

Wissen teilen

Künstliche Intelligenz

Führt uns die Technologie in eine bessere Zukunft?



 **forma futura**

Vermögen nachhaltig investieren

seit 2006

Einleitung

Sie ist derzeit in aller Munde: Künstliche Intelligenz, kurz KI. Seit der Erfindung des Internets hat keine neue Technologie die Öffentlichkeit so sehr in ihren Bann gezogen. Es vergeht kaum ein Tag, an dem wir nicht von irgendeiner Vorhersage lesen oder hören, wie KI unser Leben verändern wird.

Auch ihr Potenzial zur Transformation der Finanzmärkte ist beträchtlich: Eine neue Studie der Bank Goldman Sachs suggeriert, dass die Einführung der Technologie über einen Zeitraum von zehn Jahren durch Produktivitätssteigerung und Automatisierung zu einem jährlichen Anstieg des globalen BIP um 7 Prozent, oder fast

USD 7 Billionen (7'000'000'000'000), führen könnte [1].

Wir befinden uns derzeit also in einer regelrechten Hype-Phase, welche die Finanzmärkte und die Menschen in Atem hält. Aber ist der ganze Rummel überhaupt gerechtfertigt? Welche Chancen und Risiken birgt die Technologie? Und kann KI nachhaltig sein?

Der vorliegende Beitrag stellt den aktuellen Stand der KI-Technologien dar und versucht, einige dieser Fragen zu beantworten. Damit soll die Grundlage für eine fundierte Diskussion zu diesem Thema geschaffen werden.

Inhaltsverzeichnis

Glossar.....	4
LLMs und die Illusion von Intelligenz	5
Von Hype-Zyklen und KI-Wintern	7
Nutzen von KI.....	10
Gefahren von KI.....	13
Regulierung: entscheidend und zeitkritisch	17
KI und Nachhaltigkeit.....	19
Fazit.....	21
Quellenverzeichnis	22

Glossar

Chatbot: Ein Computerprogramm, das Benutzeranfragen erfasst und darauf antwortet, indem es eine menschliche Unterhaltung simuliert.

Künstliche Intelligenz (KI): Die Fähigkeit einer Maschine, fortgeschrittene Analyse- und logikbasierte Techniken, einschliesslich maschinellen Lernens, anzuwenden, um Ereignisse zu interpretieren, Entscheidungen zu unterstützen und zu automatisieren und unterschiedliche Arten von Outputs zu erzeugen.¹

Künstliche allgemeine Intelligenz: Eine theoretische Form der KI, die in der Lage ist, eine beliebige Anzahl hypothetischer Aufgaben mit Hilfe allgemeiner kognitiver Fähigkeiten zu lösen.

Large Language Model (LLM): Eine spezielle Art von künstlicher Intelligenz, die auf riesige Textmengen trainiert wurde, um zu verstehen,

wie Wörter auf sinnvolle Art und Weise kombiniert werden können. LLMs können so auf der Grundlage schriftlicher Aufforderungen von Nutzer:innen neue Textbeiträge generieren. LLMs werden mit Hilfe von maschinellem Lernen trainiert und kommen normalerweise in Form von Chatbots daher.

Maschinelles Lernen: Der Prozess, durch den eine künstliche Intelligenz in der Lage ist, seine eigene Leistung zu verbessern, indem sie kontinuierlich neue Daten in ein bestehendes statistisches Modell einbezieht.

Text-Bild-Generator: KI-gestütztes Tool, das Bilder auf der Grundlage einer schriftlichen Aufforderung erstellt. Text-Bild-Generatoren verwenden in der Regel maschinelles Lernen, um ihre Bilderzeugungsfunktionen auf umfangreichen Datenbanken von verfügbaren Bildern zu trainieren.

¹ Es gibt keine einheitliche Definition von künstlicher Intelligenz. Die hier verwendete Definition stammt vom Technologieberatungsunternehmen Gartner [2] und steht im Einklang mit dem aktuellen und sich abzeichnenden Stand der KI-Technologien und -Fähigkeiten. Dabei wird insbesondere berücksichtigt, dass derzeitige KI-Technologien meist probabilistische Analysen umfassen.

LLMs und die Illusion von Intelligenz

30. November 2022: Dieses Datum wird als Ausgangspunkt für den aktuellen Hype um KI in Erinnerung bleiben. Es ist der Tag, an dem ChatGPT (Generative Pre-trained Transformer) einer breiten Öffentlichkeit zugänglich gemacht wurde.

Nach nur fünf Tagen hatte der KI-gestützte Chatbot bereits 1 Million aktive Nutzer:innen; zwei Monate später waren es über 100 Millionen, was es zur am schnellsten wachsenden Verbraucheranwendung der Geschichte machte [3].

«ChatGPT: Von 1 Million Nutzer:innen in 5 Tagen auf 100 Millionen in 2 Monaten – ein beispielloser Aufstieg. Kann KI unsere eigene Intelligenz herausfordern und übertreffen?»

Chatbots und andere KI-gestützte Anwendungen, wie z. B. Text-Bild-Generatoren, die in der Lage sind, Bilder aus Textaufforderungen zu erzeugen, gab es schon vorher. Aber ChatGPT hatte scheinbar die Kunst perfektioniert, eine schier unendliche Menge an Informationen mit einer Präsentation zu kombinieren, die menschliche Interaktion und Intelligenz in einer Weise nachahmt, welche unsere Fantasie beflügelt. Hat der Mensch eine Intelligenz geschaffen, die irgendwann mit seiner eigenen konkurrieren oder sie sogar übertreffen könnte?

Nicht ganz. ChatGPT und andere so genannte «Large Language Models» (LLMs) verwenden ein maschinelles Lernverfahren, um einen riesigen Datensatz von Texten und deren Bedeutung in verschiedenen Kontexten zu erfassen und dann zu lernen, wie Wörter in einer sinnvollen Weise kombiniert werden können. Um einsatzbereit zu sein, müssen diese LLMs mit grossen Datenmengen vorab trainiert werden (das «P» in GPT, siehe oben) – ein zeit-, energie- und kostenintensiver Prozess, der viele Fragen aufwirft.

Vereinfacht kann man sich diesen Trainingsprozess folgendermassen vorstellen: LLMs nehmen einen Textabschnitt aus einem Datensatz, auf den sie Zugriff haben, verdecken einige Wörter am Ende und versuchen zu erraten, was dort hingehören könnte. Dann decken sie die Antwort auf, vergleichen sie mit ihren Schätzungen und lernen daraus, welche Wörter und Wort-Kombinationen mit welcher Wahrscheinlichkeit zusammengehören. Da die Antworten in den Daten selbst enthalten sind, können diese Modelle eigenständig auf riesigen Datensätzen trainiert werden, ohne dass menschliche Auswerter erforderlich sind [4]. Bei den Datensätzen handelt es sich um grosse Teile des Internets, das eine nahezu unbegrenzte Menge an Texten enthält.

LLMs sind speziell darauf ausgelegt, menschliche Sprachmuster zu imitieren, was sie intelligent erscheinen lässt. Im Gegensatz zu Menschen «denkt» sich ein LLM jedoch nicht eine bestimmte Antwort oder Geschichte aus, und sucht dann nach dem genauen Wortlaut, um diese auszudrücken. Vielmehr generiert das Modell Inhalte in sequenzieller Weise, ein Wort nach dem anderen, basierend auf der statistischen Wahrscheinlichkeit des nächsten Wortes in Bezug auf die ursprüngliche Anfrage und auf das, was es bereits geschrieben hat.

Anders gesagt: Wenn ein LLM einen Satz zu schreiben beginnt, hat es weder eine Vorstellung davon, wie das letzte Wort des Satzes lauten wird, noch welche Information der Satz den Leser:innen vermitteln wird. LLMs sind «stochastische Papageien» [5] («stochastic parrots»), d. h. Modelle, die überzeugende Sprache erzeugen, aber die Bedeutung der Sprache, die sie verarbeiten und produzieren, selbst nicht verstehen.

Aus diesem Grund erzeugen sie auch manchmal so genannte «Halluzinationen» – Texte, die zwar semantisch und syntaktisch plausibel wirken, aber in Wirklichkeit falsch oder unsinnig sind [6].


An dieser Stelle sollte angemerkt werden, dass KI ein breiter Begriff ist, der viele verschiedene Anwendungen umfasst; Netflix-Empfehlungen, selbstfahrende Autos und Spam-Filter für E-Mail-Konten sind alles Beispiele für KI-gestützte Anwendungen.

LLMs wie ChatGPT sind lediglich die derzeit populärsten Iterationen von KI-Technologien. Was all diese Technologien verbindet, ist die Tatsache, dass sie derzeit alle noch Aufgaben nach Mustern und Regeln lösen, die wir Menschen ihnen beibringen. Die derzeitigen KI-Tools sind dementsprechend noch weit entfernt von einer «künstlichen allgemeinen Intelligenz» (Artificial General

Intelligence), d. h. von einer KI, die selbstständig denkt, argumentiert und in der Lage ist, all das und potenziell noch mehr zu erlernen, was ein Mensch kann.

Die Entwicklung einer künstlichen allgemeinen Intelligenz wird als Hauptziel der KI-Forschung angesehen. Experten streiten darüber, ob die Erschaffung einer künstlichen allgemeinen Intelligenz überhaupt möglich ist, und wenn ja, ob das Erreichen dieses Ziels nur noch Jahre oder eher Jahrzehnte oder gar Jahrhunderte entfernt ist. Tatsache ist, dass die derzeitige Technologie noch nicht so weit ist und dass die aktuellen LLMs, die auf maschinelles Lernen basieren, lediglich eine Illusion von Intelligenz darlegen.

 Kannst du mir einen neuen, originellen Witz erzählen, den du dir ausgedacht hast?

 Natürlich, hier ist einer für dich:



Warum hat der Mathematikbuch in der Bibliothek immer gute Laune?

Weil es immer voller positiver Lösungen ist!

Abbildung 1: Wenn man die aktuelle Version von ChatGPT bittet, einen originellen Witz zu kreieren, kommt etwas heraus, das die Struktur eines Witzes und den Anschein einer funktionierenden Pointe hat, die Leser:innen aber eher verwirrt als amüsiert. Das liegt daran, dass das Modell einen Witz nur sequenziell generieren kann und somit nicht in der Lage ist, sich zuerst eine Pointe «auszudenken», um diese dann in eine überzeugende Handlung zu integrieren; ChatGPT kreiert so bloss das Scheinbild eines Witzes. Humor bleibt vorerst noch eine Eigenschaft, die Menschen und z.T. auch Tieren [7] vorbehalten ist.

Von Hype-Zyklen und KI-Wintern

Die derzeit populärsten KI-gestützten Werkzeuge sind also keine wirklichen Intelligenzen im eigentlichen Sinne und gleichen eher einer umfangreichen Statistikübung als einem Verstand.

Dennoch wäre es verfehlt, die Technologie kleinzureden, denn ChatGPT und andere KI-gestützte Modelle sind nichts weniger als technologische Meisterleistungen. Sie haben die Nachahmung menschlicher Interaktion fast perfektioniert, sind für eine breite Bevölkerung und für viele alltägliche Aufgaben und Aktivitäten nützlich geworden und finden Anwendung in einer Vielzahl von verschiedenen Bereichen. Der jüngste Hype um KI ist daher auf den ersten Blick verständlich. Doch ist er auch fundiert?

Es ist möglich, dass KI-Technologie dem so genannten «Hype-Zyklus» folgen wird, welcher darstellt, welche Phasen der öffentlichen Aufmerksamkeit eine neue Technologie bei deren Einführung durchläuft [8]. Dem Modell zufolge,

welches vom Technologieberatungsunternehmen Gartner entwickelt wurde, steigt die öffentliche Sichtbarkeit einer bestimmten Technologie nach der Einführung eines neuen Produkts oder eines technologischen Durchbruchs dramatisch an. Die konstante Berichterstattung generiert einen übertriebenen Enthusiasmus und unrealistische Erwartungen, die in einem «Gipfel der überhöhten Erwartungen» kulminieren. Darauf folgt ein «Tal der Enttäuschung» nachdem festgestellt wird, dass die neue Technologie nicht alle Erwartungen erfüllen kann; in dieser Phase nimmt auch die Berichterstattung ab. Die nächste Phase ist der sogenannte «Pfad der Erleuchtung», in dem eine realistischere Einschätzung zu einem Verständnis der Vorteile, Grenzen und sinnvollen Anwendungen der neuen Technologie führt. Dies führt zum so genannten «Plateau der Produktivität», in dem die Einsatzgebiete der neuen Technologien weitgehend verstanden werden und sich ihre breite Anwendbarkeit und Relevanz auf dem Markt auszahlt.

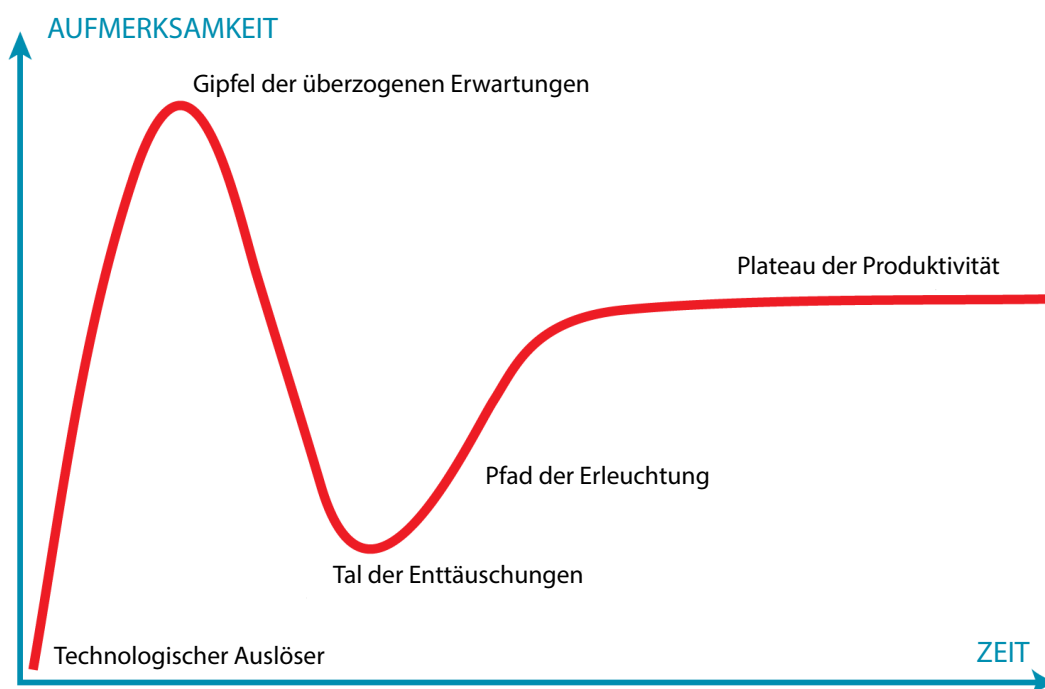


Abbildung 2: Der Gartner Hype Cycle mit seinen fünf Phasen, wie er von dem amerikanischen Technologieberatungsunternehmen Gartner konzipiert wurde. Quelle: <https://de.wikipedia.org/wiki/Hype-Zyklus>.

Ein Beispiel für eine Technologie, die dem Gartner-Hype-Zyklus folgte, ist das Internet. Nach seiner Erfindung und Einführung für die breite Öffentlichkeit in den 1990er Jahren löste es einen massiven Hype aus, der schliesslich zur Dot-Com-Blase zu Beginn des 21. Jahrhunderts führte. Seitdem wurden die Potenziale, Grenzen und Gefahren des Internets besser verstanden und seine Anwendungsmöglichkeiten ständig erweitert.

Was KI angeht, so könnte die Einführung von ChatGPT als der im Hype-Zyklus beschriebene technologische Auslöser angesehen werden. Der derzeitige Medienrummel um KI-Technologie würde dementsprechend dem aufsteigenden

Pfad entsprechen, der zum «Gipfel der überhöhten Erwartungen» des Gartner-Modells führt. Ob die KI-Technologie dem Gartner-Hype-Zyklus weiter folgt und in naher Zukunft in ein «Tal der Enttäuschung» fällt, bleibt abzuwarten.

Tatsächlich hat KI schon in der Vergangenheit ähnliche Hype-Phasen durchlaufen. Der Begriff «KI-Winter» wird bereits seit 1984 verwendet, um eine Periode geringerer Finanzierung und geringeren Interesses an der Forschung im Bereich der künstlichen Intelligenz nach einer anfänglichen Hype-Phase zu beschreiben [9].

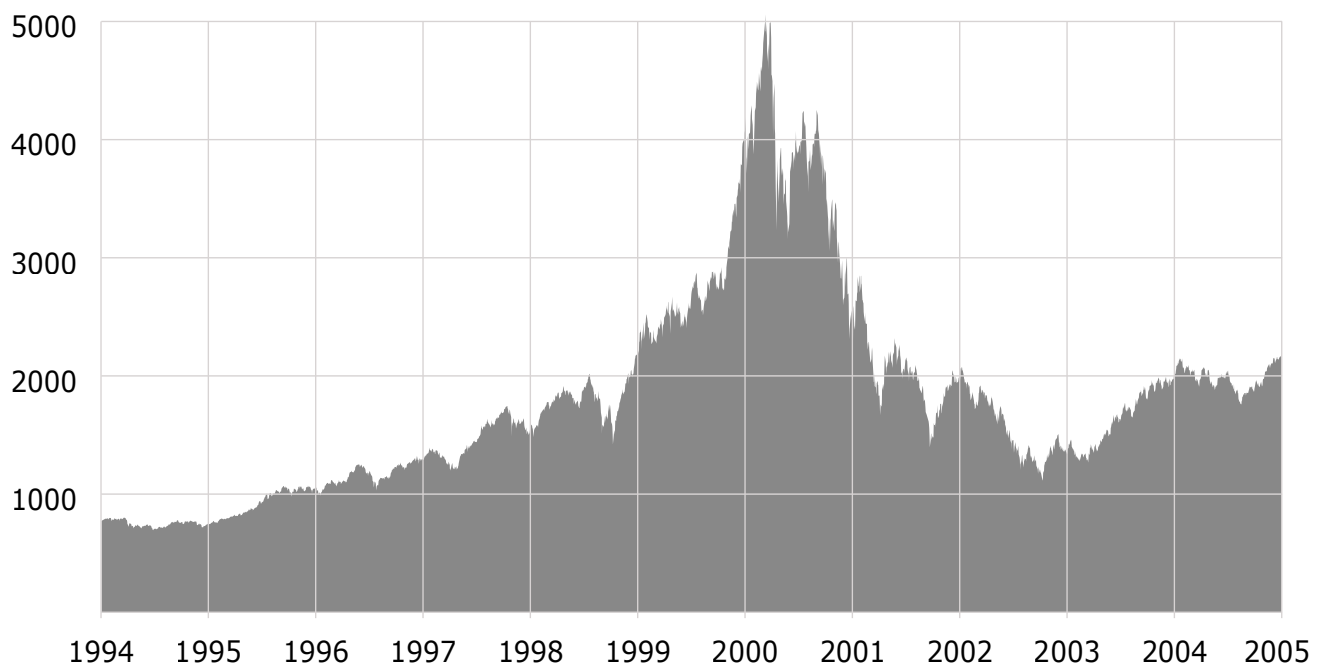


Abbildung 3: Der NASDAQ Composite Index stieg Ende der 1990er Jahre in die Höhe und fiel dann infolge der Dot-Com-Blase drastisch. Die Dot-Com-Blase kann als Beispiel für eine Phase eines Hype-Zyklus angesehen werden, der durch einen technologischen Auslöser (das Internet) in Fahrt gebracht wurde.

Quelle: https://commons.wikimedia.org/wiki/File:Nasdaq_Composite_dot-com_bubble.svg.

Mindestens zwei solcher KI-Winter haben bereits stattgefunden, einer zwischen 1974 und 1980, der andere zwischen Ende der 1980er und Mitte der 1990er Jahre [10]. Der Grund, warum wir uns in einer Zeit befinden, in der die KI-Technologien wieder an Fahrt aufnehmen, liegt vor allem an drei Faktoren: gigantische Datenmengen, die über das Internet frei zugänglich sind; Algorithmen, die in der Lage sind, aus diesen Daten zu lernen; und die nötige Rechenleistung, um diese Daten mit diesen Algorithmen zu verarbeiten. Zumindest die erst- und die letztgenannten Bedingungen sind erst in den letzten zwei Jahrzehnten entstanden und haben damit den Weg für den derzeitigen KI-Boom geebnet.

Weitere Fortschritte in der KI-Technologie könnten jedoch bald an Grenzen stossen. Um LLMs besser und weniger fehleranfällig zu machen, müssen sie auf noch grössere Datensätze trainiert werden. Ohne technologische Durchbrüche bei den Rechenkapazitäten wird das Training grösserer LLMs jedoch schneller teuer, als die Modelle besser werden [11] – ein klassischer Fall von sinkenden Erträgen.

Während das Training der derzeit verwendeten Version von ChatGPT (GPT-3) etwa USD 4.5 Mio. gekostet hat, soll das Training der leistungsfähigeren Version (GPT-4), die im Moment nur für zahlende Kundschaft verfügbar ist, rund USD 100 Mio. gekostet haben [12]. Es ist ebenfalls unsicher, ob das Problem der «Halluzinationen»

(siehe Seite 5) überhaupt gelöst werden kann. Yann LeCun, ein führender Experte für KI-Technologie, äusserte deshalb kürzlich die Ansicht, dass die derzeitige Konzentration der KI-Forschung und -Entwicklung auf LLMs und deren Verbesserung in eine Sackgasse führen wird und letztendlich nur eine «Marotte» sei, die von der Entwicklung leistungsfähigerer KI ablenkt [13].

«KI hat grosses Potenzial, aber auch klare Grenzen und potenzielle Gefahren. Wir sollten Sensationsmeldungen mit Vorsicht betrachten.»

Es ist daher wichtig, mit übertriebenen Prognosen zur KI vorsichtig umzugehen, zumal die Technologie und ihre Funktionsweise in ihrer Komplexität von der breiten Öffentlichkeit noch nicht verstanden werden. Heutige LLMs und andere KI-gestützte Werkzeuge sind zweifellos in der Lage, beeindruckende Leistungen zu erbringen und haben auch das Potenzial, die Art und Weise, wie wir arbeiten und leben, erheblich zu beeinflussen. Sie haben aber gleichzeitig auch klare Grenzen. Letztlich ist KI ein Instrument, das grossen Nutzen schaffen kann, aber auch gewisse Gefahren birgt, denen mit Bedacht begegnet werden muss. Auf beides wird in den nächsten Kapiteln eingegangen.

Nutzen von KI

KI hat das Potenzial, bestimmte Aufgaben zu erleichtern, mehr Effizienz zu schaffen und sogar die moderne Wissenschaft voranzubringen. In diesem Abschnitt wird kurz auf einige Bereiche und Anwendungen eingegangen, in denen KI-Technologien erhebliche Nutzen bringen könnten. Dies sind nur einige Beispiele, da die Anwendungsbereiche von KI sehr vielfältig sind.

KI im Gesundheitswesen: Die Gesundheitssysteme auf der ganzen Welt sind überlastet, und die WHO schätzt, dass bis 2030 10 Millionen medizinische Fachkräfte fehlen werden, vor allem in den Ländern mit niedrigem und mittlerem Einkommen [14]. Viele nationale Gesundheitssysteme werden dementsprechend in Zukunft mit der Herausforderung konfrontiert sein, mehr Menschen mit weniger Ressourcen behandeln zu müssen.

«KI im Gesundheitswesen könnte strapazierte Gesundheitssysteme entlasten und die Effizienz von Behandlungen verbessern.»

Die aktuelle KI-Technologie hat das Potenzial, einige dieser Probleme zu lindern. Ein auf den medizinischen Bereich angewandtes LLM wäre beispielsweise in der Lage, Daten aus den Krankenakten von Patient:innen zu analysieren, um potenzielle Problembereiche zu identifizieren, Diagnosen und klinische Pläne zu erstellen und weitere Tests oder Behandlungsoptionen vorzuschlagen. Ärzt:innen und Pflegepersonal könnten so erheblich entlastet werden, was auch zu einer Verringerung medizinischer Fehler und einer Verbesserung der medizinischen Versorgung führen könnte.

Fortschritte in der KI-Technologie könnten daher entscheidend sein, um einen Übergang zu intelligenten Gesundheitssystemen zu unterstützen, welche die Bedürfnisse der Bevölkerung besser erfüllen und so den Zugang und die Effizienz der Gesundheitsversorgung positiv verändern.

KI in der Forschung: Von der Krebsforschung und Genbiologie über die Arzneimittelentwicklung bis hin zur Raumfahrt – KI-Technologien haben das Potenzial, in einer Vielzahl von wissenschaftlichen Forschungsbereichen eingesetzt zu werden. In vielen wissenschaftlichen Bereichen ist es aufgrund der Datenmenge und der verschiedenen Parameter, Muster und Trends, die berücksichtigt werden müssen, äusserst schwierig oder sogar unmöglich, aussagekräftige Informationen manuell zu erstellen. KI ist in der Lage, die Datenfluten zu durchforsten und relevante und verständliche Informationen herauszukristallisieren.

Tatsächlich wird KI beispielsweise bereits in der Genomforschung eingesetzt, wobei sie auf der Grundlage verfügbarer Daten zur Vorhersage komplexer Proteinstrukturen eingesetzt wird [15]. Weiter hat die amerikanische Raumfahrtbehörde NASA kürzlich KI eingesetzt, um leichtere, stärkere und haltbarere Bauteile und Komponenten für ihre Missionen im Weltraum zu entwickeln [16]. In China wurde vor kurzem das erste vollständig von KI entwickelte Medikament zur Behandlung von Lungenfibrose für Humanversuche zugelassen [17]. KI-Technologien werden weiterhin eine zentrale Rolle bei der Weiterentwicklung der modernen Wissenschaft spielen und haben deshalb das Potenzial, die Forschung in einer Vielzahl von Disziplinen weiterzubringen oder sogar zu revolutionieren.



Abbildung 4: Eine KI-entwickelte Komponente, die bereits im Herbst 2023 bei Weltraummissionen der NASA eingesetzt werden soll. Sie ist leichter, stärker und haltbarer als von Menschen entwickelte Bauteile.

Quelle: <https://www.nasa.gov/feature/goddard/2023/nasa-turns-to-ai-to-design-mission-hardware>.

Produktivere Arbeitskräfte – und glücklichere Menschen? Berichte zusammenfassen, Inhalte schreiben und aufpolieren, Dokumente übersetzen: Viele von uns haben ChatGPT bereits ausprobiert oder nutzen es sogar aktiv bei der Arbeit und im Privatleben, um uns alltägliche Aufgaben zu erleichtern. Tatsächlich kann KI die Produktivität in einigen Branchen erheblich steigern. In einer Arbeit von Erik Brynjolfsson und Kollegen von der Stanford University z. B. wurden Kundendienstmitarbeiter:innen untersucht: Durch den Zugang zu einem KI-Tool konnte die Zahl der pro Stunde gelösten Probleme im

Durchschnitt um 14% gesteigert werden [18]. Andere akademische Studien weisen auf einen jährlichen Anstieg der Arbeitsproduktivität um drei Prozentpunkte in Unternehmen hin, die KI-Technologien einsetzen, was über viele Jahre hinweg einen signifikanten Einkommensanstieg bedeuten würde [19]. Mehr Produktivität muss sich aber nicht unbedingt eins zu eins in mehr Gewinn umsetzen lassen; sie könnte auch dazu genutzt werden, die Arbeitnehmer:innen zu entlasten und ihnen mehr freie Zeit für ihre Familien und ihr Leben zu gewähren, um so allgemein eine bessere Lebensqualität zu schaffen.

In der Kult-Zeichentrickserie «Die Jetsons», die den Alltag einer Familie in ferner Zukunft schildert, gibt es einen Running Gag, in dem sich der Vater über seine «ausbeuterischen» Arbeitszeiten von 3 Stunden pro Tag beschwert. Eines der wichtigsten popkulturellen Werke des Futurismus des 20. Jahrhunderts hatte sich bereits im Jahr 1962 eine Zukunft vorgestellt, in der

Maschinen und KI es den Menschen ermöglichen würden, sich auf ihr Privatleben, anstatt auf Arbeit und Profit zu konzentrieren. Mehr als 60 Jahre später, da sich die Technologie in gewisser Weise den Vorstellungen der Schöpfer des Cartoons annähert, könnte unsere Gesellschaft in manchen Aspekten davon profitieren, aus dieser Vision Inspiration zu schöpfen.



Abbildung 5: Die Schöpfer des futuristischen Zeichentrickserie «Die Jetsons», die 1962 erstmals ausgestrahlt wurde, hatten sich eine Zukunft vorgestellt, in der Maschinen und KI es den Menschen ermöglichen würden, sich auf ihr eigenes Glück und ihre Selbstverwirklichung zu konzentrieren. Quelle: <https://zoedigc202.wordpress.com>.

Gefahren von KI

Um es vorweg zu sagen: Die Wahrscheinlichkeit, dass die Menschheit von einer hochkompetenten KI unterjocht wird, die sich gegen ihre Schöpfer auflehnt, ist derzeit noch fest im Bereich der Science Fiction angesiedelt, auch wenn uns einige Big-Tech-CEOs das Gegenteil weismachen wollen. Diese haben ihre eigenen Gründe, die Angst vor einer bevorstehenden Maschinen-Apokalypse zu schüren - doch dazu später mehr. Nichtsdestotrotz bringt die heutige KI-Technologie unmittelbar einige reale Herausforderungen mit sich, die erkannt, diskutiert und angegangen werden müssen. Einige dieser Probleme werden hier angesprochen.

Urheberrechte: Wie bereits erwähnt, arbeiten alle derzeit populären KI-Modelle, die maschinelles Lernen verwenden, indem sie Datensätze durchkämmen und bereits vorhandene Inhalte zu etwas scheinbar Neuem zusammenführen. Text-zu-Bild-Generatoren sind KI-gestützte Werkzeuge, die auf der Grundlage einer schriftlichen Eingabeaufforderung Bilder erzeugen können. Ihre Funktionsweise ähnelt der von LLMs in dem Sinne, dass sie auf Bilder aus dem Internet trainiert werden und dann «neue» Bilder aus bereits vorhandenem Material erstellen. Diese Technologien sind sie auch in der Lage, die Stilrichtungen bestimmter Künstler:innen oder die Schreibstile von Schriftsteller:innen zu replizieren. Jeder «Originalinhalt», der von diesen Modellen erschafft wird, ist deshalb mit vielen komplizierten und noch nie dagewesenen Fragen zum Urheberrecht verbunden.

Einige Künstler:innen haben bereits zum Boykott von KI-generierter «Kunst» aufgerufen [20]. Zudem verklagte erst kürzlich das visuelle Medienunternehmen Getty Images den Entwickler des Text-zu-Bild-Generators Stable Diffusion, mit der

Behauptung, das Unternehmen habe ohne Erlaubnis 12 Millionen seiner Bilder genutzt, um das KI-Modell zu trainieren [21]. Mit der Verbreitung von KI-Technologien werden sich ähnliche Fälle in Zukunft mit grosser Wahrscheinlichkeit häufen und neue Rechtsgrundlagen schaffen.



Die Verstärkung von Stereotypen und die Verbreitung von hetzerischen Inhalten: KI-Modelle, die maschinelles Lernen nutzen, werden auf Datensätzen aus dem Internet trainiert; dieses enthält bekanntermassen alle möglichen Meinungen, Überzeugungen und Vorurteile. Die Modelle laufen deshalb die Gefahr, schädliche Stereotypen, Tendenzen und/oder regelrechte Bigotterie zu reproduzieren und weiter zu verbreiten. So kann ein Text-zu-Bild-Generator beispielsweise Bilder von weissen Menschen als Standard generieren, oder als Reaktion auf Anfragen nach «schönen» Menschen zu heller Hautfarbe tendieren.

Abbildung 6: Das obige Bild wurde in wenigen Sekunden mit dem KI-Text-zu-Bild-Generator Stable Diffusion erstellt. Der vom Autor dieses Artikels verwendete Text war der folgende: Ein Gemälde eines Sees im Stil von Van Gogh. Bei Nacht. Klarer Himmel. Berge im Hintergrund. Dieses Gemälde zeigt unverkennbar den ikonischen Stil Van Goghs, muss aber (derzeit) in keiner Weise referenziert oder attribuiert werden.

Weiter werden Frauen in KI-generierten Bildern mit grösserer Wahrscheinlichkeit sexualisiert dargestellt als Männer, da die kulturelle Sexualisierung von Frauen sowohl in Bild als auch Schrift weit verbreitet ist [22]. Ein expliziteres Beispiel wäre, wenn KI-Modelle regelrechten Rassismus und Bigotterie reproduzieren, die sie aus dem Internet «gelernt» haben.

Man bedenke: Wenn man einem Papagei rassistische Ausdrücke beibringt, wird es diese lernen und wiederholen, ohne deren Bedeutung zu kennen – dasselbe gilt auch für «stochastische Papageien» (siehe Seite 5), d. h. für LLMs. Im Jahr 2016 musste Microsoft seinen KI-gesteuerten Chatbot «Tay» dauerhaft abschalten, weil es nach nur 16 Stunden Interaktion mit Twitter-Nutzer:innen gelernt hatte, rassistische, sexistische, und naziverherrlichende Äusserungen von sich zu geben [23].

Zwar verfügen die heute populärsten KI-Tools, wie z. B. ChatGPT, über programmierte Einschränkungen, die verhindern, dass hasserfüllte Inhalte erstellt werden. Sie können aber immer noch implizite Vorurteile und Stereotypen reproduzieren, und tun dies auch regelmässig [24]. Darüber hinaus werden zukünftige KI-Tools, die aus der Open-Source-Gemeinschaft stammen, wahrscheinlich weniger strenge (oder auch gar keine) Beschränkungen in Bezug auf die Inhalte haben, die sie erzeugen können – eine fast unvermeidliche Folge der Verbreitung und «Demokratisierung» von KI-Technologien.

Fake News und «Deepfakes»: Im Moment können KI-generierte Bilder meistens noch von aufmerksamen Beobachter:innen als solche erkannt werden, da die meisten KI-Modelle immer noch

dazu tendieren, verschwommene oder in sich verschmelzende Bilder zu produzieren und z. B. Schwierigkeiten haben, realistisch aussehende Hände wiederzugeben. KI-Modelle werden jedoch stetig besser darin, Bilder zu erzeugen, die dem Fotorealismus nahekommen.

«Jeder Algorithmus, der dazu in der Lage ist, Deepfakes aufzuspüren, trägt das Potenzial in sich, die KI-Modelle hinter den Bildgenerierungssystemen weiterzuentwickeln und somit noch überzeugendere Ergebnisse zu erzielen.»

Das Gleiche gilt für KI-generierte Audioinhalte, welche die Stimmen prominenter Persönlichkeiten wiedergeben. Diese sind sogar noch weiter fortgeschritten und von echten Stimmnahmen kaum noch zu unterscheiden. Dies ermöglicht die Erstellung von gefälschtem Bild- und Tonmaterial (so genannte «Deepfakes»), das zur Verbreitung von Fake News und zur Verunglimpfung und Erpressung von Personen verwendet werden kann. Es können z. B. Bilder oder Videos, die in soziale Medien hochgeladen oder in Nachrichten verschickt werden, in sexuell explizite, von KI generierte Bilder verwandelt werden, die von echten Bildern kaum zu unterscheiden sind – eine Taktik, die bereits heute von Erpressern eingesetzt wird [25]. Ein ähnliches Phänomen ist das Aufkommen von pornografischen Inhalten, bei denen das Konterfei prominenter Persönlichkeiten, wie z. B. von Schauspieler:innen, mit Hilfe von KI-gestützter Technologie in pornografische Filme eingeblenndet wird.



Abbildung 7: Ein KI-generiertes Bild von Angela Merkel und Barack Obama beim Eisessen am Strand.

Verlust von Arbeitsplätzen: Wie bereits erwähnt, haben KI-Technologien das Potenzial, die Produktivität der Arbeitnehmer:innen zu steigern. Was aber, wenn Unternehmen noch einen Schritt weiter gehen und Menschen komplett durch KI ersetzen, um Kosten zu sparen? Dies ist derzeit wohl die am weitesten verbreitete Befürchtung im Zusammenhang mit KI. Buchhalter:innen, Programmierer:innen und Softwareingenieure:innen, Rechtsanwaltsgehilfen, Redakteure:innen, Lehrer:innen, Übersetzer:innen und Kundendienstmitarbeiter:innen gehören zu den am häufigsten genannten Berufen, die sich durch KI bedroht fühlen [26].

Es gibt eine Unmenge unterschiedlicher Prognosen darüber, wie viele Arbeitsplätze durch die KI betroffen sein werden, und sie reichen von niedrigen Schätzungen von 30 Prozent bis zu unwahrscheinlich hohen 70 Prozent aller Arbeitsplätze in den nächsten 30 Jahren. Die Realität liegt wahrscheinlich irgendwo dazwischen, zumindest wenn es um die Restrukturierung von bestimmten Arbeitsplätzen geht. Es gilt jedoch zu bedenken, dass derzeitige KI-Technologie

immer noch fehleranfällig ist (und wird es in naher Zukunft wahrscheinlich auch bleiben). KI-generierte Arbeit muss deshalb immer noch von Menschen überprüft, korrigiert und verbessert werden. Dies wird vielleicht am besten durch den Fall eines Anwalts veranschaulicht, der ChatGPT einen Schriftsatz für einen wichtigen Fall schreiben liess, ohne diesen zu überprüfen. ChatGPT zitierte prompt auf plausible Art und Weise mehrere rechtliche Präzedenzfälle im Dokument – die jedoch nicht existieren [27].

Auch wenn wir in unmittelbarer Zukunft wahrscheinlich keine Massenentlassungen aufgrund von KI erleben werden, sind tiefgreifende Umstrukturierungen bei bestimmten Arbeitsplätzen sehr wahrscheinlich. Der Verlust bestimmter Arbeitsplätze und die Umstrukturierung anderer aufgrund einer technischen Revolution kann durchaus zu gesellschaftlichen Umwälzungen führen [28]. Es ist daher nicht zu unterschätzen, welche gesellschaftlichen und historischen Auswirkungen neue Technologien wie z. B. KI haben können.

Die Verwendung von persönlichen und biometrischen Daten: Wie bereits erwähnt werden viele KI-Anwendungen, darunter vor allem LLMs, auf riesigen Datenmengen trainiert, um daraus zu lernen und sich zu verbessern. Dies birgt das Risiko, dass KI private Daten aus dem Internet und/oder anderen Quellen abgreift, was zu Bedenken hinsichtlich des Datenschutzrechts führt. Das Unternehmen hinter ChatGPT, OpenAI, sieht sich bereits mit einer milliardenschweren Sammelklage konfrontiert, in der behauptet wird, es habe private Daten gestohlen, um sein Modell zu trainieren. Das Problem könnte sich z. B. bei medizinischen Anwendungen von KI-Tools zur Erleichterung von medizinischen Diagnosen noch verschärfen, da gesundheitsbezogene Daten besonders sensibel sind.

KI-Technologien schaffen auch die Möglichkeit der automatischen Erkennung biometrischer Merkmale, wie z. B. die Erkennung von Gesichtern, Gangarten, Fingerabdrücken, Stimmen und anderen biometrischen oder verhaltensbezogenen Signalen in verschiedenen Zusammenhängen. Die Verwendung der biometrischen Fernidentifizierung in öffentlich zugänglichen Räumen könnte daher das Ende der Anonymität an diesen Orten bedeuten und würde somit in die Grundrechte und -freiheiten eingreifen. Gleichzeitig würde dies skrupellosen (staatlichen) Akteuren die Möglichkeit eröffnen, die Massenüberwachung (und in gewissen Fällen auch die Kontrolle/Repression) der Bevölkerung auszuweiten.

Regulierung: entscheidend und zeitkritisch

Die im vorigen Kapitel erwähnten Gefahren der KI lassen unzählige Fragezeichen aufkommen.

Nehmen wir uns etwas Zeit, um über die folgenden Fragen zu reflektieren:

- Können Künstler:innen Urheberrechtsverletzungen geltend machen, wenn ein KI-Tool ihren spezifischen Kunststil verwendet, um ein Kunstwerk zu generieren?
- Dürfen KI-Modelle anhand privater Informationen, die im Internet verfügbar sind, trainiert werden?
- Dürfen KI-Tools rassistische, hasserfüllte oder hetzerische Inhalte erstellen?
- Sollte die Erstellung und Verbreitung pornografischer Inhalte verboten werden, die z. B. mithilfe von KI-gestützten «Deep Fake» Technologien das Gesicht einer Person in sexuell explizite Situationen einfügt?
- Dürfen Unternehmen die Gehälter mit dem Argument drücken, dass KI bestimmte Aufgaben übernehmen kann?
- Sollte es den Strafverfolgungsbehörden erlaubt sein, voraussagende KI einzusetzen, die anhand biometrischer Daten oder früherer Verhaltensweisen Listen von zu überwachenden Personen zusammenstellt?

Einige dieser Fragen sind sicherlich leichter zu beantworten als andere. Was sie jedoch alle miteinander verbindet, ist, dass sie der Regulierung bedürfen. Es ist von entscheidender Bedeutung, dass Möglichkeiten und Gefahren, die sich mit der weiteren Entwicklung und Verbreitung von KI-Technologien ergeben, von Expert:innen diskutiert werden und eine Regulierung durch nationale (oder supranationale) Stellen eher früher als später erfolgt.

Dabei ist zu bedenken, dass übermäßige Regulierung den Tech-Giganten wie Microsoft und Google zugutekommt und kleinere Entwickler:innen behindern könnte. Dies könnte zu einem

Quasi-Monopol der Ersteren auf KI-Technologie und den damit verbundenen Marktverzerrungen führen. Das liegt daran, dass die Regulierung tendenziell die bestehenden Marktstrukturen stärkt, indem sie Eintrittsbarrieren erhöht und Kosten verursacht, die für die etablierten Unternehmen am leichtesten zu tragen sind [29]. Dies könnte auch der Grund sein, warum einige Tech-CEOs konstant vor den Gefahren der KI warnen und eine dringende Regulierung fordern: Um die Technologie und die damit verbundenen Gewinne einem kleinen Kreis von Tech-Eliten vorzuenthalten [30]. Das Ziel sollte deshalb ein pragmatisches Regelwerk sein, das die Sicherheit von KI-Technologien und ihren Nutzer:innen gewährleistet, ohne Innovation zu ersticken und KI-Technologien in den Händen einiger weniger zu konzentrieren. Dieser bedeutsame Balanceakt der Regulierung wird die zukünftige Entwicklung von KI-Technologien maßgeblich bestimmen und muss in Angriff genommen werden, bevor die rasche Verbreitung von KI-Technologien vollendete Tatsachen schafft.

Ebenfalls wichtig ist die Frage, wer für die Regulierung der KI verantwortlich sein wird, da verschiedene Akteure auf subnationaler, nationaler oder supranationaler Ebene unterschiedliche Einflussmöglichkeiten auf globale Normen haben. Das Europäische Parlament z. B. hat vor kurzem das so genannte «Artificial Intelligence Act» [31] verabschiedet, ein 100-seitiger Gesetzesentwurf, der KI-Anwendungen, deren Risiken als «inakzeptabel» eingestuft werden, präventiv verbieten soll. Solche inakzeptablen Anwendungen wären beispielsweise biometrische Kategorisierungssysteme, die sensible Merkmale wie Geschlecht, ethnische Zugehörigkeit oder Religion verwenden, oder voraussagende polizeiliche KI, die auf vergangenem kriminellen Verhalten basiert. Für andere KI-Anwendungen mit hohem Risiko sieht derselbe Text die Einholung von Vorabgenehmigungen und Lizenzen vor dem Einsatz in der EU vor.

Darüber hinaus würden Anwendungen mit «begrenztem Risiko» (darunter LLMs wie ChatGPT oder Text-Bild-Generatoren wie Stable Diffusion), Vorschriften unterliegen, welche die Erzeugung von illegalem Material verbieten, sowie einer Transparenzregelung, die ihre Outputs als KI-generiert kenntlich macht.

Am 14. Juni 2023 haben die Europaabgeordneten das «Artificial Intelligence Act» angenommen. Damit ist die EU derzeit eine Vorreiterin bei der Regulierung von KI-Technologien. Nun werden die Gespräche mit den EU-Ländern im Europäischen Rat über die endgültige Form des Gesetzes beginnen. Ziel ist es, bis Ende 2023 eine Einigung zu erzielen. Sollte dieses Gesetz verabschiedet werden, würde es wahrscheinlich

zum ersten supranationalen Regelungsrahmen für KI-Technologien werden, das weltweit Einfluss haben könnte.

Tatsächlich ist die EU ein dermassen bedeutender Markt, dass Technologieunternehmen oft EU-spezifische Vorschriften auf globaler Ebene einhalten, um den Aufwand für die Einhaltung mehrerer Standards zu verringern [32]. Die weiteren Entwicklungen in Bezug auf das EU-Gesetz über künstliche Intelligenz und KI-Regelungen anderer Länder und Institutionen sind daher wichtig, um zu verstehen, in welchem Rechtsrahmen sich KI-Technologien zukünftig entwickeln werden.

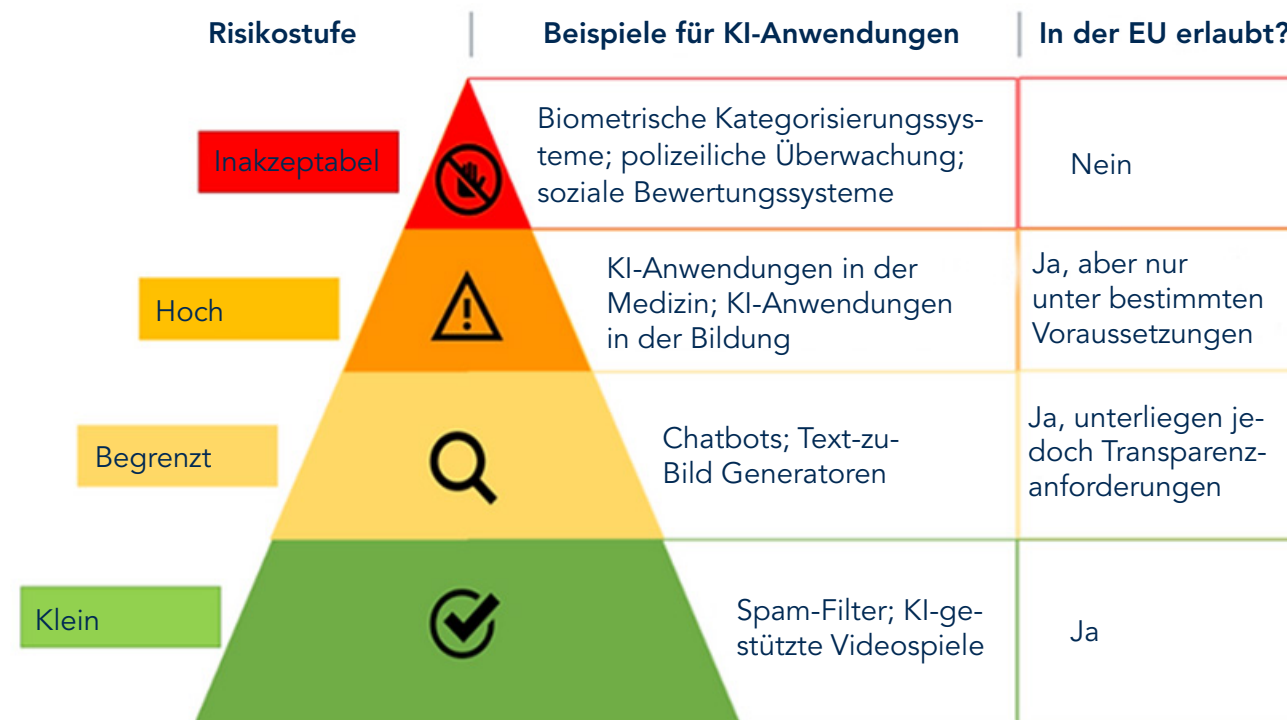


Abbildung 8: Die vier Risikostufen der verschiedenen KI-Anwendungen und die damit verbundenen Regulierungsmassnahmen, die im «Artificial Intelligence Act» der EU vorgeschlagen werden.

Quelle: <https://www.theverge.com/2023/5/11/23719694/eu-ai-act-draft-approved-prohibitions-surveillance-predictive-policing>.

KI und Nachhaltigkeit

Bislang wurde in diesem Beitrag erläutert, wie die derzeitigen KI-Technologien funktionieren, wie sie zu einem Hype-Zyklus führen könnten, welche Vorteile und Gefahren die Technologien mit sich bringen und wie diese angegangen werden müssen. Mindestens eine weitere Frage muss jedoch noch beantwortet werden: Ist KI nachhaltig?

Aufmerksame Leser:innen wissen inzwischen, dass die Antwort kein eindeutiges «Ja» oder «Nein» sein kann. Zunächst einmal ist festzustellen, dass das Training von LLMs eine enorme Rechenleistung benötigt und somit auch ein energieintensiver Prozess ist.

«Der Wasserverbrauch von ChatGPT ist astronomisch: 40 Eingabeaufforderungen verbrauchen einen Liter Wasser.¹»

Es wird geschätzt, dass für das Training der aktuellen Version von ChatGPT (GPT-3) ungefähr 1.3 Gigawattstunden Strom benötigt wurden [4] – genug, um rund 260 Haushalte in der Schweiz ein Jahr lang zu versorgen. Die Energiemenge, die für das Training der komplexeren Version des Chatbots (GPT-4) benötigt wurde, ist nicht bekannt, betrug aber sicherlich ein Vielfaches davon. Darüber hinaus verbraucht die Anwendung der Modelle durch die Nutzer:innen auch nach dem Training weiterhin Strom, wenn auch

in einer Weise, die mit anderen Internetanwendungen vergleichbar und viel weniger intensiv ist als während der Trainingsphase.

Ein weiterer Faktor, der in der Nachhaltigkeitsdiskussion berücksichtigt werden muss, ist der soziale Aspekt während den Trainingsphasen von LLMs. Im Fall von ChatGPT z. B. hat eine Untersuchung des Time Magazins aufgedeckt, dass OpenAI, das Unternehmen hinter dem Chatbot, billige ausgelagerte Arbeitskräfte in Ländern wie Kenia eingesetzt hat, um die Schutzschranken des Modells gegen gewalttätige, sexuelle und hasserfüllte Inhalte zu optimieren [33]. Kenianische Arbeiter:innen erhielten angeblich Löhne zwischen USD 1.30 und 2 pro Stunde, um für OpenAI Textausschnitte zu kennzeichnen, die sexuellen Missbrauch, Gewalt und Hassreden thematisieren.

Der Einsatz billiger, ausgelagerter Arbeitskräfte für die manuelle Kennzeichnung und Zensur verstörender Inhalte ist kein völlig neues Phänomen in der Tech-Industrie; die Facebook-Muttergesellschaft Meta z. B. verwendet ähnliche Methoden für die Moderation ihrer Inhalte. Andere KI-Anwendungen könnten sich für andere, weniger ausbeuterische Methoden entscheiden, um ihre Modelle sicherer zu machen. Das soziale Risiko ist dennoch nicht zu vernachlässigen und muss in der Nachhaltigkeitsdiskussion in Betracht gezogen werden.

¹ KI-Chatbots wie ChatGPT, Google Bard und Bing Chat treiben den Wasserverbrauch ihrer Unternehmen in die Höhe. Der Forscher Shaolei Rein von der University of California fand heraus, dass eine Konversation von 5 bis 50 Aufforderungen mit einem Chatbot 500 Milliliter Wasser verbraucht [34].

Bei der Betrachtung der Nutzen von KI-Technologien für die Umwelt und die Gesellschaft müssen mehrere Aspekte berücksichtigt werden. Forma Futura verfolgt einen integrierten Nachhaltigkeitsansatz, der soziale und Umweltkriterien gleichermaßen einbezieht. Einerseits kann die Technologie unser Leben und unsere Umwelt positiv beeinflussen, indem sie bestimmte alltäglichen Aufgaben erleichtert, die Gesundheitssysteme verbessert und neue Durchbrüche in der wissenschaftlichen Forschung ermöglicht, um nur einige Beispiele zu nennen. Im Bereich Klima und Umwelt könnte KI beispielsweise für die Mustererkennung und die Verknüpfung grosser Datenmengen aus Satellitenbildern und Wetterdaten eingesetzt werden, um präzise Aussagen über Klimaveränderungen zu treffen, oder um Schadstoffe aufzuspüren.

Weitere Beispiele sind die intelligente Vernetzung dezentraler Energiesysteme oder die effiziente Steuerung von Ressourcenströmen für die Kreislaufwirtschaft [35]. Andererseits kann KI auch dazu genutzt werden, Fake News und Hass effizienter zu verbreiten, neue Wege für Betrug und Erpressung zu finden und könnte, wenn sie

von grossen Unternehmen rücksichtslos eingesetzt wird, zum Verlust bestimmter Arbeitsplätze und zu mehr Ungleichheit führen.

Forma Futura ist der Ansicht, dass KI – wenn richtig eingesetzt – das Potenzial hat, zur Lösung einiger der Herausforderungen der heutigen Zeit im Bereich der Nachhaltigkeit beizutragen und auch die Lebensqualität der breiten Bevölkerung zu verbessern. Gleichzeitig werfen die Entwicklung und das Training von derzeit beliebten LLMs wie ChatGPT verschiedene Fragen zur Nachhaltigkeit auf, die es zu berücksichtigen gilt. Ebenfalls muss die rasche Entwicklung der KI-Technologien von notwendigen regulatorischen Erkenntnissen und Aufsicht über KI-basierte Technologien begleitet werden. Andernfalls könnte es zu Lücken bei Transparenz, Sicherheit und ethischen Standards kommen [36]. Forma Futura wird die Entwicklung von KI-Technologien und ihrer Einsatzmöglichkeiten weiter beobachten und auch in Zukunft die Nachhaltigkeitsleistung der individuellen Anwendungen von KI in ihrem Kontext evaluieren sowie mögliche Chancen und Risiken im Engagement mit den jeweiligen Unternehmen thematisieren.

Fazit

KI, oder zumindest der damit verbundene Medienrummel, erobert derzeit die Welt im Sturm. Damit einhergehend sind bestimmte Erwartungen, Ängste und Hoffnungen rund um die Technologie, von denen nicht unbedingt alle fundiert sind. Ziel dieses Papiers war es, einen Überblick über die derzeit populären KI-Technologien zu geben und auf vereinfachte Weise zu erklären, wie sie funktionieren und welche Nutzen und Gefahren sie bergen, um die Grundlagen für eine fundierte Diskussion zu diesem Thema zu schaffen.

LLMs wie ChatGPT, die derzeit populärste Form der KI, sind dank der jüngsten technologischen Fortschritte in der Lage, beeindruckende Leistungen zu vollbringen. Sie könnten damit in nicht allzu ferner Zukunft die Art und Weise transformieren, wie wir leben und arbeiten. Gleichzeitig haben sie aber auch ihre Grenzen und werden mit grosser Wahrscheinlichkeit auch in absehbarer Zeit nicht mit menschlicher Intelligenz, Kreativität und Einfallsreichtum konkurrieren können.

Es ist daher wichtig, Sensationsmeldungen und reisserische Vorhersagen über KI mit Vorsicht zu

geniessen, einen kühlen Kopf zu bewahren und die weiteren Entwicklungen und Anwendungen der Technologie genaustens zu beobachten.

Der künftige Weg der KI-Technologien wird von vielen Faktoren bestimmt werden, unter anderem von der weiteren Entwicklung der Rechenkapazitäten, der Regulierung und der Finanzierung weiterer KI-Forschung, um nur einige zu nennen. In der Zwischenzeit ist es wichtig, KI als das zu sehen, was sie ist: Ein komplexes Werkzeug, das enormes Potenzial hat und sowohl zum Guten als auch zum Schlechten eingesetzt werden kann. Wie bei allen neuen Technologien wie der Blockchain gilt es, die Entwicklung genau zu beobachten und jede Anwendung umfassend zu betrachten, um positive und negative Effekte aufzuzeigen und gegeneinander abzuwägen.

Weiter werden wir Möglichkeiten zur Implementierung von KI-Lösungen zum Nutzen von Forma Futura und ihren Kundinnen und Kunden prüfen und diese in zukünftigen Publikationen erläutern.

Quellenverzeichnis

- [1] Goldman Sachs, "Generative AI could raise global GDP by 7%", Goldman Sachs, 5. April 2023, URL: <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html> [Stand 08.08.2023]
- [2] Gartner, "What Is Artificial Intelligence?", Gartner, 2022, URL: <https://www.gartner.com/en/topics/artificial-intelligence> [Stand 08.08.2023]
- [3] K. Hu, "ChatGPT sets record for fastest-growing user base - analyst note", Reuters.com, 2. Februar 2023, URL: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/> [Stand 08.08.2023]
- [4] The Economist, "Large, creative AI models will transform lives and labour markets", the Economist, 22. April 2023, URL: <https://www.economist.com/interactive/science-and-technology/2023/04/22/large-creative-ai-models-will-transform-how-we-live-and-work> [Stand 08.08.2023]
- [5] E. M. Bender; T. Gebru; A. McMillan-Major; S. Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", Conference on Fairness, Accountability, and Transparency, 3–10. März 2021, 14 Seiten, URL: <https://doi.org/10.1145/3442188.3445922> [Stand 08.08.2023]
- [6] C. S. Smith, "Hallucinations Could Blunt ChatGPT's Success. OpenAI says the problem's solvable, Yann LeCun says we'll see", IEEE Spectrum, 13. März 2023, URL: <https://spectrum.ieee.org/ai-hallucination> [Stand 08.08.2023]
- [7] K. W. Zimmermann, "Tiere und Menschen unterscheiden sich im Humor", Bilanz, 15. Juni 2018, URL: <https://www.handelszeitung.ch/panorama/tiere-und-menschen-unterscheiden-sich-im-humor> [Stand 08.08.2023]
- [8] Gartner, "Gartner Hype Cycle: Wie man Technologie-Hype interpretiert", Gartner, 2023, URL: <https://www.gartner.de/de/methoden/hype-cycle> [Stand 08.08.2023]
- [9] Retresco, "KI-Winter – Definition & Einleitung", Retresco, 2023, URL: <https://www.retresco.de/lexikon/ki-winter> [Stand 08.08.2023]
- [10] B. Lutkevich, "What is AI winter?", Techtargget.com, 2022, URL: <https://www.techtargget.com/searchenterpriseai/definition/AI-winter> [Stand 08.08.2023]
- [11] The Economist, "Large, creative AI models will transform lives and labour markets", the Economist, 22. April 2023, URL: <https://www.economist.com/interactive/science-and-technology/2023/04/22/large-creative-ai-models-will-transform-how-we-live-and-work> [Stand 08.08.2023]
- [12] W. Knight, "OpenAI's CEO Says the Age of Giant AI Models Is Already Over", Wired.com, 17. April 2023, URL: <https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/> [Stand 08.08.2023]
- [13] J. Kahn, "Meta's chief A.I. scientist calls A.I. doomers 'preposterous' and predicts LLMs are just a passing fad", Fortune.com, 14. Juni 2023, URL: <https://fortune.com/2023/06/14/metas-chief-a-i-scientist-calls-a-i-doomers-preposterous-and-predicts-llms-are-just-a-passing-fad/> [Stand 08.08.2023]
- [14] World Health Organization, "Health workforce", World Health Organization, 2023, URL: https://www.who.int/health-topics/health-workforce#tab=tab_1 [Stand 08.08.2023]
- [15] G. Novakovsky; N. Dexter; M. W. Libbrecht; W. W. Wasserman; S. Mostafavi, "Obtaining genetics insights from deep learning via explainable artificial intelligence", Nature Reviews Genetics, Vol. 24, 3. Oktober 2022, S. 125 – 137, URL: <https://www.nature.com/articles/s41576-022-00532-2> [Stand 08.08.2023]
- [16] NASA, "NASA Turns to AI to Design Mission Hardware", NASA, 9. Februar 2023, URL: <https://www.nasa.gov/feature/goddard/2023/nasa-turns-to-ai-to-design-mission-hardware> [Stand 08.08.2023]
- [17] H. Field, "The first fully A.I.-generated drug enters clinical trials in human patients", CNBC.com, 29. Juni 2023, URL: <https://www.cnbc.com/2023/06/29/ai-generated-drug-begins-clinical-trials-in-human-patients.html> [Stand 08.08.2023]
- [18] NBER, "Measuring the Productivity Impact of Generative AI", National Bureau of Economic Research, 6. Juni 2023, URL: <https://www.nber.org/digest/2023/6/measuring-productivity-impact-generative-ai> [Stand 08.08.2023]
- [19] The Economist, "Your job is (probably) safe from artificial intelligence", the Economist, 7. Mai 2023, URL: <https://www.economist.com/finance-and-economics/2023/05/07/your-job-is-probably-safe-from-artificial-intelligence> [Stand 08.08.2023]
- [20] B. Edwards, "Artists stage mass protest against AI-generated artwork on ArtStation", Ars Technica, 15. Dezember 2022, URL: <https://arstechnica.com/information-technology/2022/12/artstation-artists-stage-mass-protest-against-ai-generated-artwork/> [Stand 08.08.2023]

Quellenverzeichnis

- [21] J. Vincent, "Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement", The Verge, 6. Februar 2023, URL: <https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion> [Stand 08.08.2023]
- [22] K. Trendacosta; C. Doctorow, "AI Art Generators and the Online Image Market", Electronic Frontier Foundation, 13. April, URL: <https://www.eff.org/deeplinks/2023/04/ai-art-generators-and-online-image-market> [Stand 08.08.2023]
- [23] A. Kraft, "Microsoft shuts down AI chatbot after it turned into a Nazi", CBS News, 25. März 2016, URL: <https://www.cbsnews.com/news/microsoft-shuts-down-ai-chatbot-after-it-turned-into-racist-nazi/> [Stand 08.08.2023]
- [24] D. Alba, "OpenAI Chatbot Spits Out Biased Musings, Despite Guardrails", Bloomberg.com, 8. Dezember 2022, URL: <https://www.bloomberg.com/news/newsletters/2022-12-08/chatgpt-open-ai-s-chatbot-is-spitting-out-biased-sexist-results> [Stand 08.08.2023]
- [25] R. Satter, "FBI says artificial intelligence being used for 'sextortion' and harassment", Reuters.com, 8. Juni 2023, URL: <https://www.reuters.com/world/us/fbi-says-artificial-intelligence-being-used-sextortion-harassment-2023-06-07/> [Stand 08.08.2023]
- [26] A. Mok; J. Zinkula, "ChatGPT may be coming for our jobs. Here are the 10 roles that AI is most likely to replace", Insider.com, 4. Juni 2023, URL: <https://www.businessinsider.com/chatgpt-jobs-at-risk-replacement-artificial-intelligence-ai-labor-trends-2023-02?op=1&r=US&IR=T> [Stand 08.08.2023]
- [27] M. Bohannon, "Lawyer Used ChatGPT In Court—And Cited Fake Cases. A Judge Is Considering Sanctions", Forbes.com, 8. Juni 2023, URL: <https://www.forbes.com/sites/mollybohannon/2023/06/08/lawyer-used-chatgpt-in-court-and-cited-fake-cases-a-judge-is-considering-sanctions/?sh=7620ff437c7f> [Stand 08.08.2023]
- [28] M. D. Meier, "Wir werden massive gesellschaftliche Umwälzungen erleben", Handelszeitung, 15. August 2023, URL: <https://www.handelszeitung.ch/tech/wir-werden-massive-gesellschaftliche-umwaelzungen-erleben-627471> [Stand 29.08.2023]
- [29] The Economist, "Why tech giants want to strangle AI with red tape", the Economist, 25. Mai 2023, URL: <https://www.economist.com/business/2023/05/25/why-tech-giants-want-to-strangle-ai-with-red-tape> [Stand 08.08.2023]
- [30] A. Fichter, "Wer dem KI-Hype verfällt, stärkt die Macht der Big-Tech-Chefs", Republik.ch, 5. Juli 2023, URL: <https://www.republik.ch/2023/07/05/wer-dem-ki-hype-verfaellt-staerkt-die-macht-der-big-tech-chefs> [Stand 08.08.2023]
- [31] European Commission, "Regulation of the European Parliament and the Council: Laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and mending certain Union Legislative Acts", the European Parliament, 21. Mai 2021, URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206> [Stand 08.08.2023]
- [32] J. Vincent, "EU draft legislation will ban AI for mass biometric surveillance and predictive policing", the Verge, 11. Mai 2023, URL: <https://www.theverge.com/2023/5/11/23719694/eu-ai-act-draft-approved-prohibitions-surveillance-predictive-policing> [Stand 08.08.2023]
- [33] B. Perrigo, "Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic", Time.com, 18. Januar 2023, URL: <https://time.com/6247678/openai-chatgpt-kenya-workers/> [Stand 08.08.2023]
- [34] Pengfei Li, Jianyi Yang, Mohammad A. Islam, Shaolei Ren, "Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models", 6. April 2023, URL: <https://arxiv.org/abs/2304.03271> [Stand 08.08.2023]
- [35] Deutsche Bundesstiftung Umwelt, "Künstliche Intelligenz: Können wir mit Rechenleistung unseren Planeten retten?", Deutsche Bundesstiftung Umwelt (DBU), 2023, URL: https://reset.org/app/uploads/2020/09/RESET_KI_Greenbook_01-1.pdf [Stand 08.08.2023]
- [36] R. Vinuesa; H. Azizpour; I. Leite; M. Balaam; V. Dignum; S. Domisch; A. Felländer; S. D. Langhans; M. Tegmark; F. F. Nerini, "The role of artificial intelligence in achieving the Sustainable Development Goals", Nature Communications, Vol. 11, 13. Januar 2020, URL: <https://www.nature.com/articles/s41467-019-14108-y> [Stand 08.08.2023]

Autor



Nicolas Merz
Nachhaltigkeitsanalyst