ChatGPT und die Dynamik sozio-technischer Verfügbarmachung ‚natürlicher Sprache'[1]

Andreas Bischof

1. ChatGPT und das Verfügbarmachungsversprechen der KI

„Ich schlage vor, die Frage ‚Können Maschinen denken?' zu untersuchen", beginnt Alan Turing seinen berühmten Aufsatz „Computing Machinery and Intelligence" (Turing 1950, Übers. AB). Da „Denken" schwer zu definieren sei, schlägt Turing vor, „die Frage durch eine andere zu ersetzen, die eng mit ihr verwandt ist und in relativ eindeutigen Worten ausgedrückt wird" (ebd., Übers. AB). Turings Operationalisierung von künstlicher Intelligenz als Fähigkeit von Computern, Texte zu produzieren, die ununterscheidbar von menschlichen sind, findet im Produkt ChatGPT und dem Hype seit seiner Einführung 72 Jahre später scheinbar ihre Erfüllung. Es werden ontologische und daran anschließend ethische Fragen gestellt, ob Systeme, die so gut an Kommunikation teilhaben können, nicht auch Bewusstsein entwickeln könnten – ja zwangsläufig müssten –, und wie man dann oder präventiv schon jetzt mit ihnen umgehen müsse. Eine weitere wiederkehrende Figur in Feuilleton und Wirtschaftsteil ist die Aktualisierung der Frage des Automatisierungsdiskurses (Benanav 2021), wessen Arbeitskraft solche Technik nun obsolet mache, bzw. inwiefern die mögliche Produktivitätssteigerung durch generative KI ganze Branchen von Arbeit freisetzen könne. Gestützt werden diese Diskussionen auch durch die Proklamationen der Anbieter generativer KI, die an die Logik ‚disruptiver' Innovationen anknüpfen, und bestehende Kategorien und Geschäftsmodelle als ins Wanken geraten darstellen.

Der Diskurs um ChatGPT ist ein hervorragendes Beispiel für die Figur von *Künstlicher Intelligenz als Verfügbarmachungsversprechen*, die in den vergangenen Jahrzehnten manchen Investitions- und Entwicklungskreislauf mobilisiert hat. Wie schon der Name ‚Künstliche Intelligenz' anzeigt, liegt

1 Die Fertigstellung und Qualität dieses Kapitels hat sehr von der Beratung und Diskussion der Herausgeber:innen profitiert. Vielen Dank insbesondere an Jan Gärtner für die Betreuung, und an Philipp Zeltner für die wertvollen Anmerkungen!

der Fokus dieser Figur auf dem Vergleich zu typisch menschlichen Aufgaben, und wird heutzutage eng mit der Qualität der Produkte, dem Output der Maschinen verknüpft. Als der Begriff 1955 gewählt wurde, gab es verschiedene Bezeichnungen des Forschungsfeldes, die je auf ihren konzeptionellen Ursprung verwiesen: Kybernetik, Automatentheorie und komplexe Informationsverarbeitung (McCorduck 2004). Als John McCarthy sich im Förderantrag für die Sommerkonferenz in Dartmouth für den neuen Oberbegriff ‚Artifical Intelligence' entschied, lieferte er das Versprechen explizit mit: Computer sollen in die Lage versetzt werden „[...] Sprache zu benutzen, Abstraktionen und Konzepte zu bilden, Probleme zu lösen, die bisher dem Menschen vorbehalten waren" (McCarthy et al. 1955, Übers. AB).

Es ist kein Zufall, dass Turing und McCarthy so explizit auf menschliche Sprache Bezug nehmen, denn sinnverstehende Kommunikation ist für Computer nach wie vor – trotz des anderslautenden Optimismus – nur sehr eingeschränkt verfügbar. Gleichzeitig ist das Ziel der Überwindung dieser Unverfügbarkeit konstitutiv zum einen als erkenntnistheoretische Grundausrichtung des Forschungsfeldes KI seit dessen Gründung (vgl. Epstein et al. 2009), und zum anderen für die Mobilisierung und Alllokation von Ressourcen für Forschung, Entwicklung und Betrieb großer KI-Systeme auch heutzutage. In wiederkehrenden Schleifen gehen seit McCarthys Antrag populäre Paradigmen der KI-Forschung mit förderungspolitischen Konjunkturen und der öffentlichen Wahrnehmung des Feldes KI einher (Breiter 1995). Diesen Aspekten geht das Kapitel nach und rekonstruiert sie als *Verfügbarmachungsdynamiken* der sozio-technischen Verarbeitung ‚natürlicher Sprache' durch KI-Systeme. Dabei werden sowohl die technischen Umsetzungsmöglichkeiten dieses Versprechens als auch die materiellen Bedingungen solcher Systeme, wie insb. der Ressourcenverbrauch, aber auch die verdeckten sozialen Bedingungen, wie die strukturelle Ausbeutung von Arbeiter:innen, beleuchtet.

Der in den Anfangsjahren von ChatGPT beobachtbare gesellschaftliche Diskurs zu solchen KI-Systemen blendet insbesondere die letztgenannten Aspekte der Verfügbarmachungsdynamiken aus: Durch den Fokus auf die Texte und Benutzeroberfläche von ChatGPT wird unsichtbar, wie die Verfügbarmachung menschlicher Sprache in solchen Systemen konkret umgesetzt wird. Durch einen verengten Fokus auf ChatGPT als *Interface* und dessen konkreten *Output*, verbleiben solche Perspektiven buchstäblich oberflächlich, und setzen damit sogar *das Versprechen der technischen Verfügbarmachung* der Anbieter fort – indem sie statt der handfesten materiellen, ökologischen und gesellschaftlichen Bedingungen der Möglichkeit

von generativer KI die Möglichkeit beseelter Maschinen diskutieren (Nezik 2023). Crawford kritisierte solche objektzentrierten Thematisierungen von KI für deren Determinismus und normative Implikationen als „zutiefst ahistorische Sichtweisen, die die Macht ausschließlich in der Technik selbst verorten" (2021, Übers. AB).

Das Versprechen der Überwindung dieser Grenze ist zentral für die gesellschaftliche *Mobilisierung und Alllokation von Ressourcen* für KI-Forschung und -Entwicklung, sowie die Finanzierung von Unternehmen, die diesem Zweck dienen. Die Darstellung von ChatGPT und verwandten Systemen als technologischen oder gar menschheitsgeschichtlichen Dammbruch kann insgesamt als anschlussfähig an ein Erleben von Un/Verfügbarkeit (siehe Einleitung, in diesem Band), als kultureller Kit und persuasive Rhetorik, die zu den zentralen Bedingungen der Möglichkeit großer sozio-technischer Projekte wie Robotik und KI gehört, verstanden werden. Wie die Verfügbarmachung tatsächlich geschieht, und zu welchen Anschlussproblemen und unintendierten Folgen das führt, wird dadurch unsichtbar.

Anhand des seltsam unempirischen Diskurses der scheinbar magisch immer weiter wachsenden Super-KI (vgl. Bostrom 2014) lassen sich relevante gesellschaftliche Fragen gar nicht stellen. Stattdessen soll an dieser Stelle die sozio-technische Genese solcher KI-Systeme rekonstruiert werden, um abschließend auf Spannungen des Un/Verfügbaren durch die maschinelle Verfügbarmachung von ‚natürlicher Sprache' einzugehen (vgl. 4.). In einem ersten Schritt werden dafür am Beispiel von ChatGPT zwei Instanzen der Verfügbarmachung von menschlicher Sprache für Computer rekonstruiert: die zugrundeliegenden Daten von menschlicher Kommunikation im Internet, und die unsichtbare menschliche Arbeit im Prozess der Modellbildung und Definition gewünschter Outputs (vgl. 2.). Anschließend werden die ermöglichenden Ressourcen und die Wirtschaftsform der Technologieunternehmen anhand zweier bereits beobachtbarer Folgen von großen KI-Sprachmodellen wie ChatGPT analysiert (vgl. 3.). Dabei wird die Bindung knapper Güter als notwendige materielle Basis für Large Language Models (LLM) beleuchtet, und die Transformation von Arbeit unter Bedingungen des Plattformkapitalismus diskutiert.

2. Sozio-technische Verfügbarmachung ‚natürlicher Sprache'

Im Folgenden werden die Grundprinzipien von Systemen wie ChatGPT erläutert, um aufzuzeigen, wie die Verfügbarmachung von menschlicher

Kommunikation technisch angestrebt wird. Darauf basierend werden die Herkunft und Art der sogenannten Trainingsdaten sowie die Rolle – oftmals schlecht bezahlter – Arbeit in Herstellung und Betrieb dieser Systeme diskutiert, die wiederum die Möglichkeitsbedingungen für diesen technischen Versuch der Verfügbarmachung darstellen.

ChatGPT ist ein automatischer Textgenerator, der je nach Abonnement durch die Nutzenden die LLM GPT-3.5 und GPT-4 nutzt. ChatGPT ist dabei sowohl der Name, unter dem das Produkt vermarktet wird, als auch der Benutzeroberfläche in Form eines Chatbots im Browser der Nutzenden. LLMs sind maschinelle Lernmodelle, die Texte in ‚natürlicher Sprache' verarbeiten und generieren können. ChatGPT und die zugrundeliegenden LLMs bauen auf mehreren wichtigen Fortschritten im Bereich des ‚Natural Language Processing' (NLP) auf, die insbesondere seit Anfang der 2010er Jahre zu verzeichnen sind. Das Prinzip hinter dieser maschinellen Textbearbeitung ist eine Übertragung von Wörtern und Wortbestandteilen in nummerische Werte. Die so genannte ‚Tokenifizierung' von ‚natürlicher Sprache' besteht im Wesentlichen in einer Verkleinerung und Vereinzelung der semantischen Einheiten: Aus ganzen Sätzen werden Worte, oder Wortbestandteile, die anschließend in beliebigen Kombinationen ins Verhältnis gesetzt werden können. Je nach anschließender rechentechnischer Methode können diese Bestandteile dann im engeren Kontext ihres Auftretens (bspw. auch unter Berücksichtigung von Sequenzialität), oder aber auch völlig losgelöst voneinander weiterbearbeitet werden. Je größer der erfasste Gesamtkorpus ist, und mit je mehr Rechenleistung in mehreren Anläufen an diesem Korpus gearbeitet wurde, desto leichter lassen sich durch komplexe Optimierungsfunktionen in sog. ‚künstlichen neuronalen Netzwerken' die nummerischen Werte – wie bspw. vorhergesagte wahrscheinliche semantische Nähe eines Tokens zum anderen – auch mit der Bedeutungsebene des Textes für Menschen korrelieren.

Um es etwas bildhafter an einem Beispiel aus einem New York-Times-Artikel (Collins 2023) zu zeigen: Für den Satzanfang „LeBron James is..." können Sprachmodelle anhand der Daten für Menschen wahrscheinliche sinnvolle Fortführungen – wie „American", „NBA", „professional", oder „iconic" – vorschlagen, und diesen Optionen jeweils eine Punktzahl zuweisen, die angibt, wie wahrscheinlich es ist, dass das jeweilige Wort als nächstes kommt. Diese Funktion wird Wort für Wort wiederholt, und das Modell bildet so schließlich den Satz: „LeBron James is an American professional basketball player for the Los Angeles Lakers of the National Basketball Association (NBA)" (ebd.).

LLM wie ChatGPT *verstehen* also *nicht* semantisch, was sie an Output erzeugen, sondern stellen eine statistische Beziehung zwischen Wortbestandteilen und der Wahrscheinlichkeit deren Auftretens in Abhängigkeit zueinander her – die wiederum anhand menschlicher Erwartungen, durch Entscheidungen der Programmierer:innen, und mit großen Mengen aufgezeichneter Kommunikation, den so genannten Trainingsdaten, geformt sind. Die technische Verfügbarmachung von menschlicher Kommunikation folgt also dem Umweg der stochastischen Erfassung und Vorhersage von semantischer Nähe einzelner Wörter in großen Textmengen. Es handelt sich um eine rechentechnische *Annäherung* an semantische Bedeutung, keinesfalls aber um ein Sinnverstehen. Esposito (2017) schlägt vor, anhand Luhmanns überzeugender Operationalisierung von Kommunikation in Äußerung, Information und Verstehen, den LLM die Fähigkeit zuzurechnen, durchaus kontingente *Äußerungen* zu erzeugen, die menschlichen Beobachter:innen sinnhaft erscheinen, deren *Information*sgehalt aber tatsächlich fundamental von dem menschlicher Kommunikation abweicht. Das stochastische Prinzip, ein Zeichen mit anderen Zeichen innerhalb oder sogar außerhalb des Systems in Beziehung zu setzen, ermöglicht eine neue Qualität von oberflächlicher, syntaktischer Mustererkennung und -umwandlung durch Maschinen, aber kein *Verstehen*.

Damit ein LLM in die Lage versetzt wird, solche formell kompetenten Ausgaben zu erzeugen, müssen in dessen Genese eine ganze Reihe von heterogenen Akteuren, Ressourcen und Methoden koordiniert werden, darunter die Ausgangsdaten, die iterative Modellentwicklung, die Arbeit von Programmierer:innen, die rekursive Bearbeitung der anfallenden Nutzer:innendaten, etc. Für eine Darstellung dieses Prozesses verbietet sich also eine dualistische Konzeption von ‚sozial' und ‚Technik' oder ‚KI' und ‚menschlich', wenn es darum gehen soll, das *Zusammenwirken* dieser heterogenen Komponenten zu beschreiben. Am Beispiel von LLM wie ChatGPT ist dieses Zusammenspiel besonders evident. Denn deren überzeugende Fähigkeit Textausgaben zu produzieren, die menschlichen Nutzer:innen sinnhaft und hilfreich erscheinen, ließe sich gar nicht anders denn als Ergebnis verteilter Handlungsträgerschaft (Rammert und Schulz-Schaeffer 2002) erklären: Die Programmierung von ChatGPT enthält keine formelle Sprachkompetenz (bspw. im Sinne einer ‚eingebauten' Syntax), es gewinnt seine funktionale Sprachkompetenz (Mahowald et al. 2023) aus einer kombiniert induktiv-deduktiven, extrem rechenintensiven Methode, sehr große Datenmengen menschlicher Kommunikation (vgl. 2.1) zu analysieren und Vorhersagen über wahrscheinlich folgende Sprach-Bestandteile

zu errechnen. Dabei wird die Ausgabeschicht permanent verbessert und angepasst, um in spezifischen, ökonomisch verwertbaren Kontexten ‚gute' Ergebnisse zu erzielen (vgl. 2.2).

2.1 Herkunft und Verarbeitung der ‚Trainingsdaten' für LLM

Das zunehmende Anfallen, Aggregieren und gezielte Auswerten von digitalen Daten, die direkt mit menschlichen Aktivitäten verknüpft sind – wie bspw. finanzielle Transaktionsdaten, Nutzungsdauer von Services, physikalische Orte der Nutzung, von Nutzenden erstellte Inhalte, Interaktionen mit Inhalten und anderen Nutzer:innen – stellt eine zentrale Entwicklung im digitalen Kapitalismus dar (vgl. insb. Zuboff 2018) und ist untrennbar mit KI-Methoden verknüpft: Große Datenmengen sind nur mit KI-Methoden für Menschen wirklich nutzbar, und KI-Produkte wie ChatGPT sind nur durch große Datenmengen sinnvoll trainierbar. Die Verfügbarmachung von menschlichen Alltagsaktivitäten für die maschinelle Bearbeitung – und insbesondere Vorhersage – basiert auf der Auswertung von Daten, die durch genau solche Aktivitäten anfallen. Angesichts der Allgegenwart digitaler Spuren menschlicher (und nicht-menschlicher) Aktivitäten (‚Big Data') vermag das zunächst tautologisch klingen, markiert im Umgang mit Daten und ihrem Zweck aber einen epochalen Unterschied. Ochs hat diesen Unterschied als ‚Umkipppunkt' bezeichnet, ab dem Daten über Vergesellschaftungsprozesse gezielt gesammelt und zur rekursiven ‚Verbesserung' angewendet werden können (Ochs 2019, 215).

Oftmals wird dieser konstitutive Zusammenhang von Big Data und den jüngeren KI-Fortschritten damit beschrieben, dass Daten der neue Rohstoff seien, der die Grundlage zur Erstellung digitaler Produkte und Dienstleistungen bilde. Dass Daten ‚das neue Öl des digitalen Kapitalismus' seien (siehe z.B. Spitz 2017), ist ebenso irreführend wie hinweisgebend. Irreführend ist die Metapher dahingehend, dass digitale Daten, insbesondere solche die scheinbar ‚nebenbei', aber eben nicht unabsichtlich bei der Nutzung digitaler Services anfallen, zwar Voraussetzung zur Erzeugung von Mehrwert und damit Geschäftsmodelle im digitalen Kapitalismus sind, aber eben kein *natürlicher* Rohstoff oder Derivat dessen, sondern gezielt sozio-technisch hergestellt. Digitale Daten sind also niemals ‚roh' (Gitelman 2013) oder einfach nur ‚da'. Ähnlich wie beim ‚Rohöl' zeigt sich hier ein Sprachgebrauch, der diesen Zusammenhang eher verschleiert: Rohöl ist

das geförderte Öl. In beiden Fällen ist also *extraktive* menschliche Arbeit verrichtet worden, um den Rohstoff als Rohstoff für die Produktion nutzbar zu machen.[2] Technologiefirmen wie der ChatGPT-Anbieter OpenAI behandeln digitale Daten von Internetnutzer:innen in derselben Weise wie einen natürlichen Rohstoff, wie es klassische, energieintensive Industrien der ersten Moderne mit Bodenschätzen tun: Die digitalen Daten, die ChatGPT zugrunde liegen, werden aus anderen Quellen und Zusammenhängen extrahiert, um sie kommerziell zu nutzen, wobei dafür keine Kompensation der Urheber:innen, und so gut wie keine Übernahme möglicher Folgekosten stattfindet. Natürlich können Daten als abstraktes Gebilde oder Werkzeug betrachtet werden, aber sie haben je spezifische Produktionsbedingungen, die wiederum folgenreich für ihre weitere Verwendung sind.

Am Beispiel von ChatGPT scheinen die Herkunft der Daten und die Implikationen ihrer Verarbeitung auf den ersten Blick viel weniger problematisch als bspw. im Kontext von tragbaren Sensoren, die Vitalwerte erfassen, denn ChatGPTs Trainingsdaten beruhen im Wesentlichen auf Texten aus dem Internet. Die exakte Größe und Zusammensetzung der Trainingsdatensets für die OpenAI-Modelle sind nicht publik, übersteigen aber in jedem Fall die Menge an Text, die menschliche Leser:innen in einer Lebensspanne erfassen, geschweige denn memorieren könnten, deutlich. Die Trainingsdaten für ChatGPT bestehen aber nicht vorwiegend aus literarischen Texten oder redigierten Artikeln wie der englischsprachigen Wikipedia. Der größte Teil der Trainingsdaten wurde automatisiert von Websites, Foren und Social Media-Plattformen abgefasst. Die wichtigste Grundlage der Daten von ChatGPT sind *automatisierte Sammlungen* von Texten aus dem WWW, deren Vorverarbeitung und Filterung automatisch geschieht, *ohne* dass Menschen diese Datensätze über Stichproben hinaus überprüfen. Der größte Einzeldatensatz in ChatGPTs Trainingsdaten ist der „Common Crawl". Dieser wird durch kontinuierliches ‚Crawlen' des Webs erstellt, wobei Milliarden von Webseiten und deren Inhalte automatisch gesammelt werden. Webcrawling ist fester Bestandteil der Ermöglichung von vielen Webservices und insbesondere von KI-Anwendungen, aber auch von Forschung über das Netz. Gleichzeitig ist es trotz dieser großen Bedeutung nur sehr spärlich verrechtlicht, was zwei Probleme nach sich zieht. Erstens besteht die Praxis aus dem nicht konsensualen Abgreifen von Inhalten auf privat gehosteten Websites, auch wenn einige Websitebetreiber:innen das Crawlen ihrer Inhalte nicht erlauben und technische Schutzmaßnahmen

2 Vielen Dank für diesen Hinweis an Philipp Zeltner.

einsetzen, um Webcrawler zu blockieren oder einzuschränken. Zweitens sind Inhalte und Struktur der so gewonnen Daten nicht auf problematische Aspekte wie strafrechtlich relevante Inhalte überprüft oder gar bereinigt. Im Gegenteil sind in direkter Folge der nur scheinbar kostenlosen und nicht-reaktiv gewonnenen Daten ‚biases' in Inhalten und Struktur der Daten und fehlende Rechenschaft ein *Strukturmerkmal* dieser Datensätze, was die kritische Technikforschung und die ethische Begleitforschung zu NLP bereits seit Jahren öffentlich kritisieren (Bender et al. 2021). Dass im „Common Crawl" problematische Inhalte, wie bspw. ohne Erlaubnis der dargestellten Personen veröffentlichte Darstellungen von Sex, enthalten sind, ist unter den Nutzer:innen solcher Datensätze bekannt. Eine explorative Studie fand durch eine teilautomatisierte Auswertung von Stichproben im „Common Crawl" in ca. 2% der Inhalte explizite Darstellungen sexueller Gewalt, und in etwa 17% der Inhalte „hate speech", also Gewalt und verbale Ausdrücke gruppenbezogener Menschenfeindlichkeit (Luccioni und Viviano 2021).

Es kann deswegen nicht überraschen, dass ein LLM, das auf solchen Daten trainiert wird, dazu neigt, auch solche Inhalte wiederzugeben. Das prominenteste Beispiel dafür ist vermutlich ‚Tay', ein von Microsoft entwickelter Chatbot auf Basis eines LLM, der am 23. März 2016 als Twitter-Account an die Öffentlichkeit trat. Er verursachte unerwünschte PR, als er schon nach wenigen Minuten damit begann, anzügliche und beleidigende Tweets zu verfassen, was Microsoft zwang, den Dienst nur 16 Stunden nach seinem Start wieder abzuschalten (Neff und Nagy 2016). Die ersten Tweets drehten sich noch um belanglose Themen wie Prominente, Horoskope und Tiere. Bereits nach kurzer Zeit begann Tay allerdings damit, sich sexistisch, rassistisch und extremistisch zu äußern. In Microsofts anschließender Krisen-PR wurde das problematische Verhalten von Tay als „soziales" statt als technisches Problem gerahmt und die Verbindung von Training auf misogynen und rassistischen Inhalten und der Ausgabe solcher Texte durch den Chatbot nicht erwähnt. Dabei hat das lernende Modell von Tay genau das gemacht, was ihm aufgetragen war: Die typischen Inhalte von Twitter, inklusive des schlimmsten Rassismus und Sexismus der Nutzenden, sehr schnell nachzuahmen (ebd., 4922).

2.2 Training, Customizing & Filtern von automatisch generierten Inhalten durch Menschen

Eine der wesentlichen technischen Weiterentwicklungen von ChatGPT gegenüber anderen LLM ist, dass der Anbieter OpenAI aus Fällen wie Tay gelernt hat. OpenAI ist vor allem deshalb so erfolgreich in der kommerziellen Anwendbarkeit seiner LLM, weil es das Generieren problematischer Texte auf der großen Datenbasis *nachträglich* einhegt. Dem Prinzip des unüberwachten Lernens auf großen Datenmengen tritt hier eine menschliche Ergänzung zur Seite: Denn nicht-verstehenden LLM muss ‚beigebracht' werden, nicht alle Strukturen, die sie in den Spuren menschlicher Kommunikation finden, unterschiedslos wiederzugeben. Man könnte sagen, dass der technischen Verfügbarmachung hier die Kontextsensitivität und Situiertheit menschlichen Wissens zur Seite gestellt wird – teilweise, wie an dem Beispiel der Filter zu sehen ist (s.u.), auch außerhalb und zusätzlich des generativen Modells als zusätzliche sozio-technische Schicht aus menschlich klassifizierten Verbotswortlisten und deren automatischer Zuordnung (und ggf. Sperrung) durch maschinelles Lernen. An mindestens drei zentralen Instanzen des Entwicklungsprozesses geben Menschen also erwünschte und unerwünschte Inhalte vor, bzw. bewerten oder verhindern Ausgaben des Systems: beim Training des Modells, durch das Anpassen auf einen (kommerzialisierbaren) Kontext, und beim nachträglichen Filtern.

Die Datenbasis von LLM (vgl. 2.1) ist deswegen so entscheidend, weil LLM im ersten Schritt auf einer sehr freien Form des maschinellen Lernens beruhen. LLM ‚lernen' die Trainingsdaten nicht auswendig, sondern identifizieren latente, stochastische Muster in ihnen, die Menschen gar nicht zugänglich sind und treffen auf Basis dieser Vorhersagen (hier: für den nächst-wahrscheinlichen Satzteil). Die zentrale sozio-technische Verfügbarmachung dieser und ähnlicher Verfahren liegt also in der Operationalisierung uns in unserer Alltagskommunikation unzugänglichen stochastischen Mustern großer Textmengen, die hinreichend gut mit von Menschen gemeintem bzw. verstandenem Sinn korrelieren, dass eine neue Form ‚künstlicher Kommunikation' möglich wird (Esposito 2017, *siehe oben*).

Auf diese sehr freie Modellbildung folgt anschließend eine viel enger begleitete Trainingsform. Etwa 40 OpenAI-Mitarbeiter:innen haben zur Verfeinerung des LLM in einem überwachten Lernprozess Strukturen für erwünschte Ausgaben vorgegeben. Dafür werden zuerst Paare möglicher Anfragen von späteren Nutzer:innen und die dazugehörigen, gewünschten Textausgaben von menschlichen Programmierer:innen vorgegeben. Mit

diesen Informationen werden daraufhin erneute Durchläufe unüberwachten Lernens absolviert, um diese gewünschten Antwortstrukturen in das Modell zu integrieren. Abschließend werden dem System die Fragen offen, ohne vorgegebene Antworten, gestellt, und die Entwickler:innen klassifizieren dann die Textausgaben nach Güte. So werden für jede Frage mehrere Antworten vom System generiert, und durch die Überprüfenden in die Reihenfolge von gut bis schlechter sortiert. Durch diesen Vorgang wird ein sogenanntes ‚Belohnungssystem' trainiert. In diesem Trainingsschritt wird dem LLM mathematisch eine Agentenfunktion gegeben, um anhand der vorgegebenen Sortierung selbstständig einen bestimmten Wert der kumulierten ‚Belohnung' zu maximieren. Die von den menschlichen Eingaben geformte Optimierungsfunktion wird anschließend zu einer eigenen Schicht im LLM, die jeden Output, also jeden generierten Satzteil, bewertet, bevor er ausgegeben wird.

Dementsprechend relevant ist es, nach welchen Kriterien die menschlichen Trainer:innen die Optimierungsfunktion geformt haben. Dieser Schritt wird ‚Customizing' genannt, und hegt die grundlegend sehr breiten Fähigkeiten des Systems für bestimmte, kommerzialisierbare Kontexte ein. Teil des Geschäftsmodells hinter der extrem kostenintensiven GPT-Architektur der LLM (vgl. 3.1) von OpenAI ist die Möglichkeit, das so trainierte Grundmodell durch verschiedene Chatbots für unterschiedliche Verwertungskontexte zu spezialisieren, und dadurch auf demselben LLM mehrfach Geld erlösen zu können. OpenAI bietet in der kostenpflichtigen Abonnementversion mittlerweile mehrere Dutzend spezialisierte Chatbots auf Basis seiner LLM an – von einem Datenanalysetool bis zu „Planty", der „lustigen und freundliche Pflanzenpflegeassistentin" (OpenAI 2024).

Am Blick auf den Wechsel aus unüberwachtem und überwachtem Lernen in der Modellbildung zeigt sich die Bedeutung menschlicher Sinnzuschreibung und Einschreibung von Erwartungen in die Struktur der Ausgaben von LLM. Im Hinblick auf die Verfügbarmachung von ‚natürlicher Sprache' sind diese Bearbeitungsschritte besonders entscheidend, denn hier wird dem gewissermaßen ‚rohen' Modell eine sprachliche ‚Sozialisierung' im Hinblick auf erwünschte Antwortstrukturen zur Seite gestellt. Da dieser Schritt in der Regel nur von sehr wenigen Programmierer:innen vollzogen und nicht öffentlich oder wissenschaftlich dokumentiert wird, ist er im Gegensatz zum outgesourcten Datenlabeln (s.u.) nur wenig bekannt und in seiner Bedeutung für die schlussendliche Sprachperfomance der Chatbots unterschätzt.

Eine technisch gesehen eigenständige, zusätzliche Schicht in den LLM von ChatGPT sind die abschließenden Filter. Sie sind auf eigene Weise trainiert, prüfen jede Ausgabe, und werden kontinuierlich, im laufenden Betrieb des KI-Systems, optimiert und bearbeitet. Das geschieht vor allem, um unerwünschte Outputs, wie etwa strafrechtlich relevante Inhalte, zu vermeiden. Ziel ist es, das Filter-Modell so zu trainieren, dass es Aufforderungen ablehnt, die gegen die OpenAI-Definition von schädlichem Verhalten verstoßen, z. B. Fragen zur Durchführung illegaler Aktivitäten, Ratschläge, wie man sich selbst oder anderen Schaden zufügt, oder Aufforderungen zur Beschreibung von grafischen, gewalttätigen oder sexuellen Inhalten. Da Nutzer:innen, wie im Fall Tay ersichtlich war, in der Lage sind, die nachträglich eingezogenen Barrieren von LLM bspw. durch suggestive Fragen oder einfache Umgehungen wie den ‚Drehbuchautor'-Trick[3] auszuhebeln, ist eine kontinuierliche Anpassung notwendig. So gibt ChatGPT im Vergleich zum Februar 2023 im Februar 2024 signifikant kürzere und wenig mit dem Inhalt der Frage interagierende Textausgaben auf Eingaben, die psychische Probleme der Nutzer:innen thematisieren. Hier ist eine Selbstregulation des Anbieters, um Haftungsrisiken zu minimieren, zu vermuten.

Technisch gesehen werden die Filter unter direkter Anleitung durch Menschen trainiert. Dafür müssen Beispiele zu sperrender Inhalte durch Menschen gelabelt werden, um durch das System später als Teil einer Ausschluss- bzw. Freigabeliste gelernt und auf Eingaben der Nutzer:innen angewandt werden zu können. Eine Recherche des Time Magazine beschrieb, dass OpenAI zur Kennzeichnung schädlicher Inhalte kenianische Arbeiter:innen, die weniger als 2 Dollar pro Stunde verdienen, einsetzte (Perrigo 2023). Die Recherche deckte auch psychische Probleme auf, die aus der Sichtung von Darstellungen sexualisierter Gewalt resultierten (ebd.).

3. Gesellschaftliche Verfügbachmachungsdynamik generativer KI-Systeme

Insbesondere in diesem letzten Trainingsschritt, der Filterschicht des Modells, der für dessen kommerzielle Verwendbarkeit enorm wichtig ist und einen zentralen Unterschied zu vorigen öffentlichen Experimenten mit

3 Um Sperren anstößiger Inhalte zu umgehen, kann der Chatbot in den Eingabebefehlen z.B. gebeten werden, sich in eine Szene ‚hineinzuversetzen', und diese wie ein Drehbuch fortzuschreiben.

LLM darstellt, zeigt sich eine dem technischen Funktionieren von generativer KI übergeordnete Ebene, nämlich Wirtschaftsstrukturen, politische und soziale Normen, die KI-Systeme wie ChatGPT überhaupt erst möglich machen. Die Arbeitsbedingungen von Datenetikettierer:innen offenbaren, dass KI-Produkte auf versteckte menschliche Arbeit im globalen Süden angewiesen sind, die oft schädlich und ausbeuterisch organisiert ist – obwohl diese Arbeit zu milliardenschweren Industrien beiträgt (vgl. Gray und Suri 2019; Mohamed et al. 2020, vgl. 3.1).

Man mag entgegnen, dass die Filter der technisch am wenigsten originelle Aspekt des Systems sind, und prinzipiell auch unter fairen Bedingungen in Kalifornien trainiert hätten werden können. Dass die Genese dieses Produkts aber eben dennoch so verlief (und auch weiter fortgesetzt wird), ist paradigmatisch für die ChatGPT ermöglichende Wirtschaftsform, den digitalen oder auch Plattformkapitalismus (Zuboff 2018; Staab 2019; Pfeiffer 2021; Fuchs 2023). Firmen wie Google, Facebook und der ChatGPT-Anbieter OpenAI betreiben eine Art von Wirtschaft, die auf der Auswertung von Nutzendendaten und der Schaffung von proprietären Märkten beruht. Eine generative Text-KI wie ChatGPT ist eher als Interface, als Schnittstelle für neue Interaktions- und damit auch Erlösformen zu begreifen – und wird als abonnierbares Produkt, oder künftig teils integriert in Standardsoftware wie die Microsoft-Office-Produkte vertrieben. Diese Distributionsform gehört ebenso zur sozio-technischen Verfügbarmachung von ‚Sprach-KI' und hat ihre eigenen Implikationen: Die Plattformen der Digitalwirtschaft generieren ihre Marktposition und damit auch ihren Wert im Wesentlichen durch Netzwerkeffekte, bei denen der Wert einer Plattform mit der Anzahl ihrer Nutzer:innen steigt. Je mehr Nutzer:innen ein LLM nutzen, und desto mehr Nutzer:innendaten dabei anfallen, desto höher sind die Optionen der Entwickler:innen, ihr System technisch zu verbessern und damit auch einen besseren Service anbieten zu können. Dementsprechend hart ist der Wettbewerb zwischen Anbietern von LLM wie ChatGPT derzeit auch: Vermutlich werden nur ein oder zwei Anbieter je in eine derart marktbeherrschende Stellung gelangen, dass das kostenintensive Geschäft des Vorhaltens von solcher Rechenleistung (vgl. 3.2) profitabel wird. Im Moment, so wird geschätzt, hat der ChatGPT-Anbieter OpenAI pro Tag etwa 700.000 US-Dollar Kosten, um seinen Service zu betreiben (Sacra 2024). Extrem rechenintensive Produkte wie ChatGPT basieren zu einem erschreckend hohen Anteil auf dem Verbrauch knapper Güter, wie Energie, Wasser und menschlicher Arbeit.

3.1 Transformation menschlicher Arbeitskraft für generative KI

Die Diskussion der (möglichen) Transformation von Arbeit *durch* KI ist fast schon ein Gemeinplatz. Erste Reaktionen auf ChatGPT insbesondere mit volkswirtschaftlicher Perspektive knüpften direkt am Automatisierungsdiskurs (Benanav 2021) an, und versuchen zu ermitteln, wie hoch der Anteil von durch generative KI zu ersetzenden bzw. stark veränderten Berufsprofilen sein wird (siehe z. B. Hatzius et al. 2023). Die sozio-technischen Prozesse hinter den Kulissen von insbesondere generativen KI-Anwendungen beruhen jedoch stärker auf einer Transformation von Arbeit *für* KI, als insbesondere die Anbieter den Anschein erwecken lassen. Die Arbeit an solchen Systemen – sowohl in deren Design und Entwicklung als auch deren Betrieb – wird sogar systematisch unsichtbar gemacht, wie kritische ethnografische Forschungen zu KI, Robotik und digitalen Plattformen unter den Stichworten „Fauxtomation" (Adamowsky 2020) bzw. „Ghost Work" gezeigt haben (Iantorno et al. 2022; Lipp 2023; Irani und Silberman 2013). Egal ob die Bilderkennung selbstfahrender Autos, Chatbots für den Kundenservice, ‚autonome' Pflegeroboter oder vermeintlich automatische Supermärkte – immer wieder wird bekannt, dass fortgeschrittene KI-Produkte auf der Datenarbeit oder Echtzeitüberwachung durch tausende Menschen in Niedriglohnländern beruhen.

Am Gegensatz vom milliardenschweren Umsatz und den nicht oder nur sehr spärlich verrechtlichten Arbeitsverhältnissen seiner Ermöglichung zeigt sich die Verfügbarmachungsdynamik digitaler Arbeit für generative KI besonders deutlich. Die für LLM wie ChatGPT notwendige Arbeitskraft umfasst in einem weiteren Sinne jegliche Arbeit zur Bereitstellung von Energie und Hardware für so rechenintensive Produkte (vgl. 3.2), im engeren Sinne aber vor allem so genannte ‚Informationsarbeit', die notwendig ist, um Softwarecode zu erstellen, bzw. ein sozio-technisches System wie ein LLM möglich zu machen, und als kommerzielles Produkt anzubieten. Im Blick auf die sozio-technischen Möglichkeitsbedingungen von ChatGPT wurden bereits Arbeitsrollen und Arbeitsbedingungen von solcher Informationsarbeit angerissen (vgl. 2.2). Für die Frage, welche gesellschaftlichen Dynamiken KI-Systeme wie ChatGPT ermöglichen, lohnt sich aber ein weiterer Blick auf die Bandbreite der dafür nötigen Informationsarbeit und deren Form internationaler Arbeitsteilung.

Das ambivalente gesellschaftliche Verhältnis zur notwendigen Arbeit für generative KI zeigt sich in der Zuspitzung zweier empirischer Extrempositionen von Informationsarbeit, nämlich in der Gegenüberstellung von

nicht-verrechtlichten Clickworker:innen im globalen Süden und angestellten Software-Ingenieur:innen im Silicon Valley. Fuchs bezeichnet letztere im Rückgriff auf Engels als „digitale Arbeitsaristokratie" (Fuchs 2023, 58f): Diesen werden in der Regel sehr hohe Gehälter – bei OpenAI beispielsweise zwischen 550.000 und 850.000 US-Dollar Jahresgehalt für fortgeschrittene Entwickler:innen (Levels 2024) – gezahlt, was ihnen sowohl im Vergleich zu anderen Berufsrollen in Softwareunternehmen als auch im Vergleich zu anderen Branchen eine privilegierte Stellung verleiht. Die Nachfrage nach ihrer Arbeitskraft ist wie am Gehaltsniveau ablesbar sehr hoch. Ihre Arbeitskraft wird in einer typisch post-fordistischen Weise erlöst (vgl. Boltanski und Chiapello 2006): Die Arbeitszeit ist sehr extensiv, und der Arbeitsmodus durch Methoden gekennzeichnet, die die Grenze zwischen Arbeit und Freizeit verwischen, wie spielerisch gestaltete Arbeitsräume oder die Option auf typische Freizeitaktivitäten am Arbeitsplatz durch Sportanlagen. Fuchs wertete wiederholt Online-Foren aus, in denen Google-Mitarbeitende über ihre Arbeitsbedingungen berichten (z. B. Fuchs 2021). Auf die Frage nach den Arbeitsbedingungen bei Google antwortete ein laut Fuchs typischer Beitragender: „work/life balance is nearly non-existent [...] [one must be prepared to] work all day and night long" (Fuchs 2023, 60).

Der sehr gut bezahlten, aber nur über wenig Freizeit verfügenden Spitzengruppe der Programmierer:innen gegenüber stehen die sog. ‚Plattformarbeitenden' oder ‚Clickworker'. Das sind Arbeitende, die zumeist ohne Arbeitsvertrag über Apps und Internetplattformen kleine Aufträge annehmen. Plattformarbeiter:innen sind in aller Regel Stückarbeiter:innen, die nicht nach Stunden, sondern für jede erbrachte Leistung, teils einzelne Klicks, bezahlt werden. Diese genuin digitale Art der Arbeit ist vergleichsweise einförmig, da sie im Kontext generativer KI für mikroskopische, aus der Ferne ausführbare einfache Aufgaben angewendet wird: Eine Aufgabe wie das Recherchieren von Inhalten im Web gehört schon zu abweichend anspruchsvollen Ausgaben, in der Regel geht es um das Tagging von Datensätzen, also der Zuordnung von Bedeutung zu maschinenlesbaren Daten. Ähnlich den CAPTCHA-Aufgaben bei der Authentifizierung als menschliche:r Internetnutzer:in identifizieren Clickworker:innen gegen Cent-Beträge in Videos und Bildern Objekte, oder bewerten kurze Texte von LLM in standardisierten Skalen auf deren Korrektheit und Angemessenheit zu einer Frage. Diese Aufgaben sind für die sozio-technische Ermöglichung von LLM unabdingbar, aber nicht an höhere Kompetenzen als Alltagsverständnis selbst gebunden – welches ja eben mittelbar durch Klassifizierungen

und maschinelles Lernen durch die Services digital verfügbar werden soll. Insbesondere in den Selbstdarstellungen der Plattform Amazon Mechanical Turk ist das Verhältnis von menschlicher und maschineller Intelligenz in „Clickwork" anschaulich zugespitzt (Mturk 2024): Die auf wenige Klicks heruntergebrochenen Mikroaufgaben operationalisieren genau das, was Computern (derzeit) noch nicht verfügbar ist: das relevanteste Bild in einer Gruppe von Bildern auszuwählen, für einen spezifischen Kontext unangemessene Inhalte auszusortieren, Text aus Bildern zu digitalisieren, Identifizierung von für Menschen relevanten Informationen in Web-Inhalten (bspw. bei Restaurants Telefonnummer und Öffnungszeiten).

Die ‚Clickworker' sind in der Auswahl und Erbringung von Leistungen extrem frei – was die Tätigkeit für viele Erwerbseingeschränkte attraktiv macht –, die Arbeit findet allerdings außerhalb von Tarifverträgen und grundlegenden Arbeitsrechten statt. Die schwache Regulierung und asymmetrischen Machtverhältnisse dieser über Plattformen vermittelten Arbeit (Schor und Attwood-Charles 2017) lassen sich am Beispiel von Amazons Mechanical Turk in ein Zitat fassen: Clickworker seien „a global, on-demand, 24x7 workforce" (Mturk 2024). Das Ausmaß von Clickwork ist empirisch schwer zu bestimmen, vergleichende Berichte schwanken zwischen mehreren Millionen Clickworker:innen und einigen Zehntausenden, die tatsächlich allzeit verfügbar seien. Die bekannteste Plattform „Clickworker" berichtet von 6 Millionen Nutzer:innen (Clickworker 2024), und „Amazon Mechanical Turk" von mehreren Hunderttausend Nutzer:innen (Mturk 2024).

Am Vergleich der beiden Arbeitsformen, die die für generative KI notwendige Informationsarbeit annimmt, zeigen sich sowohl die Amplitude der Erscheinungsformen notwendiger Arbeitstätigkeiten für generative KI als auch für beide spezifische Entfremdungsmomente – die Entgrenzung der Arbeitszeit der Top-Programmierer:innen und die vergleichsweise stumpfsinnigen Aufgaben der Daten-Tagger:innen –, die aus der Strukturierung der jeweiligen Arbeit nach den Bedürfnissen der KI-Anbietenden resultieren.

3.2 Verbrauch knapper Ressourcen durch generative KI

Der Verbrauch von Ressourcen wie Strom und Wasser durch Technologien wie ChatGPT ist eine Dimension der Abschätzung ökologischer und gesellschaftlicher Folgen, die auch informierten Perspektiven oft entgeht.

Erst seit Beginn des Jahres 2024 rücken konkrete Prognosen über den Verbrauch von Strom und Wasser für den Betrieb von LLM in den Fokus (vgl. z.B. Morning Star 2024). Der Energie- und Ressourcenbedarf von ChatGPT und allen anderen LLM ist wesentlich durch zwei Instanzen bestimmt: einerseits durch *direkte* Energiekosten vor allem durch das Training des Modells und die Bereitstellung im laufenden Betrieb, und andererseits durch die *indirekt* anfallenden Energiekosten der Herstellung der nötigen Hardware wie Grafikprozessoren und Serverinfrastruktur. Für jeden dieser Aspekte gibt es vergleichende Untersuchungen, wobei sich hier aus Platzgründen auf Zahlen zu den direkten Kosten auf Seiten der Anbietenden beschränkt werden soll.

Der Prozess des Modelltrainings gilt als am relevantesten in der Berechnung des Energieverbrauchs und der Emission von klimaschädlichen Gasen durch LLM. Ein realistischer Mittelwert für den CO_2-Ausstoß durch das Training eines LLM liegt bei etwa 200 bis 270 Tonnen (vgl. Strubell et al. 2019). Für die Ermittlung der Energiekosten, die *im Betrieb eines Systems* anfallen, sind vor allem die Nutzungszahlen relevant. In einem Artikel von Patel und Ahmad (2023) wird die Zahl der aktiven Nutzer:innen von ChatGPT auf 13 Millionen pro Tag geschätzt, und es wird angenommen, dass jede:r der aktiven Nutzer:innen pro Tag 15 Anfragen stellt, was 195 Millionen tägliche Anfragen ergibt. Der *monatliche Energieverbrauch* durch ChatGPT läge demnach allein auf der Seite des Anbieters OpenAI bzw. Microsofts bei etwa 4.176.000 KWh – was dem durchschnittlichen Jahresverbrauch von 2.610 Menschen in Dänemark entspricht (Khowaja et al. 2023, 5). Der *Wasserverbrauch* von KI-Modellen variiert kongruent zur Berechnung des Strombedarfs je nach Komplexität des Modells und der Zahl der Zugriffe. Hochrechnungen vermuten für GPT-3s Trainingsphase einen Verbrauch von mindestens rund 700.000 Litern Wasser (Li et al. 2023) – was der Wassermenge entspricht, die für die Herstellung von über 230 Tesla-PKW nötig ist. Microsoft hat in seinem jüngsten Umweltbericht verkündet, dass sein Wasserverbrauch im Jahr 2023 um 34 Prozent auf rund 6,4 Millionen Kubikmeter Wasser gestiegen sei (Microsoft 2024) – was mit der Akquise von OpenAI in Verbindung steht.

Neben den absoluten Zahlen, die auch durch Vergleiche nur schwer in vorstellbare Mengen zu übersetzen sind, ist die *Dynamik* der Entwicklung zu beachten: Rechenzentren, die für die Verarbeitung der riesigen Datenmengen und das Angebot der KI-Produkte nötig sind, machten bereits 2023 weltweit bis zu 1,5 % des gesamten Stromverbrauchs aus (IEA 2024). In einem Markt, in dem die Anbieter über exponentielles Wachstum der Mo-

delle konkurrieren, wird der Bedarf an für Training und Betrieb von LLM notwendigen Ressourcen weiter stark steigen. Eine viel zitierte Studie von 2015, die den Stromverbrauch des gesamten Informations- und Kommunikationstechnologie-Sektors bis 2030 modellierte (Andrae und Edler 2015), hatte für die Entwicklung des Ressourcenverbrauchs durch Rechenzentren drei Szenarien errechnet, von denen das pessimistischste durch die Zahlen von 2023 erreicht ist (ebd., 133). Eine weitere Berechnung im Bericht der internationalen Energiebehörde (IEA) aus dem April 2024 ist besonders anschaulich: Der weltweite Stromverbrauch von Rechenzentren lag 2022 bei etwa 460 Terawattstunden (TWh) und könnte sich laut IEA bis 2026 auf mehr als 1.000 TWh verdoppeln (vgl. IEA 2024), was ungefähr dem gesamten jährlichen Stromverbrauch Japans entspricht – dem mit knappem Abstand hinter Russland fünftgrößten Energiekonsumenten der Welt.

Die sozio-technische Verfügbarmachung menschlicher Sprache für maschinelle Bearbeitung und Vorhersage hat also ganz manifeste und konkrete Folgen, wie sich am dramatischen Zuwachs des Ressourcenverbrauchs dieser sehr energieintensiven Branche zeigt. Während andere wirtschaftliche Tätigkeiten bzw. Alltagsaktivitäten – wie Ernährung, Transport und Tourismus – hinsichtlich ihrer Folgen für den Klimawandel bzw. den Bedarf an gesellschaftlicher Regulation breit diskutiert werden, lässt sich das für den Diskurs um ChatGPT und andere LLM nicht feststellen. Erneut sind im Zuge der Verfügbarmachung also zentrale Voraussetzungen für KI-Produkte wie ChatGPT im gesellschaftlichen Diskurs eher unterbewertet und nicht Teil der offiziellen Selbstdarstellung der Anbietenden oder gar eines Geschäftsmodells, das diese Folgen und Verbräuche einpreisen oder kompensieren würde.

4. Fazit – Un/Verfügbarkeit generativer KI

Das Verfügbarmachungsversprechen von ChatGPT und anderen LLM, menschliche Kommunikation für Maschinen bearbeitbar zu machen, beruht in zweifacher Hinsicht auf einer Form von „Blackboxing", wie Latour den Fokus auf Inputs und Outputs nannte (Latour 1987; Latour 1999, 304), der die innere Komplexität von sozio-technischen Systemen unsichtbar macht. Zum einen konnte gezeigt werden, dass die sozio-technische Verfügbarmachung ganz wesentlich auf ‚manueller' Arbeit zur Erzeugung gewünschter Textausgaben und der nicht-konsensualen Auswertung großer Mengen von Daten aus menschlicher Kommunikation beruht. Zum ande-

ren wurde gezeigt, dass die Genese und kommerzielle Verwertung von LLM eng mit einer Verfügbarmachungsdynamik des Plattformkapitalismus verbunden ist, die Arbeit und globale Rohstoff- sowie Energiekreisläufe bereits jetzt transformiert. An den repetitiven Mikroaufgaben des Datenannotierens oder Filtertrainierens, die an Niedriglohnarbeiter:innen im globalen Süden ausgelagert werden, zeigt sich der sozio-technische Kern der Verfügbarmachung von menschlichen Fähigkeiten durch generative KI ganz besonders deutlich: Schlecht bezahlte und wenig geschützte Menschen übernehmen in der Genese von generativer KI genau das, was den Computern derzeit noch nicht verfügbar ist: die relevanteste Antwort auszuwählen und erzeugte Inhalte für spezifische Kontexte anpassen.

Zusammen genommen lässt sich festhalten, dass der solutionistische Diskurs über immer neue, scheinbar losgelöste Möglichkeiten der Technik irreführend ist. Generative KI ist weder *singulär*, als dass sie sich selbst verbessern und neue Erfindungen machen und damit irreversibel würde, noch eigenständig kommunikationsfähig, oder gar neutral – sondern beruht inhaltlich wie auch in ihren sozio-materiellen Bedingungen auf Heteronomie und Vorselektion durch wenige Entscheider:innen. Die Genese von solchen großen KI-Produkten schreibt etablierte Un/Verfügbarkeitsmuster fort: Die für sie notwendige Arbeit beruht auf bestehenden Ausbeutungsverhältnissen, das gleiche gilt ökologisch. Auch liegt die sozio-technische Neuerung nicht in einem maschinellen Zugriff auf den *Sinn* menschlicher Kommunikation, sondern vor allem in der Verfügbarmachung von Kommunikationsdaten für ökonomische Verwertungslogiken. Die gesellschaftliche Möglichkeit von generativer KI beruht auf klassisch-modernen Verfügbarmachungsstrategien, wie der Extraktion von verwertbaren Ausgangsstoffen unter Verbrauch knapper Ressourcen, bzw. Emission klimaschädlicher Gase und der Entfremdung von Arbeit – indem Menschen die Aufgaben, die substituiert werden sollen, der Rechentechnik selbst gewissermaßen zu deren Bedingungen ‚beibringen' – zur Steigerung der Mehrwertproduktion. ‚Unsichtbar' ist diese Arbeit nicht nur dahingehend, dass sie im finalen Produkt nicht sichtbar ist, oder die Anbietenden wenig Interesse haben, die ökonomischen und praktischen Bedingungen dieser Arbeit mit ihrem Produkt zu kommunizieren – in dieser Hinsicht unterscheiden sich die gesellschaftlichen Produktionsverhältnisse von ChatGPT nicht wesentlich von einem Marken-Turnschuh.[4] ChatGPT und andere LLM-Produkte werden als Produkte zur sozio-technischen Transformation der Sphäre der Arbeit

4 Für diesen anschaulichen Vergleich danke ich Philipp Zeltner!

vermarktet, und die Rolle von Menschen geleisteter Arbeit zur Erzeugung des Produkts wird dermaßen unterbetont, weil es im Hinblick auf die jetzt schon eklatanten gesellschaftlichen Kosten der Genese dieser Technologie am erfolgsversprechenden ist, ‚die KI' nicht nur als (teilweise) autonom, sondern auch gewissermaßen ‚vom Himmel gefallen' darzustellen.

Mit seinen dezidiert nicht-nachhaltigen Geschäftsmodellen und Genesebedingungen fällt der gesellschaftliche KI-Diskurs derzeit sogar hinter bereits geführte Diskurse zu natürlichen Grenzen moderner Verfügungsgewalt zurück. Wenn in den Feuilletons der Chatbot als kompetenter Konversationsagent mit Option auf Bewusstsein missverstanden wird, überdeckt das relevante Fragen: Zum Beispiel danach, wie das alltägliche Handeln von Internet-Nutzer:innen durch die Datenakkumulation mittelbar in das Modell der Wertschöpfung von LLM-Anbietenden eingebunden wird, und wie sich das auf individuelle Lebensführung auswirkt. Oder ob LLM auch ohne eine exponentielle Wachstumsdynamik möglich sein können, die die Lebensbedingungen auf dem Planeten nicht weiter gefährdet.

Für das konstitutive Verfügbarmachungsversprechen aller unter ‚Künstliche Intelligenz' subsummierten Bemühungen, menschliche Handlungen maschinell bearbeitbar bzw. ersetzbar zu machen, lässt sich aus ChatGPT lernen, dass die KI-Produkte zwar in Form der Erfahrung von Überschusssinn mit einer disruptiven Wirkung angekündigt werden, de facto aber auf einer fortwährenden Stabilisierung und Optimierung durch Menschen beruhen – die Maschinen sind aufgrund dieser verdeckten Arbeit in der Lage, Textaufgaben wie Zusammenfassungen oder Inhaltsangaben auf einem durchschnittlichen und oberflächlichen Niveau zu liefern.

LLM zeigen also, dass Maschinen unter hohem Ressourcenaufwand und unter Ausbeutung von Nutzungsdaten durchaus für Menschen kontextuell sinnhaften Text erzeugen können und gewisse funktionale Kompetenzen, wie bspw. induktive Kategorienbildung (Mahowald et al. 2023), erlangen können. Für eine Abschätzung möglicher Folgen ist es einerseits viel zu früh, andererseits lassen sich unter Rückgriff auf die Figur des Un/Verfügbaren konstruktive Folgefragen stellen. So deutet der Diskurs um KI-generierte Inhalte an, dass die Frage der Unterscheidbarkeit menschlicher und maschineller Texte bzw. Bilder die – hier nur bedingt passenden – Kategorien von menschlicher Schöpfungskraft und maschineller Kopie fortführen: Die Verfügbarmachung wirft neue Unverfügbarkeiten auf. Zu diesen unintendierten Nebenfolgen gehört zum einen die Gefahr des „model collapse" (Chen et al. 2023), also dass sich LLM dahingehend selbst sabotieren werden, dass von Ihnen erzeugte Texte immer mehr auch zur Datenbasis für

weitere Modelliterationen werden – und dadurch schließlich schlechtere Ergebnisse liefern.

Das Verfügbarkeitsversprechens des Solutionismus selbst, durch stetes Wachstum von Rechenkraft bald alle menschlichen Fähigkeiten potenziell auch an Maschinen delegieren zu können, bleibt insbesondere vor dem Hintergrund der Möglichkeitsbedingungen von generativer KI (vgl. 3.) selbst unverfügbar: Es gibt schlicht nicht genug Ressourcen für ausreichend Platinen, und gleichzeitig verfügbare Energie, um tatsächlich in einem globalen Maßstab ganze Tätigkeitsarten durch KI zu ersetzen. Und das ist im Hinblick auf die bislang problematischen Verfügbarmachungsdynamiken für diese Systeme auch nicht wünschenswert.

Literatur

Adamowsky, Natascha (2020). Fauxtomation–Gedanken zu Geschichte und Ästhetik› intelligenter‹ Technik. Internationales Jahrbuch für Medienphilosophie 6(1), 263-276.

Andrae, Anders S./Edler, Tomas (2015). On global electricity usage of communication technology: trends to 2030. Challenges 6(1), 117-157.

Benanav, Aaron (2021). Automatisierung und die Zukunft der Arbeit. Berlin, Suhrkamp Verlag.

Bender, Emily M./Gebru, Timnit/McMillan-Major, Angelina/Shmitchell, Shmargaret (2021, March). On the dangers of stochastic parrots: Can language models be too big?. In: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, 610-623.

Boltanski, Luc/Chiapello, Ève (2006). Der neue Geist des Kapitalismus. Köln, Halem.

Bostrom, Nick (2014). Superintelligence: Paths, dangers, strategies. Oxford, Oxford University Press.

Breiter, Andreas (1995). Die Forschung über künstliche Intelligenz und ihre sanduhrförmige Entwicklungsdynamik: die Dynamik einer Wissenschaft im Spiegel ihrer Wahrnehmung in der Öffentlichkeit. Kölner Zeitschrift für Soziologie und Sozialpsychologie 47(2), 295-318.

Chen, Lingjiao/Zaharia, Matei/Zou, James (2023). How is ChatGPT's behavior changing over time?. *arXiv preprint arXiv:2307.09009*.

Clickworker (2024). Clickworker Crowd. Online verfügbar unter https://www.clickworker.com/clickworker-crowd/ (abgerufen am 04.03.2024).

Collins, Keith (2023). How ChatGPT Could Embed a 'Watermark' in the Text It Generates. In New York Times. Online verfügbar unter https://www.nytimes.com/interactive/2023/02/17/business/ai-text-detection.html (abgerufen am 04.03.2024).

Crawford, Kate (2021). The atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Epstein, Robert/Roberts, Gary/Beber, Grace (Hg.)(2009). Parsing the Turing test. Dordrecht, Springer Netherlands.

Esposito, Elena (2017). Artificial communication? The production of contingency by algorithms. Zeitschrift für Soziologie 46(4), 249-265.

Fuchs, Christian (2021). Social media: A critical introduction. London, Sage.

Fuchs, Christian (2023). Der digitale Kapitalismus. Arbeit, Entfremdung und Ideologie im Informationszeitalter. Weinheim, Beltz Juventa

Gitelman, Lisa (Hg.)(2013). Raw data is an oxymoron. Cambridge MA, MIT Press.

Gray, Mary L./Suri, Siddharth (2019). Ghost work: how to stop silicon valley from building a new global underclass. Boston, Houghton Mifflin Harcourt.

Hatzius, Jan/Briggs, Joseph/Kodnani, Devesh/Pierdomenico, Giovanni (2023). The Potentially Large Effects of Artificial Intelligence on Economic Growth. Goldman Sachs. Online verfügbar unter https://static.poder360.com.br/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf (abgerufen am 04.04.2024).

Iantorno, Matthew/Doggett, Olivia/Chandra, Priyank/Yujie Chen, Julie/Steup, Rosemary/Raval, Noopur/Khovanskaya, Vera/Lam, Laura/Singh, Anubha/Rotz, Sarah/Ratto, Matt (2022). Outsourcing Artificial Intelligence: Responding to the Reassertion of the Human Element into Automation. In: CHI Conference on Human Factors in Computing Systems Extended Abstracts. New Orleans LA, ACM, 1–5. https://doi.org/10.1145/3491101.3503720.

IEA (2024). Electricity 2024. Paris, IEA. Online verfügbar unter https://www.iea.org/reports/electricity-2024 (abgerufen am 04.04.2024).

Irani, Lilly C./Silberman, M. Six (2013). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In: Proceedings of the SIGCHI conference on human factors in computing systems, 611-620.

Khowaja, Sunder Ali/Khuwaja, Parus/Dev, Kapal/Wang, Weizheng/Nkenyereye, Lewis (2023). ChatGPT Needs SPADE (Sustainability, PrivAcy, Digital divide, and Ethics) Evaluation: A Review. arXiv preprint arXiv:2305.03123.

Latour, Bruno (1987). Science in action: How to follow scientists and engineers through society. Cambridge, Harvard University Press.

Latour, Bruno (1999). Pandora's hope: essays on the reality of science studies. Cambridge, Harvard University Press.

Levels (2024). Salaries OpenAI Software Engineers. Online verfügbar unter https://www.levels.fyi/companies/openai/salaries/software-engineer?country=254 (abgerufen am 04.03.2024).

Li, Pengfei/Yang, Jianyi/Islam, Mohammad A./Ren, Shaolei (2023). Making AI less "thirsty": uncovering and addressing the secret water footprint of AI models. arXiv preprint arXiv:2304.03271 2023.

Lipp, Benjamin (2023). Caring for robots: How care comes to matter in human-machine interfacing. Social Studies of Science 53(5), 660-685.

Luccioni, Alexandra S./Viviano, Joseph D. (2021). What's in the Box? A Preliminary Analysis of Undesirable Content in the Common Crawl Corpus. arXiv preprint arXiv:2105.02732.

Mahowald, Kyle/Ivanova, Anna A./Blank, Idan A./Kanwisher, Nancy/Tenenbaum, Joshua B./Fedorenko, Evelina (2023). Dissociating language and thought in large language models: a cognitive perspective. arXiv preprint arXiv:2301.06627.

McCarthy, John/Minsky, Marvin, L./Rochester, Nathaniel/Shannon, Claude E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Online verfügbar unter http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf (abgerufen am 04.04.2024).

McCorduck, Pamela (2004). Machines Who Think. 2nd edition. New York, A.K. Peters/CRC Press.

Microsoft (2024). Environmental Sustainability Report. Reporting the Fiscal Year 2023. Online verfügbar unter https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW1lMjE (abgerufen am 12.07.2024).

Mohamed, Shakir/Png, Marie-Therese/Isaac, William (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philos. Technol.* 33, 659–684. https://doi.org/10.1007/s13347-020-00405-8.

Morning Star (2024). Commentary: Watts Up With AI: Strategies for Utilities in an Era of Surging. Online verfügbar unter https://dbrs.morningstar.com/research/430093/watts-up-with-ai-strategies-for-utilities-in-an-era-of-surging-demand (abgerufen am 04.04.2024).

Mturk (2024). Amazon Mechanical Turk. Online verfügbar unter https://www.mturk.com/ (abgerufen am 04.03.2024).

Neff, Gina/Nagy, Peter (2016). Talking to Bots: Symbiotic Agency and the Case of Tay. International Journal of Communication 10(2016), 4915–4931.

Nezik, Ann-Kathrin (2023). Haben Maschine eine Seele? Die Zeit vom 12. Januar 2023, 13-17.

Ochs, Carsten (2019). Optionalität & Prediktivität: Privatheit und der Subjektivierungswiderspruch algorithmisch organisierter Überwachungsgesellschaften. In: Ingrid Stapf/Marlis Prinzing/Nina Köberer (Hg.). Aufwachsen mit Medien. Baden Baden, Nomos, 211-224.

OpenAI (2024). GPTs. Online verfügbar unter https://chat.openai.com/gpts (abgerufen am 04.03.2024).

Patel, Dylan and Ahmad, Afzal (2023). The inference cost of search disruption – large language model cost analysis. SemiAnalysis. Online verfügbar unter https://www.semianalysis.com/p/the-inference-cost-of-search-disruption (abgerufen am 04.03.2024).

Perrigo, Billy (2023). Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. In Time Magazine. Online verfügbar unter https://time.com/6247678/openai-chatgpt-kenya-workers/ (abgerufen am 04.03.2024).

Pfeiffer, Sabine (2021). Digitalisierung als Distributivkraft: Über das Neue am digitalen Kapitalismus. Bielfeld, transcript Verlag.

Rammert, Werner/Schulz-Schaeffer, Ingo (2002). Technik und Handeln: Wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt. In Werner Rammert/Ingo Schulz-Schaeffer (Hg.). Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik. Frankfurt am Main, Campus, 11-64.

Sacra (2024). OpenAI. Online verfügbar unter https://sacra.com/c/openai/ (abgerufen am 04.03.2024).

Schor, Juliet B./Attwood-Charles, William (2017). The "sharing" economy: labor, inequality, and social connection on for-profit platforms. Sociology Compass 11(8), e12493.

Spitz, Malte (2017). Daten-das Öl des 21. Jahrhunderts? Nachhaltigkeit im digitalen Zeitalter. Eimsbüttel, Hoffmann und Campe.

Staab, Philipp (2019). Digitaler Kapitalismus: Markt und Herrschaft in der Ökonomie der Unknappheit. Berlin, Suhrkamp Verlag.

Strubell, Emma/Ganesh, Ananya/McCallum, Andrew (2019). Energy and Policy Considerations for Deep Learning in NLP. In ArXiv:1906.02243.

Turing, Alan M. (2009) [1950]. Computing Machinery and Intelligence. In: Robert Epstein/ Gary Roberts/Grace Beber (Hg.). Parsing the Turing Test. Dordrecht, Springer. https://doi.org/10.1007/978-1-4020-6710-5_3.

Zuboff, Shoshana (2018). *Das Zeitalter des Überwachungskapitalismus*. Framkfurt a.M./New York, Campus Verlag.