

Human in Command bei Künstlicher Intelligenz

Prof. Dr. Doris Aschenbrenner, Hochschule Aalen

1. Einleitung

Das Thema „künstliche Intelligenz“ ist seit der Deep-Learning-Revolution (Sejnowski, 2018) weltweit in aller Munde, nachdem das Forschungsfeld zuvor, im so genannten „KI-Winter“ (Umbrello, 2021) unter einem Mangel an Forschungsfinanzierung litt. Spätestens mit dem Aufkommen der Transformer (Vaswani, 2017) und ihrer Anwendung in den großen Sprachmodellen (Large Language Models, LLM) sind die Produktnamen Chat-GPT, GPT-4 oder BERT auch einem großen Teil der Öffentlichkeit bekannt. Dies geht mit einer steigenden kritischen Beschäftigung der Gesellschaft und der Politik einher. In einer deutschen Studie aus dem April 2023 hielten sich bei den Befragten die Befürwortung, Ablehnung und neutrale Haltung gegenüber der Technologie allerdings ungefähr die Waage (Fox, 2023). Es zeigt sich jedoch eine große Skepsis in Bezug auf die Handlungsfähigkeit der Politik: auf die Frage, ob die Politik genug gegen die Risiken für die Menschheit von KI unternehme, sind 21,6% unentschieden und antworten 23,5% mit "eher nein" und 47,2% sogar mit „nein, auf keinen Fall“ (Fox, 2023).

Dieser Artikel gibt einen aktuellen Überblick über die gesellschaftspolitischen Diskussionen im Kontext von Künstlicher Intelligenz, listet lesenswerte Quellen für eine tiefere Beschäftigung auf und beschäftigt sich dabei insbesondere mit den Machtverhältnissen die sich durch die verstärkte Verwendung von KI in der Gesellschaft ergeben. Der Fokus ist Kontrolle durch Menschen (Human in Command), das heißt explizit über Regulierung, Gesetzgebung und später konkret über Entwicklung von Systemen, die die jeweiligen Machtverhältnisse mitdenken.

2. Hegemonie Staat und KI

Die Regierung der Bundesrepublik Deutschland hat im November 2018 die Nationale Strategie für Künstliche Intelligenz „KI-Strategie“ beschlossen (Bundesregierung, 2018) und das Parlament hat 2017 eine Enquete Kommission „Künstliche Intelligenz“ (Bundestag, 2021) einberufen, die 2020 einen Abschlussbericht vorgelegt hat. Alle Ministerien haben eigene Förderinitiativen für das Thema, während der Zusammenschluss „Civic Coding“ (Coding, 2023) durch die Ministerien in den Ressorts Familie, Umwelt und Arbeit im Kontext von Human in Command besonders zu betonen ist, der sich mit dem Einsatz von KI für das Gemeinwohl beschäftigt. Im Ressort Arbeit gibt es zudem das Observatorium Künstliche Intelligenz in Arbeit und Gesellschaft (BMAS, 2023), welches sich mit einigen Publikationen zur „menschzentrierten KI“ mit einem Fachgremium geäußert hat (Dicks, et al., 2021) (Peters, et al., 2022). Darauf wird später nochmals detailliert eingegangen werden. Ähnliche Initiativen gibt es in anderen Ländern, beispielsweise die Forschungsagenda der Niederlande (van den Bosch, 2019).

Es gibt verschiedenen Initiativen von transnationalen Organisationen wie der WHO (WHO, 2023) oder der UNESCO (UNESCO, 2023), bzw. der OECD (OECD, 2019). Letztere ist besonders bedeutend, da deren Definition die Definition von KI-System im sogenannten EU AI Act maßgeblich beeinflusst hat.

Der EU AI Act ist eine Verordnung und damit ein verbindlicher Rechtsakt, den alle EU-Länder in vollem Umfang umsetzen müssen. Am 14. Juni 2023 wurde durch das EU-Parlament ein Entwurf beschlossen, der nun als Basis für die Verhandlungen der Mitgliedsstaaten und der europäischen Kommission als Teil eines langwierigen weiteren Prozesses dient (WEF, 2023) . Er wird als Sammlung der „strengsten

Regeln für KI weltweit“ bezeichnet (FT.com, 2023). Die bis zuletzt umstrittene Definition von künstlicher Intelligenz findet sich hier wie folgt:

„Der Begriff ‚KI-System‘ in dieser Verordnung sollte klar definiert und eng mit der Tätigkeit internationaler Organisationen abgestimmt werden, die sich mit künstlicher Intelligenz befassen, um Rechtssicherheit, Harmonisierung und hohe Akzeptanz sicherzustellen und gleichzeitig genügend Flexibilität zu bieten, um künftigen, rapiden technologischen Entwicklungen in diesem Bereich Rechnung zu tragen. Darüber hinaus sollte die Begriffsbestimmung auf den wesentlichen Merkmalen der künstlichen Intelligenz wie ihren Lern-, Schlussfolgerungs- oder Modellierungsfähigkeiten beruhen und sie von einfacheren Softwaresystemen und Programmierungsansätzen abgrenzen. KI-Systeme sind mit verschiedenen Graden der Autonomie ausgestattet, was bedeutet, dass sie zumindest bis zu einem gewissen Grad unabhängig von menschlicher Kontrolle agieren und in der Lage sind, ohne menschliches Eingreifen zu arbeiten. Die Bezeichnung ‚maschinenbasiert‘ bezieht sich auf die Tatsache, dass KI-Systeme von Maschinen betrieben werden. Durch die Bezugnahme auf explizite oder implizite Ziele wird betont, dass KI-Systeme gemäß expliziten – von Menschen festgelegten – Zielen oder gemäß impliziten Zielen arbeiten können. Die Ziele des KI-Systems können sich – unter bestimmten Umständen – von dem eigentlich vorgesehenen Verwendungszweck unterscheiden. Der Verweis auf Vorhersagen umfasst auch Inhalte, die in dieser Verordnung als eine Form von Vorhersage in Bezug auf eines der möglichen von einem KI-System generierten Ergebnisse hervorgebracht werden. Für die Zwecke dieser Verordnung sollten ‚Umgebungen‘ als Kontexte verstanden werden, in denen KI-Systeme betrieben werden, während die von einem KI-System erzeugten Inhalte – also Vorhersagen, Empfehlungen oder Entscheidungen – auf der Grundlage von Eingaben aus dem genannten Umfeld als Reaktion auf die Ziele des Systems entstehen. Durch solche Ergebnisse wird das genannte Umfeld wiederum beeinflusst, auch dadurch, dass ihm neue Informationen zugeführt werden.“ (Euparl.europa.eu, 2023)

Zentraler Aspekt der Regelung ist die Klassifizierung von Anwendungen entlang eines Risikoklassenmodells (Commission, Regulatory framework proposal on artificial intelligence, 2022), welches „unakzeptable“, „hohe“, „eingeschränkte“ und „minimale“ Risiken unterscheidet. Anwendungen der höchsten Risikoklasse dürfen nicht eingesetzt werden. Hier gab es insbesondere eine lebendige Debatte zum Einsatz biometrischer Massenüberwachung im öffentlichen Raum, die beispielsweise KI-basierte Gesichtserkennung einsetzt. Die Einigung des Parlaments sieht ein weitgehendes Verbot derartiger Anwendungen vor. Ausnahmen sind Fahndungen nach Täter*innen bei schweren Straftaten mit gerichtlicher Anordnung im Einzelfall (Borak, 2023). Dies ist allerdings, wie auch bei anderen Themen der Gesetzgebung der inneren Sicherheit ein Einfallstor dafür diese Maßnahmen trotzdem in großen Maßstab durchzuführen, welches bewusst offengelassen wurde.

Die KI-Modelle mit vielseitigen Einsatzmöglichkeiten (general purpose AI) wie Chat-GPT wurden als Hochrisiko-Anwendung eingestuft, aber nicht verboten. Es wurden hohe Anforderungen an Leistung, Interpretierbarkeit, Korrigierbarkeit, Sicherheit und IT-Security gestellt. Bei generativer KI müssen weitere Transparenzaufgaben erfüllt werden und insbesondere Sicherheitsvorkehrungen für die Erzeugung von falschen Inhalten („Fake News“) treffen (Krempel, EU-Abgeordnete: ChatGPT & Co. sollen besonders streng reguliert werden, 2023). Der aktuellen Debatte zu Urheberrechten wird ebenfalls Rechnung getragen. KI-Modelle, die Texte, Bilder, Musik oder Videos auf Basis vorhandener, oft urheberrechtlich geschützter Werke generieren, müssen diese Ansprüche anerkennen und ihre Modelle gegebenenfalls auch bereinigen.

Der EU AI Act wird, abgesehen von der chinesischen Gesetzgebung (Sheehan, 2023), das international erste große KI-Gesetz sein – die USA mit ihrer „Blueprint for an AI Bill of Rights“ (Whitehouse.gov, 2022) kommen bei der Regulierung ihrer Digitalindustrie weiterhin nur schleppend voran. Momentan befinden sich Kommission, Rat und Parlament in den Trilog-Verhandlungen. Dabei gibt es Befürchtungen von 118 zivilgesellschaftlichen Organisationen unter dem Dach der EDRI (European Digital Rights), dass die Entwürfe von Parlament und Rat KI-Unternehmen ein Schlupfloch bieten könnten, die Bestimmungen für Hochrisiko-Systeme zu umgehen (EDRI, 2023).

4. Hegemonie Wirtschaft und KI

Natürlich haben Unternehmen den politischen Prozess aktiv mitgestaltet. So gab es beispielsweise einen offenen Brief von 150 europäische Unternehmer*innen (Abbou, 2023) wie Heineken, Siemens oder Airbus, in dem sie sich kritisch zu den Mehrkosten für die Regulierungsmaßnahmen äußern (Krempf, 2023). Sie deuten eine „Produktivitätslücke“ zwischen „beiden Seiten des Atlantiks“ an und beschreiben einen Abzug hochinnovativer Unternehmen in weniger regulierte Regionen, womit höchstwahrscheinlich die USA gemeint sind. Der stockende Regulierungsprozess der USA wird massiv durch Interessensgruppen der Privatwirtschaft beeinflusst, wie beispielsweise der durch verschiedene bekannte Technologie-Persönlichkeiten unterschriebenen offenen Brief, der fordert, KI-Experimente für mindestens sechs Monate einzustellen (Futureoflife.org, 2023). Dass die Unterzeichnenden die Risiken für KI derart fürchten, sei ein Trugschluss, so Kritiker*innen. Ein derartiges Moratorium ermögliche den bislang etablierten Firmen den Ausbau ihrer Vorherrschaft und schütze sie in einem hoch dynamischen Markt vor Konkurrenz, einigen Unterzeichnenden wird daher Interessenskonflikt vorgeworfen (Dealbreuin, 2023). Damit erklären sich die anhaltende Weltuntergangsmetaphern von OpenAI und anderen sowie die starke Beeinflussung der Debatte im anhaltenden Präsidentschaftswahlkampf der USA. Einige der einflussreichen Sprecher*innen der Technologiekonzernen des Silicon Valley beraten aktuell den Kongress zu einer KI-Regulierung (Spiegel.de, 2023), die politische Mehrheitslage ist aber kompliziert. Allerdings verkünden Umfragen eine Mehrheit in der amerikanischen (Yougov.com, 2023) aber auch der deutschen (Fox, 2023) Öffentlichkeit zu Gunsten eines solchen Moratoriums.

Ein weiterer Faktor, der die Machtverhältnisse in Bezug auf die Privatwirtschaft darstellt, ist ein Überblick der Investitionen Künstliche Intelligenz. GPT-3 soll mindestens 4,6 Millionen Dollar gekostet haben (Martineau, 2020) und OpenAI hatte angegeben, dass sie pro Tag mindestens 700.000 US-Dollar für die Aufrechterhaltung des Betriebs ausgeben (Petereit, 2023). OpenAI-CEO Sam Altman bezeichnet die Entwicklungskosten für den Nachfolger GPT-4 auf mehr als 100 Millionen Dollar (Haas, 2023). Die Skalierung dieser Modelle ist ein nicht nur auf OpenAI beschränktes Problem: Je größer das Modell, desto teurer auch sein Betrieb. Anfang dieses Jahres schätzte die Bank Morgan Stanley, dass die Hälfte der Google-Suchanfragen, die von einem aktuellen GPT-ähnlichen Programm verarbeitet werden, das Unternehmen zusätzlich sechs Milliarden Dollar pro Jahr kosten könnte (economist.com, 2023). Zur Analyse des Anstiegs der Menge an Trainingsdaten sowie der Limitierung der verfügbaren Datensätze siehe (Villalobos, 2022). Daher wird der Ansatz „größer ist besser“ immer mehr in Frage gestellt, ein Fokus der Forschung ist daher, effizientere (aber auch ggf. spezialisiertere) Modelle zu generieren.

Bevor weitere Kosten als nur die monetären Aufwände zur Sprache kommen, sollen diese Summen im Verhältnis der Forschungsausgaben betrachtet werden. Ursprünglich wollte Deutschland bis 2025 fünf Milliarden Euro in Künstliche Intelligenz investieren, die Presse kritisierte, dass im Mai 2023 allerdings davon lediglich 1,3 Milliarden ausgegeben wurden (Voß, 2023). Im August wurde der „Aktionsplan KI“ vorgestellt, nach dem die öffentlichen Forschungsmittel für KI in den nächsten zwei Jahren auf fast eine Milliarde Euro verdoppelt wurden und damit insgesamt 1,6 Milliarden Euro bereitstellen. Hiermit sollen 150 Labore eingerichtet, Rechenzentren ausgebaut und Datensätze zur Verfügung gestellt werden. Die europäische Ebene möchte aus den Programmen „Digitales Europa“ und „Horizont Europa“ jährlich 1 Milliarde Euro in KI investieren und noch in diesem Jahrzehnt KI-Investitionen von mehr als 20 Milliarden Euro pro Jahr mobilisieren. Ein weiteres Ziel ist, dass über ein Viertel aller Industrie- und Haushaltsroboter in Europa produziert werden (Kommission, 2023).

Beide Ebenen mobilisieren damit trotz angespannter Haushaltslage große Summen, die indirekt auch den deutschen und europäischen Firmen zugutekommen. Ein Vergleich mit den OpenAI-Investitionen aus dem ersten Absatz hinkt aber: Die US-Regierung investierte 2022 rund 3,3 Milliarden US-Dollar, die

privaten Mittel werden im gleichen Zeitraum allerdings auf 47,7 Milliarden geschätzt (aiindex.stanford.edu, 2023).

Es bleibt die realistische Einschätzung, dass nur sehr vereinzelte deutsche bzw. europäische Forschungsinstitutionen überhaupt inhaltlich mithalten können. Wurde das Thema, die besten Köpfe halten zu können bereits 2017 medial diskutiert (Sample, 2017), so bringt der 2023 erschienene Artikel zu „Survival Strategies for Depressed AI Academics“ (Togelius, 2023) die Probleme auf den Punkt: Wissenschaftler*innen in öffentlichen Institutionen können aus verschiedenen Gründen nicht mehr mit ihren Kolleg*innen aus der Privatwirtschaft mithalten, da die aufgewendeten Mittel für Anzahl und Bezahlung der Beschäftigten, sowie die zur Verfügung stehenden Rechenkapazitäten in keinem Verhältnis stehen. Und auch durch die massiven Anstrengungen von Deutschland oder der EU, die oben dargestellt wurden, wird dies auf absehbare Zeit nur gravierender. Eines der Fazits des Papers zu „Survival Strategies“ ist ein Aufruf zu einer engeren Zusammenarbeit zwischen Hochschulen und Industrie – verständlich aber damit endgültig (wie allerdings in den Ingenieurwissenschaften üblich) ein Aufgeben zugunsten der Verwertungslogik.

Zu den weiteren Kosten für Künstliche Intelligenz gibt es aktuell ein verstärktes Bewusstsein für die ökologischen Kosten. So gibt es etwa Datenzentren, die genauso viel Elektrizität brauchen wie 50.000 Haushalte (Monserrate, 2022). Das Trainieren eines KI-Modells kann mehr als 626 Pfund CO₂-Äquivalent emittieren, was die fünffache Menge der Emissionen eines durchschnittlichen amerikanischen PKW entspricht (Hao, 2019). Die Forschungscommunity hat dieses Problem erkannt und ermutigt Forscher*innen sowie KI-Entwickler*innen, in der Praxis zusammen mit ihren Algorithmen auch ihre Emissionen zu überprüfen (ein mögliches Tool dazu in (Lacoste, 2019)). Ein Beispiel für „best practices“ für Energiereduktion um den Faktor Hundert bis Tausend liefert beispielsweise auch (Patterson, 2022).

Neben den ökologischen Kosten werden auch weitere Probleme mit dem KI-Hype aktuell medial diskutiert. So werden zum Training der KI-Algorithmen hunderttausende, wahrscheinlich sogar Millionen so genannter Clickworker*innen eingesetzt. Das Thema der unzureichenden Absicherung und den Arbeitsverhältnissen wurde bereits frühzeitig von Seiten der Arbeitnehmer*innenvertretungen thematisiert (Benner, 2015), bekam aber durch die Reportage der „Time“ zu den Arbeitsverhältnissen der für ChatGPT eingesetzten Arbeitskräfte in Kenia (Perrigo, 2023) eine neue Brisanz. Die Kritik beginnt beim niedrigen Lohn (etwa 1,32 bis 2 Dollar pro Stunde) und richtet einen großen Fokus auf die traumatisierende Wirkung von Tätigkeiten, die das Durchsehen gewaltverherrlichende oder pornographisches Material beinhalten. Ausgangspunkt der Recherche war, dass sich die Clickworker-Plattform Sama acht Monate vor dem Ende des Vertrags mit OpenAI aus dem Vertrag zurückgezogen hatte. Hintergrund war, dass sie mit der Aufgabe betraut worden war, Bilder mit sexuellen und gewalttätigen Inhalten – wohl unter anderen sexualisierte Gewalt an Kindern und Vergewaltigungen – zu sammeln (Rixecker, 2023). Solche Sammlungen unerwünschter Inhalte sind nötig, um den Algorithmen unerwünschte Quellen anzutrainieren.

5. Hegemonie Individuum und KI

Die Frage, ob es nicht möglich sei, eine dem Menschen überlegene Maschine zu bauen beschäftigte den britische Mathematiker Irving John Good bereits in den 1960er Jahren (Good, 1964). Er führte aus, dass diese Erfindung gleichzeitig die letzte menschliche Erfindung und Errungenschaft sein wird. Gerade in der Science-Fiction-Literatur (und dem entsprechenden Filmgenre) hat sich die übermächtige Künstliche Intelligenz, die die Menschheit am Ende unterjocht, als beliebtes Motto etabliert.

Durch die Entwicklung von ChatGPT scheint für viele Menschen ein weiterer Schritt in Richtung dieser Dystopie gegangen zu sein. Der Turing-Test wurde von Alan Turing im Jahr 1950 entwickelt, um

festzustellen, ob eine Maschine ein dem Menschen gleichwertiges Denkvermögen besitzt (Turing, 1950). Die übliche Form ist, dass eine Person zu unterscheiden versucht, ob es sich bei einem*r Gesprächspartner*in um eine Maschine oder um einen Menschen handelt, während diese*r Gesprächspartner*in ihn oder sie versucht, davon zu überzeugen. Nachdem dies über lange Jahre hinweg nicht erfolgreich gelöst wurde, argumentieren Wissenschaftler*innen nun, dass ChatGPT in der Lage ist, diesen Test zu bestehen (Biever, 2023) und fordern neue Wege, die Mächtigkeit von künstlicher Intelligenz zu messen. Andere Wissenschaftler*innen gehen sogar so weit, den Algorithmen bereits „Bewusstsein“ zuzuschreiben (Butlin, 2023), wenn man beispielsweise Messverfahren verwendet, die sonst bei Komapatient*innen verwendet werden. Auf der Gegenseite argumentieren andere (Bishop, 2021), dass KI per Definition „dumm“ sei, bzw. übersetzen AI mit „Agency without Intelligence“ (Florida, 2023) – oder Wirkung ohne echte Intelligenz. Angesichts der Bandbreite der vielen Definitionsmöglichkeiten von Intelligenz – eine sehr anschauliche Sammlung verschiedener Definition liefern (Legg & Hutter, 2007) – wird der Streit darüber durchaus noch länger dauern. Aktuell scheint sich die Mehrheit zumindest einig zu sein, dass ChatGPT keine so genannte „Artificial General Intelligence“ (AGI) ist (verschiedene Möglichkeiten der Annäherung an diesen Begriff liefern (Goertzel, 2007)), und damit eine „starke KI“, d.h. eine dem Menschen gleichwertig oder überlegene Intelligenz besitzt. Auch die KI-Strategie der Bundesregierung (Bundesregierung, 2018) trifft die Unterscheidung zwischen „starker“ und „schwacher“ KI, während letztere „auf die Lösung konkreter Anwendungsprobleme auf Basis der Methoden aus der Mathematik und Informatik [fokussiert], wobei die entwickelten Systeme zur Selbstoptimierung fähig sind“.

Als Spiegel der aufgeheizten gesellschaftlichen Debatte betonen Politiker*innen die menschliche Letztentscheidung (Ethikrat, 2023). Auch die europäische Ebene fordert an verschiedenen Stellen des Entwurfs zum EU AI Act explizit „Human oversight“ und verlangt im Draft standardization request (Commission, Draft standardisation request to the European Standardisation Organisations in support of, 2022): “These shall include measures enabling users to understand, monitor, interpret, assess and intervene in relevant aspects of the operation of the AI system.” Analog fordert die Deutsche Normungsroadmap Künstliche Intelligenz an verschiedenen Stellen (Wahlster, 2022), den Menschen als Teil des Systems in allen Phasen des KI-Lebenszyklus zu begreifen (vgl. u.a. Empfehlung 3, Bedarf 5-19, 5-20).

Doch wie soll diese menschliche Überlegenheit (z.B. unter dem Stichwort „Human in Command“ (Aschenbrenner, 2023) oder auch „Meaningful Human Control“ (Santoni de Sio, 2018) (Cavalcante Siebert, 2023)) aussehen? Gerade bei der Einführung komplexer Systeme, die sich ggf. im Verlauf von neuem Daten-Input (zumindest teilweise) weiterentwickeln können, stellen sich einige Grundfragen, die sich grob in das Spannungsfeld „Wer hat die Kontrolle: Mensch oder Maschine?“ einordnen lassen. Konkrete Spannungsszenarien werden in der öffentlichen Debatte diskutiert: Hier gibt es zum Beispiel die automatisierte Auswahl von Bewerbungsunterlagen (s.o. Fall Amazon, vgl. (Pumhösel, 2020). Zwar ist die automatisierte Vorsortierung der Bewerbungsunterlagen das gewünschte Verhalten der Algorithmen, eine unvorsichtige Auswahl der Datenlage oder des Lernverfahrens kann aber hier zu einem „falschen“ Ergebnis mit einem sogenannten Bias führen.

Die Frage nach der Verortung der Kontrolle wird nicht zum ersten Mal gestellt. So entzündete sich diese Diskussion insbesondere bei der Einführung von Flugassistenzsystemen: Wann soll der Mensch die Kontrolle haben, wann die Maschine? Dazu hat (Fitts, 1951) den „HABA-MABA“-Ansatz („Humans are better at, Machines are better at“) beschrieben, in den Aufgaben generell für besser für eine der beiden Parteien geeignet eingeordnet wurden. Für einen hervorragenden Überblick der darauffolgenden Diskussion im Kontext der Diskussion über autonome Flugassistenz-Systeme siehe (Inagaki, Adaptive automation: Sharing and trading of control., 2003). Jüngere Arbeiten (vgl. zum Beispiel (Bradshaw, 2012)) kritisieren diesen Ansatz hauptsächlich, weil zumindest eine zeitliche oder örtliche Dimension

der Beantwortung mit einbezogen werden muss, also Fragen nicht nur nach „Wer macht was?“, sondern eher nach „Wer macht was wann bzw. wo“. Die Diskussion bewegt sich aktuell in die Richtung der sogenannten „hybrid intelligence“ (Dellermann, 2021), die eine Verschmelzung der beiden Parteien andeutet. Zentral ist dieser Ansatz auch im „Operator 4.0“-Konzept (Romero, 2016), in dem die Erweiterung der menschlichen Fähigkeiten durch Einsatz von technologischen Hilfsmitteln (also in Kontext klassischer Assistenzsysteme) diskutiert wird.

Wie kann man sich die Kontrolle am besten teilen? Zentral ist an dieser Stelle die Wahl geeigneter Autonomiestufen zwischen „komplett manuell“ und „komplett automatisch“ sowie eine dezidierte Übergabe der Kontrolle bei Unsicherheiten oder Gefahren. Aktuelle Lösungen dieses Dilemmas bieten bspw. definierte Autonomiestufen im autonomen Fahren (eine ausführliche Kritik dazu in (Inagaki & Sheridan, A critique of the SAE conditional driving automation definition, and analyses of options for improvement., 2019)) sowie Forschung zu „Handover“-Strategien, wenn beispielsweise der Mensch das Steuer wieder übernehmen muss (wie im H2020 Projekt „Mediator“ (Grondelle, 2020)) sowie die Einführung ähnlicher Autonomiestufen für Künstliche Intelligenz in der industriellen Fertigung (www.plattform-i40.de, 2021).

6. Zusammenfassung

Die Einführung bzw. verstärkte Nutzung von Algorithmen mit Ansätzen der künstlichen Intelligenz in weiten Teilen der Gesellschaft beeinflusst Staat, Wirtschaft und Individuum. Dieser Beitrag hat die aktuellen Spannungsverhältnisse in einem Überblick zusammengefasst. Die Frage nach der Kontrolle und Kontrollierbarkeit von automatisierten Systemen bleibt wichtig und muss bei der Neuentwicklung von Systemen grundsätzlich bedacht werden. Die Herausforderung besteht darin, in Zukunft KI-Systeme zu entwickeln, in denen Mensch und Maschine produktiv miteinander arbeiten und sich gegenseitig verstehen.

Literaturverzeichnis

- Abbou, D. e. (2023). *Open letter to the representatives of the European*. Google Docs. Retrieved from <https://drive.google.com/file/d/1wrtxfvcD9FwfNfWGDL37Q6Nd8wBKXckn/view>
- aiindex.stanford.edu. (2023). *The AI Index report - Measuring trends in Artificial Intelligence*. Stanford HAI. Retrieved from <https://aiindex.stanford.edu/report/>
- Aschenbrenner, D. &. (2023). Human in Command in Manufacturing. *IFIP International Conference on Advances in Production Management Systems* (pp. pp. 559-572). Cham: Springer Nature Switzerland.
- Benner, C. (. (2015). *Crowdwork-zurück in die Zukunft? Perspektiven digitaler Arbeit*. Perspektiven digitaler Arbeit.
- Biever, C. N. (2023). ChatGPT broke the Turing test-the race is on for new ways to assess AI. *Nature*, pp. 619(7971), 686-689.
- Bishop, J. M. (2021). Artificial intelligence is stupid and causal reasoning will not fix it. *Frontiers in Psychology*, 11, 2603.
- BMAS. (2023). *Observatorium Künstliche Intelligenz in Arbeit und Gesellschaft*. Berlin: Denkfabrik. Retrieved from <https://www.ki-observatorium.de/>
- Borak, M. (2023). *EU Parliament approves AI Act amid heated biometrics debates*. Biometricupdate.com. Retrieved from <https://www.biometricupdate.com/202306/eu-parliament-approves-ai-act-amid-heated-biometrics-debates>

- Bradshaw, J. M. (2012). Human-agent-robot teamwork. *IEEE Intelligent Systems*, 27(2), 8-13.
- Bundesregierung. (2018). *Strategie Künstliche Intelligenz der Bundesregierung*. Berlin: Bundesregierung. Retrieved from <https://www.bundesregierung.de/resource/blob/997532/1550276/3f7d3c41c6e05695741273e78b8039f2/2018-11-15-ki-strategie-data.pdf>
- Bundestag. (2021). *Enquete-Kommission „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale“*. Berlin: Bundestag. Retrieved from https://www.bundestag.de/webarchiv/Ausschuesse/ausschuesse19/weitere_gremien/enquete_ki
- Butlin, P. L. (2023). *Consciousness in Artificial Intelligence: Insights from the science of consciousness*. arXiv preprint. doi:10.48550/arXiv.2308.08708
- Cavalcante Siebert, L. L. (2023). Meaningful human control: actionable properties for AI system development. *AI and Ethics*, 3(1), 241-255.
- Coding, C. (2023). *Civic Coding - Innovationsnetz KI für das Gemeinwohl*. Berlin: Denkfabrik.
- Commission, E. (2022). *Draft standardisation request to the European Standardisation Organisations in support of*. European Commission. Retrieved from <https://artificialintelligenceact.eu/wp-content/uploads/2022/12/AIA-%E2%80%93COM-%E2%80%93Draft-Standardisation-Request-5-December-2022.pdf>
- Commission, E. (2022). *Regulatory framework proposal on artificial intelligence*. European Commission. Retrieved from https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_DE.html
- Dealbreuin, J. (2023). *Elon Musk and Others Push for Moratorium on AI, Citing Safety Risks*. 247wallst.com. Retrieved from <https://247wallst.com/technology-3/2023/04/10/elon-musk-and-others-push-for-moratorium-on-ai-citing-safety-risks/>
- Dellermann, D. C. (2021). *The future of human-AI collaboration: a taxonomy of design knowledge for hybrid intelligence systems*. arXiv preprint.
- Dicks, M., Peters, R., Altepost, A., Aschenbrenner, D., Burmester, M., Gerst, D., . . . Wittlich, M. (2021). *Demokratische Technikgestaltung in der digitalen Transformation. Impulspapier zum Fachdialog „MTI – Arbeiten mit Künstlicher Intelligenz“*. Berlin: BMAS. Retrieved from https://www.denkfabrik-bmas.de/fileadmin/Downloads/KI-O/210303_KIO_Demokratische_Technikgestaltung_Impulspapier_barrierefrei.pdf
- economist.com. (2023). *The bigger-is-better approach to AI is running out of road*. The Economist. Retrieved from <https://www.economist.com/science-and-technology/2023/06/21/the-bigger-is-better-approach-to-ai-is-running-out-of-road>
- EDRi. (2023). *EU legislators must close dangerous loophole and protect human rights in the AI Act*. European Digital Rights. Retrieved from <https://edri.org/our-work/civil-society-statement-eu-close-loophole-article-6-ai-act-tech-lobby/>
- Ethikrat, D. (2023). *Mensch und Maschine - Herausforderungen durch Künstliche Intelligenz*. Berlin: Deutscher Ethikrat. Retrieved from <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf>

- Euparl.europa.eu. (2023). *Abänderungen des Europäischen Parlaments vom 14. Juni 2023 zu dem Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz)*. Europäisches Parlament. Retrieved from https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_DE.html
- Fitts, P. (1951). *Human engineering for an effective air-navigation and traffic-control system*. National Research Council, Div. of.
- Floridi, L. (2023). AI as agency without intelligence: on ChatGPT, large language models, and other generative models. *Philosophy & Technology*, 36(1), 15. Retrieved from <https://link.springer.com/article/10.1007/s13347-023-00621-y>
- Fox, P. P. (2023). *So denken die Deutschen über KI. KIRA-Report*. Berlin: Zentrum für KI-Risiken & -Auswirkungen. Retrieved from https://static1.squarespace.com/static/6426ad829db93559a3ed812e/t/64388f3a82787c5eb290743/1681428318566/KIRA_Report_2023-4.pdf
- FT.com. (2023). *EU lawmakers agree tough measures over user of AI*. Financial Times. Retrieved from <https://www.ft.com/content/da597e19-4d63-4d4d-b7d1-c3405302a1e3>
- Futureoflife.org. (2023). *Pause Giant AI Experiments: An Open Letter*. future of life institute. Retrieved from <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- Goertzel, B. (2007). *Artificial general intelligence*. New York: Springer.
- Good, I. J. (1964). *Speculations Concerning the First Ultraintelligent Machine*. Retrieved from <https://web.archive.org/web/20010527181244/http://www.aiveos.com/~bradbury/Authors/Computing/Good-IJ/SCtFUM.html>
- Grondelle, E. e. (2020). *HMI Functional Requirements, D1.5 of the H2020 project MEDIATOR*. <https://mediatorproject.eu/deliverables>: Online.
- Haas, G. (2023). *ChatGPT und Co: Wie große Sprachmodelle effizienter werden können*. Trendingtopics.eu. Retrieved from <https://www.trendingtopics.eu/chatgpt-und-co-wie-grosse-sprachmodelle-effizienter-werden-koennen/>
- Hao, K. (2019). *Training a single AI model can emit as much carbon as five cars in their lifetimes*. MIT Technology Review. Retrieved from <https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/>
- Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. In E. Hollnagel, *Handbook of cognitive task design* (pp. 147–169). Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.
- Inagaki, T., & Sheridan, T. B. (2019). A critique of the SAE conditional driving automation definition, and analyses of options for improvement. *Cognition, technology & work*, pp. 569-578.
- Kommission, E. (2023). *Künstliche Intelligenz – Exzellenz und Vertrauen*. Europäische Kommission. Retrieved from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_de

- Krempl, S. (2023). *Brandbrief der Wirtschaft: KI-Verordnung gefährdet technologische Souveränität*. heise.de. Retrieved from <https://www.heise.de/news/Brandbrief-der-Wirtschaft-KI-Verordnung-gefaehrdet-technologische-Souveraenitaet-9204785.html>
- Krempl, S. (2023). *EU-Abgeordnete: ChatGPT & Co. sollen besonders streng reguliert werden*. Heise.de. Retrieved from <https://www.heise.de/news/EU-Abgeordnete-ChatGPT-Co-sollen-besonders-streng-reguliert-werden-8975888.html>
- Lacoste, A. L. (2019). *Quantifying the carbon emissions of machine learning*. arXiv preprint. doi:10.48550/arXiv.1910.09700
- Legg, S., & Hutter, M. (2007). A collection of definitions of intelligence. *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms*, pp. 17-24.
- Martineau, K. (2020). *Shrinking deep learning's carbon footprint*. MIT News. Retrieved from <https://news.mit.edu/2020/shrinking-deep-learning-carbon-footprint-0807>
- Monserrate, S. G. (2022). *The Staggering Ecological Impacts of Computation and the Cloud*. MIT Press. Retrieved from <https://thereader.mitpress.mit.edu/the-staggering-ecological-impacts-of-computation-and-the-cloud/>
- OECD. (2019). *Artificial Intelligence & Responsible Business Conduct*. OECD. Retrieved from <https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf>
- Patterson, D. (2022). *Good News About the Carbon Footprint of Machine Learning Training*. Google Research. Retrieved from <https://blog.research.google/2022/02/good-news-about-carbon-footprint-of.html>
- Perrigo, B. (2023). *Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic*. Time. Retrieved from <https://time.com/6247678/openai-chatgpt-kenya-workers/>
- Petereit, D. (2023). *ChatGPT: So viel Geld muss OpenAI täglich in den Betrieb des Chatbots stecken*. t3n.de. Retrieved from <https://t3n.de/news/openai-taegliche-kosten-betrieb-chatgpt-chatbot-1548533/>
- Peters, R., Dicks, M., Altepost, A., Aschenbrenner, D., Burmeister, M., Carolus, A., . . . Zimmerling, M. (2022). *Arbeiten mit Künstlicher Intelligenz – Perspektiven für eine menschenzentrierte Gestaltung von KI*. Berlin: BMAS. Retrieved from <https://www.bmas.de/DE/Service/Publikationen/Broschueren/a865-arbeiten-mit-kuenstlicher-intelligenz.html>
- Pumhösel, A. (2020, 03 15). Gender-Bias: Schlechtere Jobchancen für Frauen durch Algorithmen. *Der Standard*, <https://www.derstandard.de/story/2000115720676/gender-bias-schlechtere-jobchancen-fuer-frauen-durch-algorithmen>.
- Rixecker, K. (2023). *ChatGPT und Kenia: Schlecht bezahlte Arbeiter müssen strafbare Inhalte aussortieren*. t3n.de. Retrieved from <https://t3n.de/news/chatgpt-openai-inhalte-filtern-schlecht-bezahlte-arbeiter-1528694/>
- Romero, D. B.-B. (2016, September). The operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. *IFIP international conference on advances in production management systems, Springer, Cham*, pp. 677-686.

- Sample, I. (2017). 'We can't compete': why universities are losing their best AI scientists. The Guardian. Retrieved from <https://www.theguardian.com/science/2017/nov/01/cant-compete-universities-losing-best-ai-scientists>
- Santoni de Sio, F. &. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*, 5, 15.
- Sejnowski, T. J. (2018). *The deep learning revolution*. MIT press.
- Sheehan, M. (2023). *China's AI Regulations and How They Get Made*. Carnegie Endowment to international peace. Retrieved from <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117>
- Spiegel.de. (2023). *Künstliche Intelligenz zügeln? US-Politiker und Techexperten beraten Regeln*. Spiegel.de. Retrieved from <https://www.spiegel.de/netzwelt/netzpolitik/elon-musk-bill-gates-und-mark-zuckerberg-in-washington-us-politiker-und-techfirmen-bosse-beraten-regeln-fuer-kuenstliche-intelligenz-a-3afde05d-25ef-45fb-9fa6-bdbb669b517d>
- Togelius, J. &. (2023). *Choose your weapon: Survival strategies for depressed AI academics*. arXiv preprint. Retrieved from <https://arxiv.org/pdf/2304.06035.pdf>
- Turing, A. M. (1950, Oktober 1). Computing machinery and intelligence. *Mind*, pp. S. 433–460. doi:10.1093/mind/LIX.236.433
- Umbrello, S. (2021). AI Winter. In M. K. (eds.), *Encyclopedia of Artificial Intelligence: The Past, Present, and Future of AI* (pp. pp. 7-8). <https://doi.org/10.13140/RG.2.2.10932.91524>: Santa Barbara, USA: ABC-CLIO.
- UNESCO. (2023). *Ethics of Artificial Intelligence*. UNESCO. Retrieved from <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>
- van den Bosch, A. v. (2019). *Artificial Intelligence Research Agenda for the Netherlands*. Amsterdam: NWO.
- Vaswani, A. S. (2017). Attention is all you need. *Advances in neural information processing systems*.
- Villalobos, P. S. (2022). *Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning*. arXiv preprint. Retrieved from <https://arxiv.org/pdf/2211.04325.pdf>
- Voß, O. (2023). *Wo stecken die KI-Milliarden?* Tagesspiegel Background. Retrieved from <https://background.tagesspiegel.de/digitalisierung/wo-stecken-die-ki-milliarden>
- Wahlster, W. W. (2022). *Deutsche Normungsroadmap künstliche Intelligenz*. DIN e.V. DKE. .
- WEF. (2023). *The European Union's Artificial Intelligence Act - explained*. World Economic Forum. Retrieved from <https://www.weforum.org/agenda/2023/06/european-union-ai-act-explained/>
- Whitehouse.gov. (2022). *Blueprint for an AI Bill of Rights*. Office of Science and Technology Policy, The White House. Retrieved from <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>
- WHO. (2023). *WHO calls for safe and ethical AI for health*. WHO. Retrieved from <https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health>

www.plattform-i40.de. (2021). *Technologieszenario "Künstliche Intelligenz in der Industrie 4.0"*. Plattform "Industrie 4.0". Retrieved from <https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/KI-industrie-40.html>

Yougov.com. (2023). *More than 1,000 technology leaders recently signed an open letter calling on researchers to pause development of certain large-scale AI systems for at least six months world-wide, citing fears of the "profound risks to society and humanity."*. YouGov. Retrieved from <https://today.yougov.com/topics/technology/survey-results/daily/2023/04/03/ad825/2>