

Zwischen Erlösung und Verdammnis

Christoph Meinel

Geboren 1954 in Meißen,
Gründungspräsident der
German University of
Digital Science (G_UDS).
Vormals Direktor und
CEO des Hasso-Plattner-
Instituts für Digital
Engineering, Potsdam.

Maxim Asjoma

Geboren 1985 in Zwickau,
promovierter Philosoph,
Altstipendiat der
Konrad-Adenauer-Stif-
tung, Head of Public
Partnerships, Hasso-
Plattner-Institut für
Digital Engineering,
Potsdam.

Ein nüchterner
Blick auf die
„neue KI-Revolution“

**Wir leben in Zeiten technischer Wunder.
Kaum erfasst von der ersten Welle des
digitalen Wandels, rollt bereits die zweite
Welle der Digitalisierung auf uns zu:**

die KI-Revolution mit ihrer disruptiven Macht, die Einmaligkeit des Menschen an sich und alles Dagewesene grundsätzlich infrage zu stellen. Mit der Künstlichen Intelligenz (KI) kommt eine Technologie in die Welt, die „geistige“ Fähigkeiten entwickelt, über die bisher nur Menschen verfügten.

Generative KI versetzt Maschinen in die Lage, zu gewünschten Inhalten autonom Texte in menschenähnlicher Qualität zu verfassen, Illustrationen, Bilder und Medien zu kreieren sowie Musik und Videos wie von Zauberhand und in einer Güte hervorzubringen, die staunen lässt. Viele Jahre bemühten sich Experten, zu erklären, wie KI sinnvoll

zum Einsatz gebracht werden kann. Zunächst waren die Ergebnisse nur wenig erfolgreich – man denke dabei nur an sämtliche „smarten“ Dinge, den Gesundheitsbereich oder autonomes Fahren. Erst die Veröffentlichung von *ChatGPT* (Version 3.5 des *GPT*-Modells) des Softwareunternehmens *OpenAI* im November 2022 machte schlagartig klar, dass wir es mit einer ganz neuen Qualität von Technologie zu tun bekommen. Innerhalb von fünf Tagen registrierten sich über eine Million Nutzer bei dem kleinen Start-up, das über Nacht zur weltweit erfolgreichsten KI-Firma (noch vor Google, *Microsoft* und Amazon) avancierte.

Das Erfolgsgeheimnis lag in der einfachen Nutzung der Software für den Endverbraucher: Über einen beliebigen Webbrowser können wir mit der *ChatGPT*-KI chatten und jede beliebige Frage stellen („prompten“), das System gibt in Sekundenschnelle Antwort – von einem einfachen Ja oder Nein bis hin zu elaborierten Aufsätzen zu historischen Themen, etwa über den Dreißigjährigen Krieg. Die Illusion eines intelligenten Gegenübers war perfekt.

Sicherlich haben moderne KI-Verfahren des *Machine Learning* im Bereich der Bild und Spracherkennung (sowie beim Schach- und Go-Spiel) den Turing-Test schon lange bestanden, also gezeigt, dass ihre Leistung von der eines Menschen kaum mehr zu unterscheiden war beziehungsweise diese in bestimmten Aspekten sogar übertraf; im Bereich des kreativen Schreibens und der Konversation war das bislang nicht der Fall. Der schnelle Erfolg des *ChatGPT*-Modells weckte nicht nur Hoffnungen, sondern auch Ängste. So fürchteten Hollywood-Drehbuchautoren um ihre Existenz, und es formierte sich eine Reihe namhafter Apokalyptiker, die den Ängsten, dass KI den Menschen selbst überflüssig machen würde, Ausdruck verliehen; erstaunlicherweise allen voran und sicherlich auch mit Marketing-Hintergedanken die Entwickler dieser modernen KI-Verfahren selbst, wie *OpenAI*-Gründer Sam Altman, Elon Musk und KI-Pionier Geoffrey Hinton. Es ist daher höchste Zeit, diese neue Welle der Entwicklung generativer KI nüchtern zu betrachten.

Wie funktioniert „ChatGPT“?

Zunächst ist festzuhalten, dass die technischen Grundlagen eines KI-Modells wie *GPT 3.5* und inzwischen *GPT-4* bereits seit einigen Jahren in der Welt waren. Die Technologie basiert auf dem als „Deep Learning“ bezeichneten *Machine Learning*-Verfahren, bei dem Computern über komplexe mathematische Verfahren und die massenhafte Eingabe von Daten ein selbstständiges Lernen ermöglicht wird: Sie erkennen mithilfe statistischer und probabilistischer Methoden Muster in Texten und Medien und eignen sich daraus die Fähigkeit an, diese in variierender Weise zu reproduzieren.

Im Fall von *ChatGPT* stellte sich der Erfolg ein, weil *OpenAI* Verfahren entwickelt hatte, um die Inhalte des gesamten Internet zu crawlen (zu lesen und Inhalte zu extrahieren) und mit dieser riesigen Datenmenge ihr KI-Modell zu trainieren. *ChatGPT* kennt also die zu allen Themen im Internet veröffentlichten, frei verfügbaren Texte und Medien und analysiert deren Struktur sowie Satzbildung. Auf dieser Grundlage kann das System sehr gute Vorhersagen treffen, wie sich ein angefangener Satz thematisch sinnvoll und grammatisch fehlerfrei fortsetzt, und so erstaunliche Inhalte generieren, die auf den ersten Blick oft sogar als kreativ und neu erscheinen und den Eindruck erwecken, das System verstehe, worum es geht.

Ein zweiter Blick verdeutlicht, dass Letzteres nicht der Fall sein kann. *ChatGPT* kann auf Basis des bis zum Zeitpunkt X aus dem Internet gelernten digitalen Menschheitswissens erfolgreich feststellen, welche Silbe bei einem bestimmten Inhalt auf eine andere folgt, bis ein sinnvoller, vollständiger Text entsteht, oder welche Pixel neben andere gesetzt werden müssen, um ein gewünschtes Bild zu erzeugen, das dem Menschen als natürlich erscheint. Mit anderen Worten: Es gibt die statistisch plausibelste Antwort auf eine Frage, basierend auf dem, worauf es trainiert wurde. Die Qualität der KI hängt also direkt von der (historischen) Datenlage ab.

Einfache Fehler

Zum Weiterlernen benötigt das System neue Daten, um qualitativ hochwertigere Antworten zu liefern. So ist auch zu erklären, warum *GPT 3.5* auf die Frage, wer im November 2022 Bundeskanzler der Bundesrepublik Deutschland war, „Angela Merkel“ anstatt „Olaf Scholz“ antwortet, da das System bis zu diesem Zeitpunkt nur mit Daten vor der letzten Bundestagswahl trainiert worden war. Aber das KI-System machte lange Zeit oftmals auch einfache Fehler, wie die Aussage, dass die Wurzel der Zahl vier eine irrationale Zahl sei. Weiterhin fällt es *ChatGPT* schwer, Unkenntnis eines Sachverhalts auszudrücken. Wenn die Datengrundlage schlecht ist, dann erfindet beziehungsweise „halluziniert“ das System Quellen und Daten, um dennoch eine Antwort zu geben. All das zeigt, dass es sich nicht um eine Maschine handelt, die versteht, was sie antwortet, sondern dass sie lediglich das hervorbringt, was sie in den Weiten des Web aufgeschnappt und in einen statisch plausiblen Zusammenhang gebracht hat.

Damit wird klar, dass die apokalyptischen Phantasien vermutlich übertrieben sind. Auch das *GPT*-Modell ist weit davon entfernt, ein Selbstbewusstsein oder gar den Wunsch zu entwickeln, den Menschen abzulösen. In diesem Sinne kann es nach aktuellem Stand nur ein mächtiges

1 Emma Strubell / Ananya Ganesh / Andrew McCallum: Energy and Policy Considerations for Deep Learning in NLP, College of Information and Computer Sciences, University of Massachusetts Amherst, 05.06.2019, <https://arxiv.org/pdf/1906.02243.pdf> [letzter Zugriff: 30.07.2024].

2 David Patterson / Joseph Gonzalez / Quoc Le / Chen Liang / Lluís-Miquel Mun-guia: Carbon Emissions and Large Neural Network Training, 2021, <https://arxiv.org/abs/2104.10350> [letzter Zugriff: 30.07.2024].

3 „According to ChatGPT, a single GPT query consumes 1567% (15x) more energy than a Google search query. Details in comments“, www.reddit.com/r/aipromptprogramming/comments/1212kmm/according_to_chatgpt_a_single_gpt_query_consumes/?rdt=39526 [letzter Zugriff: 30.07.2024].

4 Das ist eine konservative Schätzung. Mittlerweile sollte sich die Nutzung durchschnittlich im mittleren zweistelligen Millionenbereich bewegen, vgl. Robert Brandl / Cai Ellis: „ChatGPT-Statistiken 2024. Alle aktuellen Zahlen über den Chatbot von OpenAI“, in: [tooltester](http://tooltester.com/de/blog/chatgpt-statistiken/), 19.07.2023, www.tooltester.com/de/blog/chatgpt-statistiken/ [letzter Zugriff: 30.07.2024].

Werkzeug in den Händen von (wenigen?) Menschen bleiben. Der tiefere Blick in die Technologie offenbart auch, wie leicht es sein wird, sie im Zweifel auszuschalten, indem ihr beispielsweise das Essenzielle, die „Nahrung“ in Form von aktuellen Daten, entzogen wird. Wie die nachfolgende Betrachtung zeigt, ist KI auch am Ende, wenn die komplexe Rechenzentreninfrastruktur zusammenbricht, Seltene Erden ausgehen oder – ganz simpel – der Strom abgestellt wird.

Generative KI ist äußerst energiehungrig. Damit Systeme wie *ChatGPT* trainiert werden und funktionieren können, sind eine massiv ausgebaute Infrastruktur von Datenzentren, modernste Grafikprozessoren und ein nahezu gigantischer Energiezufluss notwendig. Schon das Training eines auf ein spezifisches Thema ausgerichteten KI-Systems verbraucht das CO₂-Äquivalent von 300 Flugpassagen von New York City nach San Francisco und zurück.¹ Das Training vom Vorgängermodell *GPT-3* war noch energieintensiver und kommt mit etwa 550 Tonnen CO₂-Äquivalent auf 550 Hin- und Rückflüge.² In der Nutzung verbraucht jede *ChatGPT*-Anfrage etwa sieben Wattstunden (im Vergleich dazu eine einfache Google-Abfrage nur 0,3 Wattstunden).³ Bei täglich über zehn Millionen Abfragen⁴ ergibt das einen täglichen (!) Energieverbrauch von siebzig Megawattstunden. Das entspricht dem Energieverbrauch von 14.000 Haushalten in Deutschland. Man stelle sich nur vor, was es bedeuten würde, wenn weltweit jeder täglich *ChatGPT* nutzen würde! So verwundert es nicht, dass die Senkung des Energieverbrauchs beim maschinellen Lernen eines der wichtigsten wissenschaftlichen Ziele in der KI-Forschung ist.

Wovor wir uns dennoch fürchten sollten

Die eigentliche Gefahr der KI-Systeme liegt im instrumentellen Gebrauch durch den Menschen selbst. KI ist ein mächtiges Werkzeug, das nicht in die Hände der Falschen geraten darf. Durch Unachtsamkeit kann sie vorurteilsbehaftet trainiert werden und damit in der Anwendung zu fatalen Entscheidungen führen, wenn kein moralisch aufrichtiger oder aufmerksamer menschlicher „Aufpasser“ im Spiel ist. Sie kann aber auch aktiv missbraucht werden, um machtpolitische Interessen zu lasten der Allgemeinheit durchzusetzen. Wovor wir uns also am meisten fürchten sollten, ist eine mangelnde Bildung in der breiten Bevölkerung im Bereich KI. Wer versteht, wie *Machine Learning* funktioniert, ist gewappnet gegen unsinnige Antworten von *ChatGPT*, misstraut realistisch anmutenden Medien im digitalen Raum, insbesondere in den sozialen Medien, und entwickelt einen gesunden Menschenverstand dafür, was stimmen kann und was nicht. Unser bisheriges Versagen in der Entwicklung digitaler Kompetenzen im Sinne einer digitalen „Aufklärung“

breiter Bevölkerungsschichten wird im Umgang mit KI zur größten Herausforderung.

Auch aus anderer Richtung droht Ungemach, und das kann nicht deutlich genug betont werden. Die aktuellen gesellschaftlichen Debatten in Deutschland und Europa kreisen oft darum, dass wir die Nutzung von KI zum „Wohle“ der Gesellschaft einschränken sollten. Das Gesetz zur künstlichen Intelligenz der Europäischen Union (*EU Artificial Intelligence Act, EU AI Act*) ist aktuell der prominenteste Ausdruck dieser Sorge vor ausländischer Technologieführerschaft. Es werden vorschnell Regeln beschlossen, anstatt die Herausforderung anzunehmen und im technologischen Wettrennen optimale Lösungen zu entwickeln. So kommen schlussendlich die Lösungen anderer Akteure zum Einsatz, die vermutlich kaum nach unseren Vorstellungen und Wünschen entwickelt werden.

Dabei ist der Einsatz von KI in vielen, wenn nicht gar allen Sektoren der Gesellschaft äußerst hilfreich. Gesellschaftliche Probleme könnten angegangen werden, darunter Fachkräftemangel, Effizienzsteigerung in Wirtschaft und Verwaltung, Verbesserung der Gesundheitsvorsorge, der Sicherheit und der Bildung. Was bei der Debatte um die Gefahren von KI oft verkannt wird, ist der Schaden, der entsteht, wenn wir die Technologie nicht selbst nutzen und fortentwickeln. Nur ein kleines Beispiel aus unserer aktuellen Forschung kann bereits illustrieren, wie generative KI im Bereich Bildung beitragen kann, die Ausbildung zu verbessern.

Der Reflex, zu verbieten

Eine Sorge vieler Pädagogen und Hochschullehrer besteht darin, dass KI das Prüfungswesen zerstört, weil nicht mehr zwischen Schülerleistung und KI-generierter Leistung unterschieden werden kann. *ChatGPT* wurde deshalb kurz nach Erscheinen etwa an den öffentlichen Schulen von New York verboten. Auch an deutschen Universitäten gab es diesen Reflex, anstatt die Situation als Chance zu begreifen, Prüfungsparadigmen zu überarbeiten und bessere zu entwickeln. Anstatt zu beklagen, dass eine antiquierte Prüfungstechnik wegen möglicher KI-Hilfen nicht mehr funktioniert, müssen wir uns im Bildungswesen auf die Nutzung geeigneter KI-Werkzeuge einstellen, um bessere Lernformate und (KI-basierte) Prüfungen zu entwickeln und so der Gefahr zu entgehen, unmündig vor neuer Technologie zu stehen.

Gute Pädagogen nutzen bereits heute generative KI im Unterricht, um ihren Schülern zu vermitteln, wie KI funktioniert und wo ihre Grenzen liegen. Dabei ist das Versprechen von *OpenAI*, dass nun jeder Mann über einfache *Prompts* in seinem Webbrowser oder auf seinem

Smartphone die Ergebnisse erhalte, die er sich wünscht, zumindest naiv. Gutes „Prompten“ – also inhaltlich sinnführendes Anfragen und Nachhaken bei einer KI – ist nicht einfach. Nicht umsonst entwickelt sich aktuell ein neues Kompetenzfeld, bei dem es darum geht, sinnvolle *Prompts* für generative KI-Systeme zu entwickeln und anzuwenden. Es ist ein echtes Handwerk, gut zu prompten, und dabei können *MicroApps* – Anwendungen mit Schnittstellen zur generativen KI – bei der Entwicklung relevanter *Prompts* helfen, wie sich anhand erster und erstaunlicher, bisher unveröffentlichter Ergebnisse aktueller Forschungsprojekte meines [Ch. Meinel, Anm. d. Redaktion] Doktoranden Mohammed Elhayany im Kontext digitaler Lernplattformen wie *openHPI.de* und *German-UDS.academy* zeigen lässt.

„MicroApps“ für bessere Bildungsqualität

Elhayany hat eine *MicroApp* „Multiple Choice Generator“ (MCG) zur automatisierten Generierung von Multiple-Choice-Aufgaben entwickelt, die Lehrkräfte signifikant dabei unterstützt, gute Testaufgaben für Prüfungen zu entwickeln. Der MCG liefert auf Grundlage vorgegebener Parameter (Anzahl der Fragen, Anzahl der richtigen Antworten et cetera) inzwischen bessere Ergebnisse als menschliches Lehrpersonal. In der Studie wurden Lehrkräfte – eingeteilt in drei Gruppen – gebeten, Prüfungsfragen zu erstellen: *erstens* Lehrpersonal, das gänzlich eigenständig Prüfungsaufgaben erstellt; *zweitens* Lehrpersonal, das dazu *ChatGPT* nutzt; und *drittens* Lehrpersonal, das MCG verwendete. Anschließend wurden die erarbeiteten Prüfungsfragen von anderen Pädagogen bewertet. Das Ergebnis war eindeutig: Die zweite und die dritte Gruppe schlossen in Bezug auf Quantität und Qualität der Prüfungsfragen deutlich besser ab, wobei die dritte die besten Ergebnisse erzielte. Insbesondere im Bereich der MINT-Fächer lassen sich so komplexe Prüfungsfragen in Sekundenbruchteilen erstellen, deren Qualität die der menschlichen Lehrkräfte erreicht oder gar übersteigt. Der größte Vorteil der *MicroApp*-Nutzung war, dass der generierte *Prompt* bei Eingabe der Prüfungsparameter transparent wurde und Pädagogen nachvollziehen konnten, wie die KI letztlich vorgegangen ist.

In einer weiteren Studie nutzten wir die *MicroApp* „AI-powered Recap Interactive Assistant“. Diese Anwendung erstellte – anhand von Textbüchern und relevantem Prüfungsstoff – Tests, die es Lernenden ermöglichten, in ein interaktives Verhältnis bei der Prüfung zu kommen. Anstatt nur die Bewertung „richtig“ oder „falsch“ auszugeben, wurde Schülern das Ergebnis der Prüfung angezeigt und (automatisiert) erklärt, warum Antworten unzureichend waren. Das System gibt weiterhin Anregungen dazu, wie man sich verbessern könnte. Über ein solches

System ist es möglich, auch in komplexen Lernkontexten Binnendifferenzierung mithilfe generativer KI-Systeme zu erreichen, um jeden Schüler an der Grenze seiner individuellen Leistungsfähigkeit fördern zu können.

Die Möglichkeiten von *MicroApps* mit Schnittstellen zur generativen KI sind immens, eine Erforschung steht noch am Anfang. Weitere Anwendungen könnten in automatisierten Debattierapps liegen, bei denen der maschinelle Counterpart die Argumente der Gegenseite transparent macht und so zu einem besseren Verständnis eines Sachverhalts beiträgt. Auch über den Einsatz interaktiver digitaler Zwillinge, also videoechter, digitaler Abbilder der Professoren, die auf Basis auf die speziellen Inhalte ihrer Vorlesungen trainierter *Large Language Models* in beliebigen Sprachen mit ihrer Umgebung interagieren können, denken wir nach. In unserer in Gründung befindlichen Digital-Universität, der *German University of Digital Science*⁵, sollen diese digitalen Zwillinge den (digital) Studierenden rund um die Uhr für Rückfragen und Erläuterungswünsche zum Lernstoff zur Verfügung stehen.

Wir sollten uns durch die scheinbaren Alleskönner-KI-Systeme nicht verunsichern lassen und vor allem nicht in apokalyptische Phantasien abtauchen, deren Zweck vermutlich in einem klugen Marketing aufseiten der KI-Entwickler besteht, sondern die vielen Herausforderungen, mit denen uns diese neue Technologie konfrontiert, selbstbewusst und souverän annehmen. Generative KI ist ein wertvolles neues Werkzeug, das dann zum vollen Nutzen der Allgemeinheit heranreift, wenn wir lernen, wie sie funktioniert und wie wir sie für unsere Zwecke effektiv gebrauchen können. Die Bildung und das lebenslange Lernen bieten einen Schlüssel dazu, diese neue Technik zu beherrschen, ihre Gefahren zu erkennen, ihren Missbrauch zu verhindern und sie zum Wohl unserer Gesellschaft einzusetzen. Noch besteht die Möglichkeit, nicht als Getriebene dieser Technologieentwicklung zu stranden.

⁵ Website German University of Digital Science, <https://german-uds.de> [letzter Zugriff: 30.07.2024].