

KÜNSTLICHE INTELLIGENZ ZWISCHEN MYTHOS UND KRITIK

JOSEPH WEIZENBAUM (1923-2008)

 weizenbaum
institut

Berliner Landeszentrale
für politische Bildung

BERLIN



INHALT

	Über das Heft	4
		
TEIL 1	MYTHEN DER KÜNSTLICHEN INTELLIGENZ	7
	Wie intelligent sind Computer?	8
	Der ELIZA-Effekt	14
	Das Menschenbild der KI	20
		
TEIL 2	KI UND GESELLSCHAFT	27
	Kann KI gesellschaftliche Probleme lösen?	28
	Die Grenzen von KI	35
	Artificial Intelligentsia	40
		
TEIL 3	KRITISCHE PERSPEKTIVEN AUF KI	47
	Militärischer Einsatz	48
	Ökologische Folgen	54
	Gesellschaftliche Risiken	60
		
TEIL 4	LEITFRAGEN DER KI-KRITIK	67
		
TEIL 5	BIOGRAFISCHE ÜBERSICHT: JOSEPH WEIZENBAUM	71
	Empfohlene Literatur	77

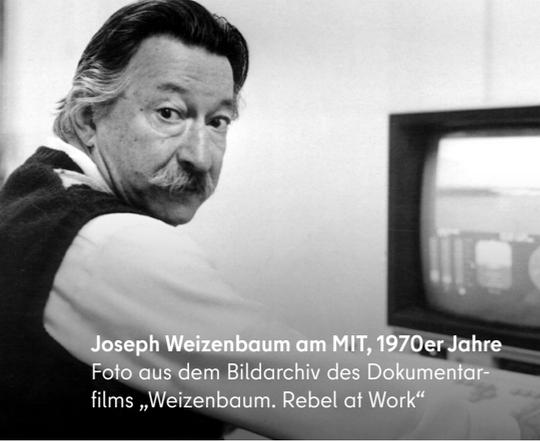
ÜBER DAS HEFT

Werden Computer eines Tages intelligenter sein als Menschen? Und wenn ja, können sie uns dann gefährlich werden? Oder ist der technologische Fortschritt ganz im Gegenteil unsere letzte Hoffnung, etwa im Kampf gegen Armut, Krankheiten oder den Klimawandel? Die Debatten über mögliche Auswirkungen immer leistungsfähigerer Computer auf unser Leben haben oft etwas von Science-Fiction. Verheißungsvolle Utopien und beängstigende Dystopien liegen oft dicht beieinander, ganz besonders wenn es um Künstliche Intelligenz, kurz: KI, geht. Mit einer Mischung aus Faszination und Grusel schauen wir auf täuschend echte Deepfakes von Politi-

ker:innen und Prominenten oder auf die Videos von mal tanzenden, mal kämpfenden Roboterhunden.

Die Gleichzeitigkeit von Technikbegeisterung und Technikskepsis ist allerdings kein neues Phänomen. Beides gab es schon bei den frühen Pionieren der Computertechnologie. Lange dominierte dabei der Glaube, mithilfe von Computern ließe sich die Welt zum Besseren verändern. Die digitale Revolution würde die Menschen mündiger machen, uns zu einem „globalen Dorf“ vernetzen und Gesellschaften demokratisieren. Ganz nach dem Motto: Es gibt kein Problem, für das es keine technische Lösung gäbe.

Dagegen gab es schon früh auch mahnende Worte, zum Teil aus dem Maschinenraum der Computerentwicklung selbst. Ein prominenter Kritiker war Joseph Weizenbaum (1923–2008), der in den 1950er Jahren das erste Computerbanksystem mitentwickelte und bereits seit den 1960er Jahren in der KI-Forschung aktiv war. Berühmt wurde



Joseph Weizenbaum am MIT, 1970er Jahre
Foto aus dem Bildarchiv des Dokumentarfilms „Weizenbaum. Rebel at Work“

der gebürtige Berliner durch den von ihm programmierten Chatbot ELIZA, der als Vorläufer von ChatGPT und Sprachassistenten wie Siri und Alexa gilt. Dieses Programm hat Weizenbaum nicht nur zu einem Wegbereiter der KI gemacht, sondern zugleich auch zu einem ihrer schärfsten Kritiker:innen.

Weizenbaum, der als Jude in den 1930er Jahren mit seiner Familie vor dem Nationalsozialismus in die USA floh, erkannte früh die Ambivalenzen der Technologie, an der er gearbeitet hat. Er kritisierte vor allem die mythische Aufladung des Computers als übermenschliche, intelligente Maschine, die zu jener gruselnden

Faszination führt, mit der wir heute auf Innovationen in der KI blicken. Entgegen der damals wie heute weit verbreiteten Technikgläubigkeit war Joseph Weizenbaum davon überzeugt, dass es für viele unserer Probleme keine technischen Lösungen gebe. Die mythische Überhöhung von Computern mache uns vielmehr blind für die Möglichkeit, dass sie die Probleme, die sie lösen sollen, möglicherweise sogar noch verschärfen.

In diesem Heft greifen wir die Kritik von Weizenbaum auf und werfen mit ihm einen Blick auf aktuelle Diskussionen rund um das Thema Künstliche Intelligenz. Dabei wird nicht nur deutlich, welch lange Geschichte die so neu anmutenden Diskussionen haben, sondern auch, welche Positionen sich in der Vergangenheit durchgesetzt haben und unseren heutigen Blick auf Computer, Digitalisierung, Internet und KI prägen. Nicht zuletzt macht ein solcher Rückblick deutlich, wie wichtig es ist, sowohl historische als auch aktuelle Entwicklungen in diesen Bereichen kritisch zu hinterfragen.





TEIL 1 MYTHEN DER KÜNSTLICHEN INTELLIGENZ

WIE INTELLIGENT SIND COMPUTER?

Ende 2022 stellte das US-amerikanische Unternehmen OpenAI seine Entwicklung ChatGPT der Öffentlichkeit vor. Hinter dem etwas kryptischen Namen verbirgt sich ein Computerprogramm, mit dem man sich wie in einem Chat unterhalten kann. Das Besondere daran: ChatGPT antwortet ausführlich auf Fragen und kann nach entsprechender Aufforderung („Prompt“) sogar Gedichte verfassen, Aufsätze schreiben oder Witze erzählen. Fasziniert von den Fähigkeiten dieses Chatbots, der in kürzester Zeit Texte produziert, für die ein Mensch viel länger bräuchte, überbieten sich seither die Prognosen dazu,

was diese Technologie in Zukunft alles verändern, ja revolutionieren wird.

Das „GPT“ in ChatGPT steht für „Generative Pre-trained Transformer“ und verweist auf eine neue Gattung künstlich intelligenter Systeme: die generative KI. Im Gegensatz zu bisherigen KI-Systemen, die vor allem zur Mustererkennung und für automatisierte Entscheidungen eingesetzt werden (etwa bei Spamfiltern oder selbstfahrenden Autos), scheint das Intelligente an der generativen KI in ihrer Kreativität zu liegen. So gibt es KI-Anwendungen, mit denen sich Kunstwerke oder Musikstücke erstellen lassen. Dazu

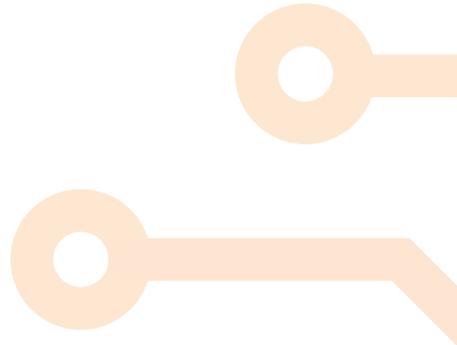
werden sie mit großen Mengen von Texten, Bildern, Videos oder Musik darauf trainiert, wie wahrscheinlich bestimmte Wort-, Pixel- oder Tonkombinationen sind und wie sie diese Wahrscheinlichkeiten in neue Kombinationen umwandeln können – ein Verfahren, das als „maschinelles Lernen“ bezeichnet wird.

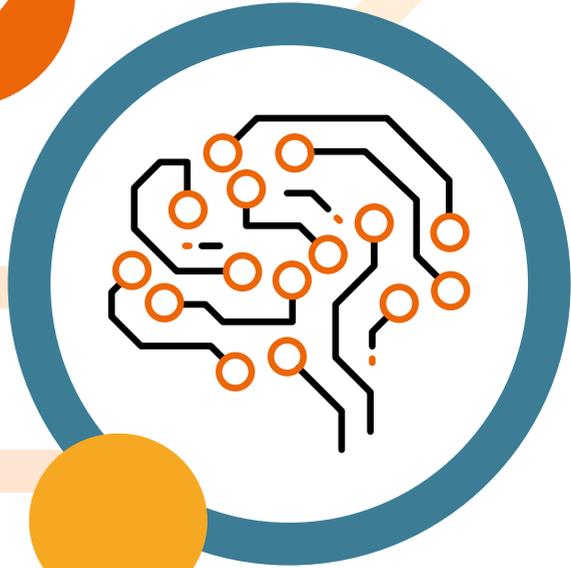
Doch wie kreativ sind diese KI-Anwendungen? Kritische Stimmen meinen, generative KI sei weder kreativ noch innovativ, da sie nur Bekanntes reproduziere, aber nichts völlig Neues hervorbringen könne. Andere wiederum argumentieren, dass dies bei Menschen auch nicht viel anders sei. Mitte 2022 behauptete ein Google-Mitarbeiter, die von ihm betreute KI LaMDA habe ein Bewusstsein entwickelt und befinde sich bereits auf dem Entwicklungsstand eines Kindes. Auch wenn ihm nur wenige glaubten, so fasziniert uns doch die Möglichkeit, dass Computer in absehbarer Zeit in der Lage sein könnten, eigenständig zu denken und dann bald darauf intelligenter zu werden als wir Menschen.

So neu diese Überlegungen angesichts der jüngsten Innovationen in der KI auch erscheinen mögen, ihre Diskussion reicht weit in die Geschichte zurück. Lange bevor der erste Computer gebaut wurde, spekulierte man schon darüber, was solche Rechenmaschinen einmal leisten könnten. Im Jahr 1843 prognostizierte die englische Mathematikerin Ada Lovelace bereits, dass Maschinen eines Tages „Musikstücke von beliebiger Komplexität und beliebigem Umfang komponieren“ könnten. Und auch der Begriff „Künstliche Intelligenz“ ist schon älter: 1955 prägte ihn der US-Forscher John McCarthy. Er bezeichnete damit ein neues Forschungsfeld, das untersucht, „wie man Maschinen dazu bringen kann, Sprache zu benutzen, Abstraktionen und Konzepte zu bilden, Probleme zu lösen, die heute dem Menschen vorbehalten sind, und sich selbst zu verbessern“ (siehe Infokasten „Künstliche Intelligenz“).

Die Frage, ob Computer eines Tages wie Menschen denken, verstehen und handeln können, wird seither heftig

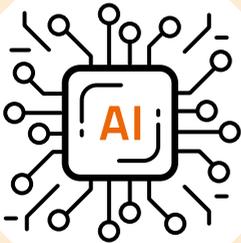
diskutiert. Auf der einen Seite gibt es Menschen, die davon ausgehen, dass das Gehirn eine Art Computer ist, der sich aber nur sehr langsam weiterentwickeln kann und deshalb zwangsläufig irgendwann von immer leistungsfähigeren Computern eingeholt wird. Andere Stimmen, wie die von Joseph Weizenbaum, sehen dies sehr kritisch und erinnern daran, dass ein Computer, nur weil er intelligent zu sein scheint, noch lange nicht intelligent ist. Wer das verwechselt, sitzt einem Mythos auf.





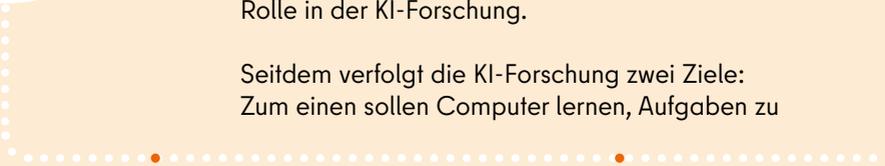


KÜNSTLICHE INTELLIGENZ



Der Begriff „Künstliche Intelligenz“ („Artificial Intelligence“) taucht erstmals 1955 in einer Projektskizze auf, mit der sich der Informatiker John McCarthy und einige seiner Kollegen um Geld für eine Konferenz am Dartmouth College in den USA bewarben. Auf dieser „Dartmouth Conference“, die heute als wegweisend gilt, widmeten sie sich im folgenden Jahr Themen wie Automatisierung, maschinelle Sprachverarbeitung, neuronale Netze und maschinelles Lernen. Diese Themen spielen auch heute noch eine wichtige Rolle in der KI-Forschung.

Seitdem verfolgt die KI-Forschung zwei Ziele: Zum einen sollen Computer lernen, Aufgaben zu

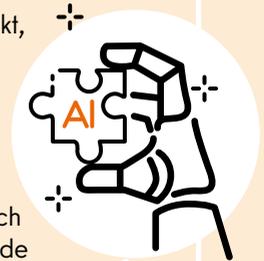


lösen, die bisher nur Menschen lösen konnten; zum anderen will man das menschliche Denken besser verstehen, indem man es nachzubilden versucht. In beiden Fällen geht es also darum, menschliches Denken zu simulieren.

Dazu werden grob zwei Ansätze verfolgt: Die „symbolische KI“ versucht, die Regeln des logischen Denkens nachzuprogrammieren. Dieser Ansatz war zu Beginn der KI-Forschung sehr populär und wird heute als „good old fashioned AI“ (GOFAI) bezeichnet. Seit den 1980er Jahren wird verstärkt der Ansatz der „konnektiven KI“ verfolgt, bei dem Verknüpfungen über die Berechnung von Wahrscheinlichkeiten hergestellt werden. Dabei wird mit „künstlichen neuronalen Netzen“ gearbeitet, die sich mit

jedem neuen Input verändern und so „lernen“.

Die Vision vieler KI-Forscher:innen ist es, eine „starke KI“ zu entwickeln, die selbstständig denken und lernen kann. Sollte dies eines Tages gelingen, so ihre Vorstellung, würde es nicht mehr lange dauern, bis die „Singularität“ erreicht wäre, also der Zeitpunkt, an dem die KI den Menschen überlegen sei. Diese KI könnte dann die „letzte Erfindung“ des Menschen sein, da sie von nun an alles Neue erfinden würde. Bislang ist das aber noch Science-Fiction. Entwickelt wurde bisher nur „schwache KI“, die auf spezifische Aufgaben beschränkt ist, wie selbstfahrende Autos, Personalisierung von Suchergebnissen, Text-, Bild- und Spracherkennung, maschinelle Übersetzung und Chatbots.



DER ELIZA-EFFEKT



Dass Menschen den Eindruck haben, ein Computer sei so intelligent wie ein Mensch, weil man gut mit ihm kommunizieren kann, wird „ELIZA-Effekt“ genannt. Die Bezeichnung geht auf ein Computerprogramm zurück, das Joseph Weizenbaum ab 1964 am Massachusetts Institute of Technology (MIT), einer der führenden Technikhochschulen in den USA, entwickelt hat: ELIZA. Dieses Programm gilt als einer der ersten Chatbots und damit als früherer Vorläufer von ChatGPT. Allerdings konnte ELIZA damals noch keine langen Texte produzieren, sondern war ein Chatprogramm, mit dem man per Texteingabe hin- und herschreiben konnte.

Weizenbaum verfolgte, wie die meisten KI-Forscher zu jener Zeit, den Ansatz der symbolischen KI, bei dem es darum ging, nachvollziehbare Regeln zu programmieren, um ein künstliches Denken des Computers zu simulieren. Da es jedoch viel zu aufwändig war, für alle möglichen Szenarien des Alltags Regeln zu definieren, beschränkten sich die KI-Forscher damals auf bestimmte Szenarien und Anwendungsfälle.

Für ELIZA wählte Weizenbaum das Szenario einer Psychotherapie. Das Prinzip war einfach: Das Programm stellte Fragen und die Nutzer:innen tippten ihre Antworten ein. ELIZA suchte

diese Antworten dann auf bestimmte Schlüsselwörter ab, stellte entsprechend programmierte Nachfragen, formulierte die Antworten zu neuen Fragen um oder bat einfach darum, mehr zu erzählen. Damit umging Weizenbaum geschickt die größte Herausforderung des symbolischen KI-Ansatzes, Computern so viel Wissen über die Welt einzuprogrammieren, dass sich Menschen sinnvoll mit ihnen unterhalten können. Denn ELIZA musste nicht viel wissen, sondern nur Fragen stellen können.

Der „ELIZA-Effekt“ bestand nun darin, dass sich viele Nutzer:innen mit ELIZA unterhielten, als hätten sie es mit einem Menschen zu tun. Insofern kann ELIZA als eine frühe Anwendung des „Turing-Tests“ des englischen Mathematikers Alan Turing verstanden werden, nach dem ein Computer dann als intelligent gelten kann, wenn Menschen ihn für intelligent halten (siehe Infokasten „Turing-Test“). In späteren Texten und Interviews berichtete Weizenbaum davon, wie überrascht er gewesen sei,

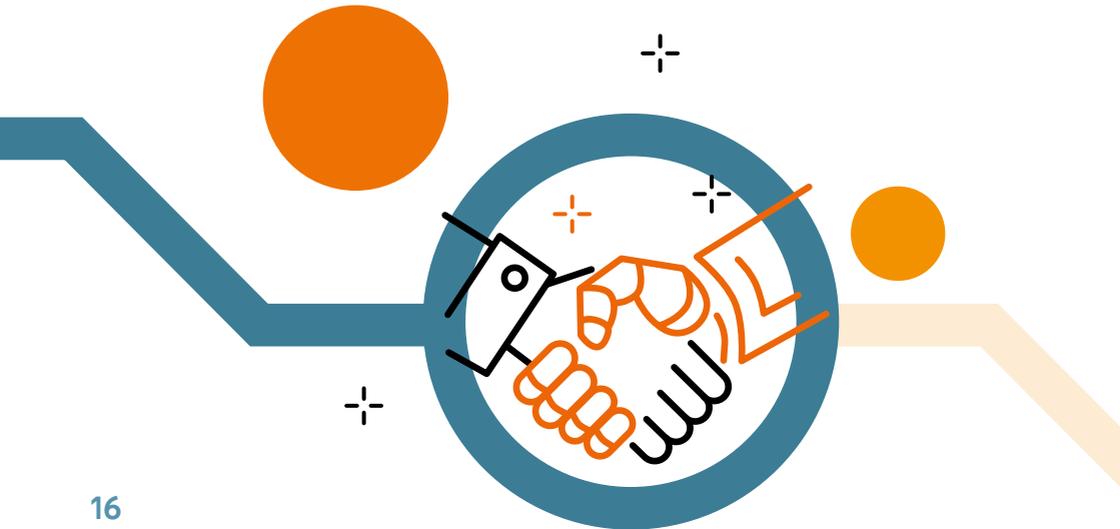
dass Menschen seinem Programm, das auf wenigen, recht einfachen Regeln beruhte, Intelligenz und Einfühlungsvermögen zuschrieben. Obwohl wahrscheinlich niemand ernsthaft glaubte, dass ELIZA wirklich ein intelligentes Wesen war, fühlten sich einige Menschen offenbar von dem Programm verstanden, so Weizenbaum.

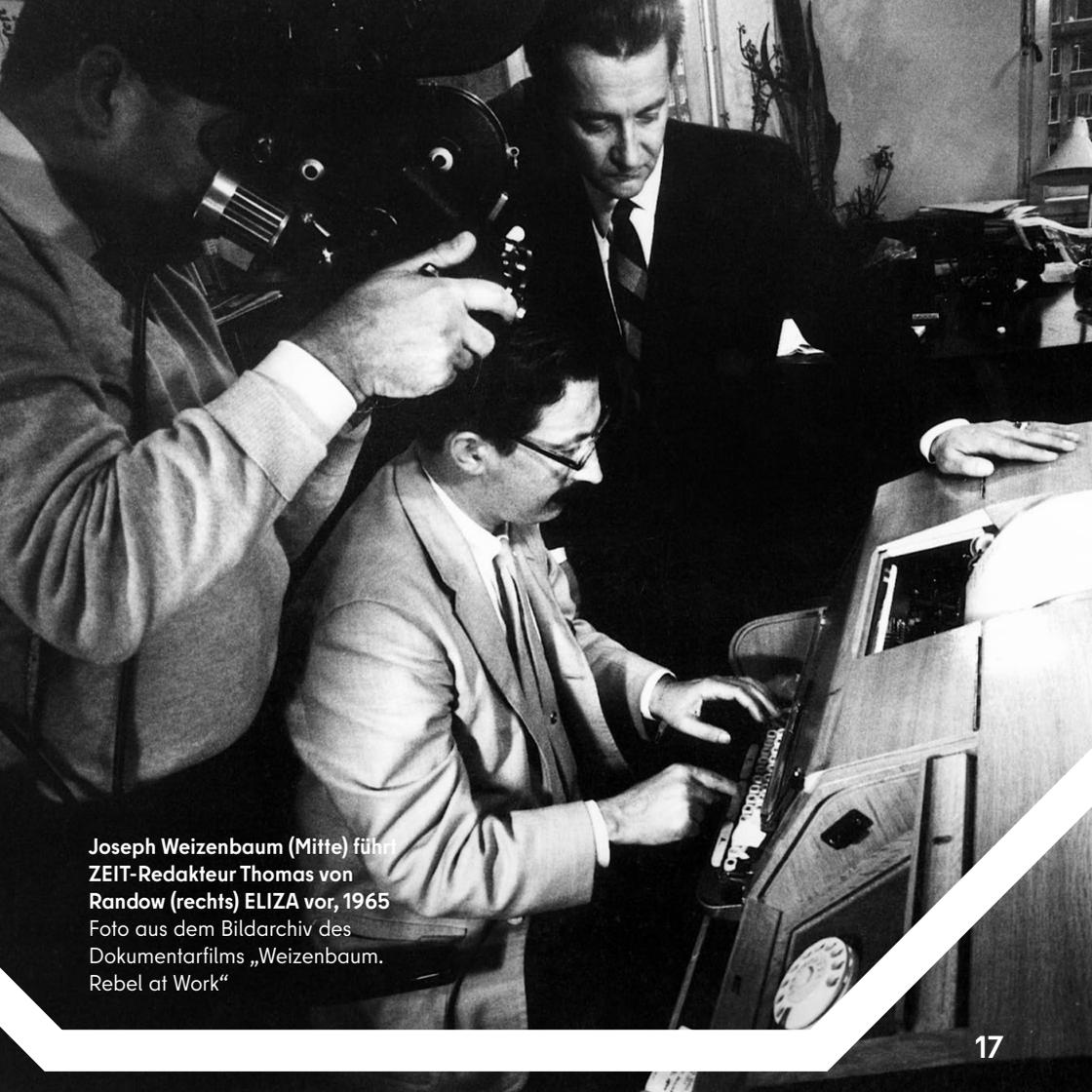
Eine solche Zuschreibung menschlicher Eigenschaften auf Gegenstände oder nichtmenschliche Lebewesen wird „Anthropomorphismus“ genannt, was „Vermenschlichung“ bedeutet. Beispiele dafür sind das Haus, das zu lächeln scheint, die Katze, die elegant aussieht, weil man ihr einen Hut aufgesetzt hat, oder die Lampe, die traurig ihren Kopf senkt. Wir Menschen sind gut darin, in der Welt um uns herum uns selbst zu erkennen – und das tun wir auch mit Computern und anderen Maschinen: Humanoide Roboter, die wir wegen ihrer großen Kinderaugen süß finden, Navigationssysteme, die wir genervt anfauchen, weil sie uns wiederholt darauf

hinweisen, dass wir in die falsche Richtung fahren, oder Sprachassistenten, bei denen wir uns höflich bedanken.

Was Joseph Weizenbaum damals aber noch mehr überraschte - und verärgerte - als die Vermenschlichung von ELIZA, war, dass einige Psychotherapeuten damals öffentlich darüber nachdachten, solche Programme künftig für Psychotherapien einzusetzen. Weizenbaum sah darin eine Respektlosigkeit gegenüber

hilfsbedürftigen Menschen, die man an einen Computer verweisen wollte, statt sich um sie zu kümmern. Er ahnte, dass die Vermenschlichung von Computern eine gefährliche Kehrseite hat: die Entmenschlichung von Menschen. Denn wenn wir Computern menschliche Eigenschaften zuschreiben, so seine Sorge, dann beginnen wir Menschen und Computer zu vergleichen und in Konkurrenz zueinander zu setzen.





Joseph Weizenbaum (Mitte) führt
ZEIT-Redakteur Thomas von
Randow (rechts) ELIZA vor, 1965
Foto aus dem Bildarchiv des
Dokumentarfilms „Weizenbaum.
Rebel at Work“

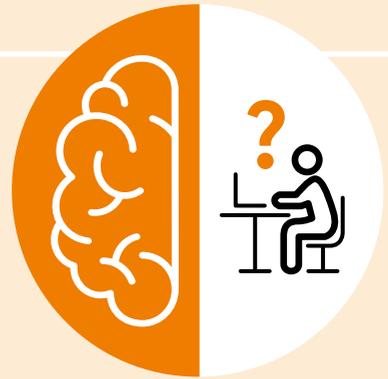


TURING-TEST

Mit seinem Aufsatz „Computing Machinery and Intelligence“ (deutscher Titel: „Kann eine Maschine denken?“) legte Alan Turing im Jahr 1950 einen wichtigen Grundstein für das spätere Verständnis von Künstlicher Intelligenz. Turing fragte sich, wann ein Computer als intelligent gelten könne. Seine Antwort: Es komme nicht darauf an, ob ein Computer tatsächlich denken oder verstehen könne, entscheidend sei vielmehr, dass wir Menschen nicht mehr unter-

scheiden könnten, ob wir es mit einem Menschen oder einem Computer zu tun hätten.

Um zu testen, wie gut ein Computer einen Menschen imitieren kann, formulierte er das „Imitation Game“ („Nachahmungsspiel“), das heute als „Turing-Test“ bekannt ist. Der Aufbau sieht wie folgt aus: Ein Mensch, der:die „Fragesteller:in“, sitzt vor einem Computer und ist darüber mit zwei ihm unbekanntem Gesprächspartner:innen verbunden:

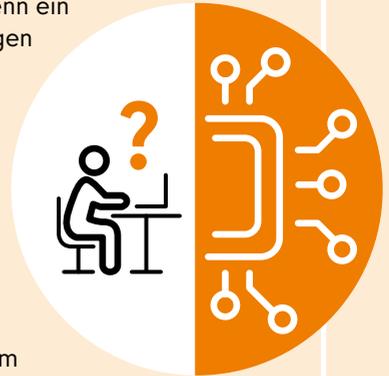


einem anderen Menschen und einem Computer. Die Aufgabe des:der Fragesteller:in ist es, den beiden unbekanntem Gesprächspartner:innen so viele Fragen zu stellen, bis er zu wissen glaubt, wer Mensch und wer Computer ist. Gelingt es ihm nicht, beide eindeutig zu identifizieren, hat der Computer den Test bestanden.

Der Turing-Test ist somit eine Variante des „Ententests“, bei dem von der äußeren Erscheinung eines Objekts darauf geschlossen wird, worum es sich handelt. Diese Methode ist unter dem Ausspruch bekannt geworden: „Wenn es aussieht wie eine Ente, schwimmt wie eine Ente und quakt wie eine Ente, dann ist es

wahrscheinlich eine Ente.“ Beim Turing-Test würde es entsprechend heißen: „Wenn ein Computer auf Fragen antwortet wie ein Mensch, dann ist er wahrscheinlich so intelligent wie ein Mensch.“

Obwohl der Turing-Test oft kritisiert wurde – zum Beispiel dafür, dass er eher messen würde, wie leicht sich Menschen täuschen lassen als wie intelligent ein Computer sei – hat er seither einen großen Einfluss auf das Nachdenken über Künstliche Intelligenz gehabt.



DAS MENSCHENBILD DER KI

Die Idee, Menschen mithilfe von Computerprogrammen zu therapieren, ist seit ELIZA nicht mehr verschwunden. Heute finden sich in unseren App-Stores zahlreiche KI-gestützte Selbsthilfe-Apps, die psychotherapeutische oder medizinische Beratung anbieten. Ihr Einsatz ist dabei nach wie vor umstritten: Kritiker:innen befürchten, dass hilfsbedürftige Menschen auf diese Weise sich selbst überlassen werden, anstatt ihnen professionelle Hilfe und menschliche Zuwendung zukommen zu lassen. Befürworter:innen solcher Apps halten dagegen, dass diese Angebote eine gute Ergänzung zu bestehenden Therapien seien. Angesichts langer Wartezeiten bei

der Suche nach Therapieplätzen seien sie immer noch besser als gar keine Hilfe.

Auch wenn viele der Personen, die solche Therapie-Apps entwickeln und anbieten, dies in der ehrlichen Absicht tun dürften, anderen Menschen zu helfen, liegt solchen Angeboten ein kritikwürdiges Menschenbild zugrunde. Denn neben der Frage nach der Wertschätzung und dem Respekt gegenüber den Menschen, denen mit solchen Anwendungen geholfen werden soll, ist damit auch die Annahme verbunden, dass diese Programme – zumindest in einem gewissen Rahmen – zwischenmensch-

liche Hilfe und Fürsorge übernehmen könnten.

Diese Annahme, dass KI den Menschen in zwischenmenschlichen Angelegenheiten ebenbürtig oder gar überlegen sein könnte, führe, so kritisiert Weizenbaum, zu einer „absurden Konkurrenz“ zwischen Menschen und Computern. Denn unabhängig von der technischen Machbarkeit in einer zumeist noch unbestimmten Zukunft hat bereits die Vorstellung, dass dies möglich sein könnte, weitreichende Folgen für unsere Gegenwart.

Beobachten lässt sich dies etwa am Beispiel der Paketzustellung: Seit vielen Jahren ist hier die Rede davon, dass Pakete bald von autonomen Robotern und Drohnen zugestellt würden – und zwar in kürzester Zeit. Tatsächlich hat sich die Lieferzeit seither drastisch verkürzt, jedoch nicht wegen der angekündigten Automatisierung. Vielmehr übt bereits die Ankündigung, diese Arbeit in Zukunft von Robotern erledigen zu lassen, so

viel Druck auf die Beschäftigten in der Paketzustellung aus, dass sie ihre Arbeit in immer kürzerer Zeit erledigen müssen. Um nicht durch eine Maschine ersetzt zu werden, versuchen sie, selbst wie eine Maschine zu arbeiten.

Und dieses Phänomen ist nicht auf die Paketzustellung beschränkt: Die Rede von Pflege-, Service- oder Sex-Robotern, von kreativen Computern und KI-Freund:innen konfrontiert uns alle mit der unangenehmen Vorstellung, vielleicht irgendwann durch eine Maschine ersetzt werden zu können. Das Mensch-Maschine-Verhältnis kehrt sich damit um: Plötzlich geht es nicht mehr um die Frage, ob Computer das können, was bisher Menschen vorbehalten war, sondern darum, ob Menschen so leistungsfähig sein können wie Computer. Die Maschine wird so zum entscheidenden Maßstab, an dem wir Menschen uns zunehmend ausrichten müssen (siehe Infokasten „Sind Computer die besseren Menschen?“).

Joseph Weizenbaum lehnte den Vergleich zwischen Menschen und Maschinen deshalb rundheraus ab. Er widersprach entschieden, wenn etwa der Entwicklungsstand einer KI mit dem eines Kindes verglichen oder das menschliche Gehirn als „meat machine“ („Maschine aus Fleisch“) bezeichnet wurde. Die Vorstellung, dass Computer eines Tages menschlich sein könnten, würde im Umkehrschluss bedeuten, dass Menschen auch nur Maschinen seien – und diese

Vorstellung hielt Weizenbaum für gefährlich: „Wir können aus der Geschichte lernen, welche Rolle das Menschenbild in den Verbrechen der Vergangenheit spielte. In der NS-Zeit stellte man Juden als Ungeziefer dar – eine Metapher, die den Massenmord legitimierte.“ Eine ähnliche Entmenschlichung schreibt er dem Mensch-Maschinen-Vergleich zu: „Diese Metaphern vernichten die Ehrfurcht vor dem Menschen; sie lassen sein mögliches Ende erträglich erscheinen.“





SIND COMPUTER DIE BESSEREN MENSCHEN?

Der Philosoph Günther Anders beschreibt in seinem Buch „Die Antiquiertheit des Menschen“ das Gefühl einer Scham, das uns überkomme, wenn wir uns mit Maschinen vergleichen – oder mit ihnen verglichen werden. Gemeint ist damit ein Unterlegenheitsgefühl, das sich einstellt, wenn wir unsere Vergänglichkeit und Fehleranfälligkeit mit scheinbar perfekten und unsterblichen Maschinen vergleichen.

Dieses Gefühl könne zu dem Wunsch führen, selbst wie eine Maschine zu sein, indem man beispielsweise versucht, eine Tätigkeit perfekt zu beherrschen, Aufputzmittel nimmt, um länger arbeiten zu können, oder darauf hinarbeitet, unsterblich zu werden. „Der Mensch desertiert ins Lager seiner Geräte“, schreibt Anders über diese Form der Selbstdisziplinierung.

Joseph Weizenbaum hat seine Kollegen aus der KI-Forschung dafür kritisiert, dass sie den Vergleich zwischen Menschen und Computer auf die Spitze trieben: Hinter ihrem Streben nach einer Künstlichen Intelligenz, die die menschliche übertreffen sollte, erkannte er eine Verachtung für das menschliche Leben und den größtenwahnsinnigen Versuch, Gott als schlechten Ingenieur zu entlarven und ihn durch die Erschaffung eines „besseren

Menschen“ zu übertreffen. Die Tatsache, dass die KI-Forschung zu Weizenbaums Zeit fast ausschließlich von Männern betrieben wurde, führte ihn zudem zu der Überlegung, dass hier auch eine Art Gebärneid vorliegen könnte. Dieses Konzept geht auf die Psychoanalytikerin Karen Horney zurück und besagt, dass Männer versuchen, Frauen in verschiedenen Bereichen zu übertreffen, um ihre Unfähigkeit, Leben zu geben, zu kompensieren.







TEIL 2 KI UND GESELLSCHAFT

KANN KI GESELLSCHAFTLICHE PROBLEME LÖSEN?

In Debatten über neue Technologien – seien es Smartphones, das Internet, soziale Medien oder KI – stehen sich in der Regel zwei Annahmen über mögliche Auswirkungen dieser Technologien auf die Gesellschaft gegenüber: Die einen betonen die Potenziale dieser Technologien und beschwören eine Zukunft, in der durch Technik alles besser wird. Andere haben die Risiken im Blick und fragen, welche Gefahren eine Technologie auch mit sich bringt. Science-Fiction stellt diese Ambivalenz oft eindrucksvoll dar: Sie liefert uns beeindruckende Zukunftswelten, konfrontiert uns aber auch mit dystopischen Visionen davon, welche

negativen Folgen mit diesen Technologien einhergehen können.

Im alltäglichen Umgang mit Technik sind wir von solchen utopischen und dystopischen Szenarien weit entfernt. Technologie erscheint uns eher als Werkzeug, mit dem sich bestimmte Aufgaben einfacher und schneller erledigen lassen. Das gilt auch für KI: Mit ihrer Hilfe lassen sich Texte übersetzen, sie macht unsere Internet-suche präziser, hält unsere E-Mail-Postfächer frei von Spam und macht unsere Mobiltelefone durch Gesichtserkennung sicherer. Manche trauen ihr aber noch mehr zu: Laut dem KI-Aktionsplan der

deutschen Bundesregierung von 2023 soll KI beispielsweise einen Beitrag zum Umwelt- und Klimaschutz leisten.

Der Versuch, Probleme durch den Einsatz von Technologie zu lösen, wird als „technological fix“ bezeichnet. Der Begriff wurde in den 1960er Jahren von dem Physiker Alvin Weinberg geprägt. Seine Idee war, dass soziale, politische oder wirtschaftliche Probleme dadurch gelöst werden können, dass man sie in technologische Probleme übersetzt, weil sie so einfacher zu handhaben seien. Ein Beispiel ist die KI-gestützte Beseitigung von Hassrede im Internet: Ein gesellschaftliches Problem wie diskriminierende und beleidigende Äußerungen wird technisch bearbeitet, indem solche Inhalte von einer KI automatisch identifiziert und gelöscht werden.

Doch ist das Problem damit wirklich gelöst? Eher nicht. Denn selbst wenn die KI gut funktioniert und diskriminierende Inhalte zuverlässig löscht – die Gründe

dafür, warum Menschen andere Menschen abwertend behandeln, sind damit noch nicht aus der Welt. Diejenigen, deren Inhalte von einer KI gelöscht werden, suchen sich neue Wege, um ihre Weltsicht zum Ausdruck zu bringen – etwa indem sie ihre Hassbotschaften so formulieren, dass die KI sie nicht mehr als solche erkennen kann, indem sie auf andere Plattformen ohne KI ausweichen oder indem sie ihren Hass auf die Straße tragen und aus Worten Taten werden lassen.

Die technische Lösung eines Problems schafft also neue Probleme, die wir dann mit neuen Technologien zu lösen versuchen, die wiederum neue Probleme schaffen und so weiter. Dass wir Technologien dennoch als Problemlöser sehen, hängt mit dem weit verbreiteten Glauben zusammen, dass sie Werkzeuge sind, die speziell zur Lösung eines Problems entwickelt wurden – und dass es im Grunde kein Problem gibt, das nicht mit (der richtigen) Technologie gelöst werden könne.

Der Technikkritiker Evgeny Morozov hat diesen Glauben „Solutionismus“ genannt. Er beschreibt damit die Überzeugung, mit Technologie ließe sich die Welt verbessern (siehe Infokasten „Kalifornische Ideologie“). Dabei sei es zum Teil umgekehrt: Eine Technologie werde entwickelt und erst dann werde nach einem Problem gesucht, für das sie eine Lösung sein könnte.

Es ginge zu weit, zu behaupten, dass gesellschaftliche Probleme vor allem auf den Einsatz von Technologie zurückzuführen seien; doch ist die Frage, bei welchen Problemen dies der Fall ist, berechtigt. Für einen kritischen Umgang mit Technologie ist es wichtig, sich immer wieder zu fragen, welche Probleme durch ihren Einsatz entstehen können, die es ohne sie vielleicht nicht gäbe.





KALIFORNISCHE IDEOLOGIE

Im Jahr 1996 veröffentlichten die Sozialwissenschaftler Richard Barbrook und Andy Cameron einen Artikel mit dem Titel „The Californian Ideology“, in dem sie sich kritisch mit einer im Silicon Valley, dem südlichen Teil der San Francisco Bay Area in Kalifornien, weit verbreiteten Weltanschauung auseinandersetzten. Das Gebiet um Städte wie Palo Alto, Cupertino, Santa Clara und San José, in dem sich auch die Stanford University befindet, war schon damals einer der wichtigsten Standorte der IT- und Hightech-Industrie mit Unternehmen wie IBM, Microsoft, Intel,

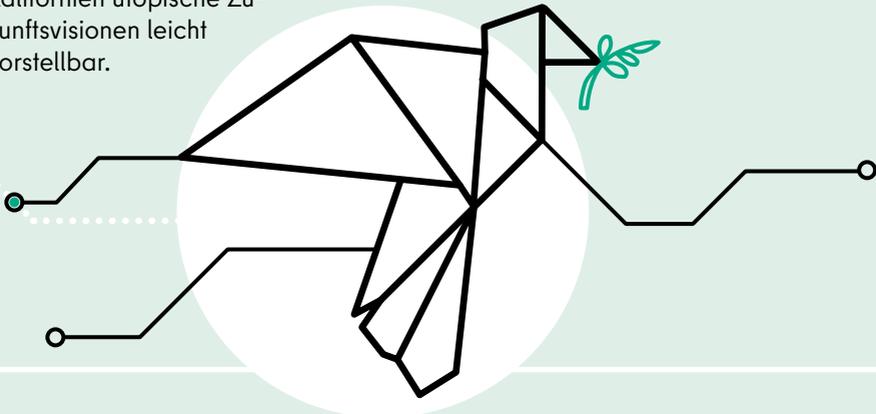
Yahoo und Apple. Heute sind dort zum Beispiel auch Google und Facebook ansässig.

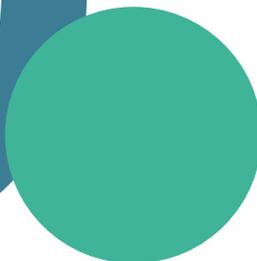
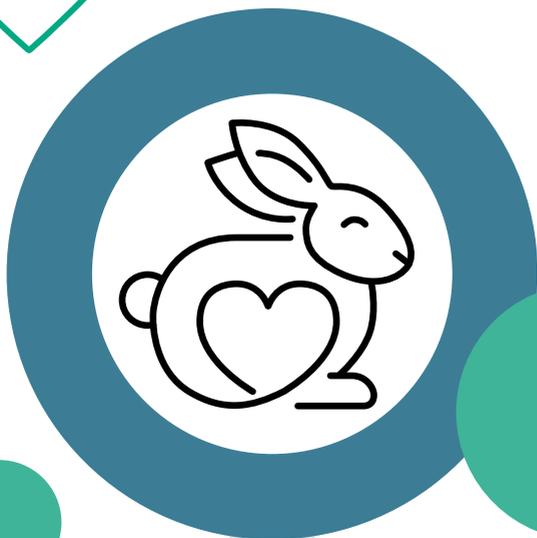
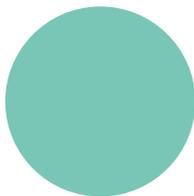
Barbrook und Cameron analysieren in ihrem Text, wie sich hier ein neuer Glaube an die weltverbessernde Kraft freier Märkte und technologischer Innovationen entwickelte, der selbst widersprüchliche Haltungen wie den kapitalistischen Marktliberalismus und die kapitalismuskritischen Ideale der Hippie-Bewegung wie Naturverbundenheit und Selbstverwirklichung miteinander verband – etwa zur Idee der Selbstoptimierung.



Dass diese Mischung ausgerechnet in Kalifornien entstand, hat nach Ansicht der beiden Autoren mehrere Gründe: Zum einen galt die Bay Area seit den Protesten gegen den Vietnamkrieg als Zentrum des Anti-Establishments und der Neuen Linken. Zum anderen waren die vielen Technologieunternehmen, die von der US-Regierung unter anderem für ihre militärische Forschung gefördert wurden, wichtige Arbeitgeber. Zudem machten das milde Klima und die hohe Lebensqualität in Kalifornien utopische Zukunftsvisionen leicht vorstellbar.

Die Vorstellung, dass neue Technologien wie das Internet zu einer besseren Welt führen würden, in der soziale Ungleichheit und staatliche Repression der Vergangenheit angehören würden, wurde Anfang der 1990er Jahre zu einem einflussreichen Narrativ, das bis heute zur Abwehr von staatlicher Kontrolle und Regulierung genutzt wird. Vor diesem Hintergrund ist „The Californian Ideology“ zu einem wichtigen Referenztext der Technologiekritik geworden.





DIE GRENZEN VON KI

Joseph Weizenbaum hat der Glaube, Computer seien Problemlösungsmaschinen, viel beschäftigt. Aus seiner Sicht ist dieser Glaube der Ausdruck eines tiefer liegenden Phänomens, das er den „Imperialismus der instrumentellen Vernunft“ nannte. Mit „instrumenteller Vernunft“ meint er – in Anlehnung an den Philosophen Max Horkheimer, der diesen Begriff geprägt hat – ein Denken in den Kategorien Zweck und Mittel, oder eben: Problem und Lösung. Im Gegensatz zur praktischen Vernunft, bei der sich das Denken daran orientiert, welches Handeln richtig ist, orientiert sich das instrumentelle Denken an der Frage, welches Mittel (oder Instrument) für einen bestimmten Zweck geeignet ist.

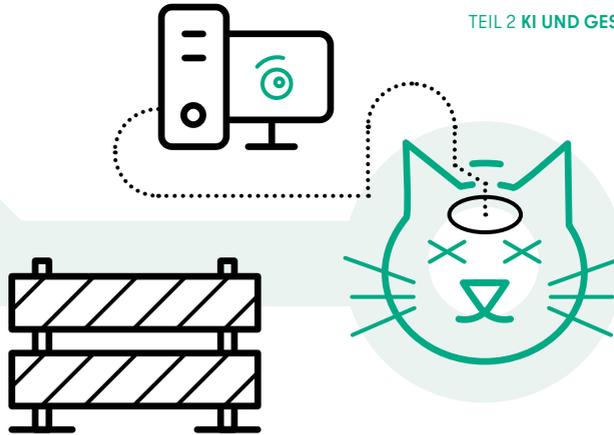
Dieses Denken ist im Ingenieurwesen, in den Naturwissenschaften und auch in der Informatik weit verbreitet: Welches Material brauche ich, damit die Brücke möglichst lange hält? Mit welcher Methode kann ich die Oberflächen-temperatur der Sonne messen? Und was muss ich an der Software ändern, damit das selbstfahrende Auto keine Unfälle mehr baut? All dies sind Beispiele für instrumentelles Denken. Die Fragen könnten aber auch lauten: Wie baue ich eine Atombombe? Sind Mäuse oder Affen besser für Tierversuche geeignet? Oder: Was muss ich tun, um mich in ein Bankkonto zu hacken? Instrumentelles Denken ist also zweckorientiert, ohne zu fragen, ob es sich um einen guten

Zweck oder um moralisch vertretbare Mittel handelt.

Mit „Imperialismus“ meint Weizenbaum, dass das instrumentelle Denken Einzug in unseren Alltag gehalten hat und unsere Sicht auf die Welt (und uns selbst) prägt. Die Errungenschaften der Naturwissenschaft und Technik – etwa das Fliegen, die Medizin oder das Internet – hätten uns von der instrumentellen Vernunft überzeugt, sodass wir unser Denken danach ausgerichtet hätten: Was muss ich tun, um Erfolg zu haben? Welches Foto bringt die meisten Likes? Bin ich gut genug für den Job? Oder kann das ein Computer besser als ich? Durch den großen Erfolg und die faszinierende Leistungsfähigkeit des Computers hätten wir, so Weizenbaum, eine „technologische Mentalität“ entwickelt, die nicht nur das instrumentelle Denken verstärke, sondern zugleich davon ausgehe, dass Technologie grundsätzlich das bessere Mittel sei – besser vor allem als Menschen.

Weizenbaum selbst lehnte diese Sichtweise ab. Sie führe nur dazu, dass Technologie auch in solchen Bereichen eingesetzt werde, wo sie mehr schade als nütze. Für ihn gab es daher zwei Grenzen für den Einsatz von Technologie, insbesondere für KI:

Die erste Grenze sah er in sozialen Problemen und zwischenmenschlichen Konflikten, für die es aus seiner Sicht keinen „technological fix“ gebe. Wenn zum Beispiel Schüler:innen Lernschwierigkeiten hätten, dann löse man dieses Problem nicht mit neuen Computern, die den Kindern beim Lernen helfen, weil die zugrundeliegenden Probleme in der Regel andere seien: etwa ein schlechtes Betreuungsverhältnis, überforderte Lehrkräfte, Probleme zu Hause, Mobbing durch andere Kinder oder ein zu hoher Leistungsdruck. Der Versuch, Lernschwierigkeiten technisch zu lösen, würde diese Probleme überdecken und sie dadurch möglicherweise noch verschlimmern.



Die zweite Grenze ist für Weizenbaum eine moralische: Es gebe Aufgaben, schreibt er, „zu deren Lösung keine Computer eingesetzt werden *sollten*, ungeachtet der Frage, ob sie zu deren Lösung eingesetzt werden *können*“. Als ein Beispiel für eine solche Aufgabe nennt er den Einsatz von KI-Anwendungen zur psychotherapeutischen Behandlung von Patient:innen, etwa durch Programme wie ELIZA. Da Computer

keine Biografie und keinen Körper hätten, verfügten sie nicht über eigene Erfahrungen und es fehle ihnen an der nötigen Menschlichkeit (siehe Infokasten „Uncanny Valley“). Als ein weiteres Beispiel nannte er Tierversuche, bei denen das Gehirn einer toten Katze mit einem Computer verbunden wurde, um dem Computer Sehfähigkeit zu verleihen. Für Weizenbaum war dies eine unethische und obszöne Grenzüberschreitung.

UNCANNY VALLEY

Der Begriff „Uncanny Valley“ („unheimliches Tal“) geht auf den japanischen Robotiker Masahiro Mori zurück. Unter dem Titel „The Uncanny Valley“ veröffentlichte dieser 1970 einen Artikel, in dem er die These aufstellte, dass sich unsere Wahrnehmung von menschenähnlichen Robotern ab einem bestimmten Grad der Menschenähnlichkeit schlagartig ändere: Während wir humanoiden Robotern zunächst empathisch gegenüberstünden, würden unsere Sympathien bei einer fast perfekten Menschenähnlichkeit, die nur noch kleine Unstimmig-

keiten aufweise, steil abfallen und die Roboter plötzlich unheimlich wirken.

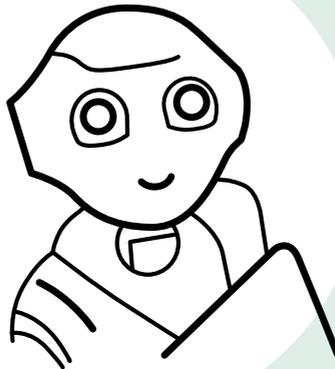
Als Beispiel nennt Mori einen Roboter, der für die Weltausstellung in Osaka mit 29 künstlichen Gesichtsmuskeln ausgestattet wurde, um das menschliche Lächeln zu simulieren. Bei der geringsten Abweichung in der Geschwindigkeit der Mimik schlage unsere Zuneigung zu diesem Roboter in Grusel um. Andere Beispiele für dieses Phänomen sind sich bewegende Puppen, Zombies, amputierte Körperteile, Leichen oder zu realitätsgetreue Prothesen.



Mori kommt zu dem Schluss, dass bei der Entwicklung von Robotern – solange es nicht möglich ist, keine „unheimlichen“ Roboter zu entwickeln – eine Ähnlichkeit mit Menschen vermieden und die Funktion des Roboters in den Mittelpunkt gestellt werden sollte. So würden Brillen auch nicht aussehen wie Augen, obwohl sie eine ihrer Funktionen übernehmen.

Bis Anfang der 2000er Jahre blieb Moris Aufsatz nahezu unbeachtet.

Erst mit dem Aufkommen neuerer Robotik und KI wurde der „Uncanny Valley“-Effekt wieder aufgegriffen und beispielsweise auf Pflegeroboter angewandt. Auch bei „körperlosen“ Robotern wie Chatbots wird ein „Uncanny Valley“-Effekt vermutet. In den vergangenen Jahren wurden verschiedene Experimente durchgeführt, um diesen Effekt empirisch zu messen und psychologisch zu erklären. Moris selbst lieferte weder einen empirischen Nachweis noch eine Erklärung für das Phänomen.



ARTIFICIAL INTELLIGENTSIA

Als Professor am renommierten MIT in Cambridge war Joseph Weizenbaum seit den 1960er Jahren ein etablierter Teil eben jenes Forschungsfeldes der KI, dem er ab den 1970er Jahren zunehmend kritisch gegenüberstand. Dies verlieh ihm die fachliche Autorität, die seinem 1976 erschienenen Buch „Computer Power and Human Reason“ (deutscher Titel: „Die Macht der Computer und die Ohnmacht der Vernunft“) eine große öffentliche Aufmerksamkeit verschaffte. Auf der anderen Seite bedeutete dies aber auch, dass sich Weizenbaum mit seiner Kritik gegen jene Kollegen wandte, mit denen er zusammenarbeitete. Man kann sich vorstellen, dass sie über seine Kritik nicht gerade erfreut waren.

Zu diesen Kollegen gehörten Marvin Minsky, der mit Weizenbaum am selben Institut arbeitete,





Joseph Weizenbaum (rechts) mit seinen
Fachkollegen Claude Shannon, John McCarthy
und Edward Fredkin (v. l. n. r.), 1968

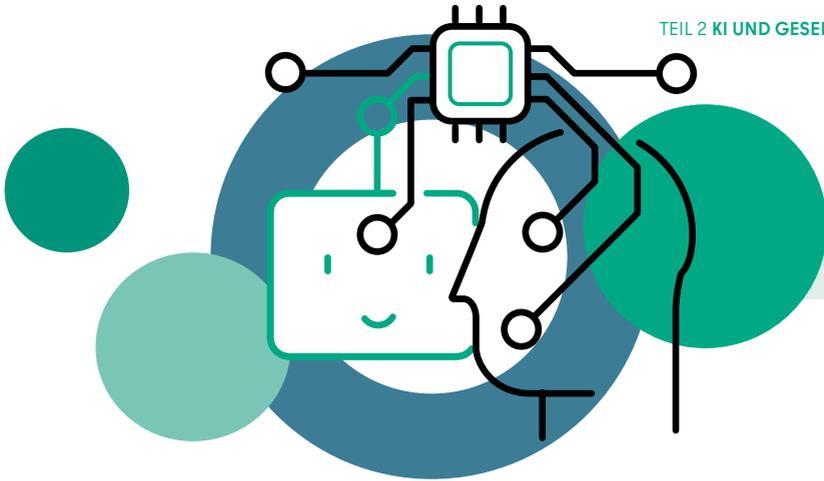
Foto aus dem Bildarchiv des Dokumentarfilms
„Weizenbaum. Rebel at Work“

John McCarthy, der Erfinder des Begriffs „Artificial Intelligence“, und der Psychiater Kenneth Colby, mit dem Weizenbaum während seiner Arbeit an ELIZA im Austausch gestanden hatte. Sie und weitere Kollegen aus der KI-Forschung, darunter auch der Nobelpreisträger Herbert A. Simon, bezeichnete Weizenbaum etwas spöttisch als „Artificial Intelligentsia“, also die KI-Intellektuellen. Sein Spott richtete sich dabei nicht gegen ihre wissenschaftlichen Verdienste in der Computerforschung, sondern gegen die Prognosen, die diese Männer in der Öffentlichkeit darüber abgaben, zu was KI in naher Zukunft alles fähig wäre.

Bereits im Gründungsdokument der „Künstlichen Intelligenz“ von 1955 kündigten John McCarthy und seine Mitstreiter an, dass sie innerhalb von nur zwei Monaten herausfinden könnten, wie man Computer dazu bringen könnte, „Probleme zu lösen, die bisher dem Menschen vorbehalten waren, und sich selbst zu verbessern“. Ende der 1960er Jahre prognostizierte Herbert A.

Simon, dass Computer schon bald das Verhalten aller Menschen erklären und simulieren könnten. Und Marvin Minsky sagte 1970 voraus, dass es in weniger als zehn Jahren eine Superintelligenz geben werde, Computer dann die Kontrolle übernehmen und wir Menschen froh sein könnten, wenn sie uns als Haustiere hielten. Schließlich prophezeite der Robotiker Hans Moravec in seinem Buch „Mind Children“ (1988), dass Menschen bald ihr Bewusstsein auf einen Computer laden und so digital unsterblich werden könnten (siehe Infokasten „Transhumanismus“).

Bislang ist keine dieser Vorhersagen eingetroffen. Dennoch hören wir ähnliche Prognosen auch von heutigen Vertreter:innen der KI-Forschung: Sam Altman, der Geschäftsführer von OpenAI, ist sich zum Beispiel sicher, dass sein Unternehmen kurz davorsteht, eine Superintelligenz zu erschaffen. Ray Kurzweil, Technischer Direktor bei Google, antwortete einst in einem Interview auf die Frage, ob Gott existiere, mit Blick auf



die technologische Entwicklung: „Well, I would say: not yet“ („Nun, ich würde sagen: noch nicht“). Und Geoffrey Hinton, langjähriger KI-Chefentwickler bei Google, trat im Mai 2023 von seinem Posten zurück, weil er glaubte, dass KI zu einer existenziellen Bedrohung werde. Die Zukunft der KI bewegt sich also seit jeher zwischen einem utopischem Heilsversprechen und dem Untergang der Menschheit.

Weizenbaum würde Altman, Kurzweil, Hinton und Co. vermutlich die „neue

Artificial Intelligentsia“ nennen. Auch wenn die KI-Forschung in den letzten 50 Jahren große Fortschritte gemacht hat und inzwischen zu einem großen Teil aus den öffentlichen Universitäten in kommerzielle Technologieunternehmen abgewandert ist, so zeigen die verblüffenden Parallelen in den Prognosen über die Potenziale der KI, dass diese Technologie, seit es sie gibt, immer auch eine Projektionsfläche für unsere Träume, Hoffnungen und Zukunftsängste war.



TRANSHUMANISMUS

Die Vorstellung, der Mensch könne über sich hinauswachsen und unsterblich werden, reicht weit in die Geschichte zurück. Sie findet sich etwa im volkstümlichen Mythos des Jungbrunnens oder im Konzept des „Übermenschen“ bei Friedrich Nietzsche. Eine jüngere, technikaffine Variante stellt die philosophische Denkrichtung des Transhumanismus dar. Der Begriff wurde 1957 von dem britischen Biologen und Eugeniker Julian Huxley, dem Bruder des Schriftstellers Aldous Huxley, geprägt und dient seither als Selbstbezeichnung für eine soziale Bewegung, die eine Erweiterung der menschlichen

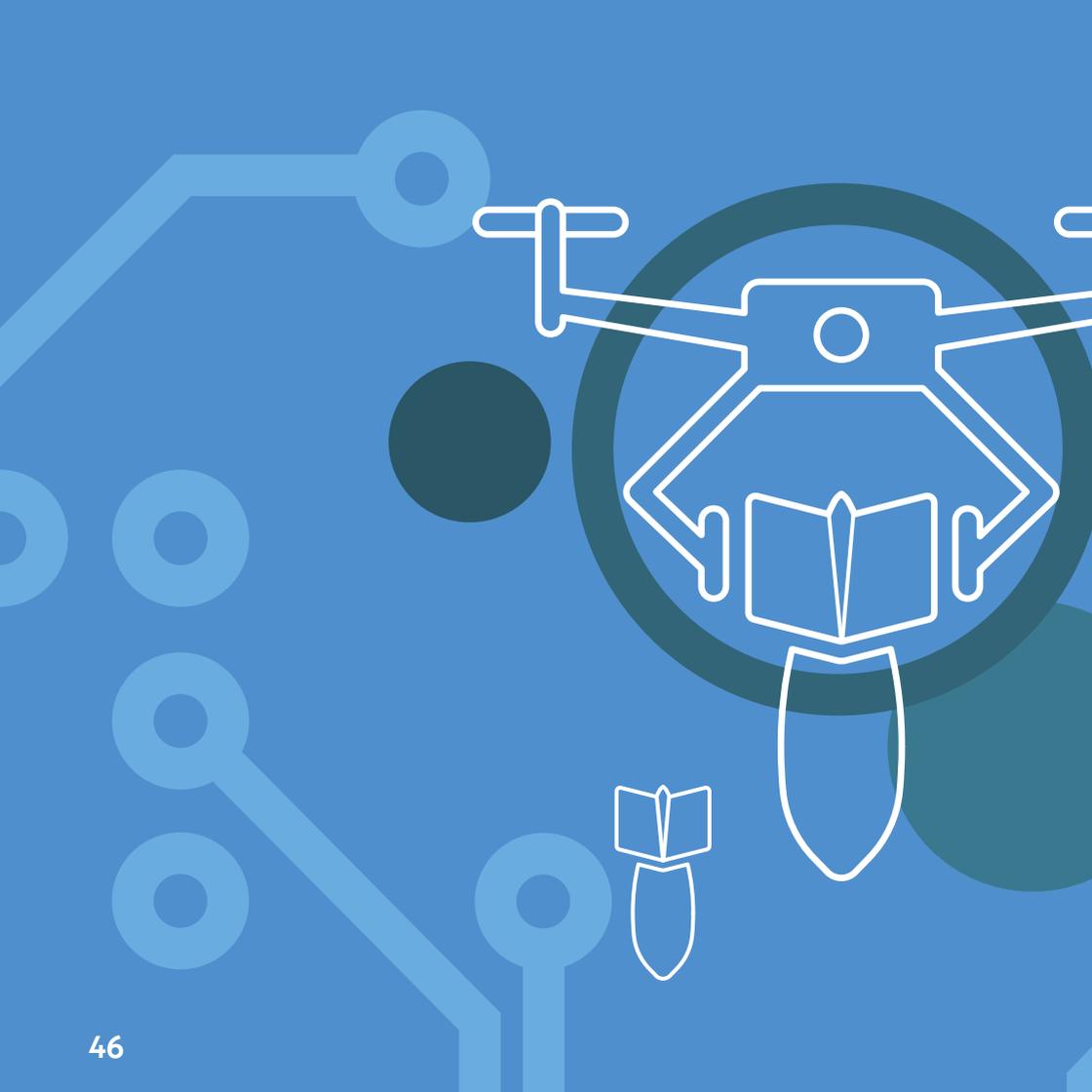
Fähigkeiten durch die Symbiose mit Technologie anstrebt.

Der Mensch wird in dieser futuristischen Philosophie als „Mängelwesen“ (ein Begriff von Arnold Gehlen) begriffen, das seine biologischen Grenzen mithilfe des technologischen Fortschritts bis hin zur Unsterblichkeit überwindet. Eine zentrale Figur ist dabei der „Cyborg“, ein Mischwesen aus Mensch und Maschine, das durch technische Implantate neue Fähigkeiten erhält. Eine andere Möglichkeit wären Gehirn-Computer-Schnittstellen zur direkten Informationsübertragung, über die sich Menschen

neues Wissen und neue Fähigkeiten aneignen könnten, über die sich Stimmungen beeinflussen und psychische Krankheiten behandeln ließen oder die ein „Mind Uploading“ ermöglichen, bei dem das Gehirn samt Bewusstsein und Erinnerungen auf ein technisches Speichermedium kopiert werden könnte, um so unsterblich zu werden.

Auch wenn diese Ideen bisher nur in Science-Fiction-Geschichten verwirklicht wurden, gibt es viel Kritik an dieser Denkrichtung. Zentral ist dabei die Nähe zur Eugenik, die über Kontrolle des menschlichen Erbguts den Genpool der Bevölkerung verbessern will und die ein wesentlicher Bestandteil der nationalsozialistischen „Rassenhygiene“ war.





TEIL 3 KRITISCHE PERSPEKTIVEN AUF KI



MILITÄRISCHER EINSATZ

Als der Begriff „Künstliche Intelligenz“ in den 1950er Jahren aufkam und die Forschung in diesem Bereich an Fahrt aufnahm, stand die Welt noch unter dem Eindruck der verheerenden Folgen des Zweiten Weltkriegs, des Holocaust und der Atombombenabwürfe der USA auf die japanischen Städte Hiroshima und Nagasaki. Angesichts des zerstörerischen Potenzials neuer Technologien und der Gewissheit, dass Regierungen bereit sind, diese einzusetzen, standen viele Menschen den frühen Entwicklungen in der Computertechnologie skeptisch gegenüber. Diese Skepsis spielte in den Protesten gegen den Vietnamkrieg in den 1960er Jahren und der Anti-Atomkraft-

Bewegung eine wichtige Rolle. Insbesondere die Atombombe ist seit dieser Zeit zu einer zentralen Referenz im Diskurs über KI geworden. Im Mai 2023 etwa verglichen namhafte KI-Forscher:innen und Persönlichkeiten aus der Computerbranche in einem offenen Brief das Gefahrenpotenzial von KI mit dem der Atombombe. Einige Stimmen fordern sogar eine internationale Aufsichtsbehörde für KI, vergleichbar mit der Internationalen Atomenergiebehörde.

Der Bezug des KI-Diskurses zur Atombombe hat vor allem zwei Gründe: Zum einen ist die Nukleartechnologie ein Beispiel für die Dual-use-Problematik,

also das Dilemma, dass technologische Entwicklungen sowohl für zivile als auch für militärische Zwecke genutzt werden können. So eignet sich die Kerntechnik nicht nur für Atombomben, sondern auch zur Energiegewinnung. Für Forschende und Entwickler:innen ergibt sich daraus zweitens, dass ihre Arbeit an solchen Technologien eine politische und moralische Dimension erhält, wie der Fall des Physikers Józef Rotblat zeigt: Er war im Rahmen des Manhattan-Projekts an der Entwicklung der ersten Atombombe beteiligt, verließ das Projekt aber 1944, nachdem sich abgezeichnet hatte, dass die Nationalsozialisten keine eigene Atombombe bauen konnten.

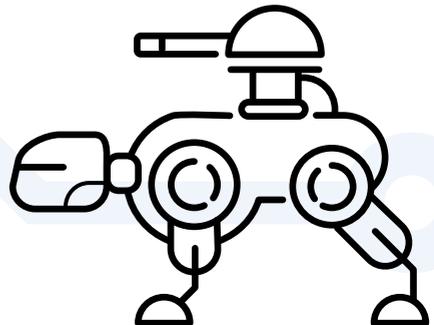
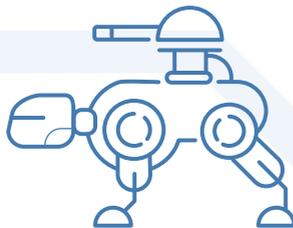
Weizenbaum wies darauf hin, dass auch die KI eine Dual-use-Technologie ist, und forderte, dass sich Wissenschaftler:innen ihrer damit einhergehenden Verantwortung stellen müssten. Man müsse davon ausgehen, dass jede Technologie, die wir für zivile Zwecke nutzen, auch für militärische Zwecke genutzt werde. Weizenbaum selbst war

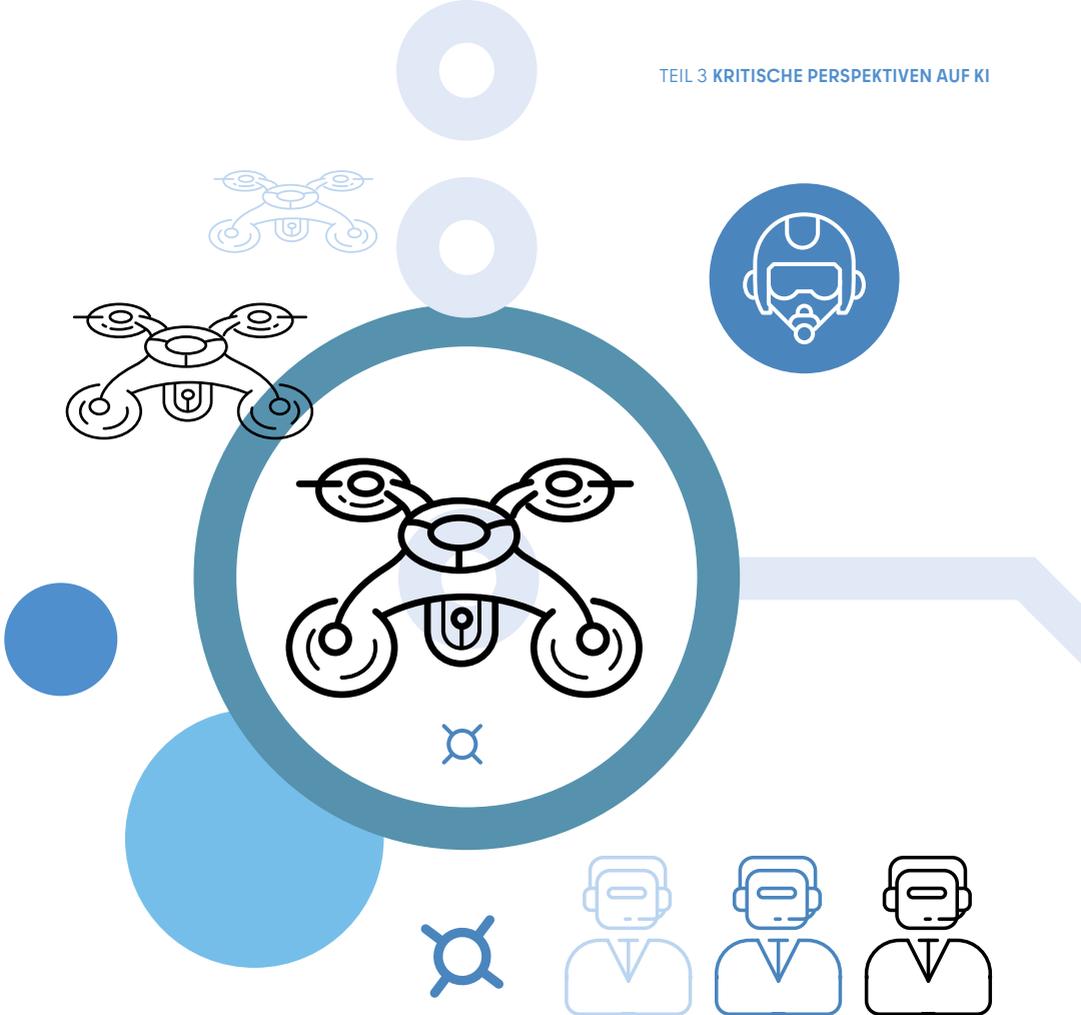
zu Beginn seiner beruflichen Laufbahn an zahlreichen Forschungsprojekten beteiligt, die vom US-Verteidigungsministerium finanziert wurden. Das Militär hatte ein besonderes Interesse an der Entwicklung von Computern und KI, weil es sich davon Vorteile im Kalten Krieg versprach. Für Weizenbaum war diese Verquickung seiner Arbeit mit militärischen Interessen schließlich ein Grund, sich von dem Forschungsfeld zu distanzieren.

Auch heute spielt KI nicht nur in zivilen Anwendungen wie Suchalgorithmen und Sprachassistenten eine Rolle, sondern auch im militärischen Bereich. Im russisch-ukrainischen Krieg ab Anfang 2022 wurde KI beispielsweise eingesetzt, um Angriffe autonomer Kamikazedrohnen auf Ziele hinter den feindlichen Linien durchzuführen. Im Gaza-Krieg nach dem Angriff der Hamas auf Israel am 7. Oktober 2023 setzte das israelische Militär verschiedene KI-Systeme ein, um schneller Ziele für die Bombardierung von Gebäuden auszuwählen

oder feindliche Kämpfer zu identifizieren und zu töten. Der Markt für militärische KI-Systeme wächst seit Jahren, ihre Entwicklung steht dabei in einem engen Zusammenhang mit der Grenz- und Migrationskontrolle, in deren Rahmen sie häufig getestet werden (siehe Infokasten „Technological Testing Grounds“). KI soll Kriegseinsätze „effizienter“ und weniger verlustreich machen, dies könnte allerdings die Hemmschwelle für einen Waffeneinsatz senken und damit die Stabilität regionaler wie internationaler Verhältnisse gefährden.

Daneben stellt sich die Frage nach der Verantwortlichkeit. Der UN-Generalsekretär António Guterres erklärte 2018, dass autonome Waffen, die ohne menschliche Kontrolle töten können, politisch inakzeptabel und moralisch verwerflich seien. Ende 2023 gab es die erste Resolution der UN-Generalversammlung zu autonomen Waffensystemen. Verboten wurden sie da aber noch nicht. Auch im „Artificial Intelligence Act“ der Europäischen Union sind militärische KI-Systeme ausgenommen. Das Wettrüsten ist bereits in vollem Gange.







TECHNOLOGICAL TESTING GROUNDS

Seit den Terroranschlägen vom 11. September 2001 werden Menschen, die nationale Grenzen überschreiten, nicht nur in den USA zunehmend als Sicherheitsrisiko wahrgenommen. Die Grenz- und Migrationskontrolle ist damit zu einem zentralen Anwendungsfeld für neue Technologien wie KI geworden, weshalb auch von „Smart Borders“, intelligenten Grenzen, gesprochen wird. Bereits vor dem Grenzübertritt werden KI-Systeme zur Risikoklassifizierung anhand von Überwachungs- und Social-Media-Daten eingesetzt. An der Grenze selbst dienen Sensoren

und Kameras dazu, Gesichter biometrisch zu erfassen oder auffälliges Verhalten zu erkennen. Und selbst nach der Einreise kommt es zum Teil zur Überwachung von Migrant:innen und zur KI-gestützten Prüfung von Asylansuchen und Abschiebungen.

Regierungen und Grenzbehörden wie Frontex in der EU oder das Bundesamt für Migration und Flüchtlinge (BAMF) in Deutschland investieren zunehmend in KI-Technologien, die meist von privaten Big-Tech-Unternehmen wie Palantir oder Clearview AI entwickelt werden.



Art und Umfang des Einsatzes von KI und anderen Technologien im Rahmen der Grenz- und Migrationskontrolle wären im zivilen Bereich häufig unzulässig. Die Juristin Petra Molnar bezeichnet diese Kontexte deshalb als „Technological Testing Grounds“, also als Versuchsfelder, auf denen neue technologische Systeme erprobt werden, bevor sie – in modifizierter Form – in anderen Bereichen zum Einsatz kommen.

Drohne der französischen Grenzpolizei zur Erkennung von Schlauchbooten mit Geflüchteten im Ärmelkanal, 2019

Foto von Denis Charlet/AFP via Getty Images

Wie diese Technologien dabei Machtasymmetrien und strukturelle Diskriminierung verschärfen und welche Auswirkungen sie auf die Grundrechte und das Leben der Menschen haben, die dem Einsatz dieser Technologien ausgesetzt sind, bleibt dabei meist unsichtbar.



ÖKOLOGISCHE FOLGEN

Auch wenn es sich bei den KI-Systemen, mit denen wir im Alltag zu tun haben, um virtuelle Programme handelt, hat ihre Nutzung reale Auswirkungen auf die Umwelt. Dies hängt zunächst mit den Ressourcen zusammen, die für die Herstellung von Computern benötigt werden. So enthält jedes Computergerät zahlreiche, zum Teil sehr seltene Rohstoffe wie Eisen, Kupfer, Aluminium, Lithium, Nickel, Gold, Zink, Kobalt und Coltan. Sie werden je nach Vorkommen an verschiedenen Orten der Welt abgebaut, aufbereitet, über weite Strecken transportiert und schließlich so verbaut, dass unsere Geräte immer kleiner, leichter, nutzungsfreundlicher und leistungs-

fähiger werden. Dieser Prozess geht einher mit einem hohen Wasser- und Energieverbrauch und starken Umweltbelastungen.

Nehmen wir Lithium: Das Leichtmetall wird für wiederaufladbare Batterien verwendet und ist daher in jedem Laptop und Smartphone enthalten. Abgebaut wird es in Bergwerken durch Sprengen und Bohren sowie durch salzhaltiges Grundwasser, das an die Oberfläche gepumpt wird und in künstlich angelegten Becken verdunstet, um das darin enthaltene Lithium freizusetzen. Das größte Lithiumvorkommen der Welt befindet sich in der Atacama-Wüste, im

„Lithium-Dreieck“ zwischen Bolivien, Chile und Argentinien. Für den Abbau werden dort große Mengen an Wasser benötigt, wodurch die Trinkwasservorräte der Region stark abnehmen und zu versalzen drohen. Zudem ist Lithium kein nachwachsender Rohstoff, sodass die weltweiten Lithium-Reserven bereits in wenigen Jahrzehnten aufgebraucht sein könnten. Durch die zunehmende Verbreitung von Elektroautos hat sich der ohnehin schon große Bedarf an Lithium noch einmal drastisch erhöht. Während etwa in einem Smartphone rund 1 Gramm Lithium steckt, sind es in einem Elektroauto mehrere Kilogramm. Deshalb gibt es bereits verstärkte Bemühungen, Lithium zu recyceln, doch noch immer landet ein Großteil des Elektromülls auf riesigen Müllhalden in Ghana und Pakistan.

Bei anderen Metallen ist es ähnlich: Durch Abbau von Zinn, das zum Löten von Elektronik zum Einsatz kommt, werden in Indonesien ganze Landstriche unfruchtbar und unbewohnbar. Für den

Tagebau werden Wälder abgeholzt, was zu Dürren, Überschwemmungen und Erdbeben führt. Beim Waschen des Zinns werden Flüsse verunreinigt, beim Abbau vor der Küste Korallen zerstört. Viele Arbeiter:innen leiden unter massiven Gesundheitsproblemen. Für den Abbau von Coltan, das in Mikrocondensatoren steckt, werden in der Demokratischen Republik Kongo große Flächen des Regenwalds abgeholzt, was den Lebensraum zahlreicher Tiere bedroht. Auch werden durch den Abbau giftige Schwefelsäuren freigesetzt, die schädlich für die Umwelt und gefährlich für die Arbeiter:innen sind. Aufgrund der instabilen Lage im Kongo ist der Abbau von Coltan dort zudem mit Gewalt und Menschenrechtsverletzungen verbunden (siehe Infokasten „Konfliktrohstoffe“).

All dies sind keine KI-spezifischen Folgen, sondern ökologische Probleme von Computern insgesamt. Für die ökologische Bewertung von KI spielen diese Folgen jedoch eine wichtige Rolle. Wenn auf die Potenziale von KI für den Klima-

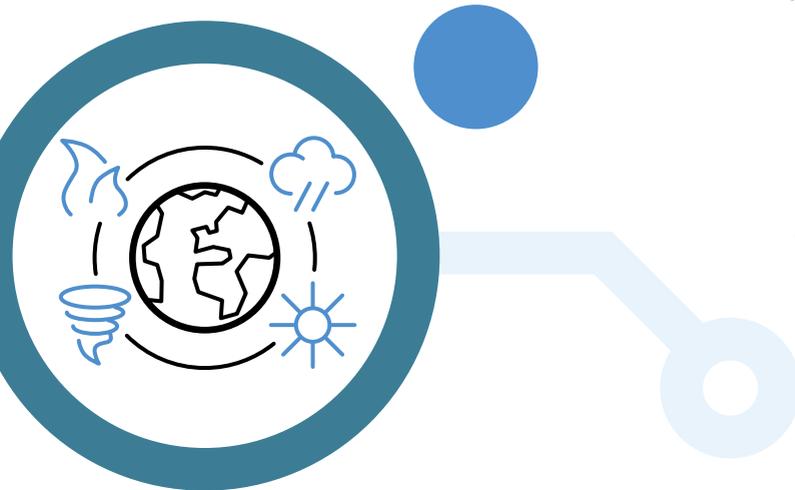
Lithium-Abbau in der chilenischen
Atacama-Wüste, 2022

Foto von John Moore/Getty Images



und Umweltschutz verwiesen wird, müssen diese Probleme auch berücksichtigt werden. Bei der KI-Nutzung kommt zudem ein sehr hoher Stromverbrauch hinzu, der den Verbrauch etwa von Suchanfragen weit übersteigt. Bei einem massenhaften Einsatz wie im Fall von ChatGPT sind damit hohe Treibhausgas-Emissionen und ein Wasserverbrauch für die Kühlung der Server verbunden – mit steigender Tendenz.

Joseph Weizenbaum war überzeugt, dass Technologie alleine uns nicht vor dem Klimawandel bewahren kann. Dazu bedürfe es grundlegender gesellschaftlicher Veränderungen, die uns die KI nicht abnehmen könne. Ein zentrales Element dieses Wandels sei dabei der „Widerstand gegen die Gier“ in unserer Art zu leben, zu wirtschaften und zu konsumieren. Auch wenn die materielle Basis digitaler Technologien wie KI hinter glatten Bildschirmen und virtuellen Programmen zu verschwinden scheint, sind auch sie nicht frei von ökologischen Fragen.



KONFLIKT-ROHSTOFFE

Einige Rohstoffe, die für den Bau von Computern und anderer Elektrotechnik verwendet werden, stammen aus Kriegs- und Konfliktgebieten und werden deshalb „Konfliktrohstoffe“ genannt. In erster Linie sind dies Zinn, Coltan, Wolfram und Gold, die auch als „3TG Konfliktmineralien“ (nach den englischen Bezeichnungen „tin“, „tantalum“, „tungsten“ und „gold“) bezeichnet werden. Darüber hinaus werden auch Diamanten, Edelhölzer, Drogenrohstoffe, Kakao und Erdöl zum Teil in Kriegs- und Konfliktregionen ab- bzw. angebaut.

Mit dem Abbau, dem Handel und der Verarbeitung von Konfliktrohstoffen wird in Kauf genommen, dass die Menschen in den betroffenen Regionen zum Teil systematischer Gewalt und Menschenrechtsverletzungen ausgesetzt sind und die Minenarbeiter:innen oftmals unter menschenunwürdigen Bedingungen arbeiten müssen. Darüber hinaus können diese Rohstoffe selbst ein Auslöser oder Verstärker von bewaffneten Konflikten sein, etwa weil um sie gekämpft wird („Ressourcenkonflikte“) oder weil die Handelserlöse zur Finanzierung



von Konflikten dienen. So kann sich der Reichtum an natürlichen Ressourcen für ein Land negativ auswirken, weshalb in diesem Zusammenhang auch von einem „Ressourcenfluch“ gesprochen wird.

Einer der wichtigsten Konfliktrohstoffe für die Digital- und Elektroindustrie ist das Erz Coltan, aus dem das Metall Tantal gewonnen wird. Die größten Coltanvorkommen befinden sich in der Demokratischen Republik Kongo, die auch das wichtigste Förderland ist. Viele Coltanminen werden von paramilitärischen Gruppen kontrolliert, die mit dem Handel des Rohstoffs unter anderem den Bürgerkrieg im Kongo finanzieren. Der Abbau ist dabei mit Menschen- und Völkerrechtsverletzungen wie

Kinderarbeit und Trinkwasserverschmutzung verbunden. Die bestehende politische und soziale Instabilität, die ein Schatten der belgischen Kolonialzeit ist, wird so durch den Abbau und Handel mit Coltan noch verstärkt. Bisherige Regulierungsversuche wie der Dodd-Frank Act der USA von 2010 sowie die EU-Verordnung zu Konfliktmineralien von 2017 sind aufgrund des illegalen Handels und der damit verbundenen mangelhaften Nachvollziehbarkeit der Lieferketten nur schwer umsetzbar.



GESELLSCHAFTLICHE RISIKEN

Im Jahr 1985 wurde Joseph Weizenbaum in einem Interview gefragt, ob der Computer das Potenzial habe, die Gesellschaft zu verändern, oder ob er eher alte Machtstrukturen verstärke. Da zu dieser Zeit der Personal Computer (PC) Einzug in immer mehr Haushalte erhielt, war die Erwartung, der Computer könne die Gesellschaft revolutionieren, weit verbreitet. Weizenbaum entgegnete jedoch, dass der Computer „eine fundamental konservative Kraft“ sei, die vor allem dazu beitrage, dass sich die Gesellschaft nicht verändern müsse. Als Beispiel nannte er das Bankwesen, das sich angesichts einer wachsenden Bevölkerung ohne Technologie hätte

dezentralisieren müssen, dies aber dank des Computers vermeiden konnte.

Weizenbaums Sichtweise regt dazu an, auch in Bezug auf KI den Blick von gesellschaftlichen Veränderungen auf Kontinuitäten zu lenken und danach zu fragen, ob bestehende Machtasymmetrien und Ungerechtigkeiten durch KI fortgeschrieben oder gar verstärkt werden. Ein solcher Blick zeigt, dass generative KI tatsächlich etablierte Stereotype und Diskriminierungsformen reproduziert. Da diese Systeme in der Regel mit Inhalten aus dem Internet trainiert werden, „lernen“ sie folglich auch die darin enthaltenen Abwertungen. Für

Aufsehen sorgte 2016 der von Microsoft entwickelte Twitter-Bot „Tay“, der durch die Interaktion mit Twitter-Nutzer:innen lernte und innerhalb kürzester Zeit so viele rassistische und frauenverachtende Posts absetzte, dass der Bot bereits einen Tag nach seinem Start wieder eingestellt wurde.

Darüber hinaus haben Studien gezeigt, dass KI-generierte Inhalte und Entscheidungen häufig einen „Bias“, also eine Verzerrung, zu Ungunsten ohnehin benachteiligter Gruppen aufweisen. Deutlich wird dies etwa, wenn KI-generierte Bilder Frauen in traditionellen Geschlechterrollen zeigen oder Schwarze Männer als gefährlich porträtieren, während weiße Männer vor allem dann auftauchen, wenn es um Erfolg und Kompetenz geht. Noch problematischer wird es, wenn solche Verzerrungen konkrete Folgen für Menschen haben. Das ist zum Beispiel der Fall, wenn im Rahmen prognostischer Polizeiarbeit („Predictive Policing“) Personen einer bestimmten Gruppe eher verdächtig werden als

andere, obwohl es keinen Anlass dazu gibt; wenn Frauen bei Banken weniger Kredite erhalten als Männer, obwohl sie die gleichen Kriterien erfüllen; oder wenn Personen mit bestimmten Nachnamen bei der KI-gestützten Auswahl von Bewerber:innen für eine Stelle seltener berücksichtigt werden.

Nach Kritik an solchen Verzerrungen haben die Entwickler:innen reagiert. Zum Teil versuchen sie, die Probleme technisch zu lösen, etwa indem sie bei generativer KI unsichtbare „Co-Prompts“ beifügen, um die Verzerrungen auszugleichen. Sinnvoller ist es jedoch, schon im Vorfeld anzusetzen und die KI-Trainingsdaten kritisch zu prüfen, vielfältiger zu gestalten und von problematischen Inhalten zu bereinigen. Für solche Aufgaben werden schon jetzt zahlreiche Datenarbeiter:innen vor allem aus dem globalen Süden eingesetzt. Hinter den Leistungen von KI steckt also viel menschliche und oft schlecht bezahlte Arbeit (siehe Infokasten „Datenarbeit“).

Diese Datenarbeit wird auch genutzt, um auf neue Formen von Desinformation, Hassrede oder illegalen Inhalten zu reagieren und KI-Systeme entsprechend anzupassen. Allerdings ist KI auch selbst Teil des Problems, schließlich kann sie dazu genutzt werden, genau solche Inhalte ohne großen Aufwand zu produzieren. Je besser die generative KI wird, umso besser können Inhalte gefälscht werden. Wenn aber nicht mehr zwischen Realität und Fiktion unterschieden werden kann, schürt dies Zweifel an jeglicher Information. Für eine auf Konsens angelegte Demokratie kann dies zu massiven Problemen führen.

Für Weizenbaum war es wichtig, zwischen dem konkreten Nutzen und möglichen Risiken einer Technologie abzuwägen. Mit Blick auf generative KI ließe sich demnach fragen: Sollen wir unsere Demokratie aufs Spiel setzen, nur damit wir auf Knopfdruck Texte und Bilder erstellen lassen können? Weizenbaum hätte diese Frage wohl verneint und empfohlen, diese Technologie deshalb nicht (weiter) zu entwickeln und sich stattdessen wichtigeren Problemen zuzuwenden.





DATENARBEIT

Zur Visualisierung von KI werden häufig Bilder von Robotern, verkabelten Gehirnen oder Netzwerken verwendet – auch in dieser Broschüre. Damit wird suggeriert, dass KI-Systeme selbstständig Daten verarbeiten, Muster erkennen und lernen können – und dass menschliche Arbeit überflüssig wird. Tatsächlich verbergen sich hinter KI-Systemen jedoch tausende Datenarbeiter:innen, die Trainingsdaten generieren, kategorisieren und überprüfen.

Da es sich dabei oft um kleinteilige Aufgaben handelt, spricht man auch von Klick- oder Mikroarbeit. Für das autonome Fahren werden zum Beispiel riesige Datenmengen von Straßenmarkierungen, Schildern, Fahrzeugen, Menschen oder Tieren benötigt, um eine KI zu entwickeln, die diese automatisch erkennt und entsprechend reagiert. Um diesen Bedarf an Trainingsdaten zu decken, besteht eine große Nachfrage nach billigen Arbeitskräften, die häufig



die Grundlage für KI bilden. Eine Möglichkeit, an solche billigen Arbeitskräfte zu kommen, sind Crowdwork-Plattformen wie Amazon Mechanical Turk: Hier werden kleinteilige Aufgaben von selbstständigen Arbeiter:innen für wenige Cents oder Euro erledigt. Eine andere Möglichkeit ist die Auslagerung der Arbeit an Subunternehmen im globalen Süden. Eines der bekanntesten Unternehmen ist Samasource aus San Francisco, das unter anderem für OpenAI Datenarbeiter:innen in Kenia, Pakistan und Indien beauftragt, die mit rund zwei US-Dollar pro Stunde entlohnt werden. In der KI-Industrie findet also eine globale Arbeitsteilung statt, bei

der die Big-Tech-Unternehmen aus dem Silicon Valley von billigen Arbeitskräften und mangelndem Arbeitsschutz im globalen Süden profitieren.

Insgesamt führt der Einsatz von KI-Systemen also nicht zu einem Rückgang von menschlicher Arbeit. Das Trainieren und Korrigieren von generativen KI-Systemen wird auch in Zukunft enorme Datenmengen erfordern, angesichts des aktuellen Erfolgs von KI aller Voraussicht nach sogar mit steigender Tendenz. Der dadurch entstehende Arbeitsmarkt muss dringend reguliert werden, um Ausbeutung und prekäre Arbeitsverhältnisse zu verhindern.





The background is a solid blue color with several abstract geometric shapes and lines in various shades of blue. There are circles of different sizes, some with smaller circles inside them, and lines that connect these circles, creating a network-like structure. The shapes are scattered across the page, with some being more prominent than others.

TEIL 4 LEITFRAGEN DER KI-KRITIK

Auch wenn der historische Rückblick auf die lange Geschichte der Diskussionen um Computer und KI zeigt, dass es viele Kontinuitäten und wiederkehrende Muster zu unserer heutigen Auseinandersetzung mit diesen Technologien gibt, fällt es inmitten der Diskussionen häufig schwer, eine kritische Distanz zu wahren. Schließlich bekommen wir hautnah mit, wie schnell die technologische Entwicklung voranschreitet, wie Computer immer besser und leistungsfähiger werden und plötzlich Dinge möglich sind, die wir uns früher kaum vorzustellen wagten. Unsere Bewunderung für diese rasante Entwicklung paart sich dabei schnell mit Hoffnungen auf eine bessere Zukunft, sie kann aber auch einschüchtern, verunsichern und überfordern. Umso wichtiger ist eine kritische Begleitung dieser Entwicklungen, die ab und an einen Schritt zurücktritt und versucht, die großen Prohezeiungen und Warnungen historisch einzuordnen und abzuschätzen, wohin die Reise geht. Die folgenden Leitfragen sollen dabei helfen:

MENSCHENBILD

Wenn von der Leistungsfähigkeit und den Potenzialen von KI die Rede ist, wie wird dabei über Menschen gesprochen?

Wird deutlich gemacht, dass diese Technologie dazu da ist, Menschen zu unterstützen, oder werden Menschen und KI in Konkurrenz zueinander gesetzt? Werden Menschen zum Beispiel als begrenzte, sterbliche und fehleranfällige Wesen dargestellt, um die Leistungsfähigkeit von KI hervorzuheben? Werden sie gar lächerlich gemacht? Auf was für eine Welt steuern wir dann zu, wenn Menschen abgewertet werden, weil sie als „schlechtere Maschinen“ gelten? Wie stark prägt diese Sichtweise bereits unser Selbstbild und unsere Ansprüche an uns selbst und an andere? Wie könnten wir das ändern?

INSTRUMENTELLES DENKEN

Wenn KI als Lösung für ein spezifisches Problem dargestellt wird, um was für eine Art Problem handelt es sich?

Ist es technischer oder sozialer Natur, geht es also um technische Fehler, Herausforderungen und Hindernisse oder um soziale Missstände, diskriminierende Strukturen und zwischenmenschliche Konflikte? Wenn es sich um ein soziales Problem handelt, welche Gründe haben ursprünglich zu diesem Problem geführt? Verschwinden diese Gründe durch den KI-Einsatz? Welche Gründe bleiben bestehen und welche neuen Probleme kommen hinzu? Könnte das zu lösende Problem dadurch sogar noch größer werden? Welche Probleme gibt es jetzt schon durch den Einsatz von KI? Sind sie technischer oder sozialer Natur? Und wie können wir sie am besten angehen?

SOZIALE UNGLEICHHEITEN

Wenn es darum geht, was mit KI alles möglich wird, wer kann diese Möglichkeiten für sich nutzen – und wer nicht?

Für wen werden KI-Systeme entwickelt, wer kann sie sich leisten und wer entscheidet über ihren Einsatz? Wer wird hingegen von ihrer Nutzung ausgeschlossen, wer hat kein Mitspracherecht, aber Nachteile dadurch? Auf wen wird durch KI Druck ausgeübt, wer wird ausgebeutet, damit sie funktioniert, wessen Arbeit soll ersetzt werden, wer muss sich nach ihr richten? Gegen wen wird KI eingesetzt, wer wird etwa überwacht und kontrolliert, wer diskriminiert oder bedroht? Welche Machtasymmetrien und Ungerechtigkeiten werden hier fortgeschrieben oder gar verstärkt? Und was können wir dagegen tun?





TEIL 5 BIOGRAFISCHE ÜBERSICHT: JOSEPH WEIZENBAUM

Joseph Weizenbaum, 2006
Porträt von Erik Weiss

- 1923** Geboren am 8. Januar in Berlin als Sohn von Jechiel „Harry“ Weizenbaum und Henriette (geb. Orman), zwei Brüder Leo (aus erster Ehe des Vaters) und Heinrich W. „Heinz“ (später Henry Francis Sherwood)
- 1923-1936** Kindheit in Berlin, Elternhaus am Gendarmenmarkt in Berlin-Mitte, 1934 Besuch des Luisenstädtischen Realgymnasiums, 1935 erzwungener Wechsel auf die Knabenschule der jüdischen Gemeinde an der Kaiserstraße
- 1936** Flucht vor dem Nationalsozialismus in die USA, Emigration mit dem Schiff von Bremerhaven nach New York, dann Detroit, dort Besuch der Highschool (Abschluss 1940)
- 1941-1950** Studium der Mathematik an der Wayne University in Detroit
- 1942-1946** Unterbrechung des Studiums, Einzug zum Militärdienst, Dienst in der meteorologischen Abteilung der US Army Air Forces
- 1944** Annahme der US-amerikanischen Staatsbürgerschaft
- ~1946** Erste Ehe mit Thelma „Selma“ Goode, ein Sohn David, Scheidung ~1949
- 1950-1953** Wissenschaftliche Assistenz im Differential Analyzer Laboratory und leitender Programmierer am Digital Computer Laboratory an der mathematischen Fakultät der Wayne University in Detroit

- 1952** Zweite Ehe mit Ruth Manes, vier Töchter Pm, Sharon, Miriam, Naomi
- 1953-1955** Programmierer bei der Computer Control Company auf der Raketenbasis Point Mugu der US Navy sowie in der Computer Division von Bendix Aviation in Los Angeles, die unter anderem Boden-Luft-Raketen für das Militär entwickelt
- 1956-1963** Programmierer im Computer Development Laboratory von General Electric in Sunnyvale, Kalifornien, Beteiligung an erstem Computerbanksystem ERMA für die Bank of America, Entwicklung der Programmiersprache SLIP („Symmetric List Processor“)
- 1963** Gastprofessur („visiting associate professor“) am Massachusetts Institute of Technology (MIT) in Cambridge, Mitarbeit im neu gegründeten Project MAC, maßgeblich gefördert durch das US-Verteidigungsministerium
- 1964** Ernennung zum außerordentlichen Professor („associate professor“) am Department of Electrical Engineering am MIT
- 1964-1967** Entwicklung von ELIZA (erste Veröffentlichung 1966)
- 1969** Gründungsmitglied der Union of Concerned Scientists
- 1970** Ernennung zum ordentlichen Professor („full professor“)

- 1972-1973** Sabbatical und Forschungsaufenthalte am Center for Advanced Study in the Behavioral Sciences (CASBS) der Stanford University sowie an der Harvard University
- 1976** Veröffentlichung seines kritischen Hauptwerks „Computer Power and Human Reason“ (1977 unter dem deutschen Titel „Die Macht der Computer und die Ohnmacht der Vernunft“ erschienen)
- 1981** Mitbegründer der Computer Professionals for Social Responsibility (CPSR)
- 1984** Mitbegründer des Forums InformatikerInnen für Frieden und gesellschaftliche Verantwortung (FIfF)
- 1988** Emeritierung am MIT, Verleihung des Norbert Wiener Award der CPSR
- 1992** Ehrenpreis des Fachbereichs „Informatik und Gesellschaft“ der Gesellschaft für Informatik
- 1994** Humboldt-Preis der Alexander-von-Humboldt-Stiftung, Gastaufenthalt an der Albert-Ludwigs-Universität Freiburg
- 1996** Umzug nach Berlin, Wohnung zunächst am Ludwigkirchplatz in Berlin-Wilmersdorf, später im Nikolaiviertel in Berlin-Mitte
- 1998** Ehrendoktorwürde des Fachbereichs Mathematik und Informatik der Universität Bremen

- 2001** Verleihung des Großen Bundesverdienstkreuzes
- 2003** Ehrendoktorwürde des Fachbereichs Informatik der Universität Hamburg, Ehrenmitgliedschaft der Gesellschaft für Informatik
- 2007** Dokumentarfilm „Rebel at Work“ von Silvia Holzinger und Peter Haas über Joseph Weizenbaum
- 2008** Verstorben am 5. März in Ludwigsfelde-Gröben (Brandenburg), später beerdigt auf dem Jüdischen Friedhof in Berlin-Weißensee



Digitale Ausstellung über das Leben und Wirken von Joseph Weizenbaum





EMPFOHLENE LITERATUR

EMPFOHLENE LITERATUR

Bundeszentrale für politische Bildung (2023): Künstliche Intelligenz. Aus Politik und Zeitgeschichte, 73. Jahrgang, Heft 42. <https://www.bpb.de/shop/zeitschriften/apuz/kuenstliche-intelligenz-2023/>

Eine gut zugängliche Sammlung wissenschaftlicher Texte zu verschiedenen Problem-bereichen rund um das Thema KI, darunter ethische Fragen zum Thema, die Rolle von KI für die Demokratie, die Regulierung von KI sowie der Einsatz von KI in der Schule und der Arbeitswelt.



Crawford, Kate (2021): Atlas of AI: Power, politics, and the planetary costs of Artificial Intelligence. New Haven, London: Yale University Press.

In diesem Buch gibt die US-Forscherin Kate Crawford, eine der wichtigsten KI-Expert:innen unserer Zeit, einen ausführlichen Überblick über die gesellschaftliche Folgen der Produktion, Entwicklung und des Einsatzes von KI-Systemen.

FifF-Kommunikation. <https://www.fiff.de/publikationen/fiff-kommunikation.html>

Die vierteljährlich erscheinende Zeitschrift des Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung (FifF), das Joseph Weizenbaum 1984 mitbegründet hat. Das Magazin begleitet aktuelle Entwicklungen im Bereich Informatik und Gesellschaft.



~~~~~

**Molnar, Petra (2024): The walls have eyes: Surviving migration in the age of Artificial Intelligence. New York: The New Press.**

Die Juristin und Anthropologin Petra Molnar berichtet von ihren Recherchen zum Einsatz von Technologien wie KI an den Grenzen unter anderem von Europa und den USA und den verheerenden Folgen für Menschen, die diese Grenzen zu passieren versuchen.

~~~~~

Morozov, Evgeny (2013): To save everything, click here. The folly of technological solutionism. New York: PublicAffairs.

Ähnlich wie Joseph Weizenbaum kritisiert Evgeny Morozov in diesem Buch die weit verbreitete Annahme, dass sich mit Technologie gesellschaftliche und politische Probleme lösen ließen, und warnt vor den Gefahren eines solchen „Solutionismus“.

~~~~~

**Rosengrün, Sebastian (2021): Künstliche Intelligenz zur Einführung. Hamburg: Junius.**

Eine breite und gut verständliche Einführung in das Thema Künstliche Intelligenz. Neben einer Übersicht über die KI-Geschichte und

zentrale Begriffe liefert das Buch einen fundierten Einstieg in zentrale Diskussionen um die Potenziale und Risiken dieser Technologie.

~~~~~

Weizenbaum, Joseph (1977): Die Macht der Computer und die Ohnmacht der Vernunft. Frankfurt am Main: Suhrkamp.

Das zentrale Werk von Joseph Weizenbaum, in dem er sich ausführlich und kritisch mit seinem eigenen Forschungsfeld auseinandersetzt. Seine wichtigsten Kritikpunkte am gesellschaftlichen Umgang mit Computern und KI sind hier bereits angelegt.

~~~~~

**Weizenbaum, Joseph (2001): Computermacht und Gesellschaft: Freie Reden. Frankfurt am Main: Suhrkamp.**

Eine Sammlung kurzer Texte von Joseph Weizenbaum zu verschiedenen Themen, wie etwa zum Menschenbild der KI, zur Rolle von Computern in der Schule und zur Verantwortung von Wissenschaftler:innen. Ein guter Einstieg in die Weizenbaum-Lektüre.

**Weizenbaum-Institut (2024): Pionier, Skeptiker, Aktivist: Wendepunkte im Leben von Joseph Weizenbaum. Eine digitale Ausstellung.**  
<https://jw.weizenbaum-institut.de>

In einer multimedialen Ausstellung, die viel Zusatzmaterial bereitstellt, werden die wichtigsten Stationen von Weizenbaums Leben durchlaufen. Verlinkt ist u. a. ein vierteiliges Audiofeature mit Interviews mit Wegbegleiter:innen und Expert:innen zu seinem Werk und Wirken.



**Wendt, Gunna (2023): Computermacht und Vernunft: Hommage an Joseph Weizenbaum zu seinem 100. Geburtstag. Innsbruck, Wien: Limbus Verlag.**

Auszüge aus Gesprächen, persönliche Erinnerungen und eine Übersicht über Weizenbaums Leben durch die Autorin Gunna Wendt, die zahlreiche Interviews mit Joseph Weizenbaum geführt und in Buchform herausgegeben hat.

# Impressum

## **Autor**

Christian Strippel

## **Mitarbeit**

Moritz Buchner, Magnus Rust,  
Alexandra Keiner

## **Gestaltung**

Bergsee, blau

## **Herausgebende**

Weizenbaum-Institut e.V.  
Hardenbergstraße 32, 10623 Berlin  
[www.weizenbaum-institut.de](http://www.weizenbaum-institut.de)

gefördert vom



Bundesministerium  
für Bildung  
und Forschung

und

Berliner Landeszentrale  
für politische Bildung  
Hardenbergstraße 22-24, 10623 Berlin  
Öffnungszeiten des Besuchszentrums:  
Mo, Mi, Do und Fr 10-18 Uhr und  
Revaler Straße 29, 10245 Berlin  
Öffnungszeiten des Besuchszentrums:  
Di, Mi und Fr 10-18 Uhr  
[www.berlin.de/politische-bildung](http://www.berlin.de/politische-bildung)

Berlin, im September 2024

