medicine* 25 (1), S. 44–56. DOI: 10.1038/s41591-018-0300-7.

Torres, Phil (2019): Facing disaster: the great challenges framework. In: *Foresight* 21 (1), S. 4–34. Online verfügbar unter https://www.emeraldinsight.com/doi/abs/10.1108/FS-04-2018-0040.

UN Habitat (Hg.) (2019): Roundtable on floating cities at UNHQ calls for innovation to benefit all. Online verfügbar unter https://unhabitat.org/roundtable-on-floating-cities-at-unhq-calls-for-innovation-to-benefit-all/, zuletzt geprüft am 14.08.2019.

Veenhoff, Sylvia (2019): Konsum 4.0 Chancen und Risiken für die Umwelt. Stakeholderdialog im BMU. Fraunhofer ISI. Berlin, 28.10.2019.

Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G. et al. (2001): The sequence of the human genome. In: *Science* 291 (5507), S. 1304–1351. DOI: 10.1126/science.1058040.

Wainberg, Michael; Merico, Daniele; Delong, Andrew; Frey, Brendan J. (2018): Deep learning in biomedicine. In: *Nature Biotechnology* 36, 829 EP -. DOI: 10.1038/nbt.4233.

Warnke, Philine; Cuhls, Kerstin; Daniel, Lea; Gheorghiu, Radu; Andreescu, Liviu; Dragomir, Bianca et al. (2019): 100 Radical Innovation Breakthroughs for the future. The Radical Innovation Breakthrough Inquirer. Fraunhofer ISI; Institutul de Prospectiva; University of Turku. Online verfügbar unter https://publica.fraunhofer.de/eprints/urn_nbn_de_0011-n-5491363.pdf, zuletzt geprüft am 26.11.2021.

Waterborne Technology Platform (Hg.) (2019): Oceans and seas as source of natural resources. Vision. EU. Online verfügbar unter https://www.waterborne.eu/vision/oceans-and-seas-as-a-source-of-natural-resources.

WBGU (2011): Welt im Wandel Gesellschaftsvertrag für eine Große Transformation. 2. veränderte Auflage. Berlin.

WBGU (2013): Welt im Wandel Menschheitserbe Meer. Unter Mitarbeit von Hans Joachim Schellnhuber, Dirk Messner, Claus Leggewie, Reinhold Leinfelder, Nebojsa Nakicenovic, Stefan Rahmstorf et al. Berlin.

WBGU (2019a): Unsere gemeinsame digitale Zukunft. Zusammenfassung. 1. Auflage. Berlin: Wissenschaftlicher Beirat d. Bundesregierung Globale Umweltveränderungen.

WBGU (2019b): Unsere gemeinsame digitale Zukunft. Hauptgutachten. Hg. v. Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen (WBGU). Berlin. Online verfügbar unter https://www.wbgu.de/fileadmin/user_upload/wbgu/publikationen/hauptgutachten/hg2019/pdf/wbgu_hg2019.pd, zuletzt geprüft am 10.11.21.

WCED (1987): Our common future. [Nairobi]: [United Nations Environment Programme].

Webb, Sarah (2018): Deep learning for biology. In: *Nature* 554 (7693), S. 555–557. DOI: 10.1038/d41586-018-02174-z.

Weber, Jutta (2020): MenschMaschine. In: Martina Heßler und Kevin Liggieri (Hg.): Technikanthropologie. Handbuch für Wissenschaft und Studium. Baden-Baden: Nomos (edition sigma), S. 318–322.

Whittaker, Meredith et al. (2018): AI Now Report 2018. Online verfügbar unter https://ainowinstitute.org/AI_Now_2018_Report.pdf, zuletzt geprüft am 15.11.2021.

Wholey, Mike (2018): Agtech and the Connected Farm. Hg. v. CB Insights. Online verfügbar unter https://www.cbinsights.com/research/briefing/ag-tech-trends-connected-farming/, zuletzt geprüft am 24.11.2021.

Williams, Adam (2019): The BIG plan to create a floating city for 10,000 people. Hg. v. New Atlas. Online verfügbar unter https://newatlas.com/big-oceanix-city/59175/, zuletzt geprüft am 14.08.2019.

Williams, Bernard (1985): Ethics and the limits of philosophy. London, New York: Routledge (Routledge Classics).

Wilson, Edward O. (1984): Biophilia. The human bond with other species. Cambridge (MA): Harvard University Press.

Wilutzky, Wendy; Stephan, Achim; Walter, Sven (2013): Situierte Affektivität. In: Achim Stephan und Sven Walter (Hg.): Handbuch Kognitionswissenschaft. Stuttgart, Weimar: J.B. Metzler, S. 552–560.

Wissenschaftlicher Dienst Bundestag (Hg.) (2012): Aktueller Begriff Internet der Dinge (19).

World Economic Forum; PwC; Stanford Woods Institute for the Environment (2018): Harnessing Artificial Intelligence for the Earth. Online verfügbar unter https://www3.weforum.org/docs/Harnessing_Artificial_Intelligence_for_the_Earth_report_2018.pdf, zuletzt geprüft am 26.11.2021.

Wyss Institute (Hg.) (o.J.): Autonomous Flying Mircorobots (RoboBees). Online verfügbar unter https://wyss.harvard.edu/technology/autonomous-flying-microrobots-robobees/, zuletzt geprüft am 14.08.2019.

YouGov (04.02.2019): Social Scoring: Zwei von fünf Deutschen würden gerne das Verhalten ihrer Mitmenschen bewerten. Online verfügbar unter https://yougov.de/news/2019/02/04/social-scoring-zwei-von-funf-deutschen-wurden-gern/, zuletzt geprüft am 18.09.2019.

Zhou, Jian; Theesfeld, Chandra L.; Yao, Kevin; Chen, Kathleen M.; Wong, Aaron K.; Troyanskaya, Olga G. (2018): Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. In: *Nature genetics* 50 (8), S. 1171–1179. DOI: 10.1038/s41588-018-0160-6.

Zou, James; Huss, Mikael; Abid, Abubakar; Mohammadi, Pejman; Torkamani, Ali; Telenti, Amalio (2019): A primer on deep learning in genomics. In: *Nature genetics* 51 (1), S. 12–18. DOI: 10.1038/s41588-018-0295-5.

Zuboff, Shoashana (2018): Der dressierte Mensch. In: *Blätter für deutsche und internationale Politik* (11), S. 101–111.

Zweck, Axel; Holtmannspötter, Dirk; Braun, Matthias; Erdmann, Lorenz; Hirt, Michael; Kimpeler, Simone (Hg.) (2015): Geschichten aus der Zukunft 2030. Ergebnisband 3 zur Suchphase von BMBF-Foresight Zyklus II. VDI Technologiezentrum. Düsseldorf: VDI Technologiezentrum GmbH (Zukünftige Technologien, Nr. 102). Online verfügbar unter https://edocs.tib.eu/files/e01fb17/898439329.pdf.

A Anhang: Eingebundene Expertinnen und Experten

A.1 Beirat

Tabelle 9: Personen im Beirat

Person	Einrichtung
Dr. Aljoscha Burchardt	Deutsches Forschungszentrum für KI
Prof. Dr. Martin Butz	Universität Tübingen
Prof. Dr. Hans Diefenbacher	Universität Heidelberg und Forschungsstätte der Evangelischen Studiengemeinschaft e.V. (FEST)
Prof. Dr. Marcus Düwell	Utrecht University/Universität Bamberg
Prof. Dr. Dr. Mathias Gutmann	Karlsruher Institut für Technologie
Prof. Dr. Jochen Hinkel	Global Climate Forum
Prof. Dr. Jutta Weber	Universität Paderborn

A.2 Interviewte im Screening

Tabelle 10: Interviewte im Screening

Person	Einrichtung
Christopher Coenen	KIT-ITAS, Karlsruhe
Prof. Dr. Dr. Joachim Funke	Lehrstuhl für Allgemeine und Theoretische Psychologie, Universität Heidelberg
Dr. Jonathan Harth	Universität Witten/Herdecke, Fakultät für Kulturreflexion
Dr. Bruno Gransche	Universität Siegen
Prof. Dr. Dr. Frank Kirchner	Leiter des Forschungsbereich Robotics Innovation Center am DFKI, Bremen
Prof. Dr. Oliver Kohlbacher	Universität Tübingen
Prof. Dr. Antonio Krüger	Leiter des Forschungsbereichs Kognitive Assistenzsysteme DFKI, Saarbrücken
Prof. Dr. Fabian Theis	Institute of Computational Biology, Helmholtz-Zentrum München
Dr. Daniel Walther	FZI House of Living Labs, Karlsruhe