

ChatGPT - ein Katalysator für welche Zukunft?

Stellungnahme der Initiative Digitaler Humanismus, Wien, März 2023

Die Veröffentlichung von ChatGPT hat weltweit Begeisterung, aber auch Ängste ausgelöst. Sie hat öffentliches Bewusstsein für die weitreichenden Auswirkungen der neuesten generativen KI geschaffen. Diese reichen von vielen nützlichen Anwendungen bis zu beunruhigenden Konsequenzen für unsere offenen demokratischen Gesellschaften und das Leben der BürgerInnen.

Diese Entwicklung bietet jedoch auch eine unerwartete, willkommene Gelegenheit, der breiten Öffentlichkeit als auch politischen EntscheidungsträgerInnen zu erläutern, was KI-Werkzeuge wie ChatGPT sind und wie sie funktionieren. Hier gilt es vorteilhafte Einsatzmöglichkeiten hervorzuheben, aber auch Bedenken hinsichtlich der beträchtlichen Risiken zu äußern. Dabei unterstreichen wir die Dringlichkeit einer breiten öffentlichen Diskussion und einer rechtzeitigen sowie angemessenen Regulierung. Zudem plädieren wir für eine faire Chance für die akademische Forschung mit ihren eigenen Schwerpunkten, unabhängig von den großen Konzernen und ihren enormen Investitionen, die jetzt in die weitere kommerzielle Nutzung fließen.

ChatGPT: einige Fakten

ChatGPT (Generative Pre-Trained Transformer), das Ende 2022 von OpenAI als dialogorientierte Version ihrer generativen KI-Modelle veröffentlicht wurde, ist ein Large Language Model (LLM). Es basiert auf einer innovativen Kombination aus sogenannten Unsupervised und Reinforcement Learning, letzteres durch Menschen.¹ Die dabei notwendigen großen Datenmengen stammen aus Quellen wie Büchern, (Nachrichten-)Artikeln, Websites oder Beiträgen bzw. Kommentaren aus sozialen Netzwerken, um seine Kernfunktion als Dialogsystem zu erfüllen. Dabei simuliert das System menschliche Gespräche–Hierbei wird anhand von Wahrscheinlichkeiten "geschätzt", welches Wort bzw. Sätze wahrscheinlich auf welches Wort bzw. Sätze folgen wird bzw. werden. Dies geschieht in Übereinstimmung mit einem bestimmten Schreibstil oder Tonfall, was die Illusion eines Gesprächs mit einem Menschen erzeugt. Obwohl diese Systeme diesen Aspekt der Sprache offensichtlich gut beherrschen, sind sie auf der funktionalen Ebene eingeschränkt, z. B. durch fehlende Argumentations- und Abstraktionsfähigkeiten oder eine sehr begrenzte Situationsmodellierung. Wir Menschen kommunizieren mit anderen Menschen in der Absicht ein bestimmtes Ziel zu erreichen. Wir neigen aber dazu, diese Absicht allen Akteuren zuzuschreiben, die Sprache produzieren. So lassen wir uns leicht dazu verleiten, menschliche Intelligenz, wie wir sie verstehen, auf Maschinen zu projizieren, die in der Lage sind, Sprache in irgendeiner Form zu imitieren.

¹ Am 14. März wurde der Nachfolger GPT-4 eingeführt, der sowohl Bild- als auch Texteingaben akzeptiert.

Es ist wichtig zu betonen, dass solche Modelle überzeugende Kombinationen von Wörtern und Sätzen bilden können, aber kein menschliches Verständnis für unsere Fragen oder auch ihre eigenen Antworten haben. Sie haben auch kein Verständnis dafür, was "Fakten" sind und neigen dazu, sachliche Fehler zu produzieren und auch zu "halluzinieren". OpenAI räumt die Grenzen seiner Modelle ein: "ChatGPT schreibt manchmal Antworten, die vernünftig klingen, aber in Wirklichkeit falsch oder unsinnig sind". Dies war der Grund für Charakterisierungen wie "stochastischer Papagei" oder, weniger freundlich, "selbstbewusster Schwätzer", was natürlich leider gelegentlich auch auf Menschen zutreffen.

OpenAI und andere Unternehmen sind inzwischen in einen harten Wettbewerb eingetreten und sammeln weiterhin Feedbackdaten von einer rasch wachsenden Zahl von NutzerInnen. Das bedeutet, dass wir alle Versuchspersonen eines riesigen laufenden Feldversuchs dieser Unternehmen sind - ohne unsere Zustimmung.

Potenzielle "gute" Anwendungen

Ein gemeinsames Merkmal potenziell als vorteilhaft angesehener Anwendungen ist, dass KI-Tools wie ChatGPT als Assistenten des Menschen eingesetzt werden können, um dessen Fähigkeit zur Zusammenarbeit und Teilhabe an der Gesellschaft zu ergänzen. Ein solcher Ansatz, der den Menschen nicht ersetzt, sondern ergänzt, wurde bereits in den Anfängen der KI erörtert. Dabei gibt es eine lange Liste potenziell nützlicher Anwendungen, die von der Unterstützung bei der Vorbereitung juristischer Schriftsätze, der Übersetzung, dem Programmieren, dem Chipdesign, der Materialwissenschaft bis hin zur Arzneimittelentdeckung und natürlich der Bildung und Ausbildung reichen. Dies kann (und wird) Innovation fördern, neue Geschäftsmöglichkeiten eröffnen und auch Produktivitätssteigerungen in vielen Wirtschaftszweigen ermöglichen.

Anzumerken ist, dass die Technologiesgeschichte zeigt, dass viele Anwendungen nicht vorhersehbar sind, da der soziale, wirtschaftliche und kulturelle Kontext, in dem neue Technologien von den NutzerInnen übernommen und angeeignet werden, sehr unterschiedlich ist. Die NutzerInnen sind also nicht nur passive VerbraucherInnen, sondern haben in der Vergangenheit bewiesen, dass sie in der Lage sind, neue Einsatzmöglichkeiten, die ihren Bedürfnissen besser entsprechen, zu entwickeln oder zu erfinden.

Potenzielle "schlechte" Anwendungen und Risiken

Gegenwärtig ist die Liste der Bedenken hinsichtlich eines möglichen Missbrauchs dieser Technologie lang, etwa

- die "industrielle" Produktion von immer überzeugenderen betrügerischen E-Mails aller Art und deren weite Verbreitung durch eine Reihe von Akteuren, einschließlich einiger Regierungen;

- die automatische Produktion von Fake News und einer großen Anzahl von Websites für gezielte Desinformationskampagnen von Einzelpersonen, Unternehmen und Staaten;
- Automatisierung der Kommunikation für Betrügereien, z. B. durch Aufforderungen für Lösegeldzahlungen;
- schnelle und sehr effiziente Produktion von maßgeschneidertem Malware-Code, einschließlich neuer Arten von Malware, die auch die Versuche sich zu wehren, "abhören" kann;
- tiefgreifende Fälschungen durch auf Bilder trainierte Systeme.

Besorgniserregend ist auch die Gefahr, die von der Nutzung dieser Systeme für Kinder bzw. Jugendliche während ihrer prägenden Schulzeit ausgeht. Dies ist die Zeit, in der die wichtigsten kognitiven Fähigkeiten des Menschen entwickelt werden. Wir sind besorgt, dass die übermäßige Nutzung dieser Werkzeuge als (selbst nicht gut verstandene) Abkürzungen zum Lernen und Üben solcher Fähigkeiten, diese ernsthaft beeinträchtigen könnte.

Wir haben auch Bedenken bezüglich des Mangels an Transparenz und Verantwortlichkeit verbunden mit dem Verlust eines gemeinsamen Bezugs darauf, was wahr/falsch oder gut/schlecht ist. Es besteht Grund zur Sorge, dass solche Sprachmodelle absichtlich missbraucht werden oder sogar versehentlich dramatische Unfälle und soziale Störungen verursachen könnten, wobei völlig offenbleibt, wer für den Schaden verantwortlich gemacht werden kann. Es ist auch wahrscheinlich, dass ein neues "Wettrüsten" im Bereich der Cybersicherheit in Gang kommt, bei dem die SicherheitsexpertInnen ständig die Präventivmaßnahmen verbessern müssen. Dies ist mit hohen wirtschaftlichen Kosten verbunden, könnte aber auch zu einer weiteren Beschneidung bürgerlicher Freiheiten führen.

Schließlich sind wir besorgt über die enorme Konzentration von Macht und Ressourcen für die künftige KI-Forschung und Entwicklung in den Händen von Big Tech, insbesondere wenn die derzeitige Entwicklung ungebremst weitergeht. Es gibt genügend historische Beispiele, die zeigen, dass die Konzentration von wirtschaftlicher Macht schnell zu einer Konzentration von politischer Macht führt und umgekehrt.

ChatGPT, Big Tech und liberale Demokratie

In den letzten zehn Jahren haben zahlreiche Studien den fragilen Zustand der liberalen Demokratien auf der ganzen Welt aufgezeigt und sind zu dem Schluss gekommen, dass diese sich auf dem Rückzug befinden oder sogar einen "steilen Abstieg" erleben; inzwischen lebt mehr als die Hälfte der Welt unter autoritären Regimen. Wirtschaftliche Ungleichheiten, die Auswirkungen der ungezügelter Globalisierung und konstitutionelle Verwerfungen gehören zu den Hauptursachen für den Niedergang. Diese sind allerdings auch mit der Rolle von Big Tech und ihren Plattformen verwoben.

Die Konzentration von wirtschaftlicher und politischer Macht in den Händen einer kleinen Elite an der Spitze einiger weniger Großunternehmen ist ein Problem. Sie können damit einen übergroßen Einfluss auf demokratische Prozesse und Institutionen ausüben, und somit zu einer Erosion des öffentlichen Raums beitragen. Ihre politische Macht beruht neben der Lobbyarbeit auf den zunehmenden Möglichkeiten über ihre Plattformen die öffentliche Meinung zu beeinflussen, zu lenken, zu manipulieren und zu polarisieren. Diese Möglichkeiten wurden und werden von einigen genutzt um demokratische Prozesse zu untergraben.

Diese Bedrohungen für die liberale Demokratie können sich durch den KI-Wettlauf zwischen diesen Tech-Giganten, der nun durch die Entwicklungen rund um ChatGPT ausgelöst wurde, noch verschärfen. Riesige Investitionen sind in ein sich rasch entwickelndes digitales Ökosystem geflossen, dessen Richtung, Umfang und weitere Investitionen allein von wenigen Unternehmen bestimmt werden.

Die Kosten für das Training eines sehr großen KI-Systems wie ChatGPT und die damit verbundenen Anforderungen an Rechenleistung und Datenmengen stellen insofern eine Gefahr dar, indem diese Werkzeuge in den Händen einiger weniger Unternehmen und einiger weniger Regierungen konzentriert sind. Niemand kann solche Tools in einer Garage bauen, und auch akademische Einrichtungen sind immer weniger in der Lage, mit diesen Unternehmen und den großzügig finanzierten Start-ups, die dann aufgekauft werden, Schritt zu halten.

Wenn der negative Kreislauf der Plattform-Ökonomie mit ihrem "Winner takes it all"-Phänomen weitergeht, wird jeder Rest der vielleicht idealistischen ursprünglichen Vision eines pluralistischen digitalen Ökosystems, welches KI-Werkzeuge entwickelt um die Bürgerinnen und Bürger zu befähigen ihre Fähigkeiten zur sinnvollen Teilhabe an einer offenen demokratischen Gesellschaft zu ergänzen und damit zu erweitern, verloren gehen. Die Vision des digitalen Humanismus, um es kurz zu sagen, würde endgültig brutal weggefegt werden.

Kurzum: Die Entwicklung von KI-Tools wie ChatGPT, die die Frage stellen, was es bedeutet, ein Mensch zu sein, darf nicht in den Händen einiger weniger mächtiger Unternehmen liegen. Solche KI muss ein öffentliches gesellschaftliches Gut sein und demokratisch gesteuert werden.

Was getan werden muss

Die zentralisierte Kontrolle dieses Experiments und die damit verbundenen Entscheidungen über die Ausrichtung der KI-Forschung stellen eine Bedrohung für die Nachhaltigkeit der liberalen Demokratie dar. Dies wird uns aktuell deutlich und unmittelbar vor Augen geführt wird.

Die Tatsache, dass diese Bedrohung bei politischen EntscheidungsträgerInnen (man beachte den jüngsten „Aufruhr“ durch ChatGPT im Europäischen Parlament bei der Debatte über den geplanten AI Act), akademischen Netzwerken, Initiativen wie Digital Humanism und anderen Gleichgesinnten Besorgnis hervorrufen, gibt Hoffnung und Motivation, in diesem wahrscheinlich begrenzten Zeitfenster zu handeln.

Die Notwendigkeit einer Regulierung und unsere Sorge, dass unregulierte KI im Großen und Ganzen eine schlechte KI sein wird, sind nicht unbemerkt geblieben. In der Europäischen Union wird der EU-AI-Act derzeit intensiv diskutiert, mit dem Ziel, diesen in den kommenden Monaten zu verabschieden. In den USA haben die Federal Trade Commission und das Justizministerium mehrere Kartellverfahren wegen monopolistischen Verhaltens angestrengt, wobei das jüngste Verfahren gegen Google am 25. Januar 2023 eingereicht wurde. In diesem Zusammenhang verfolgen die USA und die EU im transatlantischen Trade and Technology Council eine gemeinsame Roadmap für vertrauenswürdige KI und Risikomanagement. Die Umsetzung der KI-Politik muss dabei kontinuierlich überwacht und vor allem auch dynamisch aktualisiert werden.

Neben dem Bestehen auf eine gute Regulierung und deren Umsetzung ist es notwendig, die Öffentlichkeit und die politischen EntscheidungsträgerInnen zu informieren. Sie müssen dafür sensibilisiert werden, was für die Zukunft der demokratischen Institutionen und Prozesse auf dem Spiel steht und wie groß die Gefahr ist, dass die Bürgerinnen und Bürger zu Spielfiguren in diesem auf wenige beschränkten Wettlauf um Profit und Marktanteile werden. Der öffentliche Raum für offene Diskussion und Teilhabe läuft Gefahr, von Inhalten, die bewusst auf Fehlinformation und völligem Unsinn beruhen, überschwemmt zu werden, womit schließlich das Gefühl einer demokratischen, kollektiven Zugehörigkeit untergraben wird.

Wir, die Initiative Digitaler Humanismus, AkademikerInnen und ForscherInnen im Bereich der Informatik und der Sozial- und Geisteswissenschaften, die sich aus der Perspektive des Digitalen Humanismus mit den aktuellen Entwicklungen der KI und ihren gesellschaftlichen und kulturellen Auswirkungen befassen, fühlen uns dafür verantwortlich, die breite Öffentlichkeit und die politischen Entscheidungsträger über die Chancen und Risiken zu informieren und aufzuklären, die mit ChatGPT (und ähnlichen KI-Werkzeugen) verbunden sind. Zukünftige Schwerpunkte der KI-Forschung und -Entwicklung sollten von menschlichen Anliegen und Bedürfnissen bestimmt sein, in einer Zukunft, die nicht von den gewinnorientierten Zielen großer Unternehmen dominiert wird.

Wir verpflichten uns, den Ansatz des digitalen Humanismus in unserer KI-Forschung und -Entwicklung anzuwenden, öffentlich rechenschaftspflichtig zu bleiben und offen für konstruktive Debatten zu sein.