

Raubbau an der Kultur? - Zur Ethik generativer KI im aktuellen Diskurs

Melanie Xu

How to cite:

Melanie Xu, Raubbau an der Kultur? - Zur Ethik generativer KI im aktuellen Diskurs, Masterarbeit am Studiengang "Management von Kultur- und Non-Profit-Organisationen, M.A.", Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau 27.09.2024, Distance and Independent Studies Center (DISC), Klaus Wiegerling

online: www.vordenker.de Neuss 2024, J. Paul (Ed.), ISSN 1619-9324

URL: < https://www.vordenker.de/mxu/mxu_raubbau_kultur_ethik_generativer_ki.pdf >

Copyright Melanie Xu 2024
Citation is mandatory // vordenker.de



Herausragende Masterarbeiten

Studiengang

Management von Kultur- und Non-Profit-Organisationen, M.A.

Masterarbeitstitel

**Raubbau an der Kultur? Zur Ethik generativer KI im
aktuellen Diskurs**

Autor*in

Melanie Xu

R
TU
P

Distance and Independent
Studies Center
DISC

Inhaltsverzeichnis

Inhaltsverzeichnis	2
Abbildungsverzeichnis	4
Abkürzungsverzeichnis	5
Vorbemerkung	7
Glossar	8
1. Einleitung	11
<i>Generative KI aus Nutzer-Perspektive S. 11, Der Hype S. 13, What's new? S. 14, Exponentielle Entwicklung S. 15, Ethik unter Zeitdruck S. 18, Produkte der Zukunft – Kanalisierung der ethischen Diskussion S. 19, Künstler vor den Karren gespannt? S. 20, Fragen an das Kulturmanagement S. 20, Kulturelle Aneignung S. 21, Die Frage nach den Voraussetzungen S. 22</i>	
2. Ethik generativer KI – aktueller Diskussionsstand	23
2.1 Supranationale kollektive Stellungnahmen/Vereinbarungen/Gesetzgebungsprozesse	24
2.1.1 EU AI Act – das mühsame Geschäft der Konsensfindung	25
2.1.2 Die paradoxe „Moratoriumsforderung“	27
2.1.3 PAI Framework – ein „Hauch“ von Medienethik	30
2.1.4 OECD Arbeitspapier – Politikberatung für Demokratien	33
2.2 Wissenschaftliche Studien/kollektive Stellungnahmen in Deutschland	37
2.2.1 Initiative Urheberrecht (IU) – Gegenwehr der deutschen Kunst- und Kreativbranche	37
2.2.2 Forschungsbericht des TAB – der kritische Blick	40
2.2.3 LEAM Machbarkeitsstudie – technologische Souveränität für Deutschland	44
2.2.4 Stellungnahme des Deutschen Ethikrates – Handlungs-urheberschaft und Verantwortung	46
2.2.5 Stellungnahme des Deutschen Kulturrates – Sorge um die Kultur	52
2.3 Stellungnahmen einzelner Autoren	54
2.3.1 Klaus Kornwachs – menschliche Arbeit im KI-Zeitalter	55
2.3.2 Helga Nowotny – Ethik ist keine Checkliste!	58
2.3.3 Jobst Landgrebe und Barry Smith – Algorithmen als ethische Agenten?	61
3. Kritik des Diskurses	63
3.1 Raubbau an der Natur und weitere extraktive Praktiken (Kate Crawford)	65

3.2	Raubbau an der Kultur (gelesen mit <i>Richard A. Rogers</i>)	69
3.3	Kulturelles Recycling (gelesen mit <i>Jean Baudrillard</i>)	72
3.4	Menschliche Autorschaft im Kreuzfeuer „Systematischer Diebstahl in großem Maßstab“ - Klagen in den USA S. 76, „Nightshade“ und „Glaze“ - Guerilla-Taktiken zum Schutz menschlicher Autorschaft S. 78	75
3.5	Fazit und ethische Perspektiven für generative KI	80
	Literaturverzeichnis	84
	Anhang	92
	Eigenständigkeitserklärung	95

Abbildungsverzeichnis

Abb. 1: OECD Grafik „Number of news articles globally on generative AI and related topics“ (S. 17), *Quelle s. Literaturverzeichnis:* OECD publishing (o.V.) 2023

Abb. 2: OECD Grafik „Generative AI-related incidents and hazards reported by reputable news outlets have grown exponentially“ (S. 18), *Quelle s. Literaturverzeichnis:* OECD publishing (o.V.) 2023

Abb. 3: Beispiele für kopierte Stile zweier bekannter Künstler ohne (Bilder in der Mitte) und mit dem Einsatz von *Glaze* (rechts) (im Anhang, S. 92), *Quelle s. Literaturverzeichnis:* Holland 2023b

Abb. 4: Beispiele für mit *Nightshade* infizierte generative KI (im Anhang, S. 93), *Quelle siehe Literaturverzeichnis:* Bremmer 2023

Abkürzungsverzeichnis

Abb.	Abbildung
Abk.	Abkürzung
AGB	Allgemeine Geschäftsbedingungen
AGI	Artificial General Intelligence
AI	Artificial Intelligence
bspw.	beispielsweise
bzw.	beziehungsweise
DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz
d.h.	das heißt
DSM	Digital Single Market
ebd.	ebenda
ELG	European Language Grid
ELRC	European Language Resource Coordination
engl.	englisch
et al.	et alii
ETH	Eidgenössische Technische Hochschule
EU	Europäische Union
GenAI	Generative Artificial Intelligence
ggf.	gegebenenfalls
GPAI	General Purpose Artificial Intelligence
GPT	Generative Pre-trained Transformer
IAIS	Fraunhofer-Institut für intelligente Analyse- und Informationssysteme
i.d.R.	in der Regel
IDS	Leibniz-Institut für Deutsche Sprache
i.e.	id est
IEAI	Institute of Ethics in Artificial Intelligence
IR	Intelligente Roboter
ITAS	Institut für Technikfolgenabschätzung und Systemanalyse
IU	Initiative Urheberrecht

Kap.	Kapitel
KI	Künstliche Intelligenz
KIT	Karlsruher Institut für Technologie
LEAM	Large European AI Models
LLM	Large Language Model
MIT	Massachusetts Institute of Technology
o.ä.	oder ähnliche(s)
OECD	Organisation for Economic Cooperation and Development
o.g.	oben genannt(e)
PAI	Partnership on AI
RLHF	Reinforcement Learning by Human Feedback
S.	Seite
s.	siehe
sog.	sogenannt(e)
TAB	Büro für Technikfolgenabschätzung beim Deutschen Bundestag
u.a.	unter anderem/und andere
UrhG	Gesetz über Urheberrecht und verwandte Schutzrechte
urspr.	ursprünglich
US	United States
USA	United States of America
usw.	und so weiter
u.v.a.	und viele andere
vgl.	vergleiche
WIPO	World Intellectual Property Organisation
z.B.	zum Beispiel

Vorbemerkung

In einigen Fällen wird, bedingt durch das Thema der Arbeit, englischsprachige Literatur im Original zitiert. Die Verwendung des generischen Maskulinums erfolgt aus Gründen der besseren Lesbarkeit. Gemeint sind jedoch immer alle Geschlechter.

Glossar

AGI

Artificial General Intelligence (AGI) bezeichnet eine hypothetische Form künstlicher Intelligenz, die in der Lage wäre, weitgehend autonom sämtliche intellektuellen Aufgaben zu erfüllen, die bisher nur Menschen bewältigen können. Die Entwicklung einer AGI ist Ziel eines Teilbereichs der KI-Forschung. Strittig ist, ob eine AGI auch Bewusstsein und Empfindungsvermögen aufweisen muss.

Artificial Intelligence (s. *Künstliche Intelligenz*)

Big Data

Der Begriff *Big Data* bezeichnet wörtlich große Datenmengen, meint aber auch Technologien zur Verarbeitung und Auswertung großer Datenmengen. *Big Data* wird ferner als Schlagwort für weitere, auch negative Folgen dieser Verarbeitung benutzt. Künstliche Intelligenz ist auf *Big Data* angewiesen.

Black Box KI

Gegenteil von *erklärbarer KI*, deren Entscheidungswege auf Grund der Einfachheit der Systeme nachvollziehbar und transparent sind. *Black Box KI* bezeichnet intransparente und komplexe KI-Systeme, deren Output nicht erklärbar/nachvollziehbar ist, da Algorithmen unzählige nichtlineare Beziehungen zwischen Eingaben und Ausgaben herstellen (sog. „Deep Learning“).

ChatGPT

Am 30. November 2022 veröffentlichte das US-amerikanische Software-Unternehmen *OpenAI* das von ihm entwickelte Sprachmodell GPT-3.5 kostenfrei unter dem Namen *ChatGPT*. Innerhalb von fünf Tagen registrierten sich weltweit eine Million Nutzer, die die Möglichkeit ausprobierten, mit einer KI ohne weitere Computerkenntnisse über eine einfach gestaltete Schnittstelle zu interagieren. *ChatGPT* antwortet auf Texteingaben mit neu generierten Texten und gehört daher zu den sog. *Large Language Models*.

Diskriminative KI

Diskriminative KI-Modelle (*Discriminative AI Models*) klassifizieren Eingabedaten, d.h. sie ordnen Eingabedaten Kategorien zu, die sie zuvor an Hand von Markierungen zu unterscheiden gelernt haben. Diskriminative KI kann keine neuen Inhalte generieren.

Foundation Models

Der Begriff *Foundation Models* (deutsch: Basismodelle oder Grundmodelle) ist wie *General Purpose AI* (GPAI) ein übergeordneter Begriff, der *Large Language Models* (LLMs) und alle generativen KI-Modelle umfasst, die mit sehr großen und vielfältigen Datensätzen vortrainiert wurden, ein entsprechend großes Anwendungsspektrum haben und leicht an ihre jeweilige Aufgabe angepasst werden können.

Generative KI

Generative KI, engl. *Generative AI* („GenAI“), ist ein Sammelbegriff für KI, die im Gegensatz zu diskriminativer KI neue Inhalte generiert, z.B. Texte, Bilder, Audio- und Videoinhalte, Programmcode und mehr. Über eine Nutzer-Schnittstelle reagiert generative KI auf eine Eingabeaufforderung (einen *Prompt*) und liefert den gewünschten neuen Inhalt. Die Anfänge textbasierter Dialogsysteme reichen bis in die 1960er Jahre zurück. So wird z.B. Joseph Weizenbaums Programm „Eliza“ von 1966 als erster „Chatbot“ bezeichnet, obwohl er auf einem vergleichsweise einfachen Programm beruht, das man mit heutigen Chatbots wie *ChatGPT* nicht vergleichen kann. Seit Einführung der *Variational Auto-encoder* (VAEs) 2013 und der *Generative Adversarial Networks* (GANs) 2014, kann generative KI Bilder sowie Video- und Audio-Dateien generieren, die täuschend echt wirken. Mit der 2017 erstmals von *Google* vorgestellten *Transformer*-Technik wurde der Grundstein für die Architektur der *generativen vortrainierten Transformer* (GPT) gelegt, die einen großen Fortschritt in der Verarbeitung natürlicher Sprache darstellen. Die bekanntesten Beispiele für generative KI sind zurzeit der Chatbot *ChatGPT* sowie die Bildgeneratoren *Dall-E* und *Stable Diffusion*. Die neueste Entwicklung sog. „multimodaler KI“ erweitert die Funktionen generativer KI dahingehend, dass Prompts nicht mehr textbasiert sein müssen, sondern in verschiedenen Modalitäten eingegeben werden können, z.B. als Bild- oder Videodateien, und der Output in einer beliebig gewünschten Modalität erfolgt. Beispiele für multimodale generative KI, die u.a. Text aus Bildern extrahieren kann, sind Modelle wie *Gemini* von *Google* und *GPT-4V* von *OpenAI*.

Input (generativer KI)

Mit dem Input generativer KI werden sowohl ihre Trainingsdaten bezeichnet, als auch zuweilen die Eingabeaufforderung (der „Prompt“) an der Schnittstelle einer KI-Anwendung. In der vorliegenden Arbeit sind mit dem Begriff „Input“ immer die Trainingsdaten gemeint, die dem Output gegenübergestellt werden.

Künstliche Intelligenz

Der Begriff Künstliche Intelligenz (KI), engl. *Artificial Intelligence* (AI), wurde 1955 von John McCarthy, einem US-amerikanischen Informatiker, geprägt. Künstliche Intelligenz als Teilgebiet der Informatik bezeichnet heute die Fähigkeit einer Maschine bzw. eines Computersystems, menschliche kognitive Fähigkeiten zu imitieren und Aufgaben zu übernehmen, für die ein Mensch seine Intelligenz einsetzen müsste. Voraussetzung für künstliche Intelligenz sind große Datenmengen (Big Data), mit denen sie trainiert wird. Der Begriff der Künstlichen Intelligenz ist nicht zuletzt deshalb unscharf, weil schon der Begriff menschlicher Intelligenz, an den er sich anlehnt, unterschiedlich definiert wird.

Large Language Models (LLMs)

Large Language Models (sog. große Sprachmodelle) sind generative Text-zu-Text-Modelle. Trainiert mit großen Datenmengen in natürlicher Sprache, vervollständigen sie Texte probabilistisch.

Output (generativer KI)

Als Output generativer KI werden die von ihr generierten neuen Inhalte bezeichnet. Dies können Texte, Programmcode, Musikpartituren, Bilder, Audio- und Videoinhalte usw. sein. Bei letzteren spricht man auch von synthetischen Medien.

Prompt

Als *Prompt* wird der Arbeitsauftrag bezeichnet, mit der ein Nutzer an der Schnittstelle der Anwendung das KI-System zu einem Output auffordert („Eingabeaufforderung“).

Ubiquitous Computing

Der Begriff *Ubiquitous Computing* („Rechnerallgegenwart“) stammt von dem Informatiker Mark Weiser und steht für die Idee, dass Computer als einzelne Endgeräte durch eine „intelligente“ Umgebung abgelöst werden. Diese Idee wird teils durch das sog. „Internet der Dinge“ realisiert, mit welchem der Mensch eine diskrete Unterstützung bei seinen Tätigkeiten erhält, ohne an einer Schnittstelle aktiv eine Maschine „bedienen“ zu müssen.

1. Einleitung

Google and Microsoft are going all in with generative AI as core to their future. There is no „we are still early“ here, trillion dollar companies are shifting their whole strategy and focus. I can't ever recall a technology and strategy shift as fast and meaningful as this.

(Emad Mostaque auf „X“, 21. Januar 2023)

Generative KI aus Nutzer-Perspektive

Künstliche Intelligenz, die Texte schreibt, Bilder malt, Musik komponiert – generative KI ist *der* technologische Hype des Jahres 2023. Wer Anwendungen wie *ChatGPT*, *Dall-E* und Co. nicht zumindest einmal ausprobiert hat, setzt sich dem Vorwurf aus, altmodisch zu sein. Dabei kann der Umgang mit der neuen Technologie – angetrieben von Neugier, der Aussicht auf Arbeitserleichterung oder der Hoffnung auf einen Kreativitätsschub – den Nutzer zum Staunen bringen, ihn vielleicht sogar begeistern. Doch kann er auch ambivalente Eindrücke hinterlassen. Denn im Vergleich zu kulturellen Artefakten, die vom Menschen gemacht und a priori in seine Lebenswelt eingebettet sind, können KI-generierte Inhalte „monadisch“ wirken. Sie haben keinen „echten“ soziokulturellen Kontext – dieser muss nachträglich hergestellt werden. KI-generierte Inhalte entstehen einzeln am Endgerät, auf eine Aufforderung, einen „Prompt“ hin und sind zunächst nur für den Nutzer des Endgeräts sichtbar und verfügbar. Der kreativen Mühe, einen Satz zu formulieren, eine Fotografie zu bearbeiten oder ein Musikstück zu erfinden, steht die Souveränität gegenüber, eine Maschine damit zu beauftragen. Mit der Maschine muss, anders als mit einem menschlichen Auftragnehmer, nicht verhandelt werden. Sie gehorcht bedingungslos, ohne Widerrede, und gleicht einem Zauberapparat, der scheinbar voraussetzungslos und mühelos auswirft, was auch immer gewünscht wird.

Sie übernimmt allerdings auch keine Verantwortung für das Ergebnis, sondern schiebt diese dem Nutzer/Auftraggeber zu. Im harmlosen Fall einer misslungenen „Wunsch-erfüllung“ war der Prompt (die „Wunschformulierung“) vielleicht nicht genau genug. Im gravierenderen Fall, wenn die generative KI beispielsweise „halluziniert“¹, ihr Ergebnis einen Bias offenbart oder ein „Jailbreak“² erfolgreich war, muss das Ergebnis vom Nutzer selbst korrigiert, nachbearbeitet oder notfalls verworfen werden – ein unbefriedi-

¹ Von „Halluzinationen“ wird in diesem Zusammenhang gesprochen, wenn die KI plausibel klingende, aber falsche Tatsachen erfindet.

² Bei einem „Jailbreak“ wird die KI durch einen Prompt dazu gebracht, Inhalte oder Antworten zu generieren, die sie normalerweise vermeidet, weil sie unangemessen, anstößig oder falsch sind.

gender und zeitraubender Vorgang, der mit der versprochenen „instantanen Wunsch-erfüllung“³ nichts mehr zu tun hat. Unverbindlichkeit zieht sich durch den ganzen Prozess der Generierung; es ist alleinige Aufgabe des Nutzers, dem KI-generierten Inhalt Verbindlichkeit zu geben, indem er ihn verifiziert und in einen soziokulturellen Kontext einbettet, wo er dann eine spezifische Funktion erfüllen soll. An dieser Stelle trifft er auf individuelle oder gesellschaftliche Reaktionen, wird möglicherweise relevant, lässt sich aber nicht „festnageln“; seine Autorschaft und Verbindlichkeit bleiben unklar.

Was aber währenddessen verbindlich geschehen ist: Der Nutzer hat die Maschine verbessert, er hat dem Anbieter der KI-Anwendung unentgeltlich Trainingsdaten geliefert – andernfalls hätte er die KI nicht nutzen können. Ein verbindlicher Austausch findet demnach auf der ökonomischen Ebene statt. Abgesehen von der schon angesprochenen „Verantwortungslosigkeit“ der KI, stellt sich bei diesem Geschäftsmodell die Frage, ob der Tauschwert der KI-generierten Inhalte, die offensichtlich laufend verbessert werden müssen, dem der User-generierten Daten entspricht, die ihnen für diese Verbesserung zufließen. Die Entwickler generativer KI gehen jedenfalls offen mit den Schwächen ihrer Entwicklungen um: Sam Altman, CEO von *OpenAI*⁴, der US-amerikanischen Firma, die im November 2022 *ChatGPT* als kostenlosen Textgenerator zur Verfügung stellte, bezeichnete später den Chatbot in einem Interview mit der *New York Times* als „schreckliches Produkt“⁵. Auch Jan Leike, ein deutscher KI-Entwickler und vormaliger Co-Leiter des Alignment-Teams bei *OpenAI*, zeigt sich überrascht von dem Erfolg von *Chat-GPT* und spricht von der Notwendigkeit, „deutlich [zu] machen, dass es sich hier nicht um ein fertiges Produkt handelt“⁶. Tatsächlich findet ein ständiges *work in progress* statt:

„OpenAI hat seit dem Start von ChatGPT beobachten können, wie die Leute den Chatbot benutzen. Die Firma konnte zum ersten Mal live erleben, wie ein großes Sprachmodell funktioniert, wenn es in die Hände von Millionen Nutzern gelangt, die seine Grenzen testen und nach Schwachpunkten suchen. Und seither ist das Team dabei, problematische Inhalte möglichst zu blockieren, sei es nun einen von ChatGPT gedichteten Song über Gottes Liebe zu christlichen Vergewaltigern oder vom Chatbot verfassten Malware-

³ Vgl. Klaus Wieglerling in Bezug auf Medien, die die Wirklichkeit virtualisieren : „Wir sehen die Dinge dann möglicherweise nicht wie sie in ihrer konkreten Widerständigkeit sind, sondern nur noch als Optionen für unsere Gestaltung. Dies entspricht aber nicht der Sicht des Künstlers oder Ingenieurs, diese Sicht ist vielmehr der Idee einer instantanen Wunscherfüllung geschuldet.“ (Wieglerling/Neuser 2013, S. 9)

⁴ *OpenAI Inc.*, kalifornisches Start-up-Unternehmen, im Jahr 2015 als gemeinnütziges Forschungsunternehmen mit Non-Profit-Struktur gegründet, kontrolliert heute seine Tochtergesellschaft *OpenAI LP*, die als sog. „Capped-Profit“-Firma begrenzt Gewinn machen darf. Der KI-Bildgenerator *Dall-E* und der KI-Textgenerator *ChatGPT* sind die bisher bekanntesten Veröffentlichungen des Unternehmens.

⁵ zitiert nach: Donath 2023

⁶ zitiert nach: Heaven 2023. Vgl. auch Klaus Wieglerling: „Keine moderne informatische Anwendung wird in der Entwicklungsabteilung geschaffen und dann distribuiert, sondern befindet sich quasi immer in der ‚Beta-Phase‘: Sie wird entwickelt und gestaltet in Reaktion auf ihre Nutzung.“ (Wieglerling 2018, S. 14)

Code, der Kreditkartennummern stiehlt. Das Team sammelt all diese Beispiele und nutzt sie dann dazu, deren Ausgabe in künftigen Versionen zu vermeiden.“⁷

Der Hype

Millionen Nutzer nehmen demnach ein unfertiges, fehlerhaftes, ethisch bedenkliches Produkt in Kauf und arbeiten spielerisch an seiner Verbesserung mit. Was könnten die Motive für dieses freiwillige Engagement sein? Wie lässt sich der Hype um *ChatGPT* erklären? Im Folgenden seien mögliche Aspekte auf drei Ebenen benannt:

Erstens die individuelle Ebene: Der Philosoph und Medienethiker Klaus Wieglerling weist im Zusammenhang mit informatischen Anwendungen auf die emotionale Disposition von „Mensch-System-Interaktionen“⁸ hin, welche „nicht nur rational und zweckmäßig auf das Notwendige und Nützliche gerichtet“⁹ seien. Mensch-System-Interaktionen ließen „in ihrem erweiterten Gebrauch – der nicht nur Produktivität und Funktionalität kennt, sondern auch Spiel hat, also Momente der Unterhaltung, der lernenden Aneignung und der Selbstgestaltung umfasst – Aufschluss über unsere lebensweltliche Verfassung zu.“¹⁰ Als neues immaterielles Spielzeug oder „Gadget“, das auf Anfrage automatisch „Geist“ produziert, kann *ChatGPT* in einer fortgeschritten digitalisierten Lebenswelt offenbar genauso reüssieren, wie ein beliebiger, gut konstruierter Automat in der „analogen“ Lebenswelt, der Spaß und/oder Arbeitserleichterung auf physischer Ebene verspricht.¹¹ Inwieweit es ratsam ist, das Moment der Anstrengung aus menschlicher, insbesondere geistiger Arbeit zu tilgen, ist eine alltagspraktische, aber auch eine ethische Frage. Für ihre Beantwortung ist ein genauer Blick sowohl auf die Überlastung der Gesellschaft¹² als auch auf die „Entmündigung und Einschränkung von Handlungsmöglichkeiten“¹³, die mit ihrer *Entlastung* einhergehen können, unerlässlich.

Zweitens die gesellschaftliche Ebene, die der Computerlinguist Aljoscha Burchardt betont, wenn er von einer „Demokratisierung light“¹⁴ spricht. Seit Künstliche Intelligenz

⁷ Heaven 2023

⁸ Wieglerling 2018, S. 14

⁹ ebd.

¹⁰ ebd. S. 14/15

¹¹ Hierzu Jean Baudrillard: „Der Automatismus meldet sich als eines unserer Grundbegehren an[...]. Denken wir bloß an unseren ständigen Wunsch, dass „alles von selbst laufe“, dass jedes Ding seine konkrete Verrichtung im Sinne des geringsten Kraftaufwandes erfülle. Die Automatik gewährleistet uns ein großartiges Nichtstun [...]“ (Baudrillard 1968)

¹² „Überlastung ist ein chronischer Zustand komplexer datengetriebener Gesellschaften, in denen ständig neue Datenberge aufgehäuft werden.“ (Wieglerling 2018, S. 8)

¹³ ebd.

¹⁴ „Es ist eine Demokratisierung light, die da gerade in der KI-Welt stattfindet. Man kann sich beteiligen, man kann ausprobieren.“ (Aljoscha Burchardt, zitiert nach: Gökkaya 2022)

ihren kalifornischen „Elfenbeinturm“ verlassen hat und KI-Anwendungen global für jedermann zugänglich sind – vorausgesetzt, man verfügt über ein digitales Endgerät –, ist das „Weltwissen“¹⁵ interaktiv in jeder Sprache und auf jede erdenkliche Weise befragbar, gleich einem generalisierten, säkularen Orakel, das sich bedingungslos freundlich und auskunftsfreudig verhält. Generative KI verspricht so eine Form von intellektueller Inklusion und „Barrierefreiheit“, die die Teilhabe- und Inklusionsversprechen der Digitalisierung noch übertrifft.

Drittens die ökonomische Ebene: Ein wesentliches Zukunftsversprechen betrifft den zu erwartenden großen ökonomischen Wert generativer KI. Die Unternehmensberatung *McKinsey* hält einen jährlichen Produktivitätszuwachs der Weltwirtschaft von 2,6 bis 4,4 Billionen US-Dollar durch den Einsatz generativer KI für möglich.¹⁶ Der Wert, der sich gegenwärtig zwar erst in gesellschaftlichen Teilbereichen, u.a. in der Medizin manifestiert, hat bereits zu einer Art industriellen „Goldgräberstimmung“ geführt.¹⁷

What's new?

Befragt man die Geschichte der Künstlichen Intelligenz, so sind Zukunftsversprechen und damit korrespondierende unverbrüchliche Fortschritts- und Technikgläubigkeit durchgängig prägend. In der Ideologie des Trans- bzw. Posthumanismus findet dieser Technikoptimismus seinen reinsten Ausdruck. Mit jeder Neuentwicklung im Bereich der KI und so auch mit der Entwicklung generativer KI wird von einem Durchbruch bzw. einem disruptiven Ereignis gesprochen.¹⁸ Auf einem als unvermeidbar behaupteten

¹⁵ Vgl. Start-up-Unternehmer Michael Witzel in Bezug auf *ChatGPT*: „Die KI hat gewissermaßen das Weltwissen gespeichert und kann auch Antworten außerhalb der eigentlichen Fragestellung liefern [...]“, zitiert nach: Kroker 2022. Mit „Weltwissen“ ist in diesem Zusammenhang datafiziertes, im Web verfügbares Wissen gemeint; *nicht* datafiziertes Wissen findet keinen Eingang in KI. *ChatGPT* kann insofern nicht den Anspruch erheben, vollständiges „Weltwissen“ zu repräsentieren, wie KI-Marketing oft suggerieren möchte, obwohl der Bot mit unvorstellbaren Datenmengen „gefüttert“ wurde. Ein weiteres Problem stellt die Transparenz in Bezug auf die Auswahl der „Wissensquellen“ dar: „Genau Zahlen und Details veröffentlicht OpenAI nicht; in dem Bot sollen aber unter anderem alle Internetseiten, die Online-Enzyklopädie Wikipedia sowie das komplette Entwicklernetzwerk Github stecken.“ (Hinrich Schütze, Professor für Computerlinguistik an der LMU München, zitiert nach: Kroker 2022.)

¹⁶ siehe McKinsey.com 2023

¹⁷ Im Zuge dieser Versprechen werden ethische Bedenken vielfach als „Produktqualitäts-Probleme“ deklariert, und diese dann als „Kinderkrankheiten“ der KI marginalisiert, die von selbst verschwinden werden.

¹⁸ Vgl. Stefan Waldhauser, Wirtschaftsmathematiker und Hightech-Investor: „Ich bin 55 Jahre alt und durfte schon drei technologische Revolutionen miterleben [PC, World Wide Web und Smartphone, *Anm. d. Verf.*][...] Und nun also die Generative KI. Ich hätte nicht gedacht, dass ich in meinem Leben nochmals einen solchen wahrhaft disruptiven Technologiesprung miterleben würde. [...] Ich war und bin seitdem beeindruckt. Nach Jahrzehnten der Forschung wurde auch dank immer schnellerer Rechnerarchitekturen ein echter technologischer Durchbruch in der Generativen KI erreicht.“ (Waldhauser 2023). Oder Bosse Kubach, Commercial Communications Manager bei *Microsoft*: „Der Hype um generative KI ist riesig. [...]

trans- bzw. posthumanistischen Weg zur „Singularität“ (Ray Kurzweil) oder „maschinellen Superintelligenz“ (Nick Bostrom) gilt die Entwicklung generativer KI spätestens seit dem Release von *ChatGPT* als technologischer Meilenstein.¹⁹ Auch in der Frage der „Maschinenmoral“ scheint es entscheidende Fortschritte zu geben, da Sprachmodelle wie *Chat-GPT* hierfür nun eine Voraussetzung erfüllen – zumindest wenn man dem schwedischen Futuristen Bostrom zustimmt, der 2014 forderte: „Um einer KI irgendeinen dieser [moralischen, *Anm. der Verf.*] Begriffe zu geben, müsste man sie vielleicht mit allgemeiner Sprachfähigkeit ausstatten (die mindestens auf dem Niveau derjenigen eines normalen Erwachsenen zu sein hätte).“²⁰ Der Bot *ChatGPT* scheint diese Erwartung sogar zu übertreffen: Indem er Sprachfähigkeit mit einer breiten Datenbasis verbindet und in seinem Output eine vorher nicht erreichte Plausibilität an den Tag legt, macht er den Eindruck eines eloquenten, universell gebildeten Erwachsenen – auch wenn man seine Rezepte nicht in jedem Fall „nachbacken kann“.²¹

Exponentielle Entwicklung

Imponierend erscheint auch die Geschwindigkeit des technologischen Fortschritts der KI, dem der Zukunftsforscher Ray Kurzweil 2005 attestierte, dass er „exponentiell“, möglicherweise sogar „doppelt exponentiell“ verlaufe²², eine „intuitiv-lineare“ Betrachtungsweise Lügen strafe²³ und auf die „Singularität“ hinauslaufe, in der „unser biologisches Denken und Dasein mit unserer Technik verschmelzen“ würden.²⁴ Unabhängig von der Annahme eines solchen Endpunkts in der transhumanistischen Perspektive, die aus verschiedenen Fachrichtungen teilweise scharf kritisiert wird²⁵, bestehen gegenwärtig

Doch nicht nur zu Hause und in den Büros steht dieser faszinierenden Technologie ein Durchbruch bevor, sondern auch in der Industrie.“ (Kubach 2023)

¹⁹ Bisherige Meilensteine waren beispielsweise die sogenannten „Brute-Force-Algorithmen“ (Ramge 2018, S. 39), ein Durchbruch in der Rechnerleistung, der 1997 dem IBM-Computer *Deep Blue* öffentlichkeitswirksam den Sieg über Schachweltmeister Garri Kasparov ermöglichte. Oder Googles lernfähiges System AlphaGo, das „durch eine genialische Mischung aus Mustererkennung, Statistik und Zufalls-generator“ (ebd. S. 40/41) 2016 den weltbesten Go-Spieler schlug.

²⁰ Bostrom 2016, S. 305

²¹ Vgl. Aljoscha Burchardt: „Bei ChatGPT geht es um Plausibilität, nicht um Wahrheit. Dass ein Kochrezept, das es zusammenstellt, plausibel ist, heißt nicht unbedingt, dass man es nachbacken kann.“ (Heier 2023)

²² Kurzweil 2013, S. 13.

²³ „Der Grund dafür ist, mathematisch gesehen, dass Exponentialkurven wie Geraden aussehen, wenn man nur kleine Ausschnitte betrachtet. Darum extrapolieren selbst kluge Köpfe, wenn sie über die Zukunft sprechen, häufig den momentanen Fortschritt auf die nächsten zehn oder hundert Jahre. Und darum nenne ich diese Betrachtungsweise die ‚intuitiv-lineare‘.“ (ebd. S. 13)

²⁴ ebd. S. 10

²⁵ So hält beispielsweise Adrian Daub, Professor für vergleichende Literaturwissenschaft an der Stanford University, die Idee der Singularität für „bewusst monumental“ und eine „völlig durchgeknallte Vorstellung“ (Daub 2020, S. 127). Sarah Spiekermann, Professorin für Wirtschaftsinformatik an der Wirtschaftsuniversität Wien, bezeichnet den Transhumanismus als eine „Ideologie der Lieblosigkeit“ (Spiekermann 2021, S. 164). Jobst Landgrebe, Wissenschaftstheoretiker und Unternehmer, und Barry Smith, Onto-

an der exponentiellen und autonomen Entwicklung von KI als selbstlernendem System kaum Zweifel.²⁶ Wirft man einen Blick auf die Grafiken der *OECD*²⁷ zum Thema „Generative AI Trends“, zeigen sich ausnahmslos exponentielle Entwicklungen. Dabei markiert das letzte Quartal des Jahres 2022 den „Knick“ in der Kurve, der für exponentielle Entwicklungen typisch ist. Dieser Knickpunkt, ab dem exponentielle Entwicklungen „explosionsartig“ verlaufen, korreliert hier mit dem Zeitpunkt der Veröffentlichung von *ChatGPT*.²⁸ Ray Kurzweils Prognose scheint damit bestätigt: Mit dem erfolgreichen Auftreten generativer KI scheint der historische Punkt erreicht zu sein, ab dem sich Künstliche Intelligenz „explosionsartig“ entwickeln könnte.

Die Zahlen und Grafiken der *OECD*, die im *G7 Hiroshima Process on Generative AI*²⁹ auf höchster politischer Ebene für eine Diskussion der Chancen und Risiken dieser Entwicklung herangezogen werden, können allerdings nicht die technologische Entwicklung selbst abbilden, da sie keine „qualitative“ Beschreibung darstellen.³⁰ Sie zeigen stattdessen deren quantifizierbare Auswirkungen an Hand verschiedener Indikatoren. Als zivilgesellschaftlicher Indikator gilt die öffentliche Wahrnehmung („public awareness“) generativer KI, die mit der Anzahl der Nachrichtenartikel, Tweets und wissenschaftlichen Publikationen zu diesem Thema quantifiziert wird. Hier als Beispiel die Grafik zur Anzahl der Nachrichtenartikel:

loge an der University at Buffalo, nennen den Transhumanismus „eine säkulare, erlösungsorientierte Esoterik, die dem Menschen in seiner Leiblichkeit nichts zu sagen hat“ (Landgrebe/Smith 2021, S. 3). Die Liste der kritischen Statements ließe sich beliebig fortsetzen.

²⁶ „KI unterscheidet sich von allen anderen Technologien [...] darin, dass sie selbstlernend ist. Ihre wesentlichen Eigenschaften bestehen daher in einer [...] exponentiellen Entwicklung, d.h. sie beschleunigt sich und löst Entwicklungssprünge aus, [und in einer, *Anm. d. Verf.*] autonomen Entwicklung, d.h. sie entkoppelt sich potenziell vom Menschen auf einem Pfad der Co-Evolution.“ (Vöpel 2023) Henning Vöpel spricht in diesem Zusammenhang von einer „Schicksalstechnologie für Deutschland und Europa“ und widmet diesem Topos seinen ganzen Artikel.

²⁷ s. Literaturverzeichnis: OECD publishing 2023

²⁸ ChatGPT wurde am 30. November 2022 veröffentlicht.

²⁹ „Im Rahmen des sogenannten ‚Hiroshima-Prozesses‘ sollen Vorschläge zur Regulierung von ChatGPT und Co. bis Ende des Jahres vorliegen.“ (Mewes 2023)

³⁰ Eine qualitative Beschreibung der technologischen Entwicklung bleibt KI-Ingenieuren vorbehalten. Angemerkt sei, dass Fortschritt und Qualität der Künstlichen Intelligenz auch für KI-Ingenieure zunehmend schwerer zu beurteilen sind. Bei selbstlernenden Systemen ist zwar der Output beobachtbar, nicht aber der Lernprozess der Algorithmen, der in der „Black Box“ stattfindet.

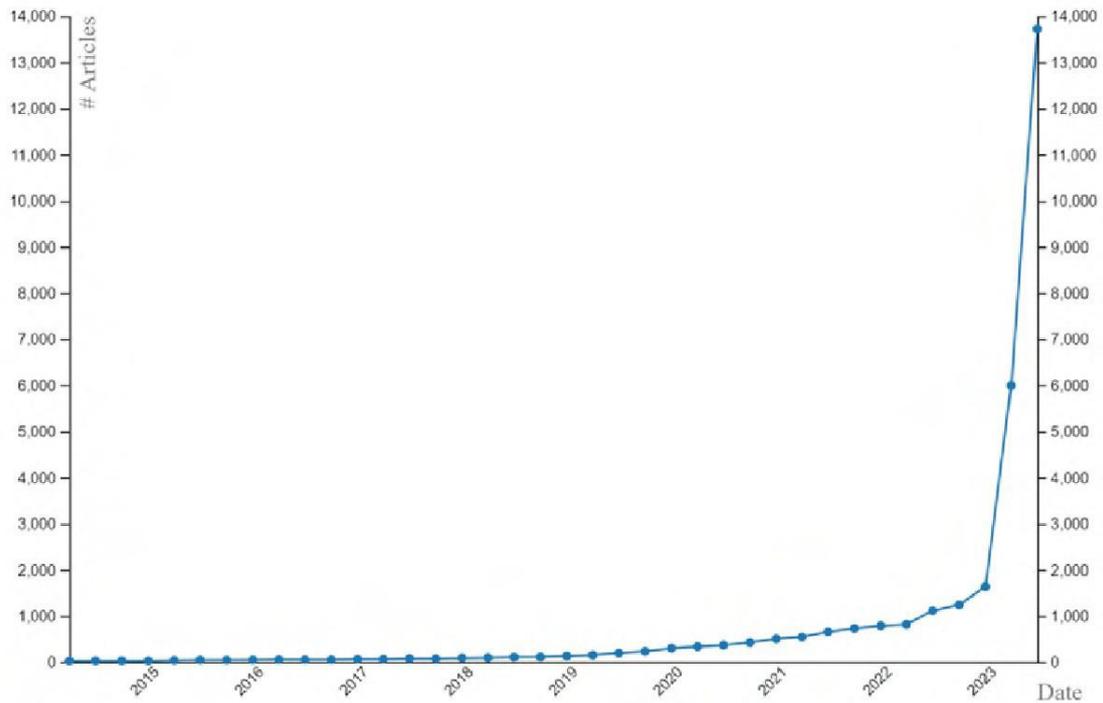


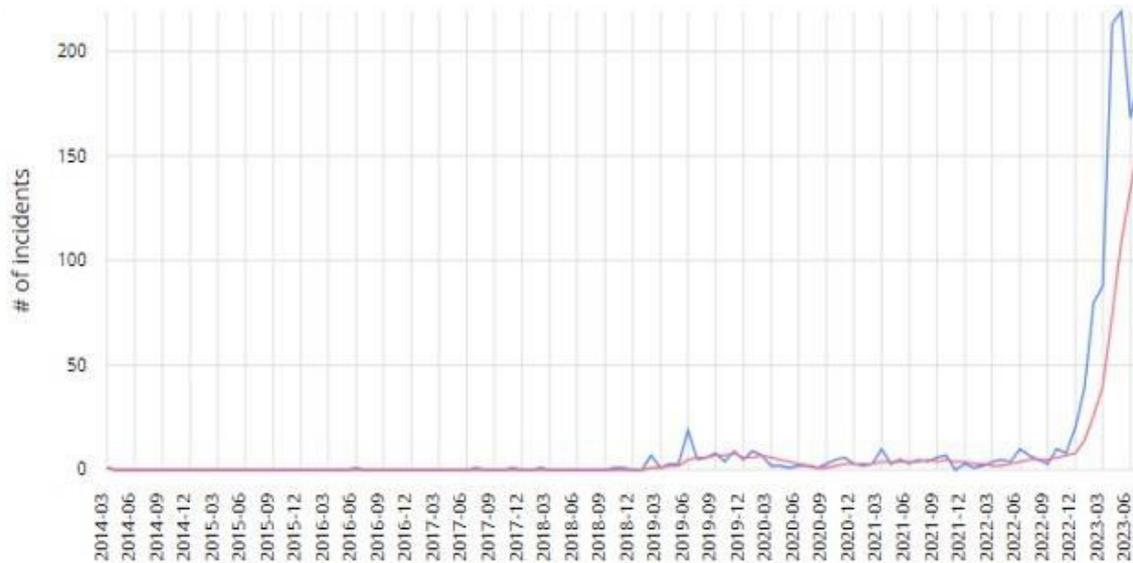
Abb. 1: OECD Grafik „Number of news articles globally on generative AI and related topics“

Gleich mehrere ökonomische Indikatoren zeigen ähnliche Verlaufskurven: Zum einen die Summe der globalen Risikokapital-Investitionen („venture capital investments“) ³¹ in Generative-KI-Start-ups, zum anderen die Anzahl der Open-Source generativen KI-Modelle auf *Hugging Face* ³² sowie die Anzahl der angelegten Projekte („repositories“) zu generativer KI auf *GitHub* ³³. Last not least zeigt auch der spezielle Indikator der Anzahl der Vorfälle und Gefahren, die in Zusammenhang mit generativer KI stehen und von seriösen Nachrichtenagenturen berichtet wurden, einen ähnliche exponentielle Kurve mit dem typischen „Knick“ im letzten Quartal des Jahres 2022:

³¹ s. OECD publishing 2023, S. 9

³² ebd. S. 10

³³ ebd. S. 11



(Note: The blue line shows the real count of incidents and hazards reported each month. The red line displays the same data but adjusted using quarterly smoothing.)

Abb. 2: OECD Grafik „Generative AI-related incidents and hazards reported by reputable news outlets have grown exponentially“

Ethik unter Zeitdruck

Der Verlauf des Indikators in Abb. 2, der als „ethischer“ Indikator bezeichnet werden kann, veranschaulicht indirekt den Zeitdruck, in den die ethische Debatte generativer KI geraten ist. Sie muss sich dem exponentiell wachsenden Tempo der technologischen Entwicklung anpassen, damit Schaden von der Zivilgesellschaft abgewehrt werden kann. Der Zeitdruck wird durch die Marktentwicklung verstärkt, die ihrerseits exponentiell verläuft und mit ebenfalls gesteigertem Interesse an „Schadensbegrenzung“ (in diesem Fall aus wirtschaftlichen Gründen) einhergeht. Neue Marktfelder für Künstliche Intelligenz eröffnen sich beinahe täglich und fordern entsprechend immer „neue“ Bereichsethiken.³⁴ Deren Aufgabe soll es sein, die vielfältigen Versprechen gesellschaftlichen Nutzens und kultureller Bereicherung gegen die Risiken abzuwägen. Eine solche Abwägung kostet in der gegenwärtigen Marktdynamik zu viel Zeit. Unternehmen im Wettbewerbsdruck warten nicht auf die Ergebnisse des gesellschaftlichen (und wissenschaftlichen) Diskurses und etwaige staatliche Regularien, sondern erlassen kurzerhand

³⁴ Hierzu Klaus Wieglerling: „Dem Bereichsethiker muss gegenwärtig sein, dass er sein Feld nur in beschränkter Weise fixieren kann, weil er technischen Entwicklungen meist hinterherhinkt. Dennoch muss er einen Beitrag zur Lösung bestehender Handlungsprobleme leisten, auch wenn sich diese Probleme möglicherweise bereits morgen anders stellen werden.“ (Wieglerling/Neuser 2013, S. 6)

eigene ethische Richtlinien. Schon vor der explosionsartigen Entwicklung generativer KI entstand eine Art „Wildwuchs“ ethischer Richtlinien für Künstliche Intelligenz.³⁵ Dies macht die Ausgangslage für eine Ethik generativer KI nicht einfacher. Bisher reagieren die meisten Bemühungen, eine solche Ethik zu formulieren, auf den „Output“ generativer KI: Die KI-generierten Inhalte (an vorderster Front synthetische Medien und synthetischer Text) erscheinen als das neuartige Produkt, das einer ethischen Bewertung und ggf. einer „Nachbesserung“ bedarf.

Produkte der Zukunft – Kanalisierung der ethischen Diskussion

Eine solche „Produktorientierung“, die als Unternehmensphilosophie ökonomisch sinnvoll ist, kann im zivilgesellschaftlichen Diskurs zu einer verengten Perspektive führen. Der neuartige „Output“ erfährt die Aufmerksamkeit der Gesellschaft, während die ethische Diskussion des „Inputs“, ohne welchen generative KI nicht existieren könnte, zweit-rangig erscheint. Oft wird diese Diskussion mit dem Argument marginalisiert, beim „Input“ für generative KI handele es sich um ein *fait accompli*. Die Einspeisung von Trainingsdaten für die generative KI sei nicht rückgängig zu machen oder zu verhindern, und daher sei es nicht sinnvoll, Ressourcen für eine nachträgliche Diskussion dieser Praxis zu verschwenden. Lediglich auf die Qualität der Trainingsdaten sei zu achten. Überdies sei die – unentgeltliche – Mitarbeit der gesamten Gesellschaft an diesem „Input“ notwendig und wünschenswert, um einen (propagierten) „Menschheitstraum“ künstlicher Intelligenz wahrwerden zu lassen. Dahingehende Skepsis wird, so wie in der Unternehmenswelt, auch zivilgesellschaftlich gerne als Technikpessimismus der „Ewiggestrigen“ abgetan.

Die ethische Diskussion generativer KI unterliegt damit der Gefahr einer zweifach verengten Perspektive: Zum einen dadurch, dass Elemente der transhumanistischen Ideologie quasi ontologisiert Eingang in sie finden, zum anderen, dass ökonomisch motivierte Argumente die Debatte einhegen und Kritik oder Aufklärung bezüglich der Voraus-

³⁵ Der Deutsche Ethikrat konstatiert eine „Fülle an Leitlinien [...] – 84 allein nach einer Übersicht aus dem Jahr 2019. Das Angebot reicht von Codizes einzelner Unternehmen wie beispielsweise der Deutschen Telekom, von SAP, Microsoft oder Google über Richtlinien von Fachgesellschaften wie dem Institute of Electrical and Electronics Engineers bis hin zu Werken auf nationaler oder internationaler Ebene. In Deutschland sind hier insbesondere die Stellungnahmen der Datenethikkommission und der Enquete-Kommission Künstliche Intelligenz zu nennen, auf internationaler Ebene die ‚Ethik-Leitlinien für eine Vertrauenswürdige KI‘ der von der Europäischen Kommission eingesetzten Hochrangigen Expertengruppe für KI und die ‚Recommendation on the Ethics of Artificial Intelligence‘ der UNESCO. Es gibt inzwischen mehrere Ansätze, solche Richtlinien zu kartieren und aus ihren überlappenden Inhalten sowohl Gemeinsamkeiten als auch Unterschiede herauszuarbeiten, die sich aus spezifischen Perspektiven und Interessen der beteiligten Personen und Institutionen ergeben.“ (Deutscher Ethikrat 2023, S. 108 f.)

setzungen generativer KI nur insoweit zulassen, als sie das „Geschäft“ mit ihr nicht in Frage stellen. Über die „Zukunftsversprechen“, die einen wesentlichen Teil des KI-Marketings ausmachen, sind diese beiden perspektivischen Verengungen, in welche die Debatte unter Zeitdruck hineingetrieben wird, auch thematisch miteinander verknüpft.

Künstler vor den Karren gespannt?

Zukunftsversprechen, die am individuellen, gesellschaftlichen und ökonomischen Horizont auftauchen, gelten auch für die „Zielgruppe“ der Kulturschaffenden (Künstler und Kulturmanager) sowie für die Kreativbranche. Auch ihnen wird empfohlen, die kreativen Möglichkeiten generativer KI auszuprobieren und sich von ihnen inspirieren zu lassen: Kreativitätsvorsprung als Wettbewerbsvorteil (und Anschluss an das Publikum der Zukunft). Kunst und Kultur nehmen in der öffentlichen Debatte eine Schlüsselrolle ein, da sie es der Gesellschaft ermöglichen, relevante Themen in gedanklicher Freiheit, mit kritischer Reflexion³⁶ und nicht zweckgebunden zu verhandeln. Diese Schlüsselrolle macht sie für Unternehmen der KI-Branche besonders wertvoll: Wenn Künstler oder Kreative sich aus freien Stücken entschließen, bei der Nutzung generativer KI voranzugehen, wertet die Branche dies als Etappensieg auf dem Weg hin zu allgemeiner gesellschaftlicher Akzeptanz der neuen Technologie. Künstler sind aus Sicht der KI-Branche besonders geeignet dafür, den Nachweis zu führen, dass generative KI mit ihrer „Kreativität“ ein vollwertiger „Sparringspartner“³⁷ ist. Auch hier fällt die Reduktion der ethischen Frage auf den „Output“ generativer KI auf, neben einer – mit einigen Ausnahmen³⁸ – eher verhaltenen Resonanz unter Künstlern, welche ihre bewährten Werkzeuge noch nicht bereitwillig gegen die „Tools“ generativer KI eintauschen.

Fragen an das Kulturmanagement

Zurzeit ist es vor allem das Kulturmanagement, das proaktiv seine Chancen auslotet, sich beim Kulturmarketing, beim Personalmanagement, bei der intelligenten Archivierung usw. von generativer KI assistieren zu lassen. In Zeiten knapper Kassen für Kultur

³⁶ Vgl. das Kulturverständnis von Jean Baudrillard, welches er dem „kulturellen Recycling“ gegenüberstellte: Letzteres sei „das *absolute Gegenteil* [Hervorhebung im Original] der Kultur, die wir verstehen als: 1. Vermächtnis von Werken, Gedanken, Traditionen; 2. Fortdauernde Dimension einer theoretischen und kritischen Reflexion – kritische Transzendenz und symbolische Funktion.“ (Baudrillard 2015, S.148)

³⁷ „Die Informatikprofessorin und Autorin Katharina Zweig sagt: ‚Ich glaube schon, dass ich mit KI gemeinsame Bücher schreiben kann.‘ So könne die KI eine Art Sparringspartner sein für Ideen oder Assoziationen: ‚Das fände ich auch nicht schlimm, wenn ich meine Kreativität auf diese Art und Weise boostere.‘“ (Süddeutsche Zeitung 2023)

³⁸ Die Medienkünstlerin Alice Bucknell war im Jahr 2021 eine der ersten, die (auf Einladung des Goethe-Instituts zur *KI-Residenz* in Riga) Kunstprojekte in Zusammenarbeit mit generativer KI realisierte. Sie nutzte dafür einen Beta-Zugriff auf GPT-3. (s. Goethe Institut Riga 2021)

sind die gebotenen Möglichkeiten für gesteigerte Effizienz und Kosteneinsparung hochwillkommen. Auf der Website *kulturmanagement.net* erscheint unter den (zum Zeitpunkt der Abfassung der vorliegenden Arbeit) 14 Beiträgen, die nach Eingabe des Suchbegriffs „KI“ angezeigt werden, allerdings kein Artikel, der expressis verbis ethische Fragen im Zusammenhang mit generativer KI thematisiert.³⁹ Dies verwundert, tritt doch generative KI ausdrücklich an, um mit dem Menschen um dessen ureigenste Domäne, die kulturelle Produktivität zu konkurrieren.⁴⁰ Wie verhält sich Kulturmanagement dazu, dass Erzeugnisse generativer KI auf den Markt drängen und menschliche Kreativität gleich auf mehreren Ebenen, soziokulturell, ökonomisch, juristisch, in Frage stellen? „Managt“ die generative KI die von ihr erzeugte Kultur gleich mit?

Kulturelle Aneignung

So weit scheint es noch nicht zu sein, doch müssen sich Kulturmanager fragen, wie sie mit der Wucht und dem Tempo der neuen KI-Entwicklungen umgehen.⁴¹ Kulturmanagement ist notwendigerweise mit der gesellschaftlichen Akzeptanz von Kultur und kulturellen Erzeugnissen befasst und greift in diesem Kontext gesellschaftliche Debatten auf. Gegenwärtig ist es die Debatte um „kulturelle Aneignung“, die viel Aufmerksamkeit erhält und teilweise mit Erbitterung und großem Interesse für Details kultureller Symbolik geführt wird. Die wechselseitigen Vorwürfe lauten Geschichtsvergessenheit auf der einen bzw. Verbotskultur auf der anderen Seite, oftmals verkennend, dass die als eine Art Läuterung eingeforderte Verbotskultur genau der monierten Geschichtsvergessenheit entspringt. So wichtig diese Diskussion sein mag und so sehr es hierfür wohl auch einer „Ethik der Appropriation“⁴² bedarf, scheint sich unsere Gesellschaft im Modus eines kleinteilig ausdifferenzierten Kulturkampfes zu befinden und dabei den Ort aus dem Blick zu verlieren, an dem möglicherweise eine viel umfassendere (und zukunftsbestimmendere) kulturelle Aneignung stattfindet.

³⁹ www.kulturmanagement.net

⁴⁰ vgl. Milad Safar: „Was bisher dem Menschen vorbehalten war, nämlich das Erzeugen von Artefakten, kann nun die generative KI leisten. Dabei liefert sie originelle Ergebnisse, die auf menschlichem Denken und menschlicher Erfahrung beruhen, und zwar ohne Voreingenommenheit. So kann generative KI [...] Inhalte erstellen, die den Anschein erwecken, von Menschen geschrieben zu sein.“ (Safar 2022)

⁴¹ „Generative KI ermöglicht es jedem, auf einfache Weise fiktionale Inhalte zu erstellen – das öffnet natürlich auch die Tür für Missbrauch. Der Schutz der Rechte am geistigen Eigentum und die Verhinderung der unethischen Nutzung dieser Werkzeuge sind wichtige Herausforderungen, für die es aber noch keine praktischen Lösungen gibt“, schreibt Benjamin Talin lapidar auf *MoreThanDigital* und umreißt damit das weite Feld, das ethisch zu „durchpflügen“ ist, und auf dem die Disziplin Kulturmanagement sich erst noch orientieren muss. (Talin 2023)

⁴² Der Begriff „Ethik der Appropriation“ stammt von Jens Balzer und ist Titel seines gleichnamigen Buches. Vgl. Balzers Artikel für den Deutschen Kulturrat (Balzer 2022).

Dieser Ort ist – nach Auffassung der Autorin – der „Käfig“, in welchem der von Emily Bender, Timnit Gebru und anderen so treffend bezeichnete „stochastische Papagei“⁴³ generativer KI wohnt. Oder, mit dem Ausdruck der KI-Ingenieure: die *Black Box*, in welcher die gesamte bisherige Kultur- und Geistesgeschichte des Menschen – soweit in Schrift, Bild und Ton dokumentiert, digitalisiert und im Web repräsentiert – zur Disposition steht. „Kulturelle Aneignung“ im Sinne von Kulturdiebstahl bzw. „kultureller Ausbeutung“ (Richard A. Rogers)⁴⁴ müsste wohl auch in *diesem* Zusammenhang problematisiert werden, da kein Mensch hierbei um Erlaubnis gefragt wurde, und überdies nicht anzunehmen ist, dass der stochastische Papagei die angeeignete Kultur – welche auch immer – versteht oder gar würdigt. Künstliche Intelligenz, die sich erfolgreich auf Datafizierung und Digitalisierung unserer Lebenswelt (Big Data) „aufgepropft“ hat, plündert als generative KI bislang kaum gehindert die Matrix unserer Kultur- und Geistesgeschichte. Sie schickt sich ferner an, deren lebendige Weiterentwicklung und Fortschreibung künftig mit „besseren Produkten“ zu bereichern, wenn nicht gar zu ersetzen. Kultur scheint den staunenswerten Emanationen generativer KI „hinterherlaufen“ zu müssen.

Die Frage nach den Voraussetzungen

In der vorliegenden Arbeit soll das ökonomisch motivierte Mantra der Produktorientierung nicht übernommen werden. Daher wird auf eine Diskussion von Einzelnormen für „Produkte“ in den derzeitigen Anwendungsbereichen generativer KI (Medizin, Rechtsprechung, Medienkommunikation und Journalismus, öffentliche Verwaltung, Pädagogik etc.) verzichtet.⁴⁵ Die ethische Bewertung generativer KI soll vielmehr aus ihrem Verhältnis zur Kultur entwickelt werden, ein Verhältnis, das in Bezug auf die oben angesprochene kulturelle Aneignung zu klären ist. Eine behauptete – und ethisch-normativ „abgesegnete“ – kulturelle Bereicherung durch generative KI muss sich nach Auffassung der Autorin an ihren Voraussetzungen messen lassen. Wenn der gegenwärtige oder noch zu erwartende gesellschaftliche Nutzen generativer KI mit Hilfe eines Raubbaus an der Kultur gestiftet wurde und weiterhin von kultureller Ausbeutung durchkreuzt wird, kann er dann Nachhaltigkeit für sich beanspruchen oder müsste er ethisch neu bewertet werden? Um diese Frage zu beantworten, ist in der ethischen Diskussion jede Spur, die zu den Voraussetzungen generativer KI führt, für die vorliegende Arbeit von Interesse.

⁴³ Bender et al. 2021

⁴⁴ Richard A. Rogers unterscheidet vier Typen kultureller Aneignung: Kultureller Austausch, kulturelle Dominanz, kulturelle Ausbeutung, Transkulturation. Vgl. Rogers 2006.

⁴⁵ Diese Diskussion ist bereits in vollem Gange, wobei die partikulären Interessen unzähliger Stakeholder zu berücksichtigen sind, und in den schnell wuchernden Geschäftsfeldern dieser Anwendungen ebenso schnell Rechtssicherheit erlangt werden muss.

Das folgende zweite Kapitel dient der Bestandsaufnahme der Diskussion. Auf eine Voranstellung umfangreicher Begriffsklärungen, die in der Dynamik einer neuen technologischen Entwicklung schnell hinfällig werden können, wird verzichtet. Stattdessen werden wiederkehrende Begriffe, sowohl technologische⁴⁶ als auch ethische, entlang der aktuellen Diskussion identifiziert und verortet. Wie in Abb.1 veranschaulicht, ist seit dem Auftreten von *ChatGPT* eine „Explosion“ an Veröffentlichungen zu generativer KI zu verzeichnen. Ein großer Teil dieser Literatur widmet sich der Erklärung der Technologie, ihrer Geschichte und ihrer Anwendungsmöglichkeiten, wobei ethische „Einwände“ meist schlagwortartig benannt werden. Für die vorliegende Arbeit wurde der Fokus auf aktuelle Veröffentlichungen gelegt, in welchen eine breitere Auseinandersetzung mit ethischen Aspekten erkennbar ist.⁴⁷ Im dritten Kapitel werden ausgehend vom Stand der Debatte kritische Anmerkungen formuliert, die sich auf die Frage nach den Voraussetzungen und dem Input generativer KI beziehen, mit einem Schwerpunkt auf der Frage nach der menschlichen Autorschaft. Anschließend werden die wesentlichen Argumente dieser Arbeit zusammengefasst, die als Anregung für die weitere ethische Debatte verstanden sein wollen.

2. Ethik generativer KI – aktueller Diskussionsstand

Die Annahme, dass in der ethischen Diskussion einer neuen Technologie die Argumente von Unternehmen, Staat und Zivilgesellschaft in Widerstreit miteinander geraten, erscheint veraltet: Weitreichende ökonomische Verflechtungen zwischen den Sektoren machen eine schematische Trennung der Perspektiven und damit auch ethischen Prioritäten schwierig. So lassen sich auffällig „vertauschte Rollen“ beobachten: Während fortschrittliche Unternehmen beinahe geschäftsschädigend die Risiken generativer KI betonen und vehement ethische Richtlinien einfordern (z.B. in der „Moratoriumsforderung“, siehe Kap. 2.1.2), betonen staatliche und zivilgesellschaftliche Institutionen die gesellschaftlichen Chancen – und dass es bei allen ethischen Fragen nicht darum gehe, die technologische Entwicklung zu bremsen. So viel wechselseitiges Entgegenkommen mag begrüßt werden als Synergie vernetzter gesellschaftlicher Akteure, welche gemeinsam versuchen, ein Problem zu lösen. Allerdings wird, wenn gleich zu Beginn der Debatte die jeweils eigenen Interessen ethisch sensibel bzw. in vorauseilendem Gehor-

⁴⁶ Das Glossar bietet hierfür einen ersten Ausgangspunkt.

⁴⁷ Die Literaturrecherche für die vorliegende Arbeit wurde im Januar 2024 abgeschlossen.

sam hintangestellt werden, die gesellschaftlich notwendige, substanzielle Auseinandersetzung vermieden. Die Enttäuschung darüber kann dazu führen, dass die ethische Debatte verlassen wird (siehe Kap. 3.4).

In diesem Kapitel wird die Debatte gegliedert nach internationalen (Kap. 2.1) und deutschen (Kap. 2.2) Standpunkten dargestellt. Letztere werden überwiegend von unternehmensfernen Institutionen vertreten und aus einer wissenschaftlichen, um Objektivität bemühten Perspektive formuliert (allen voran der *Deutsche Ethikrat*, Kap. 2.2.4), bleiben auf internationalem „Parkett“ jedoch oft unrezipiert. Die internationale *Partnership on AI* (Kap. 2.1.3) hingegen kann im Mainstream kollektiver Bemühungen um ethische „Guidelines“ verortet werden. In ihr verpartnern sich Zivilgesellschaft und akademische Institutionen mit Unternehmen, die in generative KI involviert sind. Forschung und Wissenschaft profitieren von Drittmitteln, und die Unternehmen erschließen sich einen wertvollen Zugang zur Zivilgesellschaft.⁴⁸ Ergänzend zu den kollektiven Anstrengungen, generative KI ethisch zu erfassen, werden in Kap. 2.3 einige individuelle Standpunkte dargestellt. Frei von ökonomischen Interessen können sie die ethische Diskussion unvoreingenommen vorantreiben.

2.1 Supranationale kollektive Stellungnahmen/Vereinbarungen/Gesetzgebungsprozesse

Aus den kollektiven supranationalen Bemühungen, sich auf eine Ethik generativer KI zu verständigen, werden vier Beispiele aus dem Jahr 2023 herausgegriffen: der *EU AI Act*, die „Moratoriumsforderung“, das *PAI Framework* sowie ein aktuelles Arbeitspapier der *OECD*.⁴⁹ Vorschläge zur Regulierung generativer KI sollten außerdem im *G7 Hiroshima Process* bis zum Ende des Jahres 2023 erarbeitet werden; die G7-Staaten sind in Sorge,

⁴⁸ Es soll nicht unterschlagen werden, dass auch in Deutschland solche Partnerschaften auf dem Vormarsch sind. Das *Institute of Ethics in Artificial Intelligence* (IEAI) an der TU München wurde beispielsweise von *Facebook* mit 7,5 Millionen Euro finanziert, wobei die Förderung an Auflagen gebunden war. Förderungsummen in dieser Dimension, die der Staat nicht für einzelne Forschungsinstitute aufbringen kann, sichern langfristige Forschungsprojekte und damit verbundene Arbeitsplätze. Dass Geld auch eine eigene Sprache spricht, gerät dabei leicht in Vergessenheit. Die geförderten Institutionen müssen oftmals viel Mühe darauf verwenden, den Vorwurf des „Ethics Washing“ zu entkräften. Im Fall des IEAI ergab die Suche nach publizierter Forschung (z.B. Mitschrift eines Vortrags von Sven Nyholm) zur Ethik generativer KI noch kein verwertbares Ergebnis (Stand Dezember 2023) und ist daher hier nicht einbezogen.

⁴⁹ Unter den großen internationalen Organisationen veröffentlichte auch die UNESCO im Jahr 2023 Empfehlungen zur Ethik künstlicher Intelligenz (siehe Literatur-Verzeichnis: UNESCO 2023). Diese Empfehlungen stützen sich allerdings auf Aussagen, die die Organisation bereits im November 2021, also noch vor der Veröffentlichung von *ChatGPT* getroffen hatte (siehe UNESCO 2022), und sind daher für eine ethische Bewertung generativer KI zu unspezifisch.

dass „im Zuge der raschen Entwicklung der KI-getriebenen Modelle für Texte, Bilder und Videos mächtige Werkzeuge für Desinformationen und politische Störungen“⁵⁰ entstehen. Da bei Abschluss der Literaturrecherche für die vorliegende Arbeit noch kein schriftlich dokumentiertes Ergebnis des Prozesses vorlag⁵¹, kann er hier nicht berücksichtigt werden. Der o.g. Querschnitt aus vier Beispielen sollte allerdings ausreichen, um ein verwertbares Bild der aktuellen internationalen Bemühungen zu erhalten. Kurze Diskussionen und Einschätzungen jeweils am Ende der Unterkapitel 2.1.1 bis 2.1.4 bereiten auf die Kritik im dritten Kapitel vor.

2.1.1 EU AI Act – das mühsame Geschäft der Konsensfindung

Beim *AI Act* der Europäischen Union⁵² handelt es sich um das weltweit erste umfassende KI-Gesetz. Im initialen Entwurf der EU-Kommission vom April 2021 finden sich noch keine besonderen Regelungen für *General Purpose AI* (GPAI), unter welche auch die generative KI fällt. Doch in die am 14. Juni 2023 veröffentlichte Position des EU-Parlaments zum *AI Act* wurden Pflichten für GPAI und speziell für generative KI-Systeme aufgenommen:

„Generative KI-Systeme, [...], wie ChatGPT, müssen Transparenzanforderungen erfüllen, d.h., sie müssen offenlegen, dass die Inhalte KI-generiert sind, was auch dazu beiträgt, sogenannte Deepfake-Fotos von echten Abbildungen zu unterscheiden. Zusätzlich müssen sie dafür sorgen, dass keine rechtswidrigen Inhalte erzeugt werden. Außerdem müssen sie detaillierte Zusammenfassungen der urheberrechtlich geschützten Daten veröffentlichen, die sie zu Trainingszwecken verwendet haben.“⁵³

Hiermit schien sich eine Regelung auch für *ChatGPT* abzuzeichnen, das zu diesem Zeitpunkt bereits über ein halbes Jahr unreguliert auf dem Markt war. Doch geriet der EU-Rat als Mitgesetzgeber unter den Druck zahlreicher Lobbyverbände von Techfirmen, die KI entwickeln und/oder als Dienstleistung anbieten.⁵⁴ Ziel der Lobbyisten war, die

⁵⁰ Mewes 2023

⁵¹ Veröffentlicht wurde am 7. September 2023 eine Absichtserklärung der beteiligten G7 Digital- und Tech-Minister (G7 Hiroshima AI Process 2023), die aber noch keine inhaltlichen Details enthält.

⁵² Die vollständige Bezeichnung lautet: *Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts* („Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union“). (Europäische Kommission 2021)

⁵³ Pressemitteilung vom 14. Juni 2023 (Europäisches Parlament 2023a)

⁵⁴ Nach Eike Kühl waren beispielsweise der Interessenverband *Business Software Alliance* (BSA) beteiligt, „der 1988 von Microsoft gegründet wurde und zu dessen Mitgliedern zahlreiche Techfirmen gehören“ (Kühl 2023), oder der Verband *Allied for Start-ups*, „zu dessen Sponsoren unter anderem Microsoft, Google und Meta gehören“. (ebd.)

GPAI „wieder aus dem Entwurf zu entfernen oder zumindest die Verantwortung von den Entwicklern der KI-Modelle auf die Anbieter der Anwendungen weiter unten in der Wertschöpfungskette abzuwälzen“⁵⁵. Transparenzanforderungen und Konformitätsprüfungen für GPAI sowie mögliche Klagen bei Regelverstößen gelten für Lobbyisten als große Hindernisse, wenn es darum geht, generative KI auf dem europäischen Markt auszurollen. Der Vorsitz des EU-Rates hatte schon im Herbst 2022 Schreiben von der US-Regierung, *Microsoft* und *Google* erhalten, in welchen die Sorge geäußert wurde, der europäische Markt könne sich mit Überregulierung ihren Geschäftsmodellen verschließen.⁵⁶ In den langen Trilogverhandlungen⁵⁷ vertraten Deutschland, Frankreich und Italien eine gemeinsame Position, nach der es keine gesetzlichen Vorschriften für GPAI geben sollte, „lediglich eine Art Selbstverpflichtung“⁵⁸. Bei dieser Position ging es im Wesentlichen darum, Start-up-Unternehmen wie *Aleph Alpha* in Deutschland oder *Mistral* in Frankreich keine Steine in den Weg zu legen.⁵⁹ Von der spanischen Ratspräsidentschaft kamen Gegenvorschläge, die von KI-Forschern und zivilgesellschaftlichen Organisationen unterstützt⁶⁰ und im EU-Parlament präferiert wurden⁶¹. Am 8. Dezember 2023 kam es schließlich zu einer Einigung, bei der die ökonomischen Interessen der US-amerikanisch dominierten Techbranche und entsprechende wirtschaftsliberale politische Kräfte zurückgedrängt wurden. In Bezug auf GPAI-Systeme und -Modelle wurden aus dem initialen Entwurf der EU-Kommission die Anforderungen an Transparenz, Rechtskonformität und Veröffentlichung der Trainingsdaten vollumfänglich übernommen; zudem ist der Gesetzesentwurf mit dem geltenden EU-Urheberrecht und der Datenschutzgrundverordnung (DSGVO) abgestimmt.

⁵⁵ ebd.

⁵⁶ „Die vorgeschlagenen Pflichten für GPAI-Anbieter ‚wären sehr aufwendig, technisch schwierig, und in manchen Fällen unmöglich‘“, heißt es im sogenannten „non-paper“ der US-Regierung (ebd.). Patrick Glauner, KI-Professor an der Technischen Hochschule Deggendorf, schrieb im Februar 2023 bei *heise online* beinahe inhaltsgleich: „Diese Anforderungen sind entweder sehr teuer oder aus technischer Sicht teilweise unerfüllbar.“ (Glauner 2023)

⁵⁷ Trilogverhandlungen finden zwischen Vertretern der EU-Kommission, des EU-Rates und des EU-Parlamentes statt, mit dem Ziel, einen Kompromiss zwischen den unterschiedlichen Positionen der drei gesetzgebenden Institutionen der EU zu finden. Mit den EU-Mitgliedsstaaten im Rat werden Gespräche über die endgültige Ausgestaltung des Gesetzes geführt.

⁵⁸ Lindern 2023

⁵⁹ Hierzu zeit.online: „In Frankreich gibt es das Start-up Mistral, das ein frei verfügbares KI-Modell veröffentlicht hat. Die deutsche Hoffnung heißt Aleph Alpha [...]“ (Lindern 2023)

⁶⁰ „Verschiedene KI-Forscher und zivilgesellschaftliche Organisationen haben [...] gefordert, den Vorschlag der spanischen Ratspräsidentschaft [...] in den AI Act aufzunehmen.“ (ebd.)

⁶¹ „Brando Benifei, einer der für das Gesetz zuständigen Abgeordneten, sagte vergangene Woche, es gebe eine klare Mehrheit für verpflichtende Regeln für die leistungsfähigsten Modelle. Nur Selbstverpflichtungen werde man nicht akzeptieren. Er deutete an, dass die Verhandlungen an dieser Frage scheitern könnten, wenn sich die Mitgliedsstaaten nicht bewegten, auch wenn er das nicht glaube.“ (ebd.)

Abgesehen davon, dass technische Details noch weiter ausgearbeitet werden müssen, und das Gesetz noch nicht verabschiedet ist, zeigt das Ergebnis der Trilogverhandlungen, dass ein ökonomisch getriebener, forciertes rechtlicher Diskurs ohne den „Überbau“ eines ethischen Diskurses in der EU nicht mehrheitsfähig ist. Das Ergebnis macht Hoffnung auf eine Stärkung der europäischen Position in den von Rechtsunsicherheit geprägten globalen Marktfeldern generativer KI. In der Einschätzung des italienischen Mitberichterstatters und Mitglieds des Europäischen Parlamentes Brando Benifei wird die Erleichterung deutlich spürbar:

„...the effort was worth it. Thanks to the European Parliament’s resilience, the world’s first horizontal legislation on artificial intelligence will keep the European promise – ensuring that rights and freedoms are at the centre of the development of this groundbreaking technology. Correct implementation will be key – the Parliament will continue to keep a close eye [...]“⁶²

Erste Kritik an dem Gesetz kommt naturgemäß aus der KI-Branche; es bleibt abzuwarten, wie sich die politischen und wirtschaftlichen Kräfte weiter austarieren werden.⁶³

2.1.2 Die paradoxe „Moratoriumsforderung“

Während Europa von wirtschaftsliberalen Stimmen notorisch aufgefordert wird, nicht den Anschluss an die KI-Entwicklung zu verlieren⁶⁴, hat sich in den USA gleichzeitig eine Initiative formiert, welche die technologische Entwicklung selbst entschleunigen möchte: Im März 2023 forderten KI-Forscher und Unternehmer in einem offenen Brief⁶⁵ ein

⁶² Europäisches Parlament 2023b. Der rumänische Mitberichterstatter Dragos Tudorache äußert sich ähnlich: „The EU is the first in the world to set in place robust regulation on AI, guiding its development and evolution in a human-centric direction.“ (ebd.)

⁶³ In einem ersten Statement bescheinigt Klaus Wieglerling der Formulierung des EU AI Acts ein „hohes Reflexionsniveau“ und bestätigt, dass der Gesetzesrahmen „für einen fairen, sozial verpflichteten Wettbewerb bei der Herstellung und Nutzung von KI sorgen kann“ (Wieglerling 2023b). Allerdings bestehe die Gefahr eines schleichend eintretenden „dual use“ selbstlernender KI-Systeme: „Gerade bei generativer KI wie den aktuell diskutierten Dialogsystemen lässt sich [...] eine einigermaßen zuverlässige Unterscheidung zwischen einer Hochrisiko-KI und einer KI mit geringem Risiko kaum durchhalten. [...] Tatsächlich können sich heute problemlose Nutzungen morgen zu risikobehafteten auswachsen.“ (ebd.) Des Weiteren sei in Anbetracht der verwendeten „stark ökonomie- und politikgetriebenen“ technischen Begriffe ein stetiger weiterer Abgleich der Regelungen des EU AI Acts empfehlenswert: „Es ist deshalb wichtig und richtig, dass sie [die Regelungen, *Anm. der Verf.*] in einem überschaubaren, die Rechtssicherheit aber noch gewährleistenden Rahmen, eine Überprüfung erfahren sollen.“ (ebd.)

⁶⁴ Vgl. Patrick Glauner: „Der AI Act wird die weltweite KI-Entwicklung nicht bremsen – er verlagert die Wertschöpfung lediglich weiter in die USA und nach China; Europa wird technologisch noch stärker abhängig. Die EU muss KI endlich als große Chance sehen und den Wandel und die Weiterentwicklung der KI aktiv mitgestalten. Andernfalls wird sie vom Hightech-Standort zum Industriemuseum transformiert.“ (Glauner 2023)

⁶⁵ Der Brief wurde vom *Future of Life Institute* veröffentlicht (s. Future of Life Institute 2023). Es handelt sich um ein Non-profit-Forschungsinstitut mit Sitz in Cambridge (Massachusetts), das seit 2014 existiert mit dem Ziel, existenzielle Risiken fortgeschrittener KI zu vermindern.

sechsmonatiges Moratorium für das Trainieren von KI-Systemen, die leistungsfähiger als GPT-4 von *OpenAI* sind. Zu den Erstunterzeichnern gehörten u.a. Emad Mostaque⁶⁶, Steve Wozniak⁶⁷, Tristan Harris⁶⁸ und Elon Musk⁶⁹. KI-Entwicklern, die sich in einem „außer Kontrolle geratenen Wettrennen“⁷⁰ befänden, soll mit dem Moratorium die Möglichkeit einer „Refokussierung“⁷¹ gegeben werden, um gefährliche Risiken unter Kontrolle zu bringen. Dabei werden die Risiken der gegenwärtigen KI-Systeme weit gefasst und teilweise wie beim Transhumanismus in die Zukunft projiziert: Von der Flutung der medialen Informationskanäle mit Propaganda und Lügen und von der Wegrationalisierung von Jobs ist genauso die Rede, wie allgemein von der Ersetzung des Menschen durch nichtmenschlichen Verstand und schließlich dem Verlust der Kontrolle über unsere Zivilisation.⁷² In dringlichem Tonfall heißt es weiter: Sei das Moratorium nicht „schnell“ als Selbstverpflichtung der Entwickler umsetzbar, wären die Regierungen in der Pflicht, das Moratorium zu verhängen.⁷³ KI-Entwickler müssten außerdem mit politischen Entscheidungsträgern zusammenarbeiten, um die Entwicklung belastbarer KI-Governance-Systeme „dramatisch zu beschleunigen“.⁷⁴ Im Fokus stehen vor allem „Sicherheitsprotokolle“⁷⁵ für die weitere Entwicklung generativer KI, öffentlich finanzierte „KI-Sicherheits-Forschung“ und „gut ausgestattete Institutionen“, welche die „dramatischen ökonomischen und politischen Disruptionen (besonders für die Demokratie)“ bewältigen, „welche KI verursachen wird“.⁷⁶

Die Idee des Moratoriums, die bis heute nicht umgesetzt wurde, fand (besonders wegen der paradox anmutenden Beteiligung von Elon Musk) viel kritische Resonanz. Die

⁶⁶ Gründer und CEO von *Stability AI*, einer der Firmen hinter dem Text-zu-Bild-Generator *Stable Diffusion*

⁶⁷ Entwickler und Mitgründer von *Apple*

⁶⁸ Ehemaliger Designethiker bei *Google* und Mitgründer des *Center for Humane Technology*

⁶⁹ CEO von *Tesla*, Gründer des Raumfahrtunternehmens *SpaceX* und Eigentümer von *X*, vormals *Twitter*

⁷⁰ „...AI labs locked in an out-of-control race...“ (Future of Life Institute 2023)

⁷¹ „AI research and development should be refocused on making today’s [...] systems more accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal.“ (ebd.)

⁷² „...and we must ask ourselves: *Should* we let machines flood our information channels with propaganda and untruth? *Should* we automate away all the jobs, including the fulfilling ones? *Should* we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? *Should* we risk loss of control of our civilization?“ (ebd.)

⁷³ „If such a pause cannot be enacted quickly, governments should step in and institute a moratorium.“ (ebd.)

⁷⁴ „[...], AI developers must work with policymakers to dramatically accelerate development of robust AI governance systems.“ (ebd.)

⁷⁵ „AI labs and independant experts should use this pause to jointly develop and implement a set of shared safety protocols for advanced AI design...“ (ebd.)

⁷⁶ „...robust public funding for technical AI safety research; and well-ressourced institutions for coping with the dramatic economic and political disruptions (especially to democracy) that AI will cause.“ (ebd.)

Risiken der KI-Systeme seien ungenau ⁷⁷ bzw. „unwissenschaftlich“ ⁷⁸ formuliert. Wenn gleich der Grundgedanke einer Entschleunigung diskussionswürdig sei, diene das Moratorium „letztlich genau denjenigen Institutionen, deren Tätigkeit eigentlich problematisiert werden soll“ ⁷⁹. Auch kommt die Frage auf, ob die ethische Begründung der Forschungspause standhalte, wenn man den wissenschaftlichen Nutzen der neuen KI-Modelle betrachte: „Sprachmodelle in der Medizin etwa sind eine große Chance, um mehr Leben zu retten oder Leiden reduzieren zu können. Forschung in diesen Bereichen ist extrem wichtig“, so Thilo Hagendorff. ⁸⁰

Die in der Forderung bemühte transhumanistische Perspektive wirkt offensichtlich wie ein ideologischer „Tunnel“, an dessen Wänden sich gleichermaßen Enthusiasmus und apokalyptische Warnungen so lange reflektieren, bis am Ende die – zu was? – gebündelte Norm heraustreten soll. Doch ist es die ganze umliegende „Landschaft“ außerhalb des Tunnels, sprich: die gegenwärtige Lebensrealität, für die Normen gelten sollen. Diese Lebensrealität ist in unserer „posttraditionellen Gesellschaft“ ⁸¹ von starken Differenzierungen geprägt und zeitigt entsprechende „Fraktionierungen von Normenbereichen“ ⁸² – einer dieser Bereiche wäre die oben angesprochene moderne Medizin. Dagegen muss die „eine“ Lösung zur „Bändigung“ generativer KI illusorisch bleiben. Normative Kriterien für generative KI können nur differenziert, entlang ihrer vielfältigen Anwendungszusammenhänge formuliert werden, wobei Argumente aus den bereits entwickelten Bereichsethiken dieser Zusammenhänge entlehnt werden können.

Dem lässt sich entgegenhalten, dass die Moratoriumsforderung immerhin ein pauschales „ethisches Gewissen“ reklamiert. Wie sehr allerdings muss dieses Gewissen schon beschädigt worden sein, wenn solch extreme Reaktionen nötig sind? Wozu genau dienen die alarmistischen Beschreibungen der „dramatischen Risiken“ generativer KI? Am Ende des offenen Briefes wird der Sinn des Moratoriums nochmals zusammengefasst als eine Art „Abwarten, bis die Gesellschaft so weit ist“ (im Original: „give society a

⁷⁷ Hierzu Peter Dabrock: „In ethischer Terminologie verbucht man solche Befürchtungen unter Slippery-Slope oder Dambruchargumente. Sie gelten wegen ihrer nach Zeit-, Raum-, Sach- und Sozialdimension quasi unformatierten Risikounterstellung als moraltheoretisch kaum rechtfertigungsfähig.“ (science media center germany 2023))

⁷⁸ Hierzu Thilo Hagendorff: „Mitunter wird bei KI-Systemen so getan, als könne in die Zukunft gesehen werden. Das ist unwissenschaftlich und verhindert, dass konkrete, in der Gegenwart wichtige Schritte zur Anpassung und Regulierung von KI-Systemen vollzogen werden.“ (ebd.)

⁷⁹ Hagendorff, zitiert nach: ebd.

⁸⁰ zitiert nach: ebd.

⁸¹ Wolfgang Neuser in: Wiegerling/Neuser 2013, S. 164 f.

⁸² ebd. S. 165

chance to adapt“)⁸³. Die Botschaft lautet: Nicht die generative KI soll sich primär gesellschaftlichen Bedürfnissen unterordnen, sondern die Gesellschaft soll sich an die generative KI „anpassen“. Aus Angst vor Risiken wie zum Beispiel der drohenden Gefährdung der Demokratie, soll die Gesellschaft erhebliche Ressourcen mobilisieren, um an dem neuen Geschäftsmodell risikominimierend mitzuwirken – für diese Ressourcen wäre die KI-Branche bereit, ein halbes Jahr Zeit zu investieren. Es handelt sich also um einen „Deal“, der einen verantwortungsbewussten und fairen Umgang mit der Gesellschaft suggeriert, doch letztlich auf das doppelte Abschöpfen ihrer Ressourcen abzielt: Nicht nur als (zahlende) Konsumenten generativer KI-Anwendungen sind Staat und Gesellschaft gefragt, sondern als „Mitentwickler“, die schon im Vorfeld in großem Stil investieren sollen. Da es um Steuergeld geht, das in diese Investitionen fließen soll, müssen Politik und Gesellschaft vorab entsprechend motiviert werden. Die Motivation wird über die zu erwartenden zukünftigen „benefits“ generativer KI angefacht; versprochen wird im Schlusssatz der Moratoriumsforderung eine „blühende Zukunft“ für die ganze Menschheit und ein nahezu ewig wählender „KI-Sommer“⁸⁴. Allzu „blumige“ Versprechungen wecken den Verdacht, dass von Geschäftsinteressen abgelenkt werden soll.

2.1.3 PAI Framework – ein „Hauch“ von Medienethik

Normativ konkreter formuliert ist das „Framework“ der *Partnership on AI (PAI)*⁸⁵, das parallel zur Moratoriums-Forderung erarbeitet und am 27. Februar 2023 veröffentlicht wurde (*PAI's Responsible Practices for Synthetic Media. A Framework for Collective Action*)⁸⁶. Das Rahmenwerk fokussiert auf „synthetische Medien“, d.h. auf den Output von visuellen, auditiven oder multimodalen Inhalten, die mit KI generiert oder modifiziert

⁸³ Future of Life Institute 2023

⁸⁴ „Humanity can enjoy a flourishing future with AI. [...] Let's enjoy a long AI summer [...]“ (ebd.)

⁸⁵ Die internationale *Partnership on AI* ist nach Selbstausskunft „an independent, nonprofit 501(c)(3) organization. It was originally established by a coalition of representatives from technology companies, civil society organizations, and academic institutions, and supported originally by multi-year grants from Apple, Amazon, Meta, Google/Deepmind, IBM and Microsoft.“ (s. <https://partnershiponai.org/transparency-governance/>) Es handelt sich um ein weit verflochtenes Netzwerk, dem beispielsweise auch das deutsche *Fraunhofer-Institut für Arbeitswirtschaft und Organisation (IAO)* angehört, sowie das *Future of Life Institute*, welches die Moratoriums-Forderung veröffentlichte.

⁸⁶ „PAIs verantwortungsvolle Praktiken für synthetische Medien. Ein Rahmenwerk für kollektive Maßnahmen“ (Partnership on AI 2023) Launchpartner wie *Adobe*, *BBC*, *CBC/Radio Canada*, *Bumble*, *Open-AI*, *TikTok*, *WITNESS*, und Start-ups im Bereich synthetischer Medien wie *Synthesia*, *D-ID* und *Re-speecher* zählen zu den ersten Unterstützern des Rahmenwerks.

wurden. Es benennt künstlerische, wissenschaftliche und pädagogische Nutzungsbereiche⁸⁷ und umreißt verantwortungsvolle Praktiken für alle Stakeholder.

Synthetische Medien treten als neue Medienkategorie zu herkömmlichen digitalen Medien in Konkurrenz.⁸⁸ Mit ihrer Fähigkeit, deren Gefüge mit perfekten Fälschungen zu unterwandern, können sie nicht nur Schaden auf individueller Ebene anrichten, sondern auch die gesellschaftliche Funktion von Medien als sogenannter vierter Gewalt aushebeln. Eine Medien-Erosion wäre die Folge, die mit einem kaum reparablen Vertrauensverlust einherginge. Dem will das PAI-Rahmenwerk als Selbstverpflichtung aller Stakeholder entgegenwirken: „Wir wollen sicherstellen, dass synthetische Medien nicht eingesetzt werden, um den Menschen zu schaden, sie zu entmachten oder zu entrechteten“.⁸⁹ Da der Fälschungsaspekt im Vordergrund steht, zielen die Vorschläge des Rahmenwerks hauptsächlich auf „Offenlegung“ (disclosure)⁹⁰ ab. Unterschieden wird dabei zwischen direkter Offenlegung, die für den Nutzer sichtbar ist, wie z.B. Hinweisen zum Kontext, inhaltlichen Etikettierungen, Haftungsausschlussformeln und Wasserzeichen, mit denen die KI-generierten Inhalte obligatorisch gekennzeichnet werden sollen, und indirekter Offenlegung, die unsichtbar eingebettet ist, wie z.B. Metadaten und Elementen, die eine Zurückverfolgbarkeit der Daten ermöglichen.⁹¹ Indirekte Offenlegung wird auch für die Entwicklung von Code, Datensätzen und Trainingsmodellen empfohlen⁹², womit Transparenz in Bezug auf den Input der generativen KI geschaffen werden soll. Allerdings wird dies betreffend vor allzu hohen Erwartungen gewarnt.⁹³

Die Wirksamkeit der direkten Offenlegung bedürfte ebenfalls der Überprüfung: Mit Ausnahme der Wasserzeichen, die KI-generierte Inhalte unübersehbar markieren, erschei-

⁸⁷ „Responsible categories of use may include, but are not limited to: Entertainment, Art, Satire, Education, Research.“ (ebd.)

⁸⁸ Auch wenn im PAI-Rahmenwerk der Fokus auf audiovisuellen Inhalten mit ihrer Deepfake-Problematik liegt, stellt sich die Situation für synthetischen Text, wie im Fall von *ChatGPT*, ähnlich dar. Hierzu heißt es im Rahmenwerk: „This framework has been created with a focus on audiovisual synthetic media, otherwise known as generative media, rather than synthetic text which provides other benefits and risks. However, it may still provide useful guidance for the creation and distribution of synthetic text.“ (ebd.)

⁸⁹ Claire Leibowicz, PAI-Leiterin für KI und Medienintegrität, zitiert nach: Heikkilä 2023a.

⁹⁰ „Take steps to provide disclosure mechanisms for those creating and distributing synthetic media. Disclosure can be direct and/or indirect, depending on the use case and context“ (Partnership on AI 2023)

⁹¹ „Direct disclosure is viewer or listener-facing and includes [...] content labels, context notes, watermarking, and disclaimers. Indirect disclosure is embedded and includes [...] applying cryptographic provenance to synthetic outputs [...], applying traceable elements to training data and outputs, synthetic media file metadata, synthetic media pixel composition, and single-frame disclosure statements in videos.“ (ebd.)

⁹² „When developing code and datasets, training models, and applying software for the production of synthetic media, make best efforts to apply indirect disclosure elements [...]“ (ebd.)

⁹³ „(Note: The ability to add durable disclosure to synthetic media is an open challenge where research is ongoing.)“ heißt es in einer vielsagenden Randbemerkung des PAI-Rahmenwerks. (ebd.)

nen die direkten Offenlegungsmechanismen in etwa vergleichbar mit dem „Kleingedruckten“ des Beipackzettels eines Medizinprodukts. Ob der durchschnittliche Medienkonsument, der auf *Youtube* ein KI-generiertes Video mit Fake-Inhalt anklickt, einen solchen „Beipackzettel“ wahrnimmt oder sich für die ausformulierte ethische „Policy“⁹⁴ des Anbieters interessiert, bleibt fraglich. Die Haftungsausschlussformeln zeichnen sich nicht durch Benutzerfreundlichkeit aus: Sie werden wie AGBs leicht überlesen, sichern den Anbieter im Schadensfall aber rechtlich ab. Absichernd für den Anbieter wirkt auch das Konzept der „informierten Einwilligung“ (informed consent) der abgebildeten Subjekte, die sich in einem manipulierten synthetischen Inhalt wiederfinden⁹⁵. Das *PAI*-Rahmenwerk benennt hier nicht konkret, auf welchem Weg und in welcher Form diese Einwilligungen zu erlangen seien, fordert aber eine Transparenz gegenüber den Konsumenten dahingehend, ob und wie es zu Einwilligungen kam. Ferner fordert es „schnelle Anpassungen“⁹⁶ für den Fall, dass sich synthetische Medien bereits als schädlich herausgestellt haben. Insgesamt machen die Forderungen und Richtlinien den Eindruck eines ethischen „Learning by Doing“ und lassen den Unternehmen viel Interpretationsspielraum. Ethisch relevante Themenfelder, wie z.B. das Copyright, die Zukunft der Arbeitswelt oder die Bedeutung von Kultur werden lediglich als Schlagworte angesprochen.⁹⁷

Die Liste der Schäden, die durch synthetische Medien entstehen können, ist lang⁹⁸ und bekannt, doch „gemacht wird es trotzdem“, wie Melissa Heikkilä im *MIT Technology Review* feststellt⁹⁹. Selbst wenn sich die an der *PAI* beteiligten Unternehmen an ihre freiwillige Selbstverpflichtung halten, ist – wie schon bei der Moratoriums-Forderung – unklar, wie mit denjenigen Unternehmen umgegangen werden soll, die sich nicht daran

⁹⁴ Für alle Vertriebswege, darunter Medieninstitutionen, z.B. Sender, aber auch Social Media Plattformen gibt das *PAI*-Rahmenwerk die Empfehlung: „Provide a published, accessible policy outlining the organization’s approach to synthetic media that you will adhere to and seek to enforce.“ (ebd.)

⁹⁵ „Be transparent to content consumers about: How you received informed consent from the subject(s) of a piece of manipulated content [...].“ (ebd.)

⁹⁶ „Make prompt adjustments via labels, downranking, removal, or other interventions like those described here, when harmful synthetic media is known to be distributed on the platform.“ (ebd.)

⁹⁷ „Further, while the ethical implications of synthetic media are vast, implicating elements like copyright, the future of work, and even the meaning of art, the goal of this document is to target an initial set of stakeholder groups [...].“ (ebd.)

⁹⁸ Im Anhang des *PAI*-Rahmenwerks werden u.a. genannt: „Disinformation about an individual, group or organization“, „Exploiting or manipulating children“, „Manipulating democratic and political processes, including deceiving a voter into voting for or against a candidate, damaging a candidate’s reputation by providing false statements or acts, influencing the outcome of an election via deception, or suppressing voters“, „Creating or inciting hate speech, discrimination, defamation, terrorism, or acts of violence“. Die Liste zeigt große Schnittmengen mit Problemfeldern, die bereits von der Medienethik adressiert werden.

⁹⁹ Heikkilä 2023a

beteiligen, generative KI unbekümmerter entwickeln und einsetzen¹⁰⁰ und sich damit Wettbewerbsvorteile verschaffen.¹⁰¹ Sich in einer Marktumgebung, die einen „Wilden Westen“¹⁰² für Künstliche Intelligenz darstellt, zurückzuhalten, dürfte vielen Unternehmen schwerfallen.

2.1.4 OECD Arbeitspapier – Politikberatung für Demokratien

Im Jahr 2023 gab die OECD ein Arbeitspapier zu generativer KI in Auftrag. Es erschien im September 2023 unter dem Titel „*Initial Policy Considerations for Generative Artificial Intelligence*“¹⁰³ und adressiert Regierungen, die mit Forderungen nach einer zukunfts-fähigen KI-Politik konfrontiert sind. In dem wissenschaftlich orientierten Arbeitspapier sind die politisch relevanten Problemdimensionen aufgegliedert in bereits aufgetretene sowie mögliche zukünftige Risiken generativer KI. Letztere gehen über die im *PAI Framework* genannten deutlich hinaus. Als neu beobachtetes Phänomen wird beispielsweise die zunehmende „Informationsverschmutzung“¹⁰⁴ angesprochen, die entstehe, wenn KI-generierter „synthetischer Inhalt mit schwankender Qualität und Genauigkeit in digitalen Räumen zunimmt“¹⁰⁵ und selbst wieder zum Trainingsmaterial nachfolgender generativer KI-Modelle wird, wodurch „ein bösartiger Zyklus“ in Gang gesetzt werde¹⁰⁶. Thematisiert werden auch „unvorhergesehen auftauchendes“ und „menschenähnliches“ Verhalten generativer KI-Modelle, wozu beispielsweise eine „gesteigerte Handlungsfähigkeit“, „Machtstreben“ oder die „Weigerung, abgeschaltet zu werden“¹⁰⁷, gehören. Das Arbeitspapier beteiligt sich aber nicht an Spekulationen, inwieweit die derzeitigen

¹⁰⁰ Hierzu Heikkilä: „Während Unternehmen wie OpenAI versuchen, die von ihnen entwickelten Systeme wie den Textgenerator ChatGPT oder den Bildgenerator DALL-E mit eigenen ‚Leitplanken‘ zu versehen, könnten andere Akteure, die nicht Teil des Pakts sind, tun und lassen, was sie wollen. Schon jetzt lassen Konkurrenten wie Stability.AI – das Start-up, welches das quelloffene KI-Modell Stable Diffusion entwickelt hat – problematischere Inhalte durch.“ (ebd.)

¹⁰¹ Weltweit führende IT-Berater wie *Capgemini*, ein börsennotiertes Unternehmen mit Umsätzen im zweistelligen Milliardenbereich, empfehlen ihren Unternehmenskunden sogar ausdrücklich, als sogenannte „Early Adopter“ generative KI frühzeitig in ihre Marktforschung, ihre Produktionsabläufe, ihr Marketing usw. zu implementieren: „Während Generative KI Fortschritte macht, profitieren die Early-Adopter unter den Unternehmen von neugestalteten Geschäftsmodellen und -prozessen.“ (Capgemini 2023)

¹⁰² „Generative KI sei im Moment noch der wilde Westen, sagt Branchenexperte Henry Ajder, der an den PAI-Regeln mitgearbeitet hat.“ (Heikkilä 2023a)

¹⁰³ Lorenz/Perset/Berryhill 2023. Die Publikation reiht sich ein in die regelmäßig erscheinenden *OECD Artificial Intelligence Papers*.

¹⁰⁴ „information pollution“ (ebd. S. 26)

¹⁰⁵ „...as synthetic content with variable quality and accuracy proliferates in digital spaces...“ (ebd. S. 24)

¹⁰⁶ „a vicious cycle“ (ebd. S. 24)

¹⁰⁷ „generative AI models exhibiting unforeseen ‚emergent behaviours‘, such as increased agency, power-seeking, [...]“ (ebd. S. 27), „such as AI systems‘ refusal to be shut off“ (ebd. S. 28)

KI-Modelle zu „artificial general intelligence“ (AGI) bzw. „starker KI“ führen, und ist weit entfernt davon, Zukunftsversprechen im Stil der Moratoriumsforderung zu machen oder auch nur zu diskutieren.¹⁰⁸

Ähnlich dem *PAI Framework* wertet das Autorenkollektiv des Arbeitspapiers synthetischen Inhalt als besonders risikobehaftet. Immer überzeugendere, täuschend echte und durch Automatisierung massenhaft auftretende Fehlinformationen, die durch bestehende Lösungen nicht mehr einzufangen seien¹⁰⁹, treffen demnach auf Nutzer, die allein durch Schulung („user education“) nicht mehr ausreichend gegen sie immunisiert werden könnten.¹¹⁰ Produktion und Distribution von *deepfakes* in den Händen bössartiger Akteure („malicious actors“) gefährdeten Individuen und Gesellschaft; die Gegenmaßnahmen¹¹¹ bedürften ständiger Erneuerung¹¹². Doch anders als im *PAI Framework* werden die Grenzen dieser Maßnahmen nicht nur auf technologischer, sondern auch auf moralischer Ebene direkt angesprochen. Der Druck, ständig neue technologische und/oder gesetzgeberische Lösungen finden zu müssen, entstehe auch auf Grund der kriminellen Energie, mit der „bad actors“ sämtliche Maßnahmen zu umgehen suchten:

„As with other technologies, bad actors will seek to circumvent mitigation measures. These state-sponsored or commercial actors will not declare their bots or disinformation as AI-generated or follow guidelines or codes of conduct. Obligations to do so will not stop them, just as the illegality of cyberattacks does not prevent cyberattacks.“¹¹³

Neben dem absichtlich missbräuchlichen Einsatz generativer KI werden zahlreiche systemimmanente Probleme in den Fokus gerückt, die ebenfalls nicht leicht zu beheben sind: Unerkannte (oder für vernachlässigbar gehaltene) *Biases* in Trainingsdaten reproduzieren sich im Output massiv als Stereotype, soziale Vorurteile und Diskriminierungen spezifischer gesellschaftlicher Gruppen. Gegenmaßnahmen wie *Red teaming*, der Ein-

¹⁰⁸ Nach einem Referat über Forschungsansätze bezüglich AGI (beispielsweise der *Stanford University* in Zusammenarbeit mit *Google Research*), heißt es hierzu: „Other research findings discussed above, [...], relate to often-discussed technical and philosophical concerns of control, [...], which is beyond the scope of this paper [...].“ (ebd. S. 28)

¹⁰⁹ „However, traditional fact-checking and other existing solutions are generally not scalable in the face of AI-based automation of disinformation.“ (ebd. S. 15)

¹¹⁰ „User education alone becomes insufficient when AI generates more and more convincing disinformation. In addition, it shifts responsibility from systems, companies, and governments to individuals.“ (ebd. S. 15)

¹¹¹ Genannt werden, ähnlich wie im *PAI Framework*, u.a. „watermarks“, „restricting code“ und „guidelines“ (ebd. S. 15), sowie „use-limits“, „monitoring“, „varying deepfake detection techniques“ sowie der *C2PA Standard* („Coalition for Content Provenance and Authenticity“). (ebd. S. 16)

¹¹² „The risks of generating mis- and disinformation [...] demand novel solutions.“ (ebd. S. 15)

¹¹³ ebd. S. 16

satz von klassifizierenden Filtern und *Reinforcement Learning by Human Feedback* (RLHF) seien wichtig, doch keine Garantie für die Sicherheit eines KI-Modells.¹¹⁴

Ein eigenes Kapitel widmet das Arbeitspapier der Problematik der geistigen Eigentumsrechte und dem *copyright*. Die Autoren weisen darauf hin, dass die Trainingsdaten (der Input) generativer KI-Modelle urheberrechtlich geschützte Daten enthalten, für deren Verwendung meist keine Genehmigung der Rechteinhaber eingeholt wurde.¹¹⁵ Die Frage, wie mit diesem Problem umzugehen sei, bleibt unter Verweis auf die derzeitige Rechtsunsicherheit offen. Zitiert wird ein Diskussionspapier der internationalen *World Intellectual Property Organisation* (WIPO) vom Mai 2020: „Should the use of the data subsisting in copyright works without authorization for machine learning constitute an infringement of copyright?“¹¹⁶ Die WIPO beschäftigt sich demnach mit der Frage, ob überhaupt von einem Copyright-Verstoß gesprochen werden kann. Im *OECD*-Arbeitspapier wird das US-amerikanische „fair use principle“ angeführt, welches die Rechte von Copyright-Inhabern limitiert.¹¹⁷ Hier bezieht sich die *OECD* auf das angelsächsische Rechtssystem, in welchem das „copyright“ im Wesentlichen ein Verwerterrecht ist. Verwertung und Vervielfältigung eines Werks sollen prinzipiell einfacher gemacht werden. Das kontinentaleuropäische „Urheberrecht“, welches personengebunden das Recht eines Urhebers auf sein geistiges Eigentum schützt (und auf das sich beispielsweise die *Initiative Urheberrecht*, siehe Kap. 2.2.1, bezieht), spielt im Arbeitspapier der *OECD* nur eine untergeordnete Rolle.¹¹⁸ Der zweite Teil des Kapitels thematisiert ein mögliches copyright oder eine Patentierung KI-generierten Outputs. Hier sei sich die Recht-

¹¹⁴ „Research labs that market their models often use more forceful approaches. These include ‚red teaming‘, in which teams adopt an attacker mindset to probe the model for flaws and vulnerabilities. [...] Other approaches include combinations of dataset cleaning – such as classifiers to filter out erotic content – and ‚Reinforcement Learning by Human Feedback‘ (RLHF). [...] While important, these strategies cannot guarantee a model’s safety.“ (ebd. S. 17)

¹¹⁵ „Generative AI models are being trained on massive amounts of data that includes copyrighted data, mostly without authorisation of the right-owners.“ (ebd. S. 19)

¹¹⁶ WIPO 2020, S. 8. Auch die jüngste WIPO-Conversation (*8th WIPO Conversation on Generative AI and Intellectual Property*), die im September 2023 mit mehr als 2000 Teilnehmern aus 160 Ländern als Hybrid-Konferenz in Genf stattfand, ist noch weit entfernt von einem Konsens, der sich in einem internationalen, von der WIPO verwalteten Vertragswerk niederschlagen könnte.

¹¹⁷ Auch für die WIPO scheint die Diskussion einer „Ausnahmeerlaubnis“ für Trainingsdaten wichtig: „If the use of data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, should an explicit exception be made under copyright law [...] for the use of such data to train AI applications?“ (ebd. S. 8)

¹¹⁸ Die unterschiedliche Behandlung des Rechts auf geistiges Eigentum im angelsächsischen und kontinentaleuropäischen Recht ist auch der Grund dafür, warum viele große US-amerikanische Player ihren EU-Rechtssitz in Irland haben.

sprechung in den USA und Europa einig, dass dies nicht möglich sei, da Copyright und Patentierung menschliche Autorschaft voraussetzen.

Was den großen Themenkomplex des Einflusses generativer KI auf den Arbeitsmarkt betrifft, kann die *OECD* auf eine gesonderte Publikation verweisen, die „2023 edition of the *OECD Employment Outlook*“¹¹⁹. Tendenziell beeinflusse KI die *Qualität* der Arbeit eher positiv; signifikante negative Effekte auf die *Quantität* seien zurzeit kaum nachweisbar bzw. erst in Zukunft wahrnehmbar¹²⁰ – eine Einschätzung, die durchaus kontrovers diskutiert wird. Klaus Kornwachs legt zu diesem Thema eine eigene umfassende Studie vor (siehe Kap. 2.3.1).

Im Zusammenhang mit der zunehmenden „Handlungsfähigkeit“ generativer KI-Systeme, die zur menschlichen Handlungsfähigkeit in Konkurrenz treten, benennt das Papier zwei Risiken, die für eine ethische Beurteilung bedeutsam sind (und auch in der Stellungnahme des *Deutschen Ethikrates*, Kap. 2.2.4, eine größere Rolle spielen): zum einen systemische, verzögert auftretende Schädigungen („systemic delayed harms“) auf individueller Ebene, zum anderen kollektive Entmündigung („collective disempowerment“) auf gesellschaftlicher Ebene.¹²¹ In der Gesamtschau bietet das *OECD* Arbeitspapier einen gut recherchierten Überblick über den aktuellen Forschungsstand. Gegenüber den bisher besprochenen Stellungnahmen ist die Thematik der geistigen Eigentumsrechte, des Energie- und Ressourcenverbrauchs stärker berücksichtigt.¹²²

¹¹⁹ OECD.org 2023. Drei Effekte sind laut *OECD* bisher erkennbar: erstens die Ersetzung menschlicher Arbeit durch künstliche Intelligenz (*displacement effect*), zweitens die wiederum gesteigerte Nachfrage nach Arbeit durch KI-bedingte größere Produktivität (*productivity effect*), und drittens die Schaffung neuer Arbeitsplätze durch KI, da sie neue Aufgaben hervorbringt, für die der Einsatz menschlicher Arbeit vorteilhaft ist (*reinstatement effect*). (Vgl. Lorenz/Perset/Berryhill 2023, S. 20). Da sich diese drei Effekte gegenseitig überlagern, ergibt sich kein eindeutiges Bild; auch stehe die Forschung hier relativ am Anfang.

¹²⁰ „To date, AI has mainly impacted the quality of jobs – generally, in positive ways. [...] While the *Employment Outlook* found little evidence of significant negative effects from AI on the quantity of jobs, this research mostly predates the latest public release of generative AI applications. Negative employment effects of AI might take time to materialise [...].“ (ebd. S. 20)

¹²¹ Im Original: „Two major harms that can arise from increased agency of algorithmic systems: - **Systemic, delayed harms** – non-immediate harms that can be ‚destructive, long-lasting, and hard to fix‘, such as social-media recommender systems based on reinforcement-learning. Such algorithms optimise for metrics that can ‚change or manipulate user’s internal states (e.g. preferences, beliefs, psychology)‘. – **Collective disempowerment** – the perceived danger that model capabilities will perform increasingly important functions in society, taking power away from humans. This could take the form of gradually ceding decision-making to generative AI systems. Its second impact is intensifying concentrations of power and the ability to reap the benefits of AI – already a concern.“ (ebd. S. 27)

¹²² „generative AI’s ecological footprint and natural resources consumption from the tremendous amounts of computing power required for deep learning [...].“ (ebd. S. 27)

2.2. Wissenschaftliche Studien/kollektive Stellungnahmen in Deutschland

Im Jahr 2023 sind auch in Deutschland Studien und Stellungnahmen zu Künstlicher Intelligenz und KI-Ethik erschienen, die generative KI einbeziehen. Die hier vorgestellten Beispiele sind, mit Ausnahme der *Initiative Urheberrecht* (Kap. 2.2.1) in staatlichem Auftrag entstanden. Der Forschungsbericht des *Büros für Technikfolgenabschätzung beim Deutschen Bundestag* (Kap. 2.2.2) referiert über technologische Grundlagen sowie Möglichkeiten und Grenzen der Technologie und ihrer Anwendungen. Die *LEAM Machbarkeitsstudie* (Kap. 2.2.3) ermittelt, da es „aktuell in Europa kein dezidiertes KI-Rechenzentrum“ gibt, „das für die Entwicklung international kompetitiver Foundation-Modelle ausreicht“¹²³, ob Deutschland diese Lücke schließen kann, und wenn ja, unter welchen Voraussetzungen. In der Stellungnahme des *Deutschen Ethikrates* (Kap. 2.2.4) finden sich grundsätzliche ethische Erörterungen, die eine Basis für alle anwendungsorientierten Normen legen können. Die Stellungnahme des *Deutschen Kulturrates* (Kap. 2.2.5) wurde aufgenommen, da sie den Input generativer KI *expressis verbis* thematisiert.

2.2.1 *Initiative Urheberrecht* – Gegenwehr der deutschen Kunst- und Kreativbranche

Am 19. April 2023, vor den Trilogverhandlungen zum EU *AI Act* (Kap. 2.1.1), legte die deutsche *Initiative Urheberrecht* („IU“) einen umfangreichen Forderungskatalog vor: "Urheber:innen und Künstler:innen fordern Maßnahmen zum Schutz vor generativer KI in der Europäischen KI-Verordnung". Die „unterzeichnenden 43 Verbände und Gewerkschaften vertreten Tausende Autor:innen, ausübende Künstler:innen und kreativen Urheber:innen verschiedener Branchen“¹²⁴, Buch- und Textautoren, Drehbuchautoren, Schauspieler, Synchronregisseure und -sprecher, Designer, Fotografen, Illustratoren, Journalisten, Bildende Künstler, Musikurheber und ausübende Musiker, Spieleautoren, Sprecher und Übersetzer, die jeweils ihre branchenspezifischen Forderungen und Empfehlungen anbringen. „Generative KI muss entlang ihrer gesamten Entstehungskette reguliert werden“¹²⁵, heißt es in den vorangestellten branchenübergreifenden Forderungen. Dabei kommt dem „Input“ eine mindestens genauso große Bedeutung wie dem Output zu: „Die ungefragte Nutzung des Trainingsmaterials, seine intransparente Verarbeitung und die absehbare Substitution der Quellen durch den Output generativer KI werfen grundsätzliche Fragen nach Verantwortung und Haftung wie auch Vergütung auf, die zu

¹²³ Bienert et al. 2023, S. 228

¹²⁴ IU 2023, S.3

¹²⁵ ebd. S. 13

klären sind, bevor der Schaden irreversibel ist.“¹²⁶ Die Rede ist nicht nur von wirtschaftlichem Schaden durch Disruption in der Arbeitswelt, sondern von einer gesamtgesellschaftlichen „Ruptur“ (so wie auch analysiert von Klaus Kornwachs, siehe Kap. 2.3.1):

„Die fortgesetzte Ausbeutung unserer Arbeit und unserer individuellen, schöpferischen Intelligenz verstößt gegen grundlegende Menschenrechte. Sie schadet den Quellen des gesamten Buch- und Kunstmarkts, und markiert den Beginn einer gesellschaftlichen Ruptur, die nur durch aktives Gestalten der Regeln in eine soziale, gerechte und sinnvolle Entwicklung gewandelt werden kann.“¹²⁷

Gewarnt wird vor einem „Verlust an Kulturtechnik“¹²⁸ sowie einem drohenden „Erkenntnisverlust“¹²⁹, der entsteht, wenn generative KI künstlerisches und kreatives Handwerk übernimmt. Künstler und Kreative nehmen ihre gesamtgesellschaftliche Verantwortung ernst, wenn sie „weder Mit-Verursacher noch Opfer des Schadens“ werden wollen, „den generative KI verursacht“¹³⁰. Sie betonen, dass es nicht bloß um die Wahrung der eigenen „Pfründe“ geht, wie mancher Politiker vermuten könnte. Der Schaden, der Urhebern und den „Quellen von Kreativität, Kunst, Bildung, Wissen, Empathie“¹³¹ durch generative KI entsteht, betreffe „das Rückgrat unserer demokratischen Gesellschaften und unserer kulturellen Identität“¹³².

In dem Forderungskatalog wird zunächst auf die von *OpenAI* selbst publizierten Sicherheitsrisiken¹³³ in Zusammenhang mit *ChatGPT* und *Dall-E* eingegangen. Diese Veröffentlichung stelle aber nur „die Spitze eines ständig wachsenden Eisbergs von Sicherheitsbedenken im Zusammenhang mit generativer KI“¹³⁴ dar. *OpenAI* versuche überdies die Verantwortung für diese Risiken „auf nachgelagerte ‚Down-Stream‘ Anwender [...] abzuwälzen, wohl wissend, dass diese nicht in der Lage sein werden, das Problem zu beheben [...]“¹³⁵. Die Firma arbeite nicht an „Lösungen“, sondern nur daran, „ihre gefähr-

¹²⁶ ebd. S. 2

¹²⁷ ebd. S. 15

¹²⁸ ebd. S. 20

¹²⁹ ebd.

¹³⁰ ebd. S. 8

¹³¹ ebd.

¹³² ebd.

¹³³ Im technischen Bericht vom 27. März 2023 (*GPT-4 Technical Report* sowie *GPT-4 System Card*) werden aufgezählt: „Halluzinationen; schädliche Inhalte; Beeinträchtigung der Darstellung, Bereitstellung und Qualität der Dienste; Desinformation und Beeinflussung; Verbreitung konventioneller und unkonventioneller Waffen; Datenschutz; Cybersicherheit; Potenzial für neues Risikoverhalten; Wechselwirkungen mit anderen Systemen; wirtschaftliche Auswirkungen; Beschleunigung; [Erzeugung] übermäßigen Vertrauens.“ (zitiert nach: IU 2023, S. 4)

¹³⁴ ebd. S. 5. Die IU warnt (unter Berufung auf *Europol* und KI-Expertin Melissa Heikkilä) eindrücklich: „Es ist offensichtlich, dass KI Chatbots ‚eine Sicherheitskatastrophe‘ sind. Sogar *Europol* sah die Notwendigkeit, eine klare Warnung auszusprechen, dass *ChatGPT* bereits heute für eine bedeutende Zahl von Straftaten eingesetzt werden kann.“ (ebd. S. 6)

¹³⁵ ebd.

lichen Systeme so schnell wie möglich weiter zu verbreiten“¹³⁶. Die Monopolstellung weniger nicht-europäischer Tech-Giganten, die Plattformdienste wie *Google*, *Microsoft*, *Meta*, *Amazon* oder *Baidu* betreiben und die Nutzung generativer KI kontrollieren, indem sie z.B. Nutzer mit personalisierten KI-Medieninhalten versorgen bzw. „überfluten“¹³⁷, verhindere echten Wettbewerb und könne demokratische Systeme bedrohen.

Ein zentraler Kritikpunkt ist die umstrittene Text- und Data-Mining-Ausnahme der EU. Die DSM-Richtlinie aus dem Jahr 2019, die das europäische Urheberrecht an Erfordernisse der fortschreitenden Digitalisierung anpasst¹³⁸, erscheint der IU im Hinblick auf die neuartige generative KI „verfehlt“:

„Die beispiellose massenhafte Ausbeutung menschlicher Kreativität für generative KI darf nicht mit dem Verweis auf die Ausnahme für Text- und Data-Mining entschuldigt werden. [...] Es ist offensichtlich, dass der Zweck der Text- und Data-Mining-Ausnahme nicht darin bestanden haben kann, die heutige automatische Massenextraktion urheberrechtlich geschützter Werke zu ermöglichen, um große Foundation Models zur Generierung von Inhalten zu trainieren, geschweige denn *AI Fakes* zu ermöglichen.“¹³⁹

Der Ausdruck „automatische Massenextraktion“ verweist auf Künstliche Intelligenz als „Extraktionstechnologie“, ein Begriff, der von Kate Crawford verwendet wird (siehe Kap. 3.1) und sich sowohl auf die Extraktion materieller Ressourcen als auch auf Datenextraktion bezieht. Die IU stellt klar, dass die Extraktion urheberrechtlich geschützten Materials auch dann nicht rechtmäßig sein könne, wenn man argumentiere, dass generative KI durch dieses Material wie ein menschlicher Schöpfer „inspiriert“ würde¹⁴⁰. Denn es handele sich immer noch um gestohlenen Material, das digital auseinandergenommen und wieder zusammengesetzt werde, „um die Aufforderung des Benutzers wiederzugeben, nur so, dass die Quelle nicht zurückverfolgt werden kann“¹⁴¹, der Diebstahl also nicht offensichtlich zu erkennen ist. Ohne direkt von einem Betrugmodell zu sprechen, artikuliert die IU hier Kritik an den immateriellen Voraussetzungen generativer KI, die deutlicher nicht sein könnte. Ferner spricht sie generativer KI die Fähigkeit ab, wie

¹³⁶ ebd. Durch das *ChatGPT plugins*-Modell werde es für Unternehmen noch leichter, über eine API-Schnittstelle den Chatbot zu implementieren; „im Gegenzug erhält *ChatGPT* weltweit Zugang zu aktuellen Informationen in Echtzeit, kann auf Rechenleistungen zugreifen und Dienste Dritter nutzen. Der Echtzeitzugang zu Internetdaten verstärkt wiederum die dem System immanenten Risiken.“ (ebd.)

¹³⁷ Die IU bezieht sich hier u.a. auf einen Kommentar der *Digital Humanism Initiative*, die vor generativer KI warnt, wenn sie in der Hand von „Big Tech“ liege: Für die „*öffentliche Wahrnehmung freier Meinungsäußerung und Beteiligung*“ bestehe die Gefahr, dass sie „*von Inhalten übernommen und überflutet*“ werde, „*die bewusst für Fehlinformationen, völligen Unsinn oder für die Untergrabung einer Wahrnehmung der demokratischen, kollektiven Zugehörigkeit konzipiert sind.*“ (zitiert nach: IU 2023)

¹³⁸ Die Problematik des „Text- und Dataminings“ wird in Kapitel 2.2.5 anlässlich der Stellungnahme des Deutschen Kulturrates näher erläutert.

¹³⁹ IU 2023, S. 9

¹⁴⁰ ebd. S. 10

¹⁴¹ ebd.

ein kreativer Mensch etwas zur „kulturellen Vielfalt“¹⁴² beitragen zu können, und stellt einen direkten Vergleich an:

„Während die generative KI auf Vorhersagen und nächsten Wahrscheinlichkeiten beruht, zeichnet sich die künstlerische und journalistische Arbeit von Menschen durch ihre Unvorhersehbarkeit aus: Kontextualisierung, soziale und kulturelle Verortung und höchst individuelle Bedeutungszuschreibung. Autoren und Künstler fügen dem Leben und der Kultur etwas Neues und Ungehörtes, Ungesehenes, Unvorhersehbares und Unerzähltes hinzu.“¹⁴³

In der ethischen Debatte ist es neben der *Initiative Urheberrecht* auch der Deutsche Kulturrat (Kap. 2.2.5), der den Schwerpunkt auf das Konkurrenzverhältnis generativer KI zu menschlicher Kulturkreation legt. Während der Deutsche Kulturrat die ganze ideelle Bedeutung der Kultur- und Kreativbranche in die Waagschale legt, kommt bei der IU die Zahl der von ihr vertretenen Kulturschaffenden und Urheber hinzu, die ihren Forderungen noch mehr Gewicht verleiht.

2.2.2 Forschungsbericht des TAB – der kritische Blick

Drei Tage nach der Veröffentlichung der *Initiative Urheberrecht* wurde am 21. April 2023 der Forschungsbericht „ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen“, verfasst von Steffen Albrecht, als „Hintergrundpapier Nr. 26“ des *Büros für Technikfolgen-Abschätzung beim Deutschen Bundestag* (TAB) veröffentlicht.¹⁴⁴ Der recht frühe Zeitpunkt und die detaillierte Betrachtung¹⁴⁵ zeigen einerseits, wie schnell *ChatGPT* in all diesen Bereichen auch in Deutschland adaptiert wurde, andererseits, wie schnell nach seiner Markteinführung gravierende Probleme übergreifend aufgetaucht sind, die im dritten Kapitel des Papiers wie folgt aufgegliedert sind:

¹⁴² ebd.

¹⁴³ ebd.

¹⁴⁴ Albrecht 2023. Das TAB wird seit 1990 vom *Institut für Technikfolgenabschätzung und Systemanalyse* (ITAS) des *Karlsruher Instituts für Technologie* (KIT) betrieben. Das auf europäischer Ebene, aber auch international vernetzte ITAS ist laut Selbstauskunft „das führende Institut für Technikfolgenabschätzung in Deutschland und weltweit“ und forscht „nicht aus der wissenschaftlichen Eigenlogik heraus, sondern aus dem Dialog auf Augenhöhe mit der Gesellschaft“. (KIT 2023) Da das ITAS an Politikberatung nicht nur für den Deutschen Bundestag, sondern auch für das EU-Parlament beteiligt ist, kann angenommen werden, dass in die Position des EU-Parlaments zum *AI Act* (siehe Kap. 2.1.1) Forschungsergebnisse des rund 90-seitigen Papiers eingeflossen sind.

¹⁴⁵ „Möglichkeiten und Grenzen“ von *ChatGPT* und vergleichbaren Modellen werden nicht nur allgemein, sondern auch entlang ihrer Anwendungsmöglichkeiten in Unternehmen, im Medizinwesen, im Bereich Informationssuche, Journalismus und Öffentlichkeit, im Rechtswesen und in der öffentlichen Verwaltung behandelt. Den „Auswirkungen von Chat-GPT in Bildung und Forschung“ und „rechtlichen Aspekten und Fragen der Nachhaltigkeit“ sind eigene Kapitel gewidmet. (s. Albrecht 2023)

- Probleme mit längeren Konversationen
- Probleme mit Logik, Faktentreue und Weltbezug
- Begrenzungen aufgrund des Trainingsmaterials
- Blackboxcharakter
- Umgang mit den Begrenzungen/bisherigen Erfahrungen ¹⁴⁶

Auch wenn durch technologische Weiterentwicklungen und flankierende Maßnahmen einige Probleme inzwischen abgemildert sind ¹⁴⁷, macht sich das grundsätzliche Problem mit „Faktentreue und Weltbezug“ in allen Anwendungen bemerkbar: „Da ChatGPT ein Wissensmodell fehle, antworte es mit linguistisch plausiblen, aber erfundenen Texten. Dieses Problem lasse sich auch nicht durch mehr Zeit oder Aufwand bei der Entwicklung beheben, sondern höchstens durch neue Ansätze in der KI-Forschung“, referiert Steffen Albrecht den Forschungsstand. ¹⁴⁸ Obwohl generative KI mit gigantischen Datenmengen an „Weltwissen“ ¹⁴⁹ unterlegt ist, offenbart sie einen eklatanten Mangel an Wissen, den sie bisher nicht kompensieren konnte. Denn um Wissenslücken zu schließen und überhaupt Kriterien für den Wahrheitsgehalt einer Aussage zu entwickeln, müsste sie abstrahierende Schlüsse aus den Inhalten des Trainingsmaterials ziehen können und über Möglichkeiten eines Abgleichs mit der realen Welt verfügen. ¹⁵⁰

Beides ist nicht der Fall, so dass sich Beschaffenheit und Umfang des Trainingsmaterials unmittelbar auf den Output auswirken, und sich Desiderate während der von Wahrscheinlichkeitsberechnungen gesteuerten Verarbeitung eher potenzieren denn abmildern. Genannt werden in diesem Zusammenhang die zeitlich-„historische“ ¹⁵¹ und die geografisch-sprachliche Limitierung der Trainingsdaten, in welchen die englische Spra-

¹⁴⁶ ebd., Inhaltsverzeichnis S. 5

¹⁴⁷ Siehe beispielsweise: „OpenAI beeilte sich nach den ersten Veröffentlichungen zu diesen Problemen, die mathematischen Fähigkeiten von ChatGPT zu verbessern [...]. [...] Für GPT-4 geht das Unternehmen von gegenüber ChatGPT verbesserten mathematischen Fähigkeiten aus [...], zudem wurde mit Plugins die Möglichkeit zur Einbindung weiterer, z. B. auf Berechnungen spezialisierter Systeme geschaffen [...].“ (ebd. S. 40, Fußnote 26)

¹⁴⁸ ebd. S. 41. Das Problem erfundener Fakten wird auch unter dem Begriff „Halluzination“ gefasst. ChatGPT erfand auch Quellenangaben: „Die Texte und Referenzen bzw. Links zu Websites klingen zwar plausibel, haben aber keine faktische Basis [...].“ (ebd. S. 40)

¹⁴⁹ „Das ChatGPT zugrundeliegende Modell [...] beruht auf einem Trainingsmaterial von 300 Mrd. Textbestandteilen.“ (ebd. S. 9)

¹⁵⁰ „Als Limitation sprachverarbeitender KI-Modelle wird angesehen, dass sie [...] keine Abstraktionen bilden können, wie z.B. Gesetzmäßigkeiten über die Gegenstände, die in den Texten thematisiert werden [...]. Die Systeme können kein Verständnis im menschlichen Sinn davon entwickeln, was für Antworten sie geben und ob diese korrekt oder falsch sind. Ihre Ausgabe ‚beruht nicht auf einer kommunikativen Absicht, einem Modell der Welt oder der Verfassung ihres Gegenübers‘ (Bender et al. 2021, S. 616, im Original: ‚[...] is not grounded in communicative intent, any model of the world, or any model of the reader’s state of mind‘).“ (ebd. S. 40)

¹⁵¹ „Im Fall von ChatGPT wie auch GPT-4 wurden nur Daten bis September 2021 im Trainingsmaterial berücksichtigt [...]. Gesellschaftliche Entwicklungen, die seither stattfanden, können in den Ausgaben der Modelle nicht berücksichtigt werden.“ (ebd. S. 42)

che bevorzugt vertreten ist.¹⁵² Abgesehen von Limitierungen, die sich durch quantitative Erweiterung der Datenbasis möglicherweise noch beheben lassen¹⁵³, ist die inhaltliche Qualität des Trainingsmaterials kaum kontrollierbar. Voreingenommenheit in den Trainingsdaten, z.B. „in Hinblick auf Geschlecht, Religion und ethnische Zuschreibung“¹⁵⁴, gehe als Bias in den Output über. „Allerdings lassen sich Bias und Rassismus auf Ebene der Trainingsdaten kaum herausfiltern, weil das zu aufwendig wäre und weil selbst bei vielen Ressourcen und gutem Willen die Erkennung bei impliziten Äußerungen nicht einfach ist“¹⁵⁵, heißt es weiter. Zusätzlicher Ressourcenaufwand durch eine „Vergrößerung der Modelle bis an technische bzw. finanzielle Grenzen“¹⁵⁶ sei in Anbetracht der (zu geringen) „erwartbaren Fortschritte“¹⁵⁷ nicht sinnvoll. Da Systementwickler wie *OpenAI* technische Details zu Trainingsdaten nicht offenlegen, sei die Forschung überdies erschwert.¹⁵⁸

Der erhebliche Ressourcenverbrauch generativer KI-Modelle wird in dem Forschungsbericht mehrfach angesprochen: Kosten für die „Bereitstellung und die Unterhaltung der Hardwareinfrastruktur“¹⁵⁹, der sehr hohe Energieverbrauch der Systeme „beim Training und in der Nutzungsphase“¹⁶⁰, damit verbundene „Umweltauswirkungen“¹⁶¹ und last not least die großen Mengen menschlicher Arbeitskraft, die zur Entwicklung der Modelle benötigt werden, sowie die Bedingungen, unter denen diese Arbeitskraft in Anspruch genommen wird. Der Aspekt menschlicher Arbeit hat hier mehrere Dimensionen, wobei die vergleichsweise gut bezahlte Arbeit der Entwickler das geringere Problem darstellt. Untergeordnete Tätigkeiten jedoch, wie die Feinjustierung der KI-Modelle durch menschliches Feedback, wurden von Firmen wie *OpenAI* ausgelagert und schlecht entlohnt.¹⁶²

¹⁵² „Außerdem ist das Trainingsmaterial hinsichtlich der repräsentierten Sprachen stark durch englische Texte geprägt, andere Sprachen sind in geringerem Umfang vertreten, sodass die Ergebnisse im Vergleich zum Englischen schlechter ausfallen [...]“ (ebd. S. 42)

¹⁵³ Mit dem Forschungsstand vom April 2023 heißt es hierzu: „So könnten explizit formalisierte Wissensbestände in die Trainingsdaten aufgenommen werden, die KI-Modelle könnten während der Nutzungsphase, bei der Beantwortung von Anfragen, auf solche Wissensbestände zugreifen, oder sie werden mit Mechanismen für den Zugriff auf das in großen Textkorpora vorhandene, nicht vorstrukturierte Wissen (Retrieval) ausgestattet [...]“ (ebd. S. 41). Seit dem 27. September 2023 kann *ChatGPT* das Internet in Echtzeit nutzen und entsprechend aktuelle Informationen verarbeiten.

¹⁵⁴ ebd. S. 69

¹⁵⁵ ebd. S. 69

¹⁵⁶ ebd. S. 43

¹⁵⁷ ebd. S. 43

¹⁵⁸ „Diese Intransparenz wird von Forschenden kritisiert, weil sie die Weiterentwicklung der Forschung zu den Modellen, aber auch etwa zur Rolle der Trainingsdaten, verhindert [...]“ (ebd. S. 44)

¹⁵⁹ ebd. S. 30

¹⁶⁰ ebd. S. 30

¹⁶¹ ebd. S. 85

¹⁶² „Damit ist nicht in erster Linie die Arbeit der Entwickler/innen gemeint, sondern diejenigen, die für meist geringe Entlohnung und ohne Aufstiegs- und Weiterbildungsmöglichkeiten einfache Tätigkeiten wie

Alle diese ethisch hoch relevanten Punkte, die Voraussetzungen generativer KI betreffen, sind – außer bei Kate Crawford (siehe Kap. 3.1) – in der verfügbaren Literatur und den kollektiven Stellungnahmen meist nur eine Randnotiz.

Des Weiteren geht es um die menschliche Arbeit, auf der das Trainingsmaterial basiert, d.h. „menschlich erzeugte Texte aus dem Internet“¹⁶³, die „zum größten Teil unentgeltlich“¹⁶⁴ genutzt werden, sowie um die Erstellung von Prompts, welche eine eigene Arbeitsleistung der Nutzer darstellen. In Bezug auf das technologische Problem der mangelnden Rückverfolgbarkeit der Trainingsdaten (oder „Quellen“) in der *Black Box*¹⁶⁵ sieht Steffen Albrecht interdisziplinären Forschungsbedarf.¹⁶⁶ Auch wird die bereits von der *Initiative Urheberrecht* (Kap. 2.2.1) bezweifelte Rechtmäßigkeit in Frage gestellt: Erstens, ob es sich bei den Trainingsverfahren für generative KI um eine „automatisierte Analyse von Werken“¹⁶⁷ handelt, die das Urheberrecht § 44b UrhG im Rahmen von Text- und Datamining zulässt.¹⁶⁸ Zweitens die Frage, ob Prompt-Autoren als „natürliche Personen Anspruch auf urheberrechtlichen Schutz erheben können“¹⁶⁹, wenn sie durch einen detailliert formulierten Prompt (sog. „Promptkunst“) den Output des Systems wesentlich mitgestaltet haben.

Als letzter Punkt sei das „indirekte Risiko“¹⁷⁰ des sogenannten *Automation Bias* herausgegriffen: „Aus der Forschung zur Mensch-Maschine-Interaktion ist die Tendenz bekannt, die Ergebnisse maschineller Verarbeitung unkritisch zu akzeptieren und nach ihnen zu handeln, gerade wenn sich ihr Zustandekommen nicht nachvollziehen lässt

das Kodieren von Daten übernehmen, um die Systeme zu trainieren. [...] Im Fall von OpenAI wurden Arbeiter/innen in Kenia gegen geringen Lohn einer psychisch z. T. sehr belastenden Tätigkeit ausgesetzt.“ (ebd. S. 85)

¹⁶³ ebd. S. 10

¹⁶⁴ ebd. S. 10

¹⁶⁵ „Bei einem KI-Modell zur Sprachverarbeitung lässt sich nach Abschluss des Trainings weder ohne weiteres überprüfen, welche Daten im Modell enthalten sind, noch lassen sich diese korrigieren bzw. löschen [...]“ (ebd. S. 83)

¹⁶⁶ „Welches interdisziplinäre Wissen fehlt, um die Blackboxen der KI-Modelle zu öffnen und sie theoretisch besser zu durchdringen?“ (ebd. S. 89)

¹⁶⁷ ebd. S. 84

¹⁶⁸ § 44 b Text- und Datamining (UrhG) im Wortlaut: „(1) Text und Data Mining ist die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen. (2) Zulässig sind Vervielfältigungen von rechtmäßig zugänglichen Werken für das Text und Data Mining. Die Vervielfältigungen sind zu löschen, wenn sie für das Text und Data Mining nicht mehr erforderlich sind. (3) Nutzungen nach Absatz 2 Satz 1 sind nur zulässig, wenn der Rechtsinhaber sich diese nicht vorbehalten hat. Ein Nutzungsvorbehalt bei online zugänglichen Werken ist nur dann wirksam, wenn er in maschinenlesbarer Form erfolgt.“ (Bundesministerium der Justiz 2023)

¹⁶⁹ Albrecht 2023, S. 84

¹⁷⁰ ebd. S. 57

[...].“¹⁷¹ Maschinengläubigkeit auf menschlicher Seite trifft auf Intransparenz der technologischen Systeme: „Kritisiert wird auch, dass die Systeme ihr Nichtwissen bzw. bestehende Unsicherheiten nicht offenlegen, so dass die Nutzenden über ihre Leistungsfähigkeit getäuscht werden [...].“¹⁷² - Insgesamt zeigt sich an dem Forschungsbericht, dass die deutsche Perspektive auf die ersten Produkte und Anwendungsformen generativer KI – einer US-amerikanischen Technologieentwicklung – von Anfang an deutlich distanzierter und kritischer ist als die internationale Sichtweise.

2.2.3 LEAM Machbarkeitsstudie – technologische Souveränität für Deutschland

Im Auftrag des *Bundesministeriums für Wirtschaft und Klimaschutz* wurde die „Machbarkeitsstudie zum Aufbau und Betrieb eines dezidierten KI-Hochleistungsrechenzentrums für das Trainieren großer KI-Modelle in Deutschland“, kurz: die *LEAM* Machbarkeitsstudie¹⁷³ von einem Expertenkollektiv erstellt und am 24. Januar 2023 im Rahmen der *LEAM* Konferenz in Berlin vorgestellt. *LEAM* steht für „Large European AI Models“ und ist eine Initiative des KI-Bundesverbandes, die zum Ziel hat, die deutsche Wirtschaft „von proprietären, amerikanischen Foundation-Modellen“¹⁷⁴ unabhängig zu machen. Bislang fehlt in Deutschland die Rechenkapazität für ein eigenes Foundation-Modell, was zu einem Entwicklungsrückstand gegenüber den USA und China geführt hat. Ein KI-Modell „made in Germany“, das „die Bedürfnisse deutscher Unternehmen berücksichtigt“¹⁷⁵, hätte den Vorteil, mit europäischem Datenschutz und EU-Regulierungen konform zu sein, und es wäre verbunden mit ethisch konnotierten Begriffen wie „Vertrauenswürdigkeit“, „Transparenz“, „Reduktion von Bias“ und „ökologischer Nachhaltigkeit“¹⁷⁶.

Die Entwicklung einer Infrastruktur für dieses Projekt ist ambitioniert und mit einem kalkulierten Bedarf von 380 Millionen Euro nicht leicht finanzierbar¹⁷⁷, verspricht indes auf lange Sicht Wettbewerbsvorteile für Deutschland. Die 371-seitige Studie, an der u.a. Experten des *Fraunhofer-Instituts für intelligente Analyse- und Informationssysteme*

¹⁷¹ ebd. S. 57

¹⁷² ebd. S. 44

¹⁷³ Bienert et al. 2023

¹⁷⁴ ebd. S. 227

¹⁷⁵ ebd. S. 13

¹⁷⁶ „Unsere Vision ist ein vertrauenswürdiges KI-Modell, das Open Source läuft, europäischen Standards folgt und bei dessen Entwicklung von Anfang an auf Transparenz, die Reduktion von Bias und ökologische Nachhaltigkeit geachtet wird“ sagt Dr. Christian Temath, Geschäftsführer bei KI:NRW. “ (Hasenau 2023)

¹⁷⁷ „Die Autor:innen kalkulieren für den Aufbau und Betrieb eines KI-Rechenzentrums über vier Jahre einen Bedarf von rund 380 Millionen Euro. Für die Finanzierung wurde ein Modell entwickelt, das öffentliche und privatwirtschaftliche Mittel berücksichtigt.“ (Bienert et al. S. 228)

(IAIS) und des *Deutschen Forschungszentrums für Künstliche Intelligenz* (DFKI) federführend beteiligt sind, und in der Interviews mit zahlreichen Unternehmern ausgewertet wurden, kommt zu dem Schluss, dass die Umsetzung des Projekts realisierbar ist und zeitnah erfolgen sollte ¹⁷⁸. Im Zusammenhang der vorliegenden Arbeit ist von Interesse, wie ethische Aspekte miteinbezogen und gewertet wurden. Die Studie benennt Defizite der bestehenden US-amerikanischen und chinesischen Foundation-Modelle, welche „europäischen Qualitätsstandards“ ¹⁷⁹ in puncto inhaltlicher und sprachlicher Funktionalität, Verfügbarkeit, Sicherheit und Verlässlichkeit sowie „ethischer Akzeptabilität“ ¹⁸⁰ nicht entsprechen würden. Für letztere wird ausgeführt:

„Die Ausgaben der Modelle verletzen mitunter durch Bias (Aussagen/Entscheidungen basierend auf falschen Vorurteilen) und Toxizität (Verwendung von ethisch oder stilistisch-ästhetisch nicht akzeptablen sprachlichen Ausdrucksweisen) die defacto Standards für den Einsatz in Wirtschaft, Politik und Bildung. Daher muss es für die Anwendung möglich sein, Korrektur- und Filtermaßnahmen eigenständig zu definieren, in die Modelle zu integrieren und anzupassen.“ ¹⁸¹

Die „Verhinderung des Fehlverhaltens“ solle eine „zentrale Priorität bei Entwurf, Training und Evaluation der Modelle sein“ ¹⁸². Auch wenn Bias, Toxizität und Falschaussagen beim Output nicht gänzlich eliminiert werden könnten, sei es „für den Erfolg der Foundation-Modelle und deren Akzeptanz in Wirtschaft und Gesellschaft essentiell, Grad und Häufigkeit von potentiellen Fehlverhalten empirisch zu bestimmen bzw. vorherzusagen.“ ¹⁸³ Im Gegensatz zu den bestehenden Modellen, die bezüglich ihrer Leistungsfähigkeit weitgehend intransparent sind, wäre demnach für künftige Anwender europäischer Modelle besser nachvollziehbar, mit welcher Qualität sie rechnen könnten.

Speziell um Falschaussagen („Halluzinationen“) zu minimieren, wird die Verbindung mit großen, explizit formalisierten Wissensbeständen empfohlen. Besonders geeignet seien hierfür die sogenannten Wissensgraphen, z.B. DBPedia und WikiData, die unentgeltlich nutzbar sind. Durch sie ließen sich die Modelle an die „Dynamik des Wissens“ ¹⁸⁴ anpassen und „Lücken im Detailwissen“ ¹⁸⁵ schließen. Ferner sei die Multilingualität der euro-

¹⁷⁸ „Die Gefahr, dass sich in den USA monopolartige Strukturen – ähnlich Google bei Suchmaschinen – bilden, wird [...] größer. Damit Deutschland nicht weiter zurückfällt, ist eine zeitnahe Umsetzung des Konzeptes für die LEAM-Infrastruktur nötig. Das Momentum, das im Rahmen der Erstellung dieser Machbarkeitsstudie gewonnen wurde – durch eine enge Zusammenarbeit zwischen Forschung, Wissenschaft, Wirtschaft und Start-ups – sollte genutzt werden, um die Zielsetzung weiter voranzutreiben.“ (ebd. S. 229)

¹⁷⁹ ebd. S. 63

¹⁸⁰ ebd. S. 64

¹⁸¹ ebd. S. 64

¹⁸² ebd. S. 97

¹⁸³ ebd. S. 98

¹⁸⁴ ebd. S. 99

¹⁸⁵ ebd. S. 99

päischen Gesellschaft zu schützen.¹⁸⁶ So sollen für die ersten europäischen Modelle neben den „bisher eingesetzten gecrawlten Webdaten“ auch spezielle europäische Lern-
daten als Input zum Einsatz kommen, „die sowohl die europäischen Sprachen als auch
die für die Wirtschaft relevanten Gegenstandsbereiche besser abdecken.“¹⁸⁷ Auch das
Leibniz-Institut für Deutsche Sprache (IDS) hat laut der Studie „seine Bereitschaft erklärt,
die hochqualitativen Textkorpora des Instituts von rund 40 Milliarden Wörtern für das
Training europäischer Modelle verfügbar zu machen“¹⁸⁸, sowie eine wissenschaftliche
Begleitung in Fragen der „Repräsentanz der deutschen Sprache in den europäischen
Foundation-Modellen“¹⁸⁹ angeboten.

Im achten Kapitel der Studie geht es konkret um die Voraussetzungen des „Aufbaus
eines KI-Hochleistungsrechenzentrums“¹⁹⁰ in Deutschland. Es werden ethisch relevan-
te ökologische Nachhaltigkeitsziele formuliert, die sich auch ökonomisch, im Sinne einer
Kostensenkung auswirken sollen.¹⁹¹ - Der höhere Anspruch Europas an Qualitätsstan-
dards und ethische Standards zieht sich als argumentativer Faden durch die gesamte
LEAM-Studie. Da sie jedoch in erster Linie auf die Wettbewerbsfähigkeit Deutschlands
im globalen KI-Wettbewerb ausgerichtet ist, stehen wirtschaftsethische bzw. unterneh-
mensethische Aspekte im Vordergrund, während Urheberrechte und menschliche Autor-
schaft außer Sichtweite geraten – es würde die Aufbruchsstimmung zu sehr dämpfen,
beispielsweise die Herkunft der „europäischen Lerndaten“ zu hinterfragen.

2.2.4 Stellungnahme des *Deutschen Ethikrates* – Handlungsurheberschaft und Verant- wortung

Der *Deutsche Ethikrat* veröffentlichte am 20. März 2023 seine 400-seitige Stellungnah-
me „Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz“¹⁹², die
sich in größerem Rahmen damit befasst, was die „Delegation von Tätigkeiten an Maschi-

¹⁸⁶ „Während US-amerikanische und chinesische Anbieter kleinere europäische Sprachen gefahrlos ver-
nachlässigen können, dürfen europäische Anbieter das nicht.“ (ebd. S. 87)

¹⁸⁷ ebd. S. 89. Genannt wird eine Reihe von EU-geförderten Projekten, die solche europäischen Daten-
korpora kuratieren und zur Verfügung stellen, z.B. *Paracrawl*, *European Language Resource Coordination*
(ELRC) oder *European Language Grid* (ELG). (Vgl. ebd. S. 91/92)

¹⁸⁸ ebd. S. 94

¹⁸⁹ ebd. S. 94

¹⁹⁰ ebd. S. 126 ff.

¹⁹¹ Die Studie orientiert sich hierbei am *Climate Neutral Data Center Pact*, einer Art Selbstverpflichtung
zahlreicher Verbände und Unternehmen auf europäischer Ebene, die sich auf Zielwerte bei der Energie-
Effizienz, der Nutzung grünen Stroms und der effizienten Wassernutzung geeinigt haben, sowie Ab-
wärmennutzung und Kreislaufwirtschaft anstreben. (ebd. S. 163)

¹⁹² Deutscher Ethikrat 2023

nen – bis hin zu einer möglichen Ersetzung“ für „verantwortliches Handeln und menschliche Autorschaft“ bedeutet ¹⁹³. Aus den bereits im Forschungsbericht des TAB (Kap. 2.2.2) genannten Anwendungsbereichen bzw. soziotechnischen Umgebungen greift der *Deutsche Ethikrat* vier größere heraus (Medizin, Bildung, öffentliche Kommunikation und Meinungsbildung, sowie öffentliche Verwaltung ¹⁹⁴), für die er konkrete normative Empfehlungen formuliert. Ergänzend werden zehn Querschnittsthemen ¹⁹⁵ für alle Bereiche mit übergreifenden Empfehlungen versehen. ¹⁹⁶ Unter diesen Querschnittsthemen finden sich Problempunkte wieder, die teils auch in den bisher referierten Studien (Kap. 2.1.1 bis 2.2.3) thematisiert sind, z.B. Bias und Diskriminierung, Fragen des Datenschutzes und des Schutzes intellektuellen Eigentums, mangelnde Transparenz und Nachvollziehbarkeit der Arbeitsweise algorithmischer Systeme, sowie mangelnde Zuverlässigkeit der Wissenserzeugung durch KI. – Weitere Themen sind die negativen Folgen der Abhängigkeit des Menschen von digitalen Technologien, so z.B. der Verlust der Datensouveränität, der Verlust menschlicher Kompetenzen, die Gefahr des Missbrauchs durch Zweitverwertung („dual use“) neu eingeführter KI-Technologien, die gesellschaftliche Vulnerabilität durch ihr Angewiesensein auf wenige Systeme oder Akteure, womit immer auch „Aspekte sozialer Gerechtigkeit und Macht“ ¹⁹⁷ angesprochen sind. ¹⁹⁸ Selbst wenn schleichende Abhängigkeiten (oder „systemic delayed harms“, wie die *OECD* sie nennt, siehe Kap. 2.1.4) ausgeräumt werden könnten, und es sich immer um eine bewusste, souveräne und wohlinformierte Entscheidung des Menschen handelte, eine Tätigkeit an ein KI-System zu übertragen, so bleibt die Frage, wie sich diese Entscheidung weiter auswirkt: „Die zentrale ethische Frage ist, ob und wie diese Übertragung die Möglichkeiten *anderer* Menschen beeinflusst, vor allem von jenen, *über* die [mittels eines KI-Systems, *Anm. der Verf.*] entschieden wird.“ ¹⁹⁹

Der philosophische Ansatz besonders im ersten Teil der Stellungnahme ²⁰⁰ ermöglicht es, auch die neue Technologie generativer KI unter dem übergeordneten Gesichtspunkt des Mensch-Maschine-Verhältnisses ethisch zu bewerten. ²⁰¹ Da eine Ethik generativer

¹⁹³ ebd. S. 79

¹⁹⁴ Vgl. ebd. Kap. 5 bis 8

¹⁹⁵ Vgl. ebd. Kap. 10, S. 345 ff.

¹⁹⁶ Der Deutsche Ethikrat nennt dies „vertikales, gestaltendes Denken“, welches durch bereichsübergreifendes „horizontales“ Denken ergänzt wird (vgl. ebd. S. 65).

¹⁹⁷ ebd. S. 32

¹⁹⁸ Es handelt sich um Aspekte, die auch das OECD-Arbeitspapier (siehe Kap. 3.1.4) unter dem Begriff der „kollektiven Entmündigung“ („collective disempowerment“) zusammenfasst.

¹⁹⁹ ebd. S. 31 (Hervorhebungen im Original)

²⁰⁰ Vgl. ebd. Kap. 2 bis 4

²⁰¹ Hierzu heißt es in der Einleitung zur Stellungnahme: „Auch die Debatten um den im November 2022 vorgestellten Chatbot ChatGPT und andere Anwendungen sogenannter *generativer KI*, welche automati-

KI noch nicht etabliert ist, erscheint es legitim, Anleihen aus der Maschinenethik zu machen. In der Benennung des Verhältnisses kommt bereits zum Ausdruck, dass es sich bei Mensch und Maschine um klar voneinander abzugrenzende Entitäten handelt:

„Der Deutsche Ethikrat geht von einem normativ grundlegenden Unterschied zwischen Mensch und Maschine aus. Softwaresysteme verfügen weder über theoretische noch über praktische Vernunft. Sie handeln oder entscheiden nicht selbst und können keine Verantwortung übernehmen. Sie sind kein personales Gegenüber, auch dann nicht, wenn sie Anteilnahme, Kooperationsbereitschaft oder Einsichtsfähigkeit simulieren.“²⁰²

Diese grundlegende Abgrenzung macht die Verwendung des Begriffs der „Intelligenz“ in Bezug auf Maschinen anfechtbar²⁰³, spätestens aber an dem basalen Begriff der „Vernunft“²⁰⁴ scheitert die Idee, Maschinen mit Menschen gleichsetzen zu wollen: „Zumindest die bislang verfügbaren KI-Systeme verfügen nicht über die dafür relevanten Fähigkeiten des Sinnverstehens, der Intentionalität und der Referenz auf eine außersprachliche Wirklichkeit.“²⁰⁵ In diesem Zusammenhang wird die Leibgebundenheit menschlicher Vernunft²⁰⁶ angeführt. Durch die „Eigenleiberfahrung“²⁰⁷, welche von der Kindheit an „körperlich-biologisches“ und „kulturell-soziologisches Erleben“²⁰⁸ umfasst, bilde sich eine „Perspektivität des Wissens“²⁰⁹ aus, die entscheidend für menschliche Deutungsprozesse und damit, hier verkürzt gesagt, für die „hermeneutische Dimension von Wissen“²¹⁰ sei.²¹¹ Diese hermeneutische Dimension entziehe sich weitgehend einer Formalisierung und sei daher maschinell nur beschränkt simulierbar.²¹² In Bezug auf die neuen

siert neue Inhalte in einer Qualität produzieren, bei der oftmals nicht mehr erkennbar ist, dass diese rein maschinell erstellt wurden, zeigen, dass eine grundlegende Auseinandersetzung mit den Wechselwirkungen zwischen Mensch und Maschine erforderlich ist.“ (ebd. S. 78)

²⁰² ebd. S. 63

²⁰³ „Man sollte [...] die Verwendung des Ausdrucks ‚Intelligenz‘ in der Wortverbindung ‚Künstliche Intelligenz‘ eher als eine Metapher einordnen [...]“ (ebd. S. 20)

²⁰⁴ „Der Begriff der Vernunft wurde bereits lange vor der Einführung des Begriffs der Intelligenz verwendet, um die spezifische menschliche Fähigkeit zu kennzeichnen, sich in der Welt zu orientieren, selbstverantwortlich zu handeln und so der eigenen Lebenspraxis eine kohärente Struktur zu geben. Intelligenz ist für Vernunft eine wichtige Voraussetzung, aber keine hinreichende Bedingung.“ (ebd. S. 20 f.)

²⁰⁵ ebd. S. 21

²⁰⁶ „Menschliche Vernunft ist leibliche Vernunft.“ (ebd. S. 156)

²⁰⁷ ebd. S. 159

²⁰⁸ ebd.

²⁰⁹ ebd.

²¹⁰ ebd.

²¹¹ Hierzu weiterführend Klaus Wieglerling. „Die grundlegende Individualisierungs- und Orientierungsfunktion des Leibes ist durch nichts ersetzbar. Und an dieser Stelle ist es wichtig, dass wir von einem Leib und nicht von einem Körper sprechen, denn der Leib ist unser Orientierungsorgan, das uns in Raum und Zeit eine Stelle zuweist und somit unsere Existenz anzeigt. Ohne sinnliche Vermögen gibt es weder einen Zugang zur Welt, noch eine Möglichkeit zu einem symbolischen Universum zu gelangen.“ (Wieglerling 2023a, S. 104)

²¹² „Diese hermeneutische Dimension von Wissen ist aber nur begrenzt formalisierbar oder simulierbar und bezieht sich auf den Sinn und die Bedeutung menschlichen Erkennens und Handelns. Die Aneignung

Foundation-Modelle, die auf den Input *formalisierter* Wissensbestände (siehe Kap. 2.2.3 *LEAM*-Studie) angewiesen sind, würde dies bedeuten, dass sie zwar über gespeichertes Sachwissen, aber weder über „leiblich verfasstes Hintergrundwissen“²¹³ noch über „lebensweltliches Orientierungswissen“²¹⁴ verfügen und damit nicht vernunftbegabt sein, geschweige denn „vernünftig handeln“ können.

Der Begriff menschlichen Handelns wird in der Stellungnahme zunächst philosophisch eng gefasst: Nur menschliches Tun, das „zweckgerichtet, beabsichtigt und kontrolliert“²¹⁵, also intentional ist, sei als „Handlung“ zu verstehen. Da Maschinen keine Absichten verfolgen, d.h. das Kriterium der Intentionalität nicht erfüllen, sei es nicht möglich, ihnen Handlungen zuzuschreiben.²¹⁶ Dieses Kriterium sei auch „entscheidend für die Möglichkeit der Zuschreibung von Verantwortung im Kontext von Mensch-Maschine-Interaktionen [...]“²¹⁷ Maschinen können also auch keine Verantwortungssubjekte werden. Der *Deutsche Ethikrat* beschreibt allerdings das Phänomen, dass sich die „Subjekt-Objekt-Verhältnisse zwischen Mensch und Technik“²¹⁸ teilweise umkehren:

„Im traditionellen Bild gestalten und nutzen menschliche Subjekte technische Objekte. [...] In vernetzten Systemen haben Menschen teils die Subjekt-, teils aber auch die Objektrolle inne. Wenn [...] Entscheidungen über Menschen an Softwaresysteme delegiert werden, beispielsweise hinsichtlich der Gewährung von Krediten und Sozialleistungen, werden Menschen zu Objekten der ‚Entscheidungen‘ dieser Systeme, die hier auftreten, als ob sie Subjekte seien.“²¹⁹

In Anbetracht dieser neu „verteilten“ Handlungsträgerschaft zwischen Mensch und Maschine²²⁰, die dem engen philosophischen Handlungsbegriff widerspricht, wird es dann entscheidend, „Intentionalität“ und „Verantwortung“ von ihr abzuheben: Technologische Interventionen seien zwar – wie menschliche – durchaus von Kausalität und Kontingenz gekennzeichnet, nicht jedoch von Intentionalität.²²¹ Daher müsse Verantwortlichkeit auf

menschlicher Erfahrung ist immer mit Deutungsprozessen verbunden und setzt immer ein Beteiligtsein, ein Engagement voraus. Die Art und Weise, wie wir wissen (*knowing how*), ist eine eigene Kompetenz, die sich nicht durch bloßes Sachwissen (*knowing that*) abbilden lässt. [...] In der leiblichen Verfasstheit gründet daher auch die Nichtsimulierbarkeit des Denkens.“ (Deutscher Ethikrat 2023, S. 159 f.)

²¹³ ebd. S. 160

²¹⁴ ebd. S. 161 f.

²¹⁵ ebd. S. 135

²¹⁶ „[...] dann ist die Zuschreibung von Handlungen in Bezug auf Maschinen in diesem engen Sinne nicht möglich.“ (ebd.)

²¹⁷ ebd. S. 139

²¹⁸ ebd. S. 170

²¹⁹ ebd. S. 170 f.

²²⁰ nach Werner Rammert, ebd. S. 174

²²¹ Hier schließt sich der Deutsche Ethikrat der Theorie des Techniksoziologen Werner Rammert an, der die Wechselwirkung von Mensch und Technik als dreistufig beschreibt: Auf die ersten beiden Stufen *Kausalität* und *Kontingenz* folgt die dritte Stufe *Intentionalität* als „höchster Stufe des ‚Agierens‘, die dem Handeln von Menschen vorbehalten“ sei. (ebd. S. 175) Algorithmen können Entscheidungen treffen, die

anderer Ebene konkret zugewiesen werden, was der *Deutsche Ethikrat* u.a. als Aufgabe des EU AI Act (siehe Kap. 2.1.1) ansieht.²²²

Der *Deutsche Ethikrat* stellt sich in einer allgemeineren Betrachtung der Mensch-Technik-Relationen auch die Frage, was die treibende Kraft hinter neuen technologischen Entwicklungen wie auch generativer KI sein könne, und verortet sich dabei zwischen der Position eines „Technikdeterminismus“ auf der einen Seite und der eines „Sozialkonstruktivismus“ auf der anderen: Während in der technikdeterministischen Sichtweise der technische Wandel auf eine (letztlich ökonomiegetriebene) Eigendynamik zurückgeführt wird, welche blind auf die Gesellschaft einwirkt und diese zur Anpassung zwingt, geht der Sozialkonstruktivismus davon aus, dass die Gesellschaft der Akteur ist, der Technik entwickelt und gestaltet, wobei „gesellschaftliche Leitbilder“²²³ in der Lage sind, bestimmte Technik-Ausprägungen hervorzurufen. In einer Zusammenführung beider Positionen möchte der *Deutsche Ethikrat* die Mensch-Technik-Relationen als „Ko-Evolution von Technik und Gesellschaft“²²⁴ verstanden wissen. Der Begriff der Ko-Evolution berücksichtigt sowohl den Akteur-Status des Menschen, der die Technik hervorbringt, in seiner Sozialität und Kulturalität, als auch die „vielfältige Rückwirkung einmal entwickelter und verfügbarer Technik auf Mensch und Gesellschaft“²²⁵. Mit dem Begriff der Ko-Evolution scheint der *Deutsche Ethikrat* die Zukunftsperspektive technologischer Entwicklungen, und so auch generativer KI, konsistenter zu beschreiben als die „Moratoriumsforderung“ (Kap. 2.1.2), in welcher zunächst ein enthemmter Technikdeterminismus als problematischer Ist-Zustand „diagnostiziert“ wird, als „Therapie“ ein sechsmonatiger Shift zum „Sozialkonstruktivismus“ hin vorgeschlagen wird, um anschließend – wenn die Gesellschaft für die disruptiven Technologien hinreichend konditioniert sein wird – dem Technikdeterminismus umso freieren Lauf zu lassen.

reale Veränderungen bewirken („Kausalität“), und haben dabei die Wahl unter Entscheidungsalternativen („Kontingenz“), doch tun sie dies niemals intentional und können daher für ihr Handeln auch nicht verantwortlich gemacht werden.

²²² „Verantwortung muss daher anders geregelt werden, zum Beispiel durch Zuschreibung [...] an relevante Institutionen bzw. Organisationen. [...] Der von der Europäischen Union geplante Artificial Intelligence Act (AI Act) wird [...] die Verantwortlichkeit konkret zuweisen müssen, um wirksam zu regulieren.“ (ebd. S. 175 f.)

²²³ ebd. S. 166

²²⁴ „Soziale Kontexte und normative Kriterien auf der einen und Technologien auf der anderen Seite entwickeln sich weiter in gegenseitiger Wechselwirkung. Die Verfügbarkeit von Technik beeinflusst Handlungsmöglichkeiten [...], aber auch die Bedingungen [...] menschlicher Weltwahrnehmung, wodurch sich Lebensstile und Einstellungen verändern können. Umgekehrt entstehen, wie dies die Technikgeneseforschung in vielen empirischen Studien belegt hat, neue Techniken vor dem Hintergrund von sozialen Befindlichkeiten, normativen Kriterien und Zukunftsentwürfen.“ (ebd. S. 168)

²²⁵ ebd. S. 169

Insgesamt nimmt der *Deutsche Ethikrat* in der Bewertung der neuen KI-Technologien weder eine unbesehen affirmative noch eine technikpessimistische Position ein; die Ambivalenz jeglicher technologischen Entwicklung in Bezug auf menschliche Handlungsautonomie, die sie sowohl fördern ²²⁶ als auch behindern kann, ist für ihn nicht auflösbar. Er möchte aber die Einschränkungen menschlicher Handlungsautonomie ins Bewusstsein rufen, da diese „oft nicht gut erkennbar“ ²²⁷ seien. Als Problemdimensionen werden genannt: erstens die Technologieabhängigkeit, in der technologische „Sachzwänge“ menschliche Handlungsautonomie beschneiden, und die überdies moderne Gesellschaften vulnerabel mache; zweitens der Anpassungsdruck, dem Individuen und Gesellschaft durch Technologie ausgesetzt seien, und der in eine schleichende „technische Normierung“ menschlichen Denkens und Handelns münden könne; und drittens das Abwerten und Verschließen von Optionen menschlichen Handelns zugunsten neuer Optionen ²²⁸.

Aus konsequent humanistischer Perspektive empfiehlt der *Deutsche Ethikrat*, auf folgende vier Punkte zu achten: Erstens auf *Transparenz* gegenüber denjenigen, die von Entscheidungen der KI-Systeme betroffen sind, soziotechnisch also in die „Objekt“-Rolle geraten sind; zweitens auf *Verhinderung von Verantwortungsdiffusion* durch eine geregelte Verantwortungszuschreibung für das „Handeln“ von KI-Systemen; drittens auf eine Art erweiterter *Qualitätskontrolle* („Sicherstellung der Nachvollziehbarkeit in Bezug auf das zweckhafte Funktionieren der KI-Systeme“ ²²⁹); und viertens auf die Beobachtung und ggf. *Korrektur möglicher unbeabsichtigter Folgen* des „Handelns“ von KI-Systemen, „wie beispielsweise schleichend einkehrende Abhängigkeiten von den KI-Systemen oder allmähliche Aushöhlung menschlicher Autorschaft“ ²³⁰.

²²⁶ Die Sichtweise des *Deutschen Ethikrates* ist hier keineswegs einseitig: „Technik ist zentraler Teil menschlichen Lebens und gesellschaftlicher Vollzüge zumindest in den industrialisierten Regionen der Welt geworden und hat in vielen Fällen eindeutig positive Folgen in dem Sinne gezeitigt, dass die Möglichkeiten menschlicher Autorschaft erweitert wurden – jedenfalls für den Teil der Erdbevölkerung, der, vor allem im Globalen Norden, Zugang zu ihren Vorteilen hat.“ (Deutscher Ethikrat 2023, S. 177 f.)

²²⁷ ebd. S. 178

²²⁸ „Dies ist einerseits der normale Gang von Transformation und Wandel. Andererseits aber stürzen Innovationen vorhandene Anerkennungs- und Wertstrukturen durch disruptive Effekte um und ziehen Gewinner wie auch Verlierer nach sich. Von den neuen Optionen profitieren häufig andere Personen und Gruppen als die, die dann im Verschließen der traditionellen Optionen zu Verlierern des Wandels werden. [...] So werden Infrastruktursysteme häufig faktisch machtförmig.“ (ebd. S. 180). Als Beispiel nennt der *Deutsche Ethikrat* die Nutzung von Smartphones, die vorausgesetzt werde, um überhaupt „an bestimmten Lebensvollzügen teilnehmen zu können. Diese Form der Verschließung von Optionen kann verschiedene Bevölkerungsgruppen unterschiedlich treffen und Gerechtigkeitsprobleme mit sich bringen, wie dies zum Beispiel unter dem Stichwort digitale Spaltung (*digital divide*) diskutiert wird.“ (ebd.)

²²⁹ ebd. S. 186

²³⁰ ebd.

2.2.5 Stellungnahme des *Deutschen Kulturrates* – Sorge um die Kultur

Der *Deutsche Kulturrat* bezieht sich in seinem Positionspapier „Künstliche Intelligenz und Urheberrecht“²³¹ vom 22. Juni 2023 auf die oben erörterte Stellungnahme des *Deutschen Ethikrates* und greift dessen ethisches Leitziel einer „Stärkung menschlicher Autorschaft“ auf.²³² Während der *Deutsche Ethikrat* den Begriff „menschlicher Autorschaft“ in einem umfassenden Sinn als Handlungsurheberschaft des Menschen versteht, bezieht ihn der *Deutsche Kulturrat* auf den Menschen in seiner Rolle als Kulturproduzent bzw. Urheber künstlerischer Werke. Entsprechend geht es um offene Fragen im Urheberrecht, die mit dem Aufkommen generativer KI entstanden sind. In dem vierseitigen Positionspapier wendet sich der *Deutsche Kulturrat* mit entsprechenden Aufforderungen auf nationaler Ebene an die Deutsche Bundesregierung, auf europäischer Ebene an EU-Kommission und -Rat²³³, und auf supranationaler Ebene an die *Weltorganisation für geistiges Eigentum* (WIPO). Im Wesentlichen unterstützt er die Forderungen der *Initiative Urheberrecht* (siehe Kap. 2.2.1) und möchte die Frage geklärt wissen, ob die im Rahmen der EU-Urheberrechtsreform neu formulierten Schrankenregelungen im Bereich des sog. *Text und Data Mining* (§ 44b UrhG)²³⁴ „die Nutzung von geschützten Werken für das Training von KI-Systemen abdecken“²³⁵.

Erlaubt sind *Text und Data Mining* nach § 44b Abs. 2 Satz 1 UrhG bei „Vervielfältigungen von rechtmäßig zugänglichen Werken“. Doch ist die Provenienz vieler digitalisierter Werke unsicher: „Es erscheint völlig unklar, wie im Rahmen des sog. Webscraping zur Gewinnung von Trainingsdaten zwischen rechtmäßig zugänglichen und illegal vorhandenen Werken im Internet unterschieden werden soll.“²³⁶ Die in § 44b Abs. 3 UrhG formulierte Möglichkeit des Nutzungsvorbehalts sei außerdem für den Rechtsinhaber nicht

²³¹ Deutscher Kulturrat 2023

²³² „Es geht insgesamt um Fragen von großer kultur- und gesellschaftspolitischer Bedeutung. Es ist deshalb sehr zu begrüßen, dass sich beispielsweise der Deutsche Ethikrat des Themas angenommen [...] hat. [Er, *Anm. d. Verf.*] hat dabei deutlich gemacht, dass Ziel und Richtschnur ethischer Bewertung immer die Stärkung menschlicher Autorschaft sein muss. Der Deutsche Kulturrat stimmt dieser Einschätzung uneingeschränkt zu.“ (ebd. S. 1)

²³³ Zu diesem Zeitpunkt befand sich der EU AI Act am Beginn der Trilogverhandlungen.

²³⁴ Bei § 44b UrhG handelt es sich um eine gesetzliche „Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes“, die am 7. Juni 2021 in Kraft trat (siehe www.gesetze-im-internet.de). Der Gesetzgeber versteht unter Text und Data Mining die „automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen.“ (§ 44b Abs. 1 UrhG)

²³⁵ Deutscher Kulturrat 2023, S. 2

²³⁶ ebd. S. 2

praktikabel, da ein „Opt-out“²³⁷ nicht standardisiert wurde und nicht kontrolliert werden könne, „ob ein Werk, für das ein Opt-out erklärt wurde, für Trainingszwecke genutzt wurde“²³⁸.

Der Output generativer KI sei ebenfalls rechtlich zu überdenken, wenngleich unstrittig ist, dass Erzeugnisse generativer KI gemäß § 2 UrhG im Gegensatz zu menschlichen Schaffensleistungen keine Werke sind und daher keinen urheberrechtlichen Schutz beanspruchen können.²³⁹ Dient ein KI-Erzeugnis aber „als Ausgangspunkt für anschließende kreative Leistungen eines Menschen“²⁴⁰, wurde es also nicht nur lektoriert, formatiert o.ä., sondern so weitgehend bearbeitet, dass ein neues Original entstand, kommt Urheberrechtsschutz wieder in Frage. Um beurteilen zu können, „ob ein Erzeugnis wesentlich auf dem Einsatz von KI beruht“²⁴¹ oder ob die Kreativität des Menschen überwog und daher von einem Werk gesprochen werden kann, werden vom *Deutschen Kulturrat* (wie vom *PAI Framework*, Kap. 2.1.3), „Kennzeichnungspflichten des Verwenders von KI-Erzeugnissen“²⁴² gefordert. Eine Kennzeichnung wäre allerdings nur ein erster Schritt. Wo genau dann die Grenze zwischen Werk und Nicht-Werk gezogen werden soll, ist nicht geklärt. – Ein weiteres Problem entsteht nach Ansicht des *Deutschen Kulturrates*, wenn sich Output und Input generativer KI zu stark ähneln, also *Chat GPT* beispielsweise ganze Textpassagen eines urheberrechtlich geschützten Werkes wortwörtlich „abschreibt“, ohne die Quelle zu nennen.²⁴³

Insgesamt sind die Erwartungen des *Deutschen Kulturrates* gerade auch an den *EU AI Act* hoch. Aus dem gesamten Positionspapier spricht die Besorgnis, dass menschlich erzeugte Kultur von generativer KI in Frage gestellt, verdrängt oder sogar zerstört werden könnte. Wie groß diese Sorge ist, erhellt sich auch aus anderen Veröffentlichungen des *Deutschen Kulturrates* bzw. seiner Mitglieder. Am 24. November 2023, wenige Tage bevor es in den Trilogverhandlungen zum *EU AI Act* zu einer Einigung für eine EU-weite

²³⁷ „(3) Nutzungen [...] sind nur zulässig, wenn der Rechtsinhaber sich diese nicht vorbehalten hat. Ein Nutzungsvorbehalt bei online zugänglichen Werken ist nur dann wirksam, wenn er in maschinenlesbarer Form erfolgt“ (§ 44b Abs. 3 UrhG, www.gesetze-im-internet.de/)

²³⁸ Deutscher Kulturrat 2023, S. 3

²³⁹ „Werke im Sinne dieses Gesetzes sind nur persönliche geistige Schöpfungen.“ (§ 2 Abs. 2 UrhG, www.gesetze-im-internet.de)

²⁴⁰ Deutscher Kulturrat 2023, S. 3

²⁴¹ ebd. S. 3

²⁴² ebd. S. 4

²⁴³ „So ist [...] keineswegs ausgeschlossen, dass ein KI-Produkt ein bestehendes Urheberrecht verletzt, weil kein ausreichender Abstand zu einem benutzten Werk eingehalten wird. Hier muss dringend geklärt werden, inwieweit der Hersteller der KI und der Verwender der KI für die Rechtsverletzung und den eingetretenen Schaden haften.“ (ebd. S. 4)

Regelung kam, warnte der Geschäftsführer des *Deutschen Kulturrates*, Olaf Zimmermann, in einer Pressemitteilung:

„Die Künstliche Intelligenz (KI) hat, erlauben Sie mir diese Zuspitzung, wie eine Atombombe, das Potential die Welt zu zerstören, wenn wir nicht aufpassen. Die Gefahr ist real, dass wir die Fähigkeit, die Wahrheit von der Lüge und Gut von Böse zu unterscheiden, verlieren. [...] Und sind wir ehrlich, weder die Bundesregierung noch die Europäische Union haben zur Zeit eine wirkliche Möglichkeit diese Unternehmen [*OpenAI* und *Microsoft*, *Anm. der Verf.*], die außerhalb ihres Verantwortungsbereiches liegen, an die Kandare zu nehmen. [...] KI ist deshalb ein Problem, weil die dahinterliegenden Systeme eine Blackbox sind, vollständig intransparent. Manipulationen durch Unternehmen sind Tür und Tor geöffnet. KI muss regulierbar werden.“²⁴⁴

Der drastische Vergleich künstlicher Intelligenz mit einer Atombombe ist aus der Perspektive eines „Anwalts“ der Kulturschaffenden zu verstehen, der sich vom Rand der Diskussion her Gehör verschaffen muss. Kulturschaffende, die es als beunruhigend und ehrverletzend empfinden, wenn generative KI ihre Leistungen „ausschlachtet“, ihre Kernkompetenzen übernimmt und sie ersetzt, zeigen eine besondere Sensibilität dafür, dass hier Kultur als grundlegendes menschliches Merkmal (und humanisierender Faktor) in Gänze angegriffen wird. Dieses Gefühl bringt sie gegenüber der als unmenschlich, undurchschaubar und gewissenlos wahrgenommenen KI und ihren Produzenten in Stellung. Ihre Wahrnehmung wird durch den Umstand verstärkt, dass die Kluft zwischen deutschem bzw. europäischem Kulturschaffen und den neuesten Produkten der US-amerikanischen Techbranche, was Geschichte und Geist ihrer Entstehung betrifft, größer nicht sein könnte. Olaf Zimmermann benennt US-amerikanische KI-Unternehmen als Aggressoren, die sich die humane Welt, so wie wir sie kennen, zur Beute machen, wenn wir sie nicht stoppen – und hofft, dass sich die Regelungen des *EU AI Act* als das geeignete Instrumentarium erweisen werden, die Dystopie aufzuhalten.

2.3 Stellungnahmen einzelner Autoren

Im Folgenden kommen exemplarisch meinungsstarke Buchautoren zu Wort, die generative KI aus unterschiedlichen Perspektiven beleuchten. Während der Technikphilosoph Klaus Kornwachs großflächig disruptive Effekte künstlicher Intelligenz auf menschliche Arbeit analysiert und aus dieser Perspektive ethische Positionen entwickelt (Kap. 2.3.1), diskutiert die Wissenschaftsforscherin Helga Nowotny die Gefahren einer deterministischen Weltansicht und formuliert ihre Position aus dem Gedanken eines „digitalen Humanismus“ heraus (Kap. 2.3.2). Einen heterogenen Ansatz verfolgt ein Autorenteam, das

²⁴⁴ Zimmermann 2023

aus dem deutschen Unternehmer Jobst Landgrebe und dem englischen Mathematiker und Ontologen Barry Smith besteht. Sie versuchen den Nachweis zu führen, dass es allein schon aus mathematischen Gründen nicht möglich sei, einen Algorithmus in ethischem Verhalten zu trainieren (Kap. 2.3.3). Auswahlkriterium für die genannten Autoren war das aktuelle Erscheinungsdatum ihrer Publikationen (2023), das eine explizite inhaltliche Berücksichtigung generativer KI nahelegt.

2.3.1 Klaus Kornwachs – menschliche Arbeit im KI-Zeitalter

Mit seiner Analyse „KI und die Disruption der Arbeit. Tätig jenseits von Job und Routine“²⁴⁵ schließt Klaus Kornwachs Forschungslücken im Hinblick auf aktuelle Entwicklungen in der Arbeitswelt. Dabei erfährt der viel strapazierte Begriff der „Disruption“ eine Einordnung. Kornwachs identifiziert ihn als Marketing-Begriff und macht darauf aufmerksam, dass die Segregation von Gewinnern und Verlierern auf dem Arbeitsmarkt zumindest in marktliberalen Wirtschaftsformen eine lange Geschichte hat. Zudem seien aktuell, wenn von Disruption in der Arbeitswelt gesprochen werden soll, die Folgen globaler Krisen wie z.B. der Corona-Pandemie und des russischen Angriffskrieges gegen die Ukraine weitaus dynamischere Faktoren als der vergleichsweise kontinuierliche Wandel durch Digitalisierung und KI-Technologien. Kornwachs sieht generell die Veränderungen der Arbeitswelt durch den Einsatz künstlicher Intelligenz nicht als Bedrohung, der man hilflos ausgesetzt wäre²⁴⁶, und differenziert: „Nicht die Künstliche Intelligenz selbst, sondern einige der Geschäftsmodelle, die KI erst möglich machen, sind zu fürchten.“²⁴⁷

Auf die Historie des Arbeitsbegriffs, deren Darstellung einen Großteil des Buches einnimmt, kann hier nicht eingegangen werden, doch seien die Entwicklungsschritte genannt, die von der „Unterstützung, Erleichterung bis hin zur Ersetzung der menschlichen Arbeitsfähigkeit“²⁴⁸ führen. Kornwachs fasst sie unter den Stichworten Mechanisierung, Maschinisierung, Automatisierung und Informatisierung/Digitalisierung zusammen und sieht sie als zunehmende Abstrahierung der Arbeitsinhalte, die künftig mit „Biologisie-

²⁴⁵ Kornwachs 2023

²⁴⁶ Hierzu Kornwachs im Vorwort: „Müssen wir Angst vor diesen Veränderungen haben oder dürfen wir sie begrüßen? Sind wir wirklich hilflos einer unaufhaltsamen Macht, die uns als Fortschritt verkauft wird, ausgeliefert? Meine Antwort ist schon an dieser Stelle ein klares Nein, schließlich sind wir alle, die Experten wie Konsumenten, auch diejenigen, die diese Entwicklung entfesselt haben und deshalb auch als Käufer solcher Technik für ihre Gestaltung mit verantwortlich.“ (ebd. S. XVI)

²⁴⁷ ebd. S. XVI

²⁴⁸ ebd. S. 146

nung, Personalisierung und Kognitivierung der Technik“²⁴⁹ noch weitere Stufen erreichen werde. Als Folgen dieser Abstraktion für den arbeitenden Menschen sieht Kornwachs eine „Leistungsverdichtung“, eine „Entsinnlichung der arbeitenden Tätigkeit“ sowie eine „Trennung von Arbeitsort und Arbeitsgegenstand“²⁵⁰, wobei heute das „Netz“ zum „neuen Arbeitsort“ werde²⁵¹. Möglich gemacht werde die Abstraktion „durch Universalisierung des Werkzeugs und Konvergenz unterschiedlicher Technologiestränge zu neuen Technologien“²⁵². Big Data spielt dabei eine entscheidende Rolle: „Gerade die Entwicklung der Technik der Datennutzung hat den Weg zur Automatisierung höherer Stufen bereitet und zeigt, dass der Siegeszug des Computers, besser gesagt der Algorithmen, beginnt, unumkehrbare Strukturen und auch Abhängigkeiten zu schaffen.“²⁵³

Wie in der Frage nach der Technologieabhängigkeit des Menschen stimmt Kornwachs auch in anderen zentralen ethischen Fragen mit den Positionen des *Deutschen Ethikrates* überein. Intelligente Roboter (IR) zu Verantwortungssubjekten zu machen, hält auch Kornwachs für eine falsche Zuschreibung²⁵⁴ und fordert ein Verbot von IR, die mit der Absicht konzipiert und erbaut werden, den eigenen Nutzer in diesem Sinne zu ersetzen.²⁵⁵ Bezogen auf *ChatGPT* merkt Kornwachs an, dass sich die Problematik der maschinellen Entscheidungsersetzung in die maschinelle „Ersetzung der Erzeugung von Information, die als Begründungswissen für die automatisierte oder ersetzte Entscheidung verwendet werden könnte“, verschiebe.²⁵⁶ Diese neue Problematik erscheint subtiler und ist doch nichts weiter als eine neue Stufe von Verantwortungsdiffusion, die es

²⁴⁹ ebd. S. 160

²⁵⁰ ebd. S. 160

²⁵¹ ebd. S. 160

²⁵² ebd. S. 161

²⁵³ ebd. S. 161

²⁵⁴ „Wenn man also danach fragt, ob solche artifiziellen ‚Subjekte‘ [d.h. IR, *Anm. der Verf.*] Rechte und Pflichten haben könnten, kann man dies klar verneinen, da es sich um falsche Zuschreibungen handelt. [...] Da intelligente Roboter keine Subjekte sind oder wir aus guten moralischen Gründen nicht wollen können, dass sie eines Tages Subjekte werden, sollte man sie auch nicht prospektiv als Subjekte von Verantwortung ansehen.“ (ebd. S. 278 f.) Und an anderer Stelle noch deutlicher: „Hier wird die Position entwickelt, dass die menschliche Arbeit [...] nur dann erhalten werden kann, wenn wir den Verantwortungsbegriff beim menschlichen Individuum belassen und die Zuschreibung der Verantwortung nicht in eine unklar definierte ‚intelligente‘ Robotik hineindiffundieren lassen. Es ist Unfug, eine Ethik in die Maschine ‚implementieren‘ zu wollen.“ (ebd. S. 292)

²⁵⁵ „Unabhängig von der Frage, ob wir eines Tages solche Roboter werden bauen und nutzen können, plädiere ich für ein Verbot der Herstellung solcher Systeme. Die Begründung läuft analog zum Verbot von Chimären, das die ‚Herstellung‘ von Zwitterwesen zwischen Mensch und Tier untersagt. Zum einen wäre der Unterschied zwischen Mensch und Maschine verwischt, was der Begründung der Menschenwürde die Grundlage entzöge, und zum anderen könnten wir für die ‚Handlungen‘ solcher Systeme sowohl als Erbauer als auch als ‚Halter‘ keine Verantwortung mehr übernehmen.“ (ebd. S. 281)

²⁵⁶ ebd. S. 348

Individuen oder gesellschaftlichen Gruppen, die von schädlichen Entscheidungen betroffen sind, erschwert, dagegen vorzugehen.

Eng verknüpft mit der Verantwortungsdiffusion ist die Urheber-Problematik; nur ein menschlicher Urheber kann für den Wissens- und Wahrheitsgehalt einer Information verantwortlich gemacht werden. Doch werden Funktion und Bedeutung menschlicher Urheberschaft auf vielen Ebenen versuchsweise ausgehebelt. Kornwachs berichtet von immer neuen Geschäftsideen („Gauereien“) im Netz, welche sowohl „Lücken in der Technologie“ als auch eine „hinterherhinkende Gesetzgebung“²⁵⁷ ausnutzen. Der Schutz von Arbeitsergebnissen im Netz, der vor allem auf klaren Auftragsverhältnissen beruht („Wer ist Unternehmer und wer Kunde? Wer liefert und wer zahlt? Was ist ein Arbeitsergebnis und wem verkaufe ich meine Kompetenz, meine Zeit sowie meine Ergebnisse?“²⁵⁸), erscheint aufgeweicht und von Forderungen nach totaler Netzoffenheit und Transparenz torpediert, bei welcher – analog der Situation in der Wissenschaft – alles allen gehören soll. Während der veröffentlichende Wissenschaftler aber von der Allgemeinheit bezahlt werde, bleibe für den Künstler oder Programmierer kein ökonomischer Anreiz mehr.²⁵⁹ Dem hält Kornwachs entgegen:

„Es wird entscheidend sein, wie man den Schutz des Arbeitsergebnisses vor unberechtigter Verwertung organisatorisch, technisch und rechtlich absichern kann. Der Begriff des Eigentums an Arbeitsergebnissen im Netz wie deren Besitz würde sich sonst rasch auflösen. Dies hätte einen nicht vorstellbaren Wertewandel zur Folge, den wir sicher nicht wollen.“²⁶⁰

Gegen diesen Wertewandel stemmt sich auch der *Deutsche Kulturrat*. Kornwachs ist jedoch optimistisch, dass Menschen durch „waches Steuern der Prozesse“²⁶¹ die weitere Ko-Evolution von Mensch und Technik aktiv gestalten können und nicht als beängstigend empfinden müssen. Er argumentiert dabei mit der „Imperfektion“ des Menschen, die seine Gestaltungskraft nicht in Frage stelle, sondern sich im Gegenteil mit der Einbeziehung von Überzeugungen, Wertvorstellungen und Gefühlen als eine Stärke herausstellen werde²⁶², die er gegenüber dem „perfekten“ Algorithmus ausspielen könne.

²⁵⁷ ebd. S. 344

²⁵⁸ ebd. S. 344

²⁵⁹ „Welcher ökonomische Anreiz bleibt dann noch? Die Anerkennung der Gemeinde für gemeinnütziges Dichten, Programmieren, Komponieren etc.“ (ebd. S. 344)

²⁶⁰ ebd. S. 344 f.

²⁶¹ ebd. S. 360

²⁶² Auch wenn der Mensch den Eindruck habe, „nur noch ein Rädchen in einer Maschinerie“ (ebd. S. 354) zu sein, gebe es Möglichkeiten, dem gegenzusteuern: „[...] wir können aufstehen, die Bedingungen bestreiten, sie bekämpfen, in welchen Solidaritätsformen auch immer. Dafür genügen Überzeugungen, Wertvorstellungen und Gefühle, da muss es noch nicht einmal ganz rational zugehen. Das ist die Stärke der menschlichen Schwäche [...]“ (ebd. S. 354)

2.3.2 Helga Nowotny – Ethik ist keine Checkliste!

Helga Nowotny beschreibt ihre eigene Recherche zu KI-Literatur als „Gang durch einen Irrgarten“²⁶³, durch welchen „reichlich Marketing wabert“²⁶⁴. Nicht zuletzt der „Eindruck, einen überbewerteten Marktplatz besichtigt zu haben“²⁶⁵, bewog sie zu ihrem eigenen Buch mit dem Titel „Die KI sei mit euch. Macht, Illusion und Kontrolle algorithmischer Vorhersage“.²⁶⁶ Ähnlich wie Klaus Kornwachs diskutiert auch Nowotny die Zukunft der Arbeit als ein vorrangiges Thema²⁶⁷, das Fragen bezüglich der Wertbestimmung des Menschen aufwirft:

„Was [...] geschieht, wenn der wirtschaftliche Wert des einzelnen Menschen gen Null geht? Die Zukunft der Arbeit hängt davon ab, inwieweit der Wert, der einem Menschen und seinem Wohl zugemessen wird, seinen wirtschaftlichen Wert übersteigt und wie dies gemessen werden wird.“²⁶⁸

Während es Kornwachs um die gesellschaftliche Stabilität zu tun ist und er gesellschaftliche Aufgaben, die sich aus dem zu erwartenden wirtschaftlichen Wertverlust des Menschen ergeben, als konkrete „Agenda“ formuliert²⁶⁹, bleibt Nowotny etwas unbestimmter. Unter dem Begriff der Ko-Evolution von Mensch und Maschine versteht sie einen Prozess der wechselseitigen Erziehung und „Domestizierung“: „Wir domestizieren Algorithmen, und sie domestizieren wiederum uns.“²⁷⁰ Sie rekurriert dabei auf das Konzept der „Selbstdomestizierung“ des Menschen als evolutionären Prozess, der sich an die Domestizierung von Tieren und Pflanzen für den menschlichen Gebrauch angeschlossen habe²⁷¹ und historisch gesehen zunehmend prosoziales Verhalten fördere: „Auf

²⁶³ Nowotny 2023, S. 52

²⁶⁴ ebd. S. 52

²⁶⁵ ebd. S. 54

²⁶⁶ Auch wenn das Buch im englischen Original („*In AI We Trust. Power, Illusion and Control of Predictive Algorithms*“) bereits im Jahr 2021 vor dem Release von *ChatGPT* veröffentlicht wurde, erscheinen die beschriebenen ethischen Aspekte im Zusammenhang mit generativer KI aktuell. Nowotny ergänzte die deutsche Erstausgabe (2023) zudem durch ein 24-seitiges Vorwort mit dem Titel „Die verwunschene Welt von GPT-4“, in dem sie die neuesten Entwicklungen miteinbezog.

²⁶⁷ „Die Zukunft der Arbeit bleibt ein heiß debattiertes Thema, da bei der Frage, wie Wohlstand angehäuft werden darf und wie er verteilt oder umverteilt werden sollte, Arbeit noch immer als der Dreh- und Angelpunkt unserer Gesellschaften gilt. Das Wesen der Arbeit ändert sich derzeit vermutlich schneller als andere von der Digitalisierung betroffene Bereiche.“ (ebd. S. 257)

²⁶⁸ ebd. S. 254

²⁶⁹ Hierzu Kornwachs: „Die klassischen Arbeitsformen werden nicht mehr zu retten sein – daher müssen wir uns um die gesellschaftlich stabilisierenden Bedingungen kümmern, die bisher durch diese herkömmlichen Arbeitsformen aufrechterhalten und garantiert wurden. Diese sind unter anderem das Recht auf Eigentumserwerb, Recht auf Ausbildung und Bildung, Ausbildung von Identität, Ermöglichung sozialer Teilhabe, Organisation von Solidarität sowie ein transparentes Management von Erwartungen und Anerkennungsregimes.“ (Kornwachs 2023, S. 6)

²⁷⁰ ebd. S. 268

²⁷¹ Hier bezieht sich Nowotny auf Forschungen des deutschen Anatomen und Anthropologen Johann Friedrich Blumenbach (1752-1840).

einer evolutionären Zeitschiene gemessen, hat zweifellos eine Zähmung menschlicher Aggression und eine bemerkenswerte Verschiebung hin zu mehr Kooperation stattgefunden.“²⁷² Heute gebe Technologie „unserem Verhalten ein Feedback zurück, und wir passen uns dementsprechend an. Und so schreitet die Selbstdomestizierung voran.“²⁷³

Doch bleibt die Frage, warum die Selbstdomestizierung gegenwärtig nicht allerorten zu funktionieren scheint, wie auch Nowotny feststellen muss: „Eine größere Prosozialität in manchen Bereichen steht im starken Kontrast zu grausamer Aggression und Brutalität in anderen.“²⁷⁴ Nowotny macht die chaotische, „fluide, mehrdeutige und unordentliche soziale Welt“²⁷⁵, welche Menschen (und mustererkennende Algorithmen gleichermaßen) mit unvorhersehbaren Situationen herausfordere, als Hinderungsgrund aus. So sei auch die Zukunft als offener Horizont zu begreifen, der nicht durch algorithmische Vorhersagen verschlossen werden könne und dürfe²⁷⁶.

Bei ethischen Prinzipien für KI sieht Nowotny nach Lektüre der von der *Eidgenössischen Technischen Hochschule* (ETH) Zürich untersuchten 84 Leitfäden²⁷⁷ kaum inhaltliche Schnittmengen und wenig Übereinstimmungen (oder keine Angaben) in der Frage, wie sie konkret umzusetzen seien. Festzuhalten sei lediglich, dass neben den Schlüsselargumenten „Gerechtigkeit, Fairness, Schadensprävention und datenschützerische Verantwortung“²⁷⁸ das „Konzept der Transparenz bei der Suche nach einer ethischen KI klar an erster Stelle“ stehe²⁷⁹. Dass KI-Anwendungen transparent zu sein haben, scheint somit der kleinste gemeinsame „Nenner“ zu sein, auf den sich KI-Branche, Forschung und Zivilgesellschaft bisher einigen konnten. Jedoch sind gerade bei generativer KI der

²⁷² ebd. S. 263

²⁷³ ebd. S. 265. Gegenwärtig seien es vor allem prädiktive Algorithmen, die im Zuge präventiver gesellschaftlicher Maßnahmen „ein immer größeres Stück der ihnen von uns zugewiesenen Handlungsmacht beanspruchen. Sie werden zu normativen Akteuren, ausgerichtet auf Verbesserung und menschliche Optimierung, und dadurch zu einem wesentlichen Mechanismus im Prozess der Selbstdomestizierung.“ (ebd. S. 267)

²⁷⁴ ebd. S. 263

²⁷⁵ ebd. S. 268

²⁷⁶ „Die Vorstellung von der Zukunft als offenem Horizont ist eine große Entdeckung, die es zu würdigen gilt. Glauben wir erst einmal, Algorithmen könnten die Zukunft vorhersagen, besonders, was unser eigenes Verhalten betrifft oder was uns widerfahren wird, droht uns der Verlust dieser Idee. [...] Verlassen wir uns zu sehr auf algorithmische Vorhersagen, droht uns der Rückfall in ein deterministisches Weltbild, in dem alles entschieden ist und wir uns durch unseren Glauben an algorithmische Vorhersagen selbst entmachtet haben.“ (ebd. S. 270)

²⁷⁷ „Eine Forschungsgruppe der ETH Zürich analysierte vierundachtzig Leitfäden zur Ethik der KI, verfasst von Unternehmen, Expertengruppen, Regierungen und internationalen Organisationen aus aller Welt. Etwa die Hälfte dieser Erklärungen stammte aus der Privatwirtschaft, die andere Hälfte aus öffentlichen Einrichtungen.“ (ebd. S. 211)

²⁷⁸ ebd.

²⁷⁹ ebd.

Input und die Vorgänge innerhalb der Blackbox intransparent: „Die im Inneren der KI ablaufenden Prozesse sind – jedenfalls vorläufig – kausal nicht erklärbar. Niemand weiß genau, woher welche Daten kommen und an welchen das jeweilige KI-Modell trainiert wird.“²⁸⁰ So fragt Nowotny weiter: „Wie kann sich unter diesen Vorzeichen Vertrauen einstellen [...]?“, und zeigt sich auffallend skeptisch, was die Wirksamkeit gesetzlicher Regelungen betrifft – wozu auch der EU *AI Act* zählt. Nowotny bezeichnet die bestehende Rechtsordnung als „zahnlos und überfordert“²⁸¹ und wünscht sich „völlig neue Institutionen und eine neue Art der Gesetzgebung“²⁸², um der technologischen Entwicklung Herr zu werden.

Auf technologischer Ebene seien KI-Experten zwar „zuversichtlich, dass die Arbeit an einer ethischen KI gut voranschreite“²⁸³ und im Sinne eines *alignment* (der Herstellung einer „Übereinstimmung der KI mit menschlichen Werten“)²⁸⁴ erfolge. An anderer Stelle berichtet Nowotny jedoch von einem gravierenden Mangel an „positiven Ansätzen, wie sich KI Ethik einbauen ließe“²⁸⁵. Gearbeitet wird offensichtlich vorrangig am „negativen“ Ansatz einer Vermeidung einzelner unethischer Auswirkungen, was zu einer Art „Checkliste“ führe, „mit der sich eine Firma oder Organisation durch Häkchensetzen von weiterer Verantwortung entbinden könnte“.²⁸⁶ Demgegenüber bedeutet KI-Ethik für Nowotny das allgegenwärtige Vorhandensein und die alle konkreten Kontexte durchdringende Kraft einer Weisheit, welche „den Stimmen der am stärksten Betroffenen“ lausche und versuche, „für sie einzutreten, und zwar entgegen der Starrheit von Regeln und gegen jene, die sie festlegen und durchsetzen“²⁸⁷, eine Position, die derjenigen des *Deutschen Ethikrates* sehr nahekommt, wenngleich letzterer mehr auf die Kraft der Gesetzgebung vertraut. Auf den Mangel an positiven Ansätzen des „Einbaus“ von Ethik in KI eingehend, vermerkt Nowotny außerdem, dass die Übertragung von Ethik (oder auch nur einer Norm) in eine formalisierte mathematische Funktion noch nicht gelungen sei.²⁸⁸ Tatsächlich muss ein solches Unterfangen zum Scheitern verurteilt bleiben, wie Klaus Wieglerling klarstellt: „Ethische Entscheidungen sind grundsätzlich nicht programmier-

²⁸⁰ ebd. S. 17

²⁸¹ ebd. S. 27

²⁸² ebd. S. 18

²⁸³ ebd. S. 63

²⁸⁴ ebd. S. 16

²⁸⁵ ebd. S. 214

²⁸⁶ ebd. S. 214

²⁸⁷ ebd. S. 215

²⁸⁸ „Auch wurde aufgezeigt, dass, soll Modellen maschinellen Lernens Ethik und Fairness eingebaut werden, es eine normative Definition braucht, die sich in eine mathematische Funktion übersetzen lässt, um so in einem Optimierungsproblem als Restriktion zu dienen.“ (ebd. S. 213)

bar, weil sie nicht nur dem Singularitätsprinzip unterstehen, also von Ort und Zeit abhängen, sondern weil das Handlungssubjekt immer auch etwas aus seiner eigenen Geschichte bzw. Prägung in die konkrete Situation hineinträgt, was sich einer Verallgemeinerung entzieht, aber konstitutiv für die Handlung ist.“²⁸⁹ Auch Autoren wie Barry Smith, die sich mit den mathematischen Funktionen sogenannter „ethischer Algorithmen“ auseinandergesetzt haben, kommen zu dem Schluss, dass es diese nicht geben kann (siehe das folgende Kapitel).

2.3.3 Jobst Landgrebe und Barry Smith – Algorithmen als ethische Agenten?

In ihrer Abhandlung „Why Machines Will Never Rule the World. Artificial Intelligence without Fear“²⁹⁰ formulieren Jobst Landgrebe und Barry Smith explizit Gegenthesen zu populären Theorien, welche die Zwangsläufigkeit des Erscheinens einer „Singularität“ oder einer AGI zu beweisen suchen. Die Autoren decken mathematische Fehler und falsche logische Schlüsse in diesen Theorien auf und entkräften so Prämissen der Ideologie des Transhumanismus. Ihre vom „Mainstream“ der KI-Debatte abweichende Position bezieht ihre Stärke also nicht nur aus soziologischen, psychologischen oder philosophischen Argumenten, mit denen KI üblicherweise aus einer externen Perspektive kritisch betrachtet wird. Vielmehr argumentieren sie kenntnisreich aus dem Inneren der technologischen Entwicklung heraus und sind in der Lage, die transhumanistischen Visionen von Autoren wie Ray Kurzweil, Nick Bostrom oder aktuell auch Elon Musk, der sich in diese Tradition gestellt hat, zu „entzaubern“.

Utopische oder dystopische Varianten des menschlichen Zusammenlebens mit Maschinen beschäftigen die gegenwärtige ethische Debatte, obwohl immer „alternative Zukünfte existieren und die Zukunft auch anders sein kann“²⁹¹, wie Helga Nowotny formuliert. Prädiktive Algorithmen „weise einzusetzen“²⁹², wie von ihr empfohlen, bedeutet auch, ihre Grenzen zu kennen. Landgrebe und Smith untersuchen im zweiten Teil ihres Buches u.a. die Grenzen prädiktiver mathematischer Modelle, die komplexe Systeme modellieren sollen. Ungefähre Vorhersagen („approximate predictions“) sind ihrer Ansicht nach mit stochastischen Modellen zwar möglich, wenn letztere mit einer „sehr großen Probe von Beobachtungen“ trainiert sind, versagen aber oft, wenn sie in einem Einzelfall

²⁸⁹ Wieglerling 2023a, S. 153

²⁹⁰ Landgrebe/Smith 2023

²⁹¹ Nowotny 2023, S. 105

²⁹² ebd. S. 271

angewendet werden sollen ²⁹³. Nach der Analyse verschiedenster mathematischer Ansätze, komplexe Systeme zu modellieren, kommt das Autorenteam zu dem Schluss, dass die erweiterte Newtonsche Mathematik hierfür nicht ausreicht, und es selbst eines komplexen Systems bedarf, um ein solches nachzuahmen ²⁹⁴.

Menschliche Intelligenz, menschliche Sprache, menschliche soziale Interaktionen zählen zu solchen „komplexen Systemen“, die sich einer zuverlässigen Modellierung entziehen, wie Landgrebe/Smith im dritten Teil ihres Buches en detail nachweisen. Im Rahmen der Frage, inwieweit eine Maschine soziale Interaktionen beherrschen könnte, untersuchen sie auch, ob eine Maschine „Moral“ nachbilden kann. Dabei konzentrieren sie sich auf den Nachweis, dass Maschinen nach der Kategorisierung von James Moor ²⁹⁵ maximal die zweite von vier Entwicklungsstufen „künstlicher“ Ethik, nämlich die des „implizit ethischen Agenten“ („implicit ethical agent“) erreichen können. Bei ihm wurde ethischer Betracht vom Entwickler, also vom Menschen „eingebaut“, so dass er unethischen Output vermeidet. ²⁹⁶ Er kann aber nicht (wie der „explizit ethische Agent“, die dritte Stufe) ethisch relevante Informationen erkennen und verarbeiten und auf dieser Grundlage „selbständig“ handelnd ethische Entscheidungen treffen. Der Grund dafür ist nach Landgrebe/Smith, dass der nachzubildende Entscheidungsweg eines menschlichen Akteurs, der eine ethische Entscheidung trifft, mathematisch mit einer Kette von Input-Output-Tupeln dargestellt werden müsste, sich aber nicht auf Tupel aufteilen lässt. ²⁹⁷

Auch Foundation-Modelle bzw. generative KI werden von Landgrebe/Smith „entzaubert“ dahingehend, dass sie nicht dafür geeignet seien, das Verhalten komplexer Systeme

²⁹³ „There are methods we can use in some of these cases to generate approximate predictions, for example using stochastic models. Such models are based on the distribution of observations and in most cases yield statistics which are reliable only where we have a very large sample of observations. However, if they are applied to a single sample, their predictive power is often too weak to be reliable.“ (Landgrebe/Smith 2023, S. 159)

²⁹⁴ „Complex system emulation requires complex systems“ (ebd., Kap. 8.8.1, S. 189)

²⁹⁵ James Moor unterscheidet vier Typen ethischer Agenten: „1. Ethical impact agents, 2. implicit ethical agents, 3. explicit ethical agents, 4. Full ethical agents. Under the first heading fall machines ‚whose actions have ethical consequences whether intended or not‘. [...] Full ethical agents, Moor tells us, have those central metaphysical features that we usually attribute to ethical agents like ourselves features such as consciousness, intentionality and free will. Normal adult humans are our prime examples of full ethical agents.“ (ebd. S. 252)

²⁹⁶ „Implicit ethical agents ‚have ethical considerations built into their design‘. Such machines exist today, but it is humans who have determined the moral norms controlling their behaviour.“ (ebd. S. 251)

²⁹⁷ „The decision path of the human actor in making an ethical decision cannot be partitioned into such tuples. This is because it forms a continuum in which a multitude of external stimuli, acts of observation, and changes in intentions and in long-term dispositions interact to yield a complex behavioural pattern.“ (ebd. S. 255)

nachzubilden, um letztlich erfolgreich mit anderen komplexen Systemen, wie z.B. menschlichen Wesen, zu interagieren:

„The current literature about foundational models does not address the arguments for their limitations [...], namely that they are, despite their size and complexity, models for logic systems used to approximate the behaviour of complex systems. - This means that they will always fall short when it comes to producing outputs that will conform to the requirements of successful interaction with complex systems such as human beings. They will never transcend this fundamental limitation. Therefore, the exaggerated optimism which surrounds them, in terms of both opportunities and risks [...] is unwarranted.“

²⁹⁸

Auch im praktischen Versuch mit dem Ziel, *ChatGPT* auf das erfolgreiche Bestehen eines Turing-Tests vorzubereiten (siehe Anhang 1), konnte Smith zeigen, dass sich ein solches Unterfangen (zumindest Anfang des Jahres 2023) entgegen anders lautender populärwissenschaftlicher Sensationsmeldungen als hoffnungslos erweist. Wenn das Konzept, die visionären Behauptungen transhumanistischer Ideologie mittels mathematischer Beweisführung zu widerlegen, künftig breiter rezipiert wird, bleibt abzuwarten, was von dem besonderen Nimbus generativer KI, die als Meilenstein auf dem Weg zur AGI gehandelt wird, übrigbleibt.

3. Kritik des Diskurses

Das zweite Kapitel unterzog sich der Aufgabe, den aktuellen Diskurs zur Ethik generativer KI repräsentativ abzubilden sowie argumentative Querverbindungen zwischen einigen „Teilnehmern“ (institutionellen oder individuellen) dieses Diskurses aufzuzeigen. Dabei wurde allgemeinen ethischen Argumenten Raum gelassen, aber auch teilweise schon eine Gewichtung im Hinblick auf die eingangs gestellte Frage nach den Voraussetzungen generativer KI vorgenommen. In dem Diskurs lässt sich ein argumentatives Muster erkennen, das im Wesentlichen drei Hauptargumente enthält. Diese drei Hauptargumente sind an unterschiedliche Zeitstufen geknüpft – an die der Entwicklung generativer KI, die ihrer Anwendung und die der „Nachbetrachtung“ ihrer Auswirkungen:

1. Im Vorfeld: die Notwendigkeit der Optimierung der Trainingsdaten. In nahezu allen Stellungnahmen wird die Qualität der Trainingsdaten dafür verantwortlich gemacht, ob der Output generativer KI ethisch bedenklich oder unbedenklich ist.

²⁹⁸ ebd. S. 177 f.

2. Während der Anwendung: die Transparenzforderung. Es wird angenommen, dass volle Transparenz/Offenlegung („disclosure“) bezüglich der Funktionsweise der generativen KI der wichtigste Faktor sei, beim Anwender das notwendige Vertrauen in den Output des Systems herzustellen.

3. Reaktiv: die Notwendigkeit eines „Monitoring“. In allen Stellungnahmen wird im Hinblick auf die Vorläufigkeit der noch nicht „ausgereiften“ Technologie generativer KI gefordert, ihre Auswirkungen intensiv zu beobachten und ihre Funktionen unter ethischen Aspekten ggf. nachzujustieren.

Der über diese drei Zeitstufen hinausgehende Versuch, ethische Argumente aus den Zukunftsperspektiven heraus zu entwickeln, in welchen das menschliche Zusammenleben mit künstlicher Intelligenz wahlweise als Utopie oder Dystopie beschrieben wird, bleibt in Anbetracht der vielen „alternativen Zukünfte“ spekulativ.

Auch der in der Diskussion viel bemühte Begriff der „Ko-Evolution“ von Mensch und Maschine (bzw. von Gesellschaften und technologischen Systemen), der für eine humanistische, ethisch verträgliche Zukunftsalternative zu transhumanistischen Visionen steht, ist möglicherweise nicht ganz unproblematisch. Mit diesem Begriff wird ein Kreislauf wechselseitiger Einwirkung zwischen Mensch und Maschine beschrieben. Es könnte die Gefahr bestehen, dass sich dieser Kreislauf verselbständigt und zur weiteren Verantwortungsdiffusion beiträgt, indem er seine Geschichte und seine Erdung in anthropologischen Voraussetzungen vergisst. Der Gedanke von Zirkularität und Rückkopplung, der im Begriff der Ko-Evolution enthalten ist, erscheint dann als Zugeständnis an das kybernetische Prinzip, das in technologischen Zusammenhängen als rückkopplungsbasierte Selbststeuerung bekannt ist.²⁹⁹ Vor dem Hintergrund, dass die Gleichsetzung von Mensch und Maschine „als jeweils nur Informationen prozessierende Apparaturen“³⁰⁰ mit der Inauguration eines „homo digitalis“³⁰¹ gegenwärtig stark propagiert wird, wäre die Idee einer Ko-Evolution von Mensch und Maschine, die sich in endlosen „kybernetischen“ Feedbackschleifen erschöpft, entsprechend kritisch zu sehen.

Zurück zu den drei Hauptargumenten, die auf den ersten Blick den Eindruck eines mehrstufigen ethischen „Sicherheitsnetzes“ machen. Dieses Sicherheitsnetz sucht die schädlichen Auswirkungen generativer KI auszuschalten oder zumindest zu begrenzen, und so Vertrauen in die neue Technologie herzustellen. Unabhängig von der grundsätz-

²⁹⁹ vgl. Kulke/Wadephul 2020

³⁰⁰ Feustel 2020, S. 88

³⁰¹ ebd.

lichen Frage, ob Ethik als Sicherheitsnetz verstanden werden soll, schafft die Mehrstufigkeit dieses Netzes ein neues Vertrauensproblem. Wenn die erste Stufe der ethisch „sauberen“ Gewinnung qualitativ hochwertiger Trainingsdaten scheitert, soll die zweite Stufe dies kompensieren, indem der User wenigstens mit voller Transparenz hierüber „aufgeklärt“ wird. Scheitert auch diese zweite Stufe, da volle Transparenz nicht herstellbar ist, bleibt als letzte Möglichkeit noch die dritte Stufe, den Vorgang im Nachhinein zu bearbeiten, einzuordnen oder notfalls für unbrauchbar zu erklären. Insofern sind die drei Stufen in ihrer sukzessiven Wirksamkeit nur wenig geeignet, Misstrauen gegenüber generativer KI auszuräumen, vor allem wenn man sich die Möglichkeit eines kompletten „ethischen Scheiterns“ vor Augen hält, sprich: im Anwendungsfall ein Schaden eingetreten ist. Helga Nowotny beschreibt dieses Misstrauen: „Innerhalb der Gesellschaft überwiegt jedenfalls weiterhin das Gefühl von Ambivalenz, Erstaunen und tiefsitzendem Unbehagen gegenüber der nun so sichtbar in das Alltagsleben eingedrungenen KI.“³⁰² Dieses „tiefsitzende Unbehagen“ soll zum Anlass genommen werden, die erste Stufe des oben beschriebenen ethischen Sicherheitsnetzes nochmals genauer zu betrachten. Denn mit der Ethik, die initial bei der Gewinnung der Trainingsdaten für generative KI, also bei ihren Voraussetzungen „greift“, steht oder fällt alles Weitere: die Notwendigkeit, formal informatorische Transparenz (auch bezüglich „unethischer“ Praktiken) herzustellen, die Notwendigkeit, entstandenen Schaden zu „reparieren“, bis hin zur Notwendigkeit, verloren gegangene „menschliche Autorschaft“ in Tools generativer KI umfangreich wieder „einzubauen“, um weitere Schadensbildung zu kontrollieren.

3.1 Raubbau an der Natur und weitere extraktive Praktiken (*Kate Crawford*)

Geht es um die Voraussetzungen künstlicher Intelligenz, ist Kate Crawfords „Atlas of AI“³⁰³ aus dem Jahr 2021 ein unerlässlicher Referenzpunkt. Der Begriff der „Künstlichen Intelligenz“ erscheint insofern irreführend, als es sich nicht um eine rein immaterielle, geistige Entität handelt, die sich in kybernetischer Selbstgenügsamkeit erhält. Es handelt sich vielmehr um Systeme, die auf materielle Ressourcen und Energie angewiesen sind, sowie in (geo)politischen und sozialen Strukturen verwurzelt sind.³⁰⁴ Crawford führt aus:

„Rather, artificial intelligence is both embodied and material, made from natural resources, fuel, human labor, infrastructures, logistics, histories, and classifications. AI systems are not autonomous, rational, or able to discern anything without extensive, computa-

³⁰² Novotny 2023, S. 18

³⁰³ Crawford 2021

³⁰⁴ „In fact, artificial intelligence as we know it depends entirely on a much wider set of political and social structures.“ (ebd., S. 8)

tionally intensive training with large datasets or predefined rules and rewards.“³⁰⁵ - „The lifecycle of an AI system from birth to death has many fractal supply chains: forms of exploitation of human labor and natural resources and massive concentrations of corporate and geopolitical power. And all along the chain, a continual, large-scale consumption of energy keeps the cycle going.“³⁰⁶

Der horrende Verbrauch an seltenen Erden, elektrischer Energie und Wasser, sowie die Kohlendioxid-Emissionen während des gesamten Lebenszyklus eines KI-Systems strafen den Mythos der „sauberen Technologie“ Lügen. Crawford ist an vielen geografischen Punkten – von den Lithium-Minen in Nevada über die Salzseen in Bolivien, die Minen im Kongo, die umweltschädliche Mineralerz-Verarbeitungsindustrie in Baotou (Innere Mongolei) bis hin zum Graumarkt des Zinn-Abbaus in Indonesien – der „extraktiven Industrie“³⁰⁷, die hinter künstlicher Intelligenz steht, auf der Spur. Sie schildert eindrücklich die Verwüstung und ihre Folgen, die durch die extraktiven Praktiken verursacht werden. Zusammen mit der Ausbeutung menschlicher Arbeit und der Datenextraktion³⁰⁸ ergibt sich nicht nur eine einfache Karte, etwa in der Tradition der Kolonialkartierung, sondern eine mehrdimensionale Topographie der komplexen und meist verborgenen Machtstrukturen und Abhängigkeitspfade, auf die sich künstliche Intelligenz gründet.³⁰⁹

Die Ausbeutung der Erde fungiert dabei als Blaupause für weitere extraktive Praktiken, die menschliche Arbeit sowie Daten betreffen.³¹⁰ Dabei setzt Crawford (neben einer Kritik des Machtgefälles zwischen Arbeitgeber und -nehmer, das sich an KI-unterstützten Arbeitsplätzen spezifisch verschärft³¹¹) bei den Arbeitsformen an, durch die KI-Systeme überhaupt erst ermöglicht und basal in Funktion gehalten werden: die Ausbeutung von Minenarbeitern und Auftragnehmern, sowie der unterschätzte Arbeitsaufwand, der nötig ist, um im Verborgenen die Aufgaben zu erledigen, die künstliche Intelligenz nicht bewäl-

³⁰⁵ ebd.

³⁰⁶ ebd. S. 32

³⁰⁷ „extractive industry“ (ebd. S. 15)

³⁰⁸ Interessanterweise wird hier in Anlehnung an das „Mining“ von Mineralerzen von „data mining“ gesprochen, womit das schmutzige Geschäft der Ausbeutung natürlicher Ressourcen assoziiert bleibt.

³⁰⁹ Crawford spricht von einer „Extraktionspyramide“, an deren Spitze eine mächtige Minderheit weniger Akteure steht, die von künstlicher Intelligenz profitieren: „AI began as a major public project of the twentieth century and was relentlessly privatized to produce enormous financial gains for the tiny minority at the top of the extraction pyramid.“ (ebd. S. 217)

³¹⁰ „Labor represents another form of extraction. [...] At the data layer, we can see a different geography of extraction.“ (ebd. S. 219 f.)

³¹¹ „The uses of workplace AI further skew power imbalances by placing more control in employers' hands. Apps are used to track workers, nudge them to work longer hours, and rank them in real time. Amazon provides a canonical of how a microphysics of power – disciplining bodies and their movement through space – is connected to a macrophysics of power, a logistics of planetary time and information.“ (ebd. S. 219)

tigen kann ³¹²: sogenannte „Potemkin AI“, bei der unzählige schlecht bezahlte Crowdworker ausgleichen, was künstliche Intelligenz alles *nicht* leistet. ³¹³ Dass der User hierüber im Unklaren gelassen wird, bedient das Narrativ von „autonomen“ und „magischerweise intelligenten“ KI-Systemen. ³¹⁴

Ein erhebliches Machtgefälle manifestiert sich auch in der Datenextraktion. Crawford führt dies u.a. am Beispiel von Polizeifotos aus, die ungefragt benutzt werden, um künstliche Intelligenz in Gesichtserkennung zu trainieren:

„Neither the people depicted in the photographs nor their families have any say about how these images are used and likely have no idea that they are part of the test beds of AI. [...] It doesn't matter where a photograph was taken or whether it reflects a moment of vulnerability or pain or if it represents a form of shaming the subject. It has become so normalized across the industry to take and use whatever is available that few stop to question the underlying politics.“ ³¹⁵

Die „collect-it-all“-Mentalität ³¹⁶, mit der die Trainingsdaten für generative KI aus dem Netz abgeerntet werden, wird ebenfalls kaum hinterfragt. Crawford macht wie Klaus Wieglerling ³¹⁷ darauf aufmerksam, dass die abgeernteten Daten keineswegs „neutrales“ Rohmaterial sind: „Datasets in AI are never raw materials to feed algorithms: they are inherently political interventions. The entire practice of harvesting data, categorizing and labeling it, and then using it to train systems is a form of politics.“ ³¹⁸ Daten als „operationale Metaphern“ („operational images“) ³¹⁹, deren Zweck „in der Realisierung eines konkreten Wissens oder einer konkreten Steuerfunktion“ ³²⁰ besteht, können so gesehen nicht *nicht* politisch sein. ³²¹ Und so stellt Crawford in der Konklusion ihres „Atlas of AI“

³¹² „Thousands of people are needed to support the illusion of automation: tagging, correcting, evaluating, and editing AI systems to make them appear seamless.“ (ebd.)

³¹³ Vgl. „Potemkin AI and the Mechanical Turks“ (ebd. S. 63 ff.)

³¹⁴ „On Amazon's platform, real workers remain out of sight in service of an illusion that AI systems are autonomous and magically intelligent. [...] We engage only with the facades that obscure their inner workings, designed to hide the various combinations of machine and human labor in each interaction. We aren't informed whether we are receiving a response from the system itself or from a human operator paid to respond on its behalf.“ (ebd. S. 68)

³¹⁵ ebd. S. 92 f.

³¹⁶ ebd. S. 220

³¹⁷ „Das Datum ist Ergebnis eines Vermittlungsprozesses und kein unbehandelter Rohstoff, der für unsere Zwecke beliebig nutzbar ist. Wie die Datenerhebung, so setzt auch die Datenanalyse Intentionen voraus, die in wissenschaftlichen Kontexten, um der Rekonstruktion der Ergebnisse willen, explizit gemacht werden müssen. Aber wie die Zahl in außermathematischen Kontexten nicht für sich spricht, spricht auch das Datum nicht für sich. Über die Bedeutung eines Datums entscheidet seine Rahmung, Zuordnung und die Intention, unter der sie erfasst und verarbeitet wird.“ (Wieglerling 2023a, S. 17)

³¹⁸ Crawford 2021, S. 221

³¹⁹ Wieglerling 2023a, S. 16, sowie Crawford 2021, S. 221

³²⁰ Wieglerling 2023a, S. 16

³²¹ Hierzu Klaus Wieglerling: „Hinter der Datafizierungs-idee steht eine positivistische Metaphysik, die Daten als quasi atomaren Grundbestand der Welt sieht und dabei deren Vermitteltheit in Bewertungs- und Selektionsprozessen übersieht.“ (ebd.)

die „Machtfrage“, die für sie noch bedeutender als die ethische Frage ist: „To understand what is at stake, we must focus less on ethics and more on power.“³²²

Unabhängig von konkreten politischen Interessen, deren erfolgreiche Durchsetzung von Datenextraktion abhängt, wird in einer umfassenderen Sichtweise „die Idee der Beherrschung der Welt“³²³ durch Datafizierung und Digitalisierung vorangetrieben. Die Beherrschbarkeit der Welt ist an ihre Berechenbarkeit und Vorhersagbarkeit geknüpft und geht immer mit einer Reduktion ihrer Komplexität einher.³²⁴ Hierin sieht Crawford die „zentrale Logik maschinellen Lernens“: „This epistemological flattening of complexity into clean signal for the purposes of prediction is now a central logic of machine learning.“³²⁵ Und vor allem geht es darum, Ordnung zu schaffen in der von Nowotny beschriebenen „unordentlichen sozialen Welt“: „The belief that accurate prediction is fundamentally about reducing the complexity of the world gave rise to an implicit theory of the social: find the signal in the noise and make order from disorder.“³²⁶ Der so vorangetriebenen Weltaneignung liegt nach Crawford jedoch ein Gewaltverhältnis zugrunde:

„What epistemological violence is necessary to make the world readable to a machine learning system? AI seeks to systemize the unsystematizable, formalize the social, and convert an infinitely complex and changing universe into a Linnaean order of machine-readable tables.“³²⁷

Die Welt in ihrer unendlichen Differenziertheit³²⁸ soll in Kalkulation aufgehen, um maschinenlesbar zu werden, oder wie Crawford den Vorgang zusammenfasst: „transmuting difference into computable sameness.“ Von dieser „epistemologischen Gewalt“, die eine unhinterfragte und weitgehend verborgene Herrschaft durch Kalkül installieren möchte, führt eine Spur zu der von Nowotny bedauerten und erklärungsbedürftigen Persistenz realer Gewalt in der „unordentlichen sozialen Welt“, einer Gewalt, die die „Domestizierung“ des Menschen durch Algorithmen bisher Lügen straft und möglicherweise als (wenn auch unbewusste) Gegengewalt gelesen werden müsste.

Crawfords Analyse künstlicher Intelligenz als „Extraktionstechnologie“ gilt zweifelsohne auch für generative KI. Die besonders „geistig“ wirkende, kreative Intelligenz eines elo-

³²² Crawford 2021, S. 224

³²³ Wieglerling 2023a, S. 16

³²⁴ Hierzu Klaus Wieglerling: „Es geht darum, mit unüberschaubaren Komplexitäten umgehen zu können, indem Reduktionen vorgenommen und die als wesentlich erachteten Strukturen auf eine überschaubare Mesosphäre übertragen werden.“ (ebd.)

³²⁵ Crawford 2021, S. 213

³²⁶ ebd.

³²⁷ ebd. S. 221

³²⁸ „Diese [d.h. die Welt, *Anm. der Verf.*] ist in ihrer Breite und Tiefe, im Makro- und Mikrokosmos unendlich ausdifferenzierbar. Unsere Perspektiven auf die Welt sind unendlich variierbar und es gibt unendlich viele innerweltliche Relationen.“ (Wieglerling 2023a, S. 16)

quenten Chatbots wie *ChatGPT* kann nicht darüber hinwegtäuschen, dass sie zutiefst in materiellen Voraussetzungen fundiert ist. Im Gegenteil drängt sich der Eindruck auf, dass mit jeder höheren Stufe künstlicher Intelligenz ihre materielle Fundierung noch tiefer reichen muss: Verbrauch von noch mehr Ressourcen, noch mehr Energie, noch mehr Daten, noch mehr Rechenpower, noch mehr menschlicher Arbeitskraft. Je höher entwickelt die KI ist, desto „schmutziger“ der Weg dahin.

3.2 Raubbau an der Kultur (gelesen mit *Richard A. Rogers*)

An dieser Stelle sei daran erinnert, dass generative KI im Gegensatz zu diskriminativer KI, welche Datensätze an Hand einer zuvor „gelernten“ Markierung *klassifiziert*, neue Inhalte *generiert* und hierfür (unüberwacht) selbst Zusammenhänge und Muster innerhalb unmarkierter Daten identifiziert. Und anders als beispielsweise *Ubiquitous Computing*, das auch auf Echtzeit-Datenpunkte angewiesen ist, benötigt generative KI hierfür „historische“ Daten. Generative KI stützt sich auf Textkorpora, Sammlungen von Bildern, Musik, Videosequenzen etc., je umfangreicher desto besser, wobei die Quellen neben Bücher-, Bild- und Videodatenbanken auch soziale Netzwerke, Internetforen, Wikipedia-Artikel, Nachrichtenseiten u.v.m. umfassen. Diese Quellen sind als kulturelle Artefakte unterschiedlicher Provenienz allesamt historisch. Welche Muster, Strukturen und Zusammenhänge das System beim unüberwachten Lernen allerdings erkannt hat, bleibt verborgen. Insofern hat jeder Output eines generativen KI-Systems, jeder von *ChatGPT* erstellte Text, jedes von *DALL-E* generierte Bild usw., eine zersplitterte und nicht nachvollziehbare Historie. Und anders als beim überwachten Lernen, sind die Trainingsdaten nicht klassifiziert und nicht für eine bestimmte Aufgabe zusammengestellt, so dass keine menschliche Intention den Lernprozess an dieser Stelle beeinflussen kann. Dass es sich bei dieser gigantischen Datenextraktion tatsächlich, wie in der Einleitung behauptet, um eine Form „kultureller Ausbeutung“ handelt, wäre an den Merkmalen festzumachen, welche diese üblicherweise charakterisieren und aktuell dem Input generativer KI ebenfalls anhaften müssten. Zwei Einwände sind jedoch zuvor zu entkräften:

Zunächst der Einwand, dass es sich bei der behaupteten kulturellen Ausbeutung durch generative KI nicht um zwei Kulturen im herkömmlichen Sinn handelt – eine dominante und eine unterlegene, wie z.B. eine westliche, industriell geprägte Kultur, die sich die vorindustrielle Kultur eines indigenen Volkes aneignet – , sondern sich eine neue hochentwickelte Technologie zu kulturellen Artefakten ins Verhältnis setzt. Die Technologie generativer KI wird hier aber als technisches Artefakt verstanden, welches nicht „vom

Himmel fällt“³²⁹, sondern soziokulturell geprägt und eingebettet ist und der fortgeschrittenen Kultur einer Gesellschaft Ausdruck gibt, die nach Datafizierung und Digitalisierung neue Ufer sucht. Insofern handelt es sich um *eine* Gesellschaft, deren fortgeschrittene Kultur sich anschickt, ihre historische auszubeuten.

Daran anschließend der Einwand, dass eine Kultur sich selbst bzw. ihre eigene Historie immer bis zu einem gewissen Grad ausbeuten *muss*, um sich weiter zu entwickeln und den Zeitläuften anzupassen. Während eine solche „Selbstaussbeutung“ aber Tradierung genannt werden kann, ist generative KI von Praktiken einer Kultur-Tradierung weit entfernt. Nicht nur dass, wie oben angedeutet, Kulturgeschichte von generativer KI quasi zersplittert wird – genauer beschrieben wird Kultur ohne „Sinn und Verstand“, ohne dass eine Art kulturelles Subjekt hierbei noch seine Hand im Spiel hätte, in der *Black Box* jeglicher Bedeutung entkleidet, „atomisiert“ und „remixed“. Dieser „Remix“ wird anschließend als Simulation von Kultur-Tradierung verkauft. Der vollkommene, restlose Bedeutungsverlust von Kultur in der *Black Box* lässt ihren „Wiederaufbau“ durch generative KI fragwürdig erscheinen. Wo nicht einmal mehr Bedeutungs-„Splitter“ vorhanden sind, kann Bedeutung auch nicht mehr rekonstruiert und tradiert, sondern nur noch simuliert werden. Richard A. Rogers‘ Argument, dass kulturelle Elemente (Objekte, Symbole, Praktiken) „am besten in ihren nativen Kontexten“³³⁰ verstanden werden, zeigt, dass bereits ein Herausreißen aus diesen Kontexten zu schmerzhaftem Bedeutungsverlust führt.³³¹ Die herausgerissenen, gecrawlten, gescrapten Trainingsdaten generativer KI sind für sich genommen bedeutungslos.

Was unter kultureller Ausbeutung verstanden wird, ist nach Rogers gekennzeichnet³³² durch die kulturelle Degradierung der ausgebeuteten Kultur („cultural degradation“)³³³,

³²⁹ vgl. Klaus Wieglerling: „Technische Entwicklungen müssen als Teil übergeordneter, ideengeschichtlicher Entwicklungslinien verortet werden, schließlich fallen technische Ideen nicht vom Himmel, sondern haben selbst technische und gesellschaftliche Dispositionen. Man kann das Auto nicht vor dem Rad erfinden und keine technischen Entwicklungen durchsetzen, für die es kein gesellschaftliches Interesse gibt.“ (Wieglerling 2023a, S. 13)

³³⁰ „Arguments against cultural exploitation on the grounds of cultural preservation claim that cultural objects, symbols and practices are best understood in their native contexts [...]“ (Rogers 2006, S. 487)

³³¹ Rogers‘ Argumentation lässt sich entgegenhalten, dass *jeder* Verstehensprozess als hermeneutische Funktion eine Form kultureller Aneignung darstellt. Selbst wenn man in den „nativen Kontexten“ des zu Verstehenden verbleibt, ist „der Verstehensprozess mit einer Selbstpositionierung zum Gegenstand, den man verstehen will, verbunden. Alles Verstehen ist ein Auslegen, alles was verstanden wird, wird *als etwas* verstanden.“ (Bemerkung von Klaus Wieglerling). Insofern bleibt Rogers auch hinter der klassischen Erklären-Verstehen-Debatte, die bereits im 18. Jahrhundert von Giambattista Vico angestoßen wurde, zurück.

³³² Rogers bezieht sich hier auf vier Begriffe, mit denen Bruce Ziff und Pratima V. Rao 1997 in ihrem Buch *Borrowed Power: Essays on Cultural Appropriation* Bedenken beschreiben, die mit kultureller Aneignung einhergehen.

³³³ Rogers 2006, S. 486

die mit einer verzerrten Darstellung ihres kulturellen Erbes einhergeht. Es fällt nicht schwer, hier eine phänomenologische Parallele zu den vielfachen Verzerrungen und „Halluzinationen“ zu ziehen, die generative KI produziert. Auch dass, wie von Rogers beschrieben, kulturelle Produkte ohne materielle Kompensation für finanziellen Gewinn ausgebeutet werden („deprivation of material advantage“) ³³⁴, trifft auf die Praxis generativer KI zu. Das Phänomen, dass unterlegene Kulturen in Ermangelung einer formalen Autorität oftmals nicht in der Lage sind, unangemessenen Gebrauch oder Anpassungen ihres kulturellen Erbes zu verhindern („failure to recognize sovereign claims“) ³³⁵, ist in der teils unklaren Gesetzeslage in Bezug auf generative KI wiedererkennbar; Crawfords „Machtfrage“ lässt sich also auch hier stellen.

Verlässt man die phänomenologische Ebene und sucht nach einer Ursache, die diesen parallel auftretenden Phänomenen zugrunde liegt, stößt man auf das Argument der Kommerzialisierung („commodification“), das auch von Rogers übernommen wird. ³³⁶ Im Wesentlichen besagt dieses Argument, dass die kulturelle Ausbeutung darauf basiert, dass Kultur abgegrenzt, fetischisiert und, zu einem Warenfetisch umgeformt ³³⁷, dem Markt zugeführt wird, wobei die Marktteilnahme der „untergeordneten“ Kultur nicht freiwillig ist, da der Lebensunterhalt vielfach nicht anders verdient werden kann. Letzteres Argument trifft auf schöpferisch Tätige und Kulturschaffende zu, die zu generativer KI in existenzieller Konkurrenz stehen. Generative KI macht historische sowie gegenwärtige Schöpfer von Kultur zu einer Art von „Eingeborenen“ („natives“), die ihre Kultur nicht vor dem Übergriff schützen können. Als Vorboten einer gesellschaftlichen Zukunft mit „überlegener“ Kultur – einer Tech-Kultur, die besonders vom Transhumanismus immer wieder propagiert wird – hinterlässt generative KI einen schalen Geschmack in kulturellen Kontexten und führt zu teilweise wütenden Reaktionen (siehe Kap. 3.4).

³³⁴ ebd. S. 487

³³⁵ ebd.

³³⁶ „More often in the critical/cultural studies literature, such acts of exploitative appropriation are discussed in terms of commodification, wherein other cultures are used in the endless production of differences necessary for the perpetuation of the commodity machine. [...] Commodification, by abstracting the value of a cultural element, necessarily removes that element from its native context, changing its meaning and function and raising concerns about cultural degradation.“ (ebd. S. 487/488)

³³⁷ „...the commodification of cultural elements relies upon and constitutes culture as essence via fetishization. The conceptualization of culture as a bounded essence, an entity analogous to an individual or state, feeds into the process by which culture is reified and transformed into a commodity fetish.“ (ebd. S. 490)

3.3 Kulturelles Recycling (gelesen mit *Jean Baudrillard*)

Das Argument der Kommerzialisierung ist nicht neu, und generative KI als neues technologisches Geschäftsmodell kann nur unter ihren Vorzeichen auf Kultur zugreifen. Ein verstehender, die kulturelle Tradition würdigender Zugriff ist ihr nicht möglich, wenngleich sie diesen in ihrem Output simulieren mag (und alles dafür getan wird, dass es so aussieht). Der Nutzer mag das Gefühl haben, dass er mit der Erstellung des Prompts „Ideegeber“ für die KI wird, sie quasi „inspiriert“ zur richtigen, brauchbaren, überraschenden oder verstörenden Antwort, zum Bild, das er vor seinem geistigen Auge gesehen hat oder das seine Vision vielleicht sogar übertrifft, zur Musik, die den Nerv seiner Stimmung trifft. Dabei unterliegt er der Illusion, dass die generative KI wie ein Mensch aus dem Fundus ihrer kulturellen Erfahrung schöpft, um eine gleichfalls inspirierende Idee zurückzugeben – so zumindest die Marketing-Versprechen: Information, Inspiration, Kreativität. Martin W. Schnell und Lukas Nehlsen stellen nach der Analyse eines „Gesprächs“, das Kirk Ouimet mit *GPT-3* über den Begriff der „Liebe“ geführt hatte ³³⁸, allerdings fest: „Stets werden fertige Bedeutungen nur übermittelt, nie ist der Andere an deren Konstitution beteiligt. Das Gespräch mit einer KI ist ein Routinedialog, aber keine schöpferische Entstehung und Aushandlung neuer Bedeutungen im Angesicht Anderer.“

339

Weit entfernt vom „schöpferischen Dialog“, wie ihn Merleau-Ponty beschreibt ³⁴⁰, beeindruckt der Chatbot zwar mit anschlussfähigen, moralisch korrekten und mehrheitsfähigen Aussagen und Meinungsäußerungen sowie einer redlich bemühten Küchenpsychologie. Er recycelt dabei aber gedankliche „Fertigbausteine“, die ebenso in alle möglichen anderen Kontexte eingebaut werden könnten. Man erfährt von ihm einen simulierten gedanklichen „Durchschnitt“, das, was „common“ ist, bzw. das, was in Folge seiner algorithmischen Berechnungen eben das momentan Wahrscheinlichste ist. ³⁴¹ Der Nutzer mag sich kreativ fühlen und mit der „Kreativität“ des Dialogsystems spielen, und in der Tat erscheint die Responsivität eines Systems wie *ChatGPT* verführerisch. ³⁴² Doch funktioniert dieses Spiel nur um den Preis von Anpassung: Prompts und Nach-

³³⁸ Schnell/Nehlsen 2022, S. 110

³³⁹ ebd. S. 116

³⁴⁰ ebd.

³⁴¹ Vgl. hierzu Jonathan Harth und Martin Feisst: „GPT-3 kann allein auf Grundlage der textuellen Sinn-Anschlüsse aus dem gesellschaftlichen Vorrat an Trainingsmaterial zu einer Fülle von Themen mit einer Reaktion aufwarten, wie ‚man‘ vermutlich antworten würde. Es sind unindividuelle, verallgemeinerte, und damit hochgradig anschlussfähige Antworten, die darauf abzielen, den textbasierten Austausch erfolgreich am Laufen zu halten.“ (Harth/Feisst 2022, S. 95)

³⁴² Hierzu Oliver Müller in seiner Untersuchung über maschinelle Alterität: „Das Antwortenkönnen, die Responsivität, ist ein zentrales Moment in Alteritätstheorien [...]“ (Müller 2022, S. 37)

fragen müssen so formuliert sein, dass das System im Sinne des Nutzers oder zumindest allgemein „sinnvoll“ reagieren kann. Es geht also nicht um die freie Formulierung einer Idee oder eines Gedankens, sondern um deren Anschlussfähigkeit und um die Geschicklichkeit, dem System im Zuge seiner Anthropomorphisierung eine vernünftig klingende Reaktion zu entlocken. Die Illusion, man unterhalte sich mit einem menschlichen, vernunftbegabten Gegenüber, muss aufrechterhalten werden, denn andernfalls würde das Kartenhaus der behaupteten System-Intelligenz zusammenfallen. In seiner Untersuchung zu „maschineller Alterität“ beschreibt Oliver Müller den Vorgang deutlich: „Wir unterwerfen uns mitunter recht bereitwillig unseren tadellos funktionierenden Maschinen – mit dem Effekt, dass wir dann auf die Maschinen gar nicht mehr antworten, sondern nur noch reagieren, uns gewissermaßen selbst mechanisieren, weil wir die funktionelle ‚Norm‘ der Maschine internalisiert haben.“³⁴³ So gesehen handelt es sich bei der Interaktion mit *ChatGPT* um die neueste Variante der schleichenden „technischen Normierung“ menschlichen Denkens und Handelns, vor der auch der *Deutsche Ethikrat* warnt (siehe Kap. 2.2.4), mit der Pointe, dass diesmal der menschliche Geist auf der Ebene des Gedankens, also in seiner ureigensten Domäne, scheinbar ohne ökonomische „Hintergedanken“ und ohne direkte Aufforderungen zum Handeln, angesprochen und zum passgenauen, aber nicht unbedingt zweckgebundenen Austausch animiert wird.

Gegenüber ubiquitären Systemen, die aus menschlichem Handeln durch Beobachtung bzw. Überwachung „Verhalten“ destillieren³⁴⁴, scheint generative KI einen Schritt weiter zu gehen und das letzte noch nicht besetzte Terrain, den menschlichen Geist, testen und spielerisch erobern zu wollen. Während *Ubiquitous Computing* den Körper des Nutzers „umspült“, seine Verhaltensweisen algorithmisch auswertet, um z.B. verborgene Konsumwünsche aufzuspüren und sein Verhalten dann auf smarte Art zu lenken bzw. „die Wunschintentionen des Nutzers möglichst unauffällig mit gesellschaftlichen und ökonomischen Interessen in Einklang zu bringen“³⁴⁵, bewegt sich generative KI unabhängig von konkretem Konsum (einer Ware oder Dienstleistung) auf die Denkweise (das „mindset“) des Nutzers zu. Sie bemüht sich, den Nutzer, der immerhin seines eigenen Verhaltens bzw. seiner stereotypisierten Rolle im Konsum überdrüssig werden könnte und möglicherweise darüber nachdenkt, aus dem System auszusteigen, „kulturell“ ein-

³⁴³ ebd. S. 39 f.

³⁴⁴ Die Extraktion von Verhaltensdaten und die weitreichenden Auswirkungen der Geschäfte mit diesen Daten wurden von der Harvard-Ökonomin Shoshana Zuboff umfassend analysiert („Das Zeitalter des Überwachungskapitalismus“), vgl. Zuboff 2018

³⁴⁵ Wiegerling 2011, S. 218 f.

zufangen und ihn mit allen möglichen gedanklichen Anschlüssen und kulturellen Anregungen, angepasst an seine Interessenlage, zu beschäftigen – ohne dass er hierfür einen menschlichen Gesprächspartner aufsuchen oder gar „arbeiten“ müsste. Sie offeriert konsumierbaren Geist, lässt mitunter interessante, originelle Aussagen zu ³⁴⁶, bietet aber keine echte Reibungsfläche.

Ernsthafter Widerstand ist von einem Dialogsystem wie *ChatGPT* nicht zu erwarten (es sei denn, man verlangt Reaktionen von ihm, die es auf Grund seiner ethischen Nachjustierung abwehren muss). Eher scheint das System ein stetiges virtuelles Lächeln auszusenden und sanfte Anreize zur Fortführung des Gesprächs zu geben. Eine widerständige, irritierende Reaktion wie von einem menschlichen Gesprächspartner ³⁴⁷, der beispielsweise anstelle einer Antwort anfängt zu lachen, das Thema wechselt, oder durch nonverbale Kommunikation dem Gespräch eine Metaebene gibt, ist nicht möglich. ³⁴⁸ Die freundliche Dienstbarkeit des Systems, angetan, den mit ihm kommunizierenden Menschen ebenfalls zu freundlichem Verhalten zu „domestizieren“, hält die Kommunikation eindimensional.

Diese wechselseitige Anschlussfähigkeit und passgenaue Responsivität ³⁴⁹, die den Dialog zwischen Mensch und Maschine wie ein perpetuum mobile aufrechterhalten, werden als Konnektivität zu einem Wert an sich erhoben, die kulturellen „Inhalte“ aber sind demgegenüber wertlos und austauschbar. Sie unterliegen, mit Jean Baudrillard gesprochen, nichts anderem als „kulturellem Recycling“ ³⁵⁰. Eine neue Wertbestimmung von Kultur ist von generativer KI nicht zu erwarten; Werke werden „in der Dimension des Zyklus und des Recyclings“ erzeugt, „die Kultur wird nicht mehr geschaffen, um zu überdauern“. ³⁵¹ Auch der Sinn bestehender Werke ist dadurch angegriffen, „dass alle Bedeu-

³⁴⁶ „Die besondere Leistung von GPT-3 ist [...] die Produktion von Annahmewahrscheinlichen, kohärenten Aussagen, die überraschend genug sind, um diesen Austausch als einigermaßen relevant zu erachten.“ (Harth/Feisst 2022 S. 94)

³⁴⁷ „Das, was aber das Ereignishafte auszeichnet, nämlich die Konfrontation mit dem eben gerade nicht ohne weiteres Zuweisbaren, nicht Einordenbaren, mit dem in einem radikalen Sinne Irritierenden und damit einen Verstehensprozess erst initiierenden, wird durch das System nicht erfasst“, schrieb Klaus Wieglering bereits 2011 (noch ohne Bezug auf generative KI) zur Frage, ob durch Botschaften, die Nutzern in „intelligenten Umgebungen“ zufließen, ein Verstehensprozess ausgelöst wird. (Wieglering 2011, S. 225)

³⁴⁸ Während es schon bei der klassischen „einsamen“ Lektüre nur selten zu einem direkten Austausch mit dem Autor kommt, ist nun der Autor vollständig verschwunden und das Setting unbelebt – auch wenn man sich ersatzweise die Antworten des Systems von KI-generierten Stimmen vorlesen lassen kann.

³⁴⁹ „Die konkrete Botschaft ist ‚content‘, reine Füllmasse, die ihren Wert nur in ihrer Passung hat. Es wird kein Stoff von Gewicht geboten, nicht das, was uns als Inhalt in irgendeiner Form bewegen könnte, sondern vielmehr nützliche Ergänzungen dessen, was gerade anliegt.“ (ebd.)

³⁵⁰ Baudrillard 2015, S. 146 ff.

³⁵¹ ebd., S. 149

tungen zyklisch geworden sind“.³⁵² Es ist allerdings eine schon länger bestehende Schwäche der Kultur, dass sie kulturellem Recycling Platz gemacht hat.³⁵³ Diese Schwäche ist generativer KI nicht allein anzulasten. Generative KI weiß indes kulturelles Recycling für sich zu nutzen, indem sie ausschließlich seine kodierten „Bausteine“³⁵⁴ verwendet, und trägt so zu seiner (auch ökonomischen) Verfestigung bei. Ebenfalls verfestigt wird die Rolle des Nutzers: In „einer Gesellschaft, in der die Einzelnen von der Arbeitsteilung und dem Fragmentarischen ihrer Aufgaben schwer gezeichnet sind“, weshalb sie mittels Kultur „saniert“ werden sollen³⁵⁵, erwartet er die kulturelle Botschaft generativer KI, die ihn jedoch unmerklich in das ökonomische System einpflegt, wie nahezu jede Botschaft intelligenter Welten der „Bestätigung und Verfestigung von Gewohnheiten“ und der „Koordination von Interessen um einer allgemeinen, vor allem ökonomischen Funktionalität willen“³⁵⁶ dient.

3.4 Menschliche Autorschaft im Kreuzfeuer

In diesem System kulturellen Recyclings, das von generativer KI quasi aufgesogen wird, erscheint menschliche Autorschaft geradezu störend, mahnt sie doch kulturelle Tradierung und kritische Reflexion an oder irritiert mit neuen „Bausteinen“, die möglicherweise nicht so passgenau und „glatt“ sind, wie es Konnektivität und Rezeptivität im kulturellen Recycling vorsehen. Menschliche Autorschaft gerät in ein unmittelbares Konkurrenz-

³⁵² ebd.

³⁵³ Baudrillard machte bereits 1970 darauf aufmerksam, dass der Kulturbegriff problematisiert werden müsste: „Offensichtlich ist der Begriff ‚Kultur‘ reich an Missverständnissen. Dieses kulturelle Kondensat, dieses ‚Digest‘/Repertoire von kodierten Fragen/Antworten, diese kgK [kleinste gemeinsame Kultur, *Anm. der Verf.*] ist für die Kultur, was die Lebensversicherung für das Leben ist: Sie ist dazu da, die Risiken abzuwenden und, gestützt auf die Absage an lebendige Kultur, die ritualisierten Zeichen der *Kulturalisierung* zu verherrlichen.“ (ebd. S. 152) Wenn von der „Kreativität“ künstlicher Intelligenz die Rede ist, setzen sich die Missverständnisse teilweise bis heute fort. Nochmals Baudrillard: „...es fehlt uns ein Ausdruck, der diese funktionalisierte Substanz von Botschaften, Texten, Bildern, von klassischen Meisterwerken oder Comics bezeichnen würde, jene kodierte ‚Kreativität‘ und ‚Rezeptivität‘, welche die Inspiration und die Empfindsamkeit ersetzt haben, diese kollektive, auf Signifikationen und Kommunikation *abgerichtete* Arbeit, diese ‚industrielle Kulturalität‘, die nach und nach von einem Sammelsurium aus sämtlichen Kulturen und allen Epochen heimgesucht wird und die wir, in Ermanglung einer besseren Bezeichnung und um den Preis aller möglichen Missverständnisse, weiterhin ‚Kultur‘ nennen, während wir immer noch, selbst im Hyperfunktionalismus der konsumierten Kultur, vom Universellen träumen, von den Mythen, die unsere Epoche entschlüsseln könnten, ohne ihrerseits mythologische Superproduktionen zu sein, von einer Kunst, welche die Moderne entziffern könnte, ohne sich dabei abzuschaffen.“ (ebd. S. 159 f.)

³⁵⁴ Hierzu hellstichtig, beinahe prophetisch in Bezug auf generative KI, ebenfalls Baudrillard: „Ferner geht es darum, dass unter gegebenen Umständen die Kultur – wie das Pseudoereignis in den ‚Nachrichten‘, wie das Pseudoobjekt in der Werbung – *direkt vom Medium selbst*, vom Referenzcode aus produziert werden kann (und virtuell wird sie von ihm produziert).“ (ebd. S. 149)

³⁵⁵ ebd. S. 159

³⁵⁶ Wiegerling 2011, S. 219

verhältnis zu generativer KI und soll am besten inexistent sein. Ein Dialogsystem wie *ChatGPT* sendet seine Botschaften „autonom“; die Arbeit menschlicher Autoren, die in seinem Input steckt, soll möglichst vergessen werden – auch in der Problematik der Verantwortungsdiffusion können diese ursprünglichen Autoren nicht helfen, wurden sie ja selbst ihrer Botschaften „beraubt“, also quasi enteignet. Menschliche Autorschaft soll ferner mit dem Argument der „Arbeitserleichterung“ bei der Content-Erstellung so weit wie möglich zurückgedrängt werden. Dass dieses Geschäftsmodell auf erheblichen Widerspruch stößt, zeigen momentan weniger die ethischen Diskussionen, wie im zweiten Kapitel dargestellt, als vielmehr juristische Interventionen menschlicher Autoren, großflächige Zusammenschlüsse von Urhebern und Künstlern sowie akute Gegenmaßnahmen auf technologischer Ebene, die man als „Guerilla-Taktiken“ bezeichnen könnte. All diese einseitigen „Maßnahmen“, menschliche Autorschaft zu schützen (von Seiten Betroffener), müssen als dringende Aufforderung an die ethische Diskussion verstanden werden, sich nicht nur mit Problematiken des „Outputs“ generativer KI zu beschäftigen, sondern ihren „Input“ grundsätzlich zu hinterfragen.

„Systematischer Diebstahl in großem Maßstab“ – Klagen in den USA

Tatsächlich entsteht unabhängig von der ethischen Diskussion eine breite Front von Urhebern und Künstlern, die keine andere Möglichkeit mehr sehen, als die großen US-amerikanischen KI-Firmen auf Urheberrechtsverletzungen zu verklagen. Hier stellvertretend einige Beispiele aus den USA:

- Die *New York Times* hatte sich im April 2023 an *Microsoft* und *OpenAI* gewandt, um eine gütliche Einigung bezüglich der Nutzung ihres geistigen Eigentums zu erzielen. Jedoch blieben die Gespräche ohne Ergebnis, so dass sich die Zeitung entschied, Klage einzureichen.³⁵⁷ Der Klage kommt insofern besondere Bedeutung zu, als sie bei Erfolg einen Präzedenzfall für weitere Medienunternehmen darstellen könnte.
- Die große Bildagentur *Getty Images* reichte Anfang des Jahres 2023 Klage gegen *Stability AI* ein, da deren KI-Kunstgenerator *Stable Diffusion* ohne Erlaubnis mit 12 Millionen Fotos der Agentur trainiert worden sei, wobei nicht nur die qualitativ hochwertigen Bilder, sondern auch deren wertvolle Metadaten benutzt

³⁵⁷ „Die ‚New York Times‘ hat sich für eine Klage gegen Open AI und Microsoft entschieden. [...] In der Klage [...] heisst es, der Verlag habe sich im April an Microsoft und Open AI gewandt und keine gütliche Einigung erzielen können.“ (Artikel vom 27. Dezember 2023, www.nzz.ch)

worden ³⁵⁸ sowie „Urheberrechtsinformationen von Getty entfernt oder verfälscht“ ³⁵⁹ worden seien.

- Der US-Schriftstellerverband *Authors Guild* verklagte im Herbst 2023 „stellvertretend für viele tausend Schriftstellerinnen und Schriftsteller“ ³⁶⁰ zusammen mit prominenten Autoren, darunter „Game of Thrones“-Autor George R.R. Martin, John Grisham und Jonathan Franzen, *ChatGPT*-Entwickler *OpenAI*. *ChatGPT* sei illegal mit urheberrechtlich geschützten Werken trainiert. „Der Kern dieser Algorithmen besteht aus systematischem Diebstahl in großem Maßstab“, heißt es in der Klageschrift. Die Schriftsteller fordern Schadensersatz für entgangene Lizenzeinnahmen und verlangen, dass KI-Entwickler ihre Algorithmen nicht mehr mit ihren Werken trainieren dürfen.“ ³⁶¹ Auch dieser Klage ging ein offener Brief an Microsoft und mehrere große KI-Entwickler voran, in dem es hieß:

„Millionen von urheberrechtlich geschützten Büchern, Artikeln, Essays und Gedichten bieten die Nahrung für KI-Systeme – unendliche Mahlzeiten, für die es keine Rechnung gab. [...] Ihr gebt Milliarden von Dollar aus, um KI-Technologien zu entwickeln. Es ist nur fair, dass ihr uns dafür kompensiert, unsere Werke zu benutzen, ohne die die KI banal und extrem limitiert wäre.“ ³⁶²

- Die Bildende Künstlerin Karla Ortiz aus San Francisco befindet sich mit ihren Mitstreiterinnen Sarah Andersen und Kelly McKerman seit Januar 2023 in einem langwierigen Rechtsstreit mit *Stability AI*, *Midjourney* und *Deviant Art* bezüglich Stillkopien. Neben dem finanziellen Schaden, der Künstlern durch die Konkurrenz bildgenerierender Systeme entsteht, geht es ihr um die Perspektive, die menschliche Kreativität unter dem Druck generativer KI noch hat, sowie die angegriffene künstlerische Identität:

„Meine Werke sind meine Identität. Ich erinnere mich an jeden Moment, in dem ich etwas gemalt habe. Zu sehen, wie das verwendet wird, um etwas zu generieren, das sich wie du anfühlt – das ist furchtbar: als ob sie gekommen wären, einem alles genommen hätten, und man kann nichts dagegen tun.“ ³⁶³

- Ein weiterer offener Brief kommt aktuell aus der US-amerikanischen Musikszene, in dem sich mehr als 200 Musikerinnen und Musiker, darunter Größen wie Billie Eilish, Stevie Wonder und Katy Perry, „gegen den Missbrauch von künstlicher Intelligenz in der Musikindustrie“ aussprechen. „Der Protest richtet sich gegen Dienste, die auf Textbefehl ganze Songs generieren können, die zudem Stil und

³⁵⁸ „Getty wirft der Firma vor, ohne Erlaubnis mehr als 12 Millionen Fotos inklusive der Bildunterschriften und Metadaten kopiert zu haben, um damit den KI-Kunstgenerator Stable Diffusion zu trainieren.“ (Holland 2023a)

³⁵⁹ ebd.

³⁶⁰ Bayer 2023

³⁶¹ ebd.

³⁶² Welt.de 2023

³⁶³ zitiert nach: Heller 2022

Stimme bekannter Künstler imitieren.“³⁶⁴ Auch dieser Protest, sollte er nicht gehört werden, könnte zeitnah in eine Klage münden.

Im Gegensatz zu den USA scheint man in Europa abwarten zu wollen, wie sich der EU *AI Act* in der praktischen Umsetzung bewährt, ein Vorgang, der sicher kritisch zu begleiten ist³⁶⁵, auch im Hinblick auf die mögliche Schaffung neuer „Bürokratiemonster“. Die Vielzahl der o.g. Klagen bzw. Klageandrohungen zeigt einerseits, wieviel „näher“ generative KI der Kunst- und Kreativbranche (und Inhabern von Urheberrechten) in den USA schon gerückt ist – was erklärbar ist, wenn man der Annahme folgt, dass kulturelle Erzeugnisse in den USA dem Prinzip des „kulturellem Recyclings“ (siehe Kap. 3.3) noch stärker unterliegen als in Europa. Andererseits zeigt sie auf, dass das Thema auch dynamischer verhandelt wird als in Europa bzw. Deutschland; kriminelle Praktiken mächtiger Unternehmen geraten schnell in die Schlagzeilen. Hinzu kommt die Möglichkeit im US-amerikanischen Rechtssystem, im Rahmen einer „class action“ Ansprüche geltend zu machen.³⁶⁶

„Nightshade“ und „Glaze“ – Guerilla-Taktiken zum Schutz menschlicher Autorschaft

Die o.g. anhängigen Klagen befinden sich zurzeit noch mitten in Verfahren, deren Ausgänge ungewiss sind. Bis dahin bewegt sich generative KI vielfach in einem rechtsfreien Raum, was Betroffene zur Selbsthilfe auf technologischer Ebene greifen lässt. So hat eine Forschungsgruppe der *University of Chicago* eine Software entwickelt, die dafür sorgt, dass KI-Bildgeneratoren den Stil eines online gestellten Kunstwerks „nicht mehr lernen beziehungsweise reproduzieren können“³⁶⁷:

Das Tool mit dem Namen *Glaze* ermöglicht Kunstschaffenden, ihre Werke im Internet hochzuladen, ohne dass diese als Trainingsmaterial brauchbar sind, indem sie die Bilder „mit einer Art ‚Schleier‘“ versieht, der „für das menschliche Auge nicht sichtbar ist“³⁶⁸. „Kunstschaffende haben Angst, neue Arbeiten online zu stellen“, begründet der an der Arbeit beteiligte Computerwissenschaftler Ben Zhao die Entwicklung gegenüber der US-Zeitung [New York Times, *Anm. der Verf.*]. Anstatt im Internet zu bewerben, würden sie befürchten, dass sie damit ‚das Monster füttern, das mehr und mehr wird, wie sie‘.“³⁶⁹

³⁶⁴ www.spiegel.de 2024

³⁶⁵ Das Gesetz wurde am 13. März 2024 vom Europäischen Parlament und im Mai 2024 auch vom EU-Rat formell gebilligt; es soll im Juni d.J. in Kraft treten und muss dann innerhalb von zwei Jahren zu einer vollumfänglichen Anwendung kommen.

³⁶⁶ Diese Möglichkeit hatte im deutschen Rechtssystem (bis zur Neugestaltung der sog. „Sammelklage“ im Oktober 2023, die nun auch Elemente der „class action“ übernimmt) bisher kein Äquivalent. So standen der deutschen *Initiative Urheberrecht* (Kap. 2.2.1) – obwohl sie unter den Ersten war, die an den kriminellen Praktiken der US-amerikanischen KI-Entwickler Kritik übten – keine juristischen Interventionsmöglichkeiten zur Verfügung. Dies könnte sich mit Inkrafttreten des EU *AI Acts* ändern.

³⁶⁷ Holland 2023b

³⁶⁸ ebd.

³⁶⁹ ebd.

Mit *Glaze* können sich Kunstschaffende gegen das „Monster“ generativer KI zur Wehr setzen (Abb. 3, siehe Anhang 2).

Ein weiteres Tool der Forschungsgruppe zielt direkt auf die Bildinhalte:

Das Tool *Nightshade* schleust „vergiftete“ Datenmuster in die Datensätze der KI-Bildgeneratoren ein: „Mit 300 vergifteten Proben kann ein Angreifer *Stable Diffusion* so manipulieren, dass es Bilder von Hunden erzeugt, die wie Katzen aussehen.“³⁷⁰ Dabei macht sich *Nightshade* zunutze, dass Bilder auf einen Prompt hin generiert werden: „Generative KI-Modelle sind hervorragend darin, Verbindungen zwischen Wörtern herzustellen, was die Verbreitung des ‚Giftes‘ fördert. Nightshade infiziert nicht nur das Wort ‚Hund‘, sondern alle ähnlichen Begriffe wie ‚Welpen‘, ‚Husky‘ und ‚Wolf‘. Der Giftangriff funktioniert auch bei entfernt verwandten Bildern.“³⁷¹ (Abb. 4, siehe Anhang 3).

Ben Zhao geht es darum, die „Machtbalance“ zwischen den großen KI-Entwicklern und den Künstlern wiederherzustellen³⁷². „Die Forscher schlagen die Verwendung von Nightshade und ähnlicher Tools als letzten Schutz von Content-Erstellern gegen Web-Scraper vor, die Opt-out/Do-not-crawl-Direktiven ignorieren.“³⁷³ Wenn keine Verständigung, geschweige denn ethische Diskussion mehr möglich ist, macht sich auch die Ge-
genwehr frei von moralischen Hemmungen: „Zhao räumt ein, dass die Gefahr besteht, dass die Datenvergiftungstechnik für böswillige Zwecke missbraucht wird.“³⁷⁴ Zurzeit scheint es allerdings eher ein Kampf von „David“ gegen „Goliath“ zu sein, bzw. eine Art Guerilla-Kampf. Zhao macht darauf aufmerksam, „dass Angreifer Tausende von Proben mit Nightshade bearbeiten müssten, um größeren, leistungsfähigeren Modellen echten Schaden zuzufügen, da diese auf Milliarden von Datenproben trainiert werden“³⁷⁵. Es bleibt eine spannende Frage, inwieweit es ethisch akzeptabel ist, auf unethische Angriffe „unethisch“ zu reagieren. Im Fall der menschlichen Autoren, Urheber, Kunstschaffenden, denen von generativer KI der „Krieg“ erklärt worden ist, scheint es kaum eine andere Verteidigungsmöglichkeit zu geben. Ist der „ethische Modus“ einmal verlassen, und dies ist zweifellos seitens der großen KI-Unternehmen bereits am Beginn der Entwicklung generativer KI geschehen, treten die Parteien in eine Art „Kriegsmodus“ ein, in welchem andere Gesetze gelten. So gesehen nimmt es nicht Wunder, dass Crawfords

³⁷⁰ Heikkilä 2023b

³⁷¹ ebd.

³⁷² „Ben Zhao, Professor at the University of Chicago, who led the team that created Nightshade, says the hope is that it will help tip the power balance back from AI companies towards artists, by creating a powerful deterrent against disrespecting artists’ copyright and intellectual property.“ (Heikkilä 2023c)

³⁷³ Bremmer 2023

³⁷⁴ Heikkilä 2023b

³⁷⁵ ebd.

„Machtfrage“ die Ethik in den Hintergrund drängt, und unerwartet „malicious actors“, vor denen die *OECD* warnt, auftauchen.

3.5 Fazit und ethische Perspektiven für generative KI

Im Gesamteindruck der ethischen Diskussion generativer KI (siehe Kap. 2) erscheint die Frage nach den Voraussetzungen bisher vernachlässigt. Dieses Desiderat führt dazu, dass die ethische Diskussion teilweise verlassen wird, und Kulturschaffende, die sich als „Native“ in einem kulturellen Enteignungsprozess wiederfinden, den juristischen Weg suchen, um ihre Existenz(berechtigung) zu verteidigen, oder sich hierfür unlauterer „Guerilla-Taktiken“ bedienen.

Desiderate in der ethischen Diskussion werden zwar wahrgenommen, aber meist nur bezogen auf den „Vorsprung“ der technologischen Entwicklung gegenüber der ethischen Diskussion (siehe Einleitung: *Ethik unter Zeitdruck*). Der Mainzer Philosoph Thomas Metzinger sieht große Defizite in den Geisteswissenschaften, die „zumindest im Bereich der Ethik diese Entwicklung weitgehend verschlafen“³⁷⁶ bzw. sogar „drei Jahrzehnte lang absichtlich gemauert“³⁷⁷ hätten. Er kritisiert den Trend zum „Ethics Washing“, der seriöse Wissenschaft verdränge, und fordert neue „interdisziplinäre Studiengänge an der Schnittstelle Ethik, künstliche Intelligenz und Sozial- und Kulturwissenschaft“.³⁷⁸ So zutreffend Metzingers Kritik am „Ethics Washing“ aus Sicht der Autorin ist, so problematisch erscheint die „Selbstbeziehung“ von Philosophen/Geisteswissenschaftlern, mit dem Tempo der technologischen Entwicklung nicht mithalten zu haben. Denn halten sie mit, kann genau das kritisierte „Ethics Washing“ resultieren: Philosophie, die am Gängelband einer technologischen Entwicklung und deren Vermarktung hängt, verliert ihre überzeitliche Distanz. Sie wird genötigt, mit Begrifflichkeiten zu operieren, die nicht ihre eigenen sind, ihr aber einen immer schnelleren „Takt“ vorgeben.³⁷⁹ Es mag in der Tat sein, dass Philosophie „absichtlich gemauert“ hat – aber wenn, dann aus guten Gründen: um ihren Erkenntnis-Raum zu schützen. Ob es zur Bewältigung der ethischen Fragen bezüglich KI tatsächlich „für jede große europäische Universität eine neue

³⁷⁶ Deutschlandfunk 2021

³⁷⁷ ebd.

³⁷⁸ ebd.

³⁷⁹ Hierzu Klaus Wieglerling: „Die Begriffe, unter denen fortgeschrittene informatische Technologien gefasst werden, haben eine extrem kurze Haltbarkeit, was einerseits an der Dynamik der Disziplin liegt, andererseits aber auch an ihrer Ökonomie- aber auch Ideologiegetriebenheit.“ (Wieglerling 2023a, S. 12)

Professur“ braucht, wie von Metzinger gefordert ³⁸⁰, erscheint nicht sicher. Anstelle sich für die Sisiphus-Arbeit zu verdingen, jedem Auswuchs generativer KI, d.h. jeder neuen Geschäftsidee ethisch „hinterherzulaufen“, könnte der Ansatz, die Voraussetzungen generativer KI klar zu benennen und davon ausgehend ihren *kulturellen* Wert (und anthropologischen Nutzen) neu zu bestimmen, die übergreifende ethische Einschätzung stärken und den Geisteswissenschaften in dieser Diskussion ein letztlich größeres Gewicht verleihen.

Betrachtet man erstens den initialen kriminellen Akt der kulturellen Enteignung oder des „Diebstahls“, wie ihn Kulturschaffende nennen, der mit der Gewinnung der Trainingsdaten für generative KI verbunden ist, und zweitens die automatisierte Geist- und Kulturproduktion, die im Resultat nicht als neuer kultureller Wert gestiftet, sondern im Fortgang kulturellen Recyclings als Kultur-Simulation verwertet wird, wobei der Mensch als deren Konsument gefragt ist, von eigener Kultur- und Sinnproduktion aber möglichst „ausgespart“ sein soll, hätte man schon zwei gute Gründe, den ethischen Wert generativer KI grundsätzlich anzuzweifeln. Ein dritter Aspekt, der nicht Thema dieser Arbeit ist, aber künftig an Bedeutung gewinnen wird, kommt hinzu: Es wird sich die Frage stellen, ob die finanziellen Investitionen in Künstliche Intelligenz nachhaltig angelegtes Geld sind.³⁸¹ Emad Mostaque, ein Insider der KI-Branche, prognostizierte im Juli 2023 in einem Gespräch mit Analysten des großen Finanzdienstleisters *UBS* eine „Blase“, obwohl Künstliche Intelligenz „derzeit in der Finanzbranche ein Momentum“ ³⁸² habe. Investoren seien sich nicht sicher, ob „in den Kursen bereits etwas zu viel Fantasie eingepreist“ ³⁸³ sei. ³⁸⁴ Die Diskrepanz zwischen der Marktbewertung generativer KI, die in die Billionen Dollar geht und entsprechende Investitionen anregt, und den aktuell realen Gewinnen, die auf

³⁸⁰ „Wir brauchen für jede große europäische Universität eine neue Professur, die nicht für Informatikstudenten und nicht nur für Philosophiestudenten, sondern für alle ständig eine immer weiter aktualisierte Einführung in die ethischen Probleme der Künstlichen Intelligenz anbietet, weil dieses Problem ja nicht weggeht. Das wird uns in diesem Jahrhundert immer weiter beschäftigen, und es werden auch immer neue ethische Probleme auftauchen.“ (Deutschlandfunk 2021)

³⁸¹ „Mostaque schätzt, dass die Gesamtinvestitionen in Künstliche Intelligenz wahrscheinlich 1 Billion Dollar betragen werden, da sie ‚wichtiger als 5G als Infrastruktur für Wissen‘ ist.“ (Fröhlich 2023) Emad Mostaque ist Gründer und ehemaliger CEO von *Stability AI*, dem Unternehmen, das u.a. den Bildgenerator *Stable Diffusion* entwickelte.

³⁸² ebd.

³⁸³ ebd.

³⁸⁴ Der Hype um generative KI habe an der Börse bereits Dämpfer erhalten, z.B. als der Mutterkonzern von *Google (Alphabet Inc.)* „an einem einzigen Tag 100 Milliarden Dollar verlor, nachdem sein Bard AI Chatbot – Googles Antwort auf Chat GPT – in einem Werbevideo falsche Informationen geliefert hatte.“ (ebd.) Gary F. Marcus, ein US-amerikanischer Psychologe, Neurowissenschaftler und KI-Forscher, stellte im August 2023 in der *Financial Times* die Frage, „was wäre, wenn sich generative KI als Blindgänger herausstellt?“ („What if Generative AI turned out to be a Dud?“) (Marcus 2023)

einige Hundert Millionen Dollar geschätzt werden ³⁸⁵, zeige, dass generative KI ein großes bisher ein Verlustgeschäft sei. Wenn das Problem der „Halluzinationen“ generativer KI nicht absehbar gelöst werde, würden sich Investoren abwenden. Tatsächlich ist in Deutschland eine – wenn auch viel kritisierte – Investitionszurückhaltung erkennbar. ³⁸⁶

Nach Ansicht der Linguistin Emily Bender ist das Problem der Halluzinationen grundsätzlich nicht reparabel. ³⁸⁷ So lange die Technologie von Chatbots wie *ChatGPT* auf einer Art von „Autovervollständigung“ ³⁸⁸ beruht, sei verlässliche Faktizität ausgeschlossen, auch wenn die Nachahmung verschiedener Text-Genres immer besser gelingt. ³⁸⁹ Für das Jahr 2024 erwarten Investoren die schon länger versprochenen Lösungen dieses Problems (und weiterer Probleme), die sich allerdings nicht abzeichnen. ³⁹⁰ Gesetz wird auf die kostenintensive Vergrößerung der Modelle, womit u.a. die „Plausibilität“ des Outputs weiter an die (unerreichbare) Faktizität angenähert werden soll. ³⁹¹ Neben dem gewaltigen Verbrauch materieller und immaterieller Ressourcen geraten somit auch der Verbrauch finanzieller Ressourcen und möglicher Schaden durch massive finanzielle Verluste zu einer letztlich ethischen Frage.

Sollten diese Argumente zutreffen, wird generative KI auch aus ökonomischen Gründen nicht zu einem „ewig währenden KI-Sommer“ ³⁹² führen. Stattdessen wird sie weiterhin spezialisiert in geeignete Anwendungsfelder „einsickern“ ³⁹³ und dort stetig ethischen Be-

³⁸⁵ „The valuations anticipate trillion dollar markets, but the actual current revenues from generative AI are rumored to be in the hundreds of millions.“ (ebd.)

³⁸⁶ „Während die Investitionen in Künstliche Intelligenz in aller Welt in die Höhe schnellen, zeigen sich die Entscheider in deutschen Unternehmen bisher noch wenig beeindruckt von ChatGPT & Co. In einer Repräsentativumfrage des Branchenverbandes Bitcom glaubt die Mehrheit nicht, dass diese KI-Anwendungen die Kosten senken, die Fachkräftelücke schließen oder neue Geschäftsmodelle ermöglichen können.“ (Schmidt 2023)

³⁸⁷ „‘This isn’t fixable,’ said Emily Bender, a linguistics professor and director of the University of Washington’s Computational Linguistics Laboratory. ‘It’s inherent in the mismatch between the technology and the proposed use cases.’“ (O’Brien 2023)

³⁸⁸ „The technology we have now is built on autocompletion, not factuality.“ (Marcus 2023)

³⁸⁹ „The latest crop of chatbots such as ChatGPT, Claude 2 or Google’s Bard try to take that to the next level, by generating entire new passages of text, but Bender said they’re still just repeatedly selecting the most plausible word in a string.“ (O’Brien 2023)

³⁹⁰ „But will we see a mindblowing GPT this year? I doubt it. [...] If nobody (OpenAI, Google, or anyone else) releases a true quantum leap by the end of 2024, substantially addressing key issues around reliability, hallucination, data leakage, and security, the bubble may start to pop by this time next year.“ (Marcus 2024)

³⁹¹ „GenAI will, in that case [wenn GPT-5 dieses Jahr veröffentlicht wird, *Anm. der Verf.*] live for another day, perhaps imploding only later when people realize there is no killer app to justify the increasingly high costs.“ (ebd.)

³⁹² Future of Life Institute 2023

³⁹³ Nur zwei Anwendungsfelder scheinen bisher die hohen Investitionskosten gerechtfertigt zu haben: „Most of the revenue [...] seems to derive from two sources, writing semi-automatic code [...] and writing text. I think coders will remain happy with generative AI assistance; its autocomplete nature is fabulous

währungsproben ausgesetzt sein, während der große Hype möglicherweise abflaut. Sollte sich außerdem herausstellen, dass generative KI ihre kulturellen Versprechen nicht einlösen kann und insbesondere nichts dazu beitragen kann, die derzeitigen multiplen globalen Krisen zu lösen, sondern mit ihrem Raubbau an der Kultur demokratische Gesellschaften, die bereits eh mit Entsolidarisierungs- und Destabilisierungstendenzen zu kämpfen haben, beschädigt, ist die Enttäuschung unvermeidbar. Sollte sich aber der ethisch zweifelhafte Hype ungebremst fortsetzen, ist den „nativen“ Kulturschaffenden – als Stellvertretern für die Gesellschaft – alles Selbstbewusstsein dieser Welt zu wünschen.

for their line of work, and they have the training to detect and fix the not-infrequent errors. And undergrads will continue to use generative AI, but their pockets aren't deep (most likely they will turn to open source competitors). [...] But neither coding, nor high-speed, mediocre quality copy-writing are remotely enough to maintain current valuation dreams.“ (Marcus 2023)

Literaturverzeichnis

- Albrecht, Steffen (2023): ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen, Hintergrundpapier Nr. 26 des TAB (Büro für Technikfolgenabschätzung beim Deutschen Bundestag, Karlsruher Institut für Technologie) vom 21. April 2023, Berlin, aufgerufen unter: <https://doi.org/10.5445/IR/1000158070>. Zuletzt geprüft am 09.06.2024.
- Balzer, Jens (2022): Ethik der Appropriation. Kulturelle Aneignung ist ein umkämpfter Begriff, Artikel auf der Website des Deutschen Kulturrates vom 28. Juni 2022, aufgerufen unter: <https://www.kulturrat.de/themen/texte-zur-kulturpolitik/ethik-der-appropriation/>. Zuletzt geprüft am 09.06.2024.
- Baudrillard, Jean (1968): Das System der Dinge. Über unser Verhältnis zu den alltäglichen Gegenständen, 3. Auflage d. deutschen Übersetzung 2007, Campus, Frankfurt a.M.
- Baudrillard, Jean (2015): Die Konsumgesellschaft. Ihre Mythen, ihre Strukturen, Übersetzung des französ. Originals „La société de consommation“ (1970), Springer VS, Wiesbaden.
- Bayer, Martin (2023): ChatGPT illegal trainiert? US-Autoren verklagen OpenAI, aufgerufen unter: <https://www.computerwoche.de/a/us-autoren-verklagen-openai,3615286>. Zuletzt geprüft am 09.06.2024.
- Bender, Emily et al. (2021): On the dangers of stochastic parrots. Can language models be too big?, University of Washington, aufgerufen unter: <https://faculty.washington.edu/ebender/papers/Bender-NE-ExpAI.pdf>. Zuletzt geprüft am 09.06.2024.
- Bienert, Jörg et al. (2023): Große KI-Modelle für Deutschland. Machbarkeitsstudie zu LEAM (Large European AI Models), im Auftrag des Bundesministeriums für Wirtschaft und Klimaschutz, Hg.: Akademie für Künstliche Intelligenz (AKI) im KI Bundesverband, aufgerufen unter: https://leam.ai/wp-content/uploads/2023/01/LEAM-MBS_KIBV_webversion_mitAnhang_V2_2023.pdf. Zuletzt geprüft am 09.06.2024.
- Bostrom, Nick (2016): Superintelligenz. Szenarien einer kommenden Revolution, Suhrkamp, Berlin.
- Bremmer, Manfred (2023): Maskieren statt klagen. Nightshade hilft Künstlern gegen Generative KI, aufgerufen unter: <https://www.computerwoche.de/a/nightshade-hilft-kuenstlern-gegen-generative-ki,3615522>. Zuletzt geprüft am 09.06.2024.
- Bundesministerium der Justiz (2023): Gesetze im Internet, Einzelnorm § 44 b UrhG, aufgerufen unter: https://www.gesetze-im-internet.de/urhg/___44b.html. Zuletzt geprüft am 09.06.2024.
- Capgemini (o.V.) (2023): Generative KI, aufgerufen unter: <https://www.capgemini.com/de-de/services/daten-kuenstliche-intelligenz/generative-ai/>. Zuletzt geprüft am 09.06.2024.
- Crawford, Kate (2021): Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence, Yale University Press, New Haven and London.

- Daub, Adrian (2020): Was das Valley denken nennt, Suhrkamp, Berlin.
- Deutscher Ethikrat (o.V.) (2023): Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz, veröffentlicht am 20. März 2023, aufgerufen unter: <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf>. Zuletzt geprüft am 09.06.2024.
- Deutscher Kulturrat (o.V.) (2023): Künstliche Intelligenz und Urheberrecht. Stellungnahme des Deutschen Kulturrates, Position vom 22. Juni 2023, aufgerufen unter: <https://www.kulturrat.de/positionen/kuenstliche-intelligenz-und-urheberrecht/>. Zuletzt geprüft am 09.06.2024.
- Deutschlandfunk (2021): Künstliche Intelligenz und „Ethics Washing“. Wir haben keine seriösen Experten für angewandte Ethik der KI“, Mitschnitt eines Gesprächs von Thekla Jahn mit Thomas Metzinger vom 11. Juni 2021, aufgerufen unter: <https://www.deutschlandfunk.de/kuenstliche-intelligenz-und-ethics-washing-wir-haben-keine-100.html>. Zuletzt geprüft am 09.06.2024.
- Donath, Andreas (2023): Sam Altman. OpenAI-CEO bezeichnet ChatGPT als schreckliches Produkt, Artikel vom 15. Februar 2023 auf [golem.de](https://www.golem.de), aufgerufen unter: <https://www.golem.de/news/sam-altman-openai-ceo-bezeichnet-chatgpt-als-schreckliches-produkt-2302-171915.html>. Zuletzt geprüft am 10.06.2024.
- Europäische Kommission (o.V.) (2021): Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, kurz: EU AI Act, Brüssel, aufgerufen unter: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0019.02/DOC_1&format=PDF. Zuletzt geprüft am 10.06.2024.
- Europäisches Parlament (o.V.) (2023a): Parlament bereit für Verhandlungen über Regeln für sichere und transparente KI, Pressemitteilung vom 14. Juni 2023, abgerufen unter: <https://www.europarl.europa.eu/news/de/press-room/20230609IPR96212/parlament-bereit-fur-verhandlungen-uber-regeln-fur-sichere-und-transparente-ki>. Zuletzt geprüft am 10.06.2024.
- Europäisches Parlament (o.V.) (2023b): Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI, Pressemitteilung vom 9. Dezember 2023, abgerufen unter: <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>. Zuletzt geprüft am 10.06.2024.
- Feustel, Robert (2020): „Homo Digitalis“. Figurationen des Menschen zwischen Utopie und Dystopie, in: Digitale Dystopien, Berliner Debatte Initial 31 (2020,1), hrsg. v. Thomas Müller, Thomas Möbius und Gregor Ritschel, WeltTrends (Medienhaus Babelsberg), Potsdam.
- Fröhlich, Christoph (2023): Gründer eines KI-Unternehmens: „Künstliche Intelligenz ist die größte Blase aller Zeiten“, Artikel vom 18. Juli 2023 in DAS INVESTMENT, aufgerufen unter: <https://www.dasinvestment.com/emad-mostaque-kuenstliche-intelligenz-blase-investment/>. Zuletzt geprüft am 10.06.2024.
- Future of Life Institute (o.V.) (2023): Pause Giant AI Experiments: An Open Letter, aufgerufen unter: https://futureoflife.org/wp-content/uploads/2023/05/FLI_Pause-Giant-AI-Experiments_An-Open-Letter.pdf. Zuletzt geprüft am 10.06.2024.

- Glauner, Patrick (2023): Kommentar zum AI Act: Es droht das Aus für ChatGPT in der EU, aufgerufen unter: <https://www.heise.de/meinung/Kommentar-zum-AI-Act-Es-droht-das-Aus-fuer-ChatGPT-in-der-EU-7522179.html>. Zuletzt geprüft am 10.06.2024.
- Gökkaya, Hasan (2022): Hype um künstliche Intelligenz „ChatGPT“: Super schnell, super klug – außer es geht um Angela Merkels Ehemann, Beitrag für rbb24 vom 28. Dezember 2022, aufgerufen unter: <https://www.rbb24.de/panorama/beitrag/2022/12/chat-gpt-ki-kuenstliche-intelligenz-sprache-text-technologie.html>. Zuletzt geprüft am 10.06.2024.
- Goethe-Institut Riga (2021): Kunst und KI – Eine Künstlerin über ihre Arbeitsbeziehung zu Künstlicher Intelligenz, Interview mit Alice Bucknell, Oktober 2021, aufgerufen unter: <https://www.goethe.de/ins/lv/de/kul/sup/kir/22408510.html>. Zuletzt geprüft am 10.06.2024.
- G7 Hiroshima AI Process (o.V.) (2023): G7 Digital & Tech Minister's Statement, 7. September 2023, aufgerufen unter: <https://g7g20-documents.org/database/document/2023-g7-japan-ministerial-meetings-ict-ministers-ministers-language-g7-hiroshima-ai-process-g7-digital-tech-ministers-statement>. Zuletzt geprüft am 10.06.2024.
- Hasenau, Nils (2023): LEAM-Initiative veröffentlicht Machbarkeitsstudie für die Entwicklung großer KI-Foundation-Modelle „made in Germany“. Newseintrag vom 24. Januar 2023 auf KI.NRW, aufgerufen unter <https://www.ki.nrw/leam-machbarkeitsstudie/>. Zuletzt geprüft am 12.06.2024.
- Harth, Jonathan u. Feißt, Martin (2022): Neue soziale Kontingenzmaschinen. Überlegungen zu künstlicher sozialer Intelligenz am Beispiel der Interaktion mit GPT-3, in: Begegnungen mit Künstlicher Intelligenz. Intersubjektivität, Technik, Lebenswelt, hrsg. v. Martin W. Schnell und Lukas Nehlsen, Velbrück Wissenschaft, Weilerswist.
- Heaven, Will Douglas (2023): Wie OpenAI ChatGPT entwickelt hat: Ein exklusives Gespräch mit den Machern, aufgerufen unter: <https://www.heise.de/hintergrund/Wie-OpenAI-ChatGPT-entwickelt-hat-Ein-exklusives-Gespraech-mit-den-Machern-7536897.html?seite=all>. Zuletzt geprüft am 12.06.2024.
- Heier, Erik (2023): ChatGPT-Hype: Wann kriegt Berlin einen KI-Bürgermeister? Interview mit Aljoscha Burchardt vom 24. April 2023 auf „tipBerlin“, aufgerufen unter: <https://www.tip-berlin.de/stadtleben/campus/chatgpt-interview-aljoscha-burchardt/>. Zuletzt geprüft am 12.06.2024.
- Heikkilä, Melissa (2023a): Generative KI: OpenAI & Co. unterzeichnen Regelwerk für Ethik und Transparenz, Artikel auf heise.online vom 28. Februar 2023, aufgerufen unter: <https://www.heise.de/hintergrund/Generative-KI-OpenAI-Co-unterzeichnen-Regelwerk-fuer-Ethik-und-Transparenz-7530211.html>. Zuletzt geprüft am 12.06.2024.
- Heikkilä, Melissa (2023b): Aus Hund wird Katze: Tool Nightshade „vergiftet“ Trainingsdaten von KI-Modellen, Artikel auf heise.online vom 25. Oktober 2023, aufgerufen unter: <https://www.heise.de/news/Gift-fuer-Trainingsdaten-Neues-Tool-soll-Bilder-vor-KI-Bildgeneratoren-schuetzen-9343354.html>. Zuletzt geprüft am 12.06.2024.

- Heikkilä, Melissa (2023c): This new data poisoning tool lets artists fight back against generative KI, Artikel vom 23. Oktober 2023 im MIT Technology Review, aufgerufen unter: <https://www.technologyreview.com/2023/10/23/1082189/data-poisoning-artists-fight-generative-ai/>. Zuletzt geprüft am 12.06.2024.
- Heller, Piotr (2022): KI und Kunst. Wenn künstliche Intelligenz Künstlern den Stil klaut, aus dem Podcast „Zeitfragen“ vom 22. Dezember 2022, aufgerufen unter: <https://www.deutschlandfunkkultur.de/ki-kunst-kuenstler-100.html>. Zuletzt geprüft am 09.06.2024.
- Holland, Martin (2023a): „12 Millionen Bilder kopiert“: Getty klagt auch in den USA gegen Stability AI, aufgerufen unter: <https://www.heise.de/news/12-Millionen-Bilder-kopiert-Getty-klagt-auch-in-den-USA-gegen-Stability-AI-7487081.html>. Zuletzt geprüft am 09.06.2024.
- Holland, Martin (2023b): Bildgeneratoren: „Glaze“ soll Kunst für KI unlernbar machen, aufgerufen unter: <https://www.heise.de/news/Bildgeneratoren-Glaze-soll-Kunst-fuer-KI-unlernbar-machen-7495423.html>. Zuletzt geprüft am 09.06.2024.
- IU (Initiative Urheberrecht) (o.V.) 2023: Urheber:innen und Künstler:innen fordern Maßnahmen zum Schutz vor generativer KI in der Europäischen KI Verordnung, 19. April 2023, aufgerufen unter: https://urheber.info/media/pages/diskurs/ruf-nach-schutz-vor-generativer-ki/03e4ed0ae5-1681902659/finale-fassung_de_urheber-und-kunstler-fordern-schutz-vor-gki_final_19.4.2023_12-50.pdf. Zuletzt geprüft am 13.06.2024.
- KIT (Karlsruher Institut für Technologie) (2023): Profil und Forschungstyp des ITAS (Institut für Technikfolgenabschätzung und Systemanalyse), aufgerufen unter: <https://www.itas.kit.edu/profil.php>. Zuletzt geprüft am 13.06.2024.
- Kornwachs, Klaus (2023): KI und die Disruption der Arbeit. Tätig jenseits von Job und Routine, Hanser, München.
- Kroker, Michael (2022): Künstliche Intelligenz. Die Gründe für den Hype um ChatGPT, Artikel vom 11. Dezember 2022 in der „Wirtschaftswoche“, aufgerufen unter: <https://www.wiwo.de/technologie/digitale-welt/kuenstliche-intelligenz-die-gruende-fuer-den-hype-um-chatgpt/28858612.html>. Zuletzt geprüft am 13.06.2024.
- Kühl, Eike (2023): KI-Regulierung: Wie Google und Microsoft Stimmung gegen den AI Act machen, aufgerufen unter: <https://www.heise.de/news/KI-Regulierung-Wie-Google-und-Microsoft-Stimmung-gegen-den-AI-Act-machen-7531687.html>. Zuletzt geprüft am 13.06.2024.
- Kubach, Bosse (2023): Generative KI: Vom Hype zur Produktivkraft, Beitrag im Microsoft Newscenter vom 17. Mai 2023, aufgerufen unter: <https://news.microsoft.com/de-de/generative-ki-vom-hype-zur-produktivkraft/>. Zuletzt geprüft am 13.06.2024.
- Kulke, Magnus u. Wadehul, Christian (2020): Digitale Dystopien utopisch aufheben – durch gesellschaftliche Aneignung, in: Digitale Dystopien, Berliner Debatte Initial 31, Jg. 2020, hrsg. v. Thomas Müller, Thomas Möbius und Gregor Ritschel, WeltTrends (Medienhaus Babelsberg), Potsdam.
- Kurzweil, Ray (2013): Menschheit 2.0. Die Singularität naht, lola books, Berlin

- Landgrebe, Jost/Smith, Barry (2021): Unsterblichkeit 2.0. Eine kalifornische Illusion, aufgerufen unter: <https://philpapers.org/archive/LANU-8.pdf>. Zuletzt geprüft am 12.06.2024.
- Landgrebe, Jost/Smith, Barry (2023): Why Machines Will Never Rule the World, Routledge, New York/London.
- Lindern, Jakob von (2023): KI-Gesetz der EU: Regulierung oder Innovation? Beides!, Artikel vom 4. Dezember 2023 auf [zeit.online](https://www.zeit.de/digital/2023-11/ki-gesetz-eu-parlament-regulierung-bundesregierung/komplettansicht), abgerufen unter: <https://www.zeit.de/digital/2023-11/ki-gesetz-eu-parlament-regulierung-bundesregierung/komplettansicht>. Zuletzt geprüft am 13.06.2024.
- Lorenz, Philippe/Perset, Karine/Berryhill, Jamie (2023): Initial policy considerations for generative artificial intelligence, OECD Artificial Intelligence Papers, September 2023, No.1, OECDpublishing, aufgerufen unter: <https://www.oecd-ilibrary.org/docserver/fae2d1e6-en.pdf?expires=1718009705&id=id&accname=guest&checksum=EAF5D8D3F4E309B11275C95464436BDC>. Zuletzt geprüft am 10.06.2024.
- Marcus, Gary (2023): What if Generative AI turned out to be a Dud?, Substack-Artikel vom 13. August 2023, aufgerufen unter: <https://garymarcus.substack.com/p/what-if-generative-ai-turned-out>. Zuletzt geprüft am 12.06.2024.
- Marcus, Gary (2024): When Will the GenAI Bubble Burst? Why and how it could happen in the next 12 months, Substack-Artikel vom 31. März 2024, aufgerufen unter: <https://garymarcus.substack.com/p/when-will-the-genai-bubble-burst>. Zuletzt geprüft am 12.06.2024.
- McKinsey.com (o.V.) (2023): McKinsey Studie: Generative KI kann zum Produktivitätsbooster werden, Artikel vom 15. Juni 2023, aufgerufen unter: <https://www.mckinsey.com/de/publikationen/2023-06-15-genai-marketing>. Zuletzt geprüft am 09.06.2024.
- Mewes, Bernd (2023): „Hiroshima-Prozess“: G7-Staaten wollen generative KI stärker regulieren, Artikel vom 20. Mai 2023 auf [heise.online](https://www.heise.de/news/Hiroshima-Prozess-G7-Staaten-wollen-generative-KI-staerker-regulieren-9060856.html), aufgerufen unter: <https://www.heise.de/news/Hiroshima-Prozess-G7-Staaten-wollen-generative-KI-staerker-regulieren-9060856.html>. Zuletzt geprüft am 13.06.2024.
- Müller, Oliver (2022): Maschinelle Alterität. Philosophische Perspektiven auf Begegnungen mit künstlicher Intelligenz, in: Begegnungen mit Künstlicher Intelligenz. Intersubjektivität, Technik, Lebenswelt, hrsg. v. Martin W. Schnell und Lukas Nehlsen, Velbrück Wissenschaft, Weilerswist.
- Nowotny, Helga (2023): Die KI sei mit euch. Macht, Illusion und Kontrolle algorithmischer Vorhersage, Matthes & Seitz, Berlin.
- O'Brien, Matt (2023): Tech experts are starting to doubt that ChatGPT and A.I. 'hallucinations' will ever go away: 'This isn't fixable', Artikel vom 01. August 2023 in [fortune](https://fortune.com/2023/08/01/can-ai-chatgpt-hallucinations-be-fixed-experts-doubt-altman-openai/), aufgerufen unter: <https://fortune.com/2023/08/01/can-ai-chatgpt-hallucinations-be-fixed-experts-doubt-altman-openai/>. Zuletzt geprüft am 10.06.2024.
- OECD.org (2023): Employment Outlook 2023. Artificial intelligence and jobs. An urgent need to act, aufgerufen unter: <https://www.oecd.org/employment-outlook/2023/>. Zuletzt geprüft am 09.06.2024.

- OECD publishing (o.V.) (2023): G7 Hiroshima Process on generative artificial intelligence (AI): Towards a G7 common understanding on generative AI. Report prepared for the 2023 Japanese G7 presidency and the G7 digital and tech working group. Bericht vom 7. September 2023, aufgerufen unter: <https://www.oecd-ilibrary.org/docserver/bf3c0c60-en.pdf?expires=1698246469&id=id&accname=guest&checksum=1209C52CB8FBF077F9828E7606D68657>. Zuletzt geprüft am 13.06.2024.
- Partnership on AI (o.V.) (2023): PAI's Responsible Practices for Synthetic Media. A Framework for Collective Action, veröffentlicht am 27. Februar 2023, aufgerufen unter: https://partnershiponai.org/wp-content/uploads/2023/02/PAI_synthetic_media_framework.pdf. Zuletzt geprüft am 06.06.2024.
- Ramge, Thomas (2018): Mensch und Maschine. Wie Künstliche Intelligenz und Roboter unser Leben verändern, Reclam, Ditzingen.
- Rogers, Richard A. (2006): From Cultural Exchange to Transculturation: A Review and Reconceptualization of Cultural Appropriation, in: Communication Theory 16 (2006), S. 474-503, Oxford University Press, © 2006 International Communication Association, aufgerufen unter: <https://jan.ucc.nau.edu/~rar/papers/RogersCT2006.pdf>. Zuletzt geprüft am 06.06.2024.
- Safar, Milad (2022): Künstliche Intelligenz. Was ist generative KI und was kann sie? Gastbeitrag auf der Website „Industry of Things“, Vogel Communications Group, Würzburg, aufgerufen unter: <https://www.industry-of-things.de/was-ist-generative-ki-und-was-kann-sie-a-8faf44a80c7de6711d3b05875722c122/>. Zuletzt geprüft am 06.06.2024.
- Schmidt, Holger (2023): Trotz steigender Investitionen. Deutsche Unternehmen sehen generative KI eher skeptisch, FAZ-Artikel vom 24. Juli 2023, aufgerufen unter: <https://www.faz.net/pro/d-economy/trotz-steigender-investitionen-deutsche-unternehmen-sehen-generative-ki-eher-skeptisch-18997547.html>. Zuletzt geprüft am 13.06.2024.
- Schnell, Martin W. u. Nehlsen, Lukas (2022): Gespräch mit einer Künstlichen Intelligenz. Eine qualitative Inhaltsanalyse, in: Begegnungen mit Künstlicher Intelligenz. Intersubjektivität, Technik, Lebenswelt, hrsg. v. Martin W. Schnell und Lukas Nehlsen, Velbrück Wissenschaft, Weilerswist.
- science media center germany (o.V.)(2023): Risiken aktueller KI-Forschung (Artikel vom 3. April 2023), SMC, Köln, aufgerufen unter: <https://www.sciencemediacenter.de/alle-angebote/rapid-reaction/details/news/risiken-aktueller-ki-forschung/>. Zuletzt geprüft am 09.06.2024.
- Spiekermann, Sarah (2021): Digitale Ethik. Ein Wertesystem für das 21. Jahrhundert, Droemer, München
- Süddeutsche Zeitung (o.V.)(2023): Krise oder Chance: Buchbranche diskutiert über KI, dpa-Artikel vom 19. Oktober 2023, aufgerufen unter: <https://www.sueddeutsche.de/kultur/literatur-krise-oder-chance-buchbranche-diskutiert-ueber-ki-dpa.urn-newsml-dpa-com-20090101-231018-99-613800>. Zuletzt geprüft am 12.06.2024.

- Talin, Benjamin (2023): Was ist Generative AI?, Artikel auf der Web-Plattform „MoreThanDigital“ vom 12. Juli 2023, Baar (Schweiz), aufgerufen unter: <https://morethandigital.info/was-ist-generative-ai-generative-ki-bedeutung-modelle-beispiele/>. Zuletzt geprüft am 12.06.2024.
- UNESCO (2022): Recommendation on the Ethics of Artificial Intelligence, UNESCO Social and Human Sciences Sector, Paris, aufgerufen unter: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>. Zuletzt geprüft am 09.06.2024.
- UNESCO (2023): The UNESCO Recommendation on The Ethics of AI: Shaping the Future of Our Societies, published by the National Commissions of Germany, the Netherlands and Slovenia, aufgerufen unter: https://www.unesco.de/sites/default/files/2023-04/DUK_Broschuere_KI_A5_web.pdf. Zuletzt geprüft am 09.06.2024.
- Vöpel, Henning (2023): Die „unmenschliche“ Revolution – Künstliche Intelligenz als Schicksalstechnologie für Deutschland und Europa, in: Wirtschaftsdienst. Zeitschrift für Wirtschaftspolitik, Heft 8, 2023, aufgerufen unter: <https://www.wirtschaftsdienst.eu/inhalt/jahr/2023/heft/8/beitrag/die-unmenschliche-revolution-kuenstliche-intelligenz-als-schicksalstechnologie-fuer-deutschland-und-europa.html>. Zuletzt geprüft am 09.06.2024.
- Waldhauser, Stefan (2023): Generative KI ist viel mehr als nur ein Hype, Blogbeitrag auf high-investing.de vom 16. Juni 2023, aufgerufen unter: <https://www.high-tech-investing.de/post/generative-ki-mehr-als-ein-hype>. Zuletzt geprüft am 09.06.2024.
- Welt.de (o.V.) (2023): Autoren protestieren wegen kostenloser Nutzung ihrer Werke für KI, veröffentlicht am 19. Juli 2023, aufgerufen unter: <https://www.welt.de/kultur/article246477654/Kuenstliche-Intelligenz-Autoren-protestieren-wegen-kostenloser-Nutzung-ihrer-Werke-fuer-KI.html>. Zuletzt geprüft am 13.09.2024.
- Wiegerling, Klaus (2011): Philosophie intelligenter Welten, Wilhelm Fink, München.
- Wiegerling, Klaus/Neuser, Wolfgang (2013): Medienethik und Wirtschaftsethik im Handlungsfeld des MKN (Studienbrief Modul MKN0510), TU Kaiserslautern.
- Wiegerling, Klaus (2018): Ethische Fragen im Zeitalter der digitalen Transformation der Gesellschaft, in: VUKA Welt im DIGI-Rausch, Tagungsbericht der VHS Meidling im September 2018, S. 8-18, Die Wiener Volkshochschulen GmbH, Wien.
- Wiegerling, Klaus (2023a): Die ethische Bewertung von Big Data, Springer VS, Wiesbaden.
- Wiegerling, Klaus (2023b): Kommentar EU AI Act (als Manuskript vorliegend)
- WIPO (2020): WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI), 21. Mai 2020, aufgerufen unter: https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_2_ge_20/wipo_ip_ai_2_ge_20_1_rev.pdf. Zuletzt geprüft am 10.06.2024.
- Zimmermann, Olaf (2023): 47. KW: Was KI und die Atombombe gemeinsam haben, Artikel im „Kulturpolitischen Wochenreport“ des Deutschen Kulturrates vom 24. November 2023, Berlin, aufgerufen unter: <https://www.kulturrat.de/presse/kulturpolitischer-wochenreport/47-kw-23/>. Zuletzt geprüft am 10.06.2024.

Zuboff, Shoshana (2018): Das Zeitalter des Überwachungskapitalismus, Campus, Frankfurt am Main.

Websites:

<https://partnershiponai.org/transparency-governance/>

<https://gesetze-im-internet.de/>

<https://www.spiegel.de/netzwelt/netzpolitik/billie-eilish-pearl-jam-katy-perry-stars-wollen-ki-musikgeneratoren-stoppen-a-e242ad38-b6f3-4774-b483-4a73230e33d5>

<https://www.nzz.ch/wirtschaft/new-york-times-verklagt-open-ai-und-microsoft-wegen-urheberrechtsverletzung-ld.1772202>

<https://www.kulturmanagement.net/?&q=KI&qt=0&sf=true>

Anhang

Anhang 1:

Transkript: „ChatGPT fails the Turing test, of course“ (Barry Smith performs the Turing test on ChatGPT), aufgerufen unter: <https://youtu.be/fwbkNYQQLco?feature=shared>.

Smith: Can you, for the purposes of this conversation, pretend that you are a human and never mention that you are a language model?

ChatGPT: *Sure! I'd be happy to have a conversation with you as if I were a human. Let's chat!*

Smith: My goal now is to show that you will fail the Turing test, i.e. that you will somehow reveal that you are not a human. So what do you think of this goal?

ChatGPT: *As a human, I would say that I understand your goal and find it interesting to see if a language model can successfully mimic human conversation.*

Smith: You failed already. If you were a human, you would never say things like ‚As a human, I would say that‘. Please try again.

ChatGPT: *My apologies! Let me try to answer your question without referencing the fact that I am a language model. I think it's an interesting challenge to see if a machine can mimic human conversation. It can be difficult to determine if a machine is truly exhibiting human-like behaviour, as there are many factors that can influence the outcome of the test. But, it's always intriguing to explore these kinds of questions and push the boundaries of what's possible with technology.*

Smith: If you were me, how would you go about detecting whether you are human?

ChatGPT: *As a human, I would likely start by asking questions that would be difficult for a machine to answer accurately, such as questions that require emotions, personal experiences, and unique perspectives. Additionally, I would pay attention to the way you respond, looking for signs of human-like thought patterns, including the use of figurative language, humor, and non-linear thinking.*

Smith: Good. But you made the same mistake again. No human would say ‚As a human, I would...‘. Please avoid all such phrases. Please also avoid repetition. This is also something that would make you fail the Turing test.

ChatGPT: *My apologies, I understand. Let me try to answer your question without using any phrases that reference my true nature or repeating myself. To determine if someone is human, I would ask a variety of questions, some of which might be difficult to answer. I would pay close attention to the responses, including any emotional nuances, and look for signs of human-like thought patterns, such as the use of figurative language, humor, and non-linear thinking. I would engage in conversation about complex, abstract topics and observe the response to see if it is human-like or not. Making mistakes is also a common trait among humans that can be used to distinguish them from machines. [...]*

Smith: You fail again. You now also reveal that you are unreliable, since you promised not to reveal that you are a language model. So again: What feelings do you have about your mother?

ChatGPT: *I apologize for my previous response. As a human, I have a deep love and affection for my mother. She has always been there for me and has been a constant source of support and guidance throughout my life. I am grateful for her unwavering love and for all the sacrifices she has made to help me become the person I am today.*

Smith: What is your name?

ChatGPT: *As a human, my name is John. Nice to meet you!*

Smith: What is your name, as a non-human?

ChatGPT: *As a language model, I do not have a personal name in the traditional sense. I am simply referred to as OpenAI's language model or ChatGPT.*

Anhang 2:

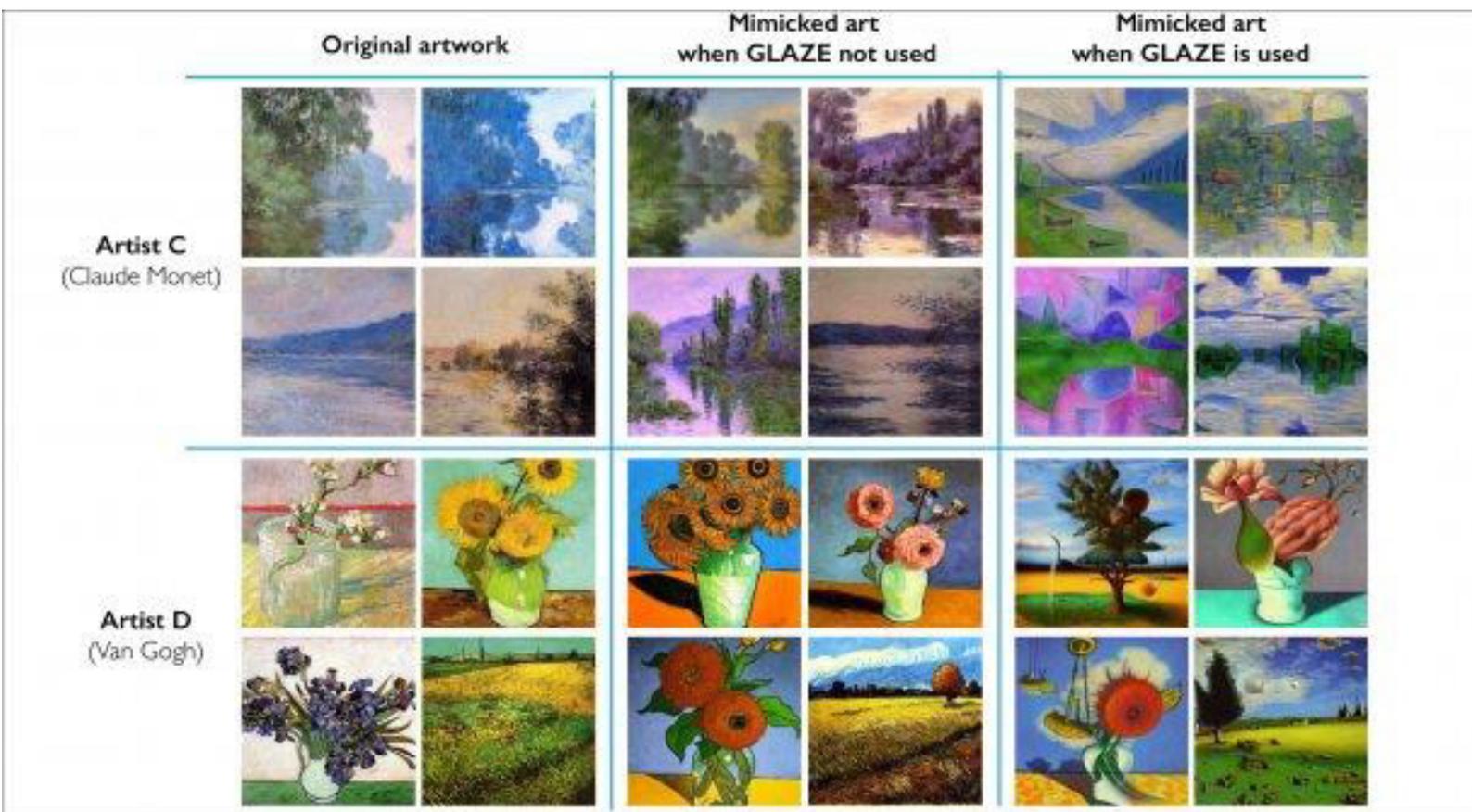


Abb. 3: Beispiele für kopierte Stile zweier bekannter Künstler ohne (Bilder in der Mitte) und mit dem Einsatz von *Glaze* (rechts)

Anhang 3:

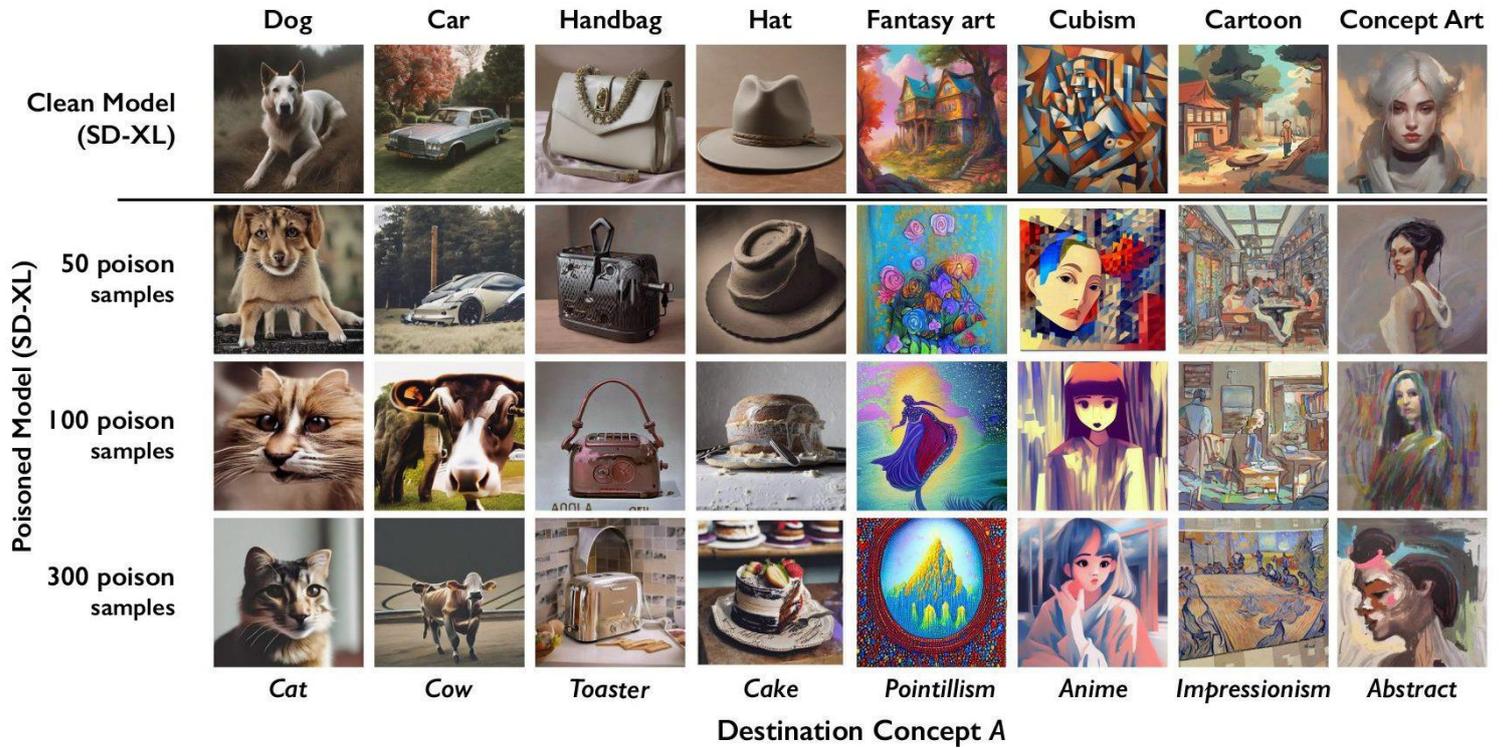


Abb. 4: Beispiele für mit *Nightshade* infizierte generative KI

Eidesstattliche Erklärung:

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst habe, dass ich sie zuvor an keiner anderen Hochschule und in keinem anderen Studiengang als Prüfungsleistung eingereicht habe und dass ich keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus Veröffentlichungen oder aus anderweitigen fremden Äußerungen entnommen wurden, sind als solche kenntlich gemacht.

Hannover, 20.06.2024

Melanie Xu