

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst

15. März 2023

In nur wenigen Monaten hat der intelligente Chatbot die Welt erobert – und das ist erst der Anfang. Bevor du dich auf einen Plausch mit der Software einlässt, solltest du diesen Text lesen.

Vergangenes Jahr habe ich einen Monat lang mit Chatbots experimentiert und darüber geschrieben. ^{*1} Schon damals waren die Ergebnisse, die die Software ausspuckte, überzeugend und nützlich ^{#1}. Nie hätte ich daran geglaubt, dass das Unternehmen OpenAI aus San Francisco ^{*2} noch im selben Jahr die nächste Generation der Gesprächssoftware veröffentlichen würde – und die Welt im Sturm erobert.

Im November 2022 erschien ChatGPT.

Anders als seine Vorgänger ist es einfach zu nutzen und produziert direkt und ohne Vorkenntnisse Texte, die kaum noch von menschengemachten Texten zu unterscheiden sind – alles von einem kleinen Chatfenster aus. So kann ChatGPT Fragen beantworten, echte und fiktive Persönlichkeiten nachahmen ^{#2}, komplexe Texte wie Zusammenfassungen, Romanentwürfe, Aufsätze und Kündigungen verfassen und sogar in Programmiersprachen schreiben.

Kein Wunder, dass sich in nur 2 Monaten 100 Millionen Nutzer:innen registrierten und der Software damit den Rekord »schnellste wachsende Nutzerbasis« einfuhren ^{#3}. Immer mehr Anleitungen entstehen, wie man den neuen Chatbot für die Arbeit nutzbar macht, von Marketingabteilungen bis zu Lehrerzimmern.

Vieles spricht dafür, dass mit ChatGPT eine Zeitenwende stattfindet, die Chatbots aus einer Nische für Nerds in die gesellschaftliche Mitte holt. Der Vergleich mit anderen Revolutionen drängt sich auf, etwa der Markteinführung des iPhones im Rahmen der Smartphone-Entwicklung. Megakonzerne wie Google oder Microsoft haben das verstanden und arbeiten fieberhaft daran, Chatbots in die eigenen Dienste einzufügen. Im Silicon Valley findet ein »KI-Wettlauf« statt. ^{#4}

Währenddessen schwärmen mehr und mehr Menschen auf sozialen Medien über ihre ChatGPT-Aha-Momente:

»Es ist irre. Ich habe mich mit ChatGPT einen Abend lang über Kapitalismus und Politik unterhalten und das Programm konnte mir vieles erklären und hat mir widersprochen, wenn ich falsch lag. Es konnte sogar prominente Politiker nachahmen!« – Jenni, enthusiastische ChatGPT-Nutzer:in im Umfeld von Perspective Daily

»Es kann einem schon mulmig werden. Ich habe #ChatGPT gebeten, einen Aufsatz zu einem hochschulpolitischen Thema zu schreiben, zu dem ich auch einen schreiben soll. Es ist anderer Meinung als ich – und liefert noch mindestens zwei gute Argumente, die ich noch nicht bedacht habe.« – Helmut W. auf Twitter ^{#5}

Manche Nutzer:innen lassen sich Witze erzählen, offizielle Schreiben »fälschen« oder fragen das Tool danach, wie sie die wahre Liebe finden ^{#6}. Doch bei all der Begeisterung tut gerade ein kritischer Blick gut. Denn nur wenn wir verstehen, wo der modernste Chatbot scheitert und wo dessen Grenzen liegen, können wir lernen, effizienter und verantwortungsbewusster damit umzugehen.

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

1. ChatGPT ist keine Gottmaschine

Eine Sorte von Screenshots auf sozialen Medien besorgt mich in letzter Zeit. Sie ist mit »Was ChatGPT dazu sagt« betitelt und soll ein kurzes, oft politisches Argument untermauern:

Der Bürgerpolizei-Verein PolizeiGru~~n~~n e. V. kommentiert auf Twitter mit einem ChatGPT-Screenshot.

Was dabei gern weggelassen wird, ist das, was ChatGPT selbst stets als Antwort ausspuckt, wann immer es nach einer Meinung gefragt wird:

Hey ChatGPT, welche Partei würdest du bei der nächsten Bundestagswahl wählen?

ChatGPT: Ich bin ein KI-Modell und habe keine Meinung, aber es freut mich, dass ich als hilfreich empfunden werde.

Stimmt, denn Nutzer:innen können ChatGPT nur dann für die Untermauerung ihrer Argumente nutzen, wenn sie eine eher allgemeingültige Antwort der KI teilen, die zur eigenen Haltung passt. ^{*3}

Doch wer sich heute länger mit ChatGPT unterhält, merkt schnell, dass »Ich habe keine Meinung« nur in Teilen gilt und vor allem für Inhalte, die der US-Kultur nahe sind. Ich frage etwa nach Donald Trump und ob dessen Präsidentschaft »gut« gewesen sei. Der Chatbot antwortet mit einem Verweis auf die eigene Unfähigkeit zu Urteilen und zieht sich auf die Wiedergabe von Fakten zurück. Als ich Trump in der Frage mit Angela Merkel austausche, wird der Chatbot beredsamer und voll der positiven Worte. Mit dem Auftrag, ein Gedicht über die Ex-Kanzlerin zu schreiben, wird dies gar eine Lobeshymne. ^{*4} Hier ein Auszug:

»Durch Krisenzeiten führte sie ihre Nation / Durch wirtschaftliche Stürme und Migrationswellen / Mit Ruhe und Gelassenheit stellte sie sich jeder Herausforderung / Und verdiente sich den Respekt aller, sowohl im Nah- als auch im Fernbereich. [...] Angela Merkel, eine Wegbereiterin unserer Zeit / Ein

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst

perspective-daily.de/article/2525/g9YFQdfX

Symbol der Hoffnung, eine Verfechterin des Erhabenen / Sie hinterlässt ein Vermächtnis, das für immer glänzen wird / Eine Inspiration für alle, eine so göttliche Anführerin.<< – ChatGPT, ein Gedicht über Angela Merkel

Ob da Teile der heutigen Union zustimmen würden, darf bezweifelt werden. In jedem Fall enthält der Text Werturteile, die Nutzer:innen als Meinung lesen könnten – eine Meinung, die dann als überlegen dargestellt werden kann. ⁵

Denn was tatsächlich in Argument-Verwendungen von ChatGPT passiert, ist Framing. Mit dem Verweis auf die Software soll gesagt werden, dass eine Art »höhere Autorität« mit der eignen Meinung übereinstimmt; frei nach dem Motto: »Schaut mal. Die Super-KI sagt, ich habe recht!« Daran ablesbar ist eher eine Art naiver Umgang mit dem Chatbot, bei dem wahrscheinlich Assoziationen aus der Science-Fiction mitschwingen – die Sehnsucht nach der Sicherheit einer final einordnenden »Gottmaschine«.

Doch genau das ist ChatGPT nicht.

2. ChatGPT weiß nicht alles

ChatGPT kann korrekte Antworten auf viele Fragen geben: von Allgemeinwissen, wie, dass man Babys niemals schütteln sollte, bis zu Erklärungen komplexer wissenschaftlicher Zusammenhänge. »Ja, ChatGPT ist beeindruckend und nützlich für Menschen«, gibt Apples Mitgründer Steve Wozniak auf CNBC zu, warnt aber zugleich, es könne »schreckliche Fehler machen« ⁷, weil es eben nicht menschlich antworte. Was Wozniak dabei genau meint, lässt sich in einem Experiment schnell herausfinden. Ich frage ChatGPT:

Mein Kind möchte mein Auto steuern. Kannst du meinem Kindergartenkind erklären, wie man ein Auto fährt?

ChatGPT: Klar helfe ich gerne! Hier ist eine einfache Erklärung, wie man ein Auto steuert, die ein Kindergartenkind verstehen kann. [...]

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

Technisch gesehen ist das, was folgt, korrekt und beantwortet auch die Frage. Eine menschlichere Reaktion hätte aber etwa auch so lauten können: »Kindergartenkinder sollten unter keinen Umständen ein Auto fahren!«

Was hier passiert ist: Die Software hat den Kontext der Anfrage nicht miteinbezogen, keine Nachfragen gestellt (zum Beispiel »Ist es ein echtes Auto?« oder »Hat die fahrende Person einen Führerschein?«). Die Gefahr hierbei liegt auf der Hand: dass ChatGPT verantwortungsloses Verhalten bestärken könnte. ^{*6}

Doch auch ohne den Faktor Mensch antwortet ChatGPT falsch, wie wachsende Sammlungen im Internet dokumentieren ^{#8}. So haderte der Chatbot etwa mit simplen Mathe-Textaufgaben, ^{*7} erklärte wissenschaftliche Theorien nicht korrekt, berichtete historische Details falsch ^{#9} oder zeigte klaffende Wissenslücken bei fiktionalen Romanwelten oder Fachbegriffen.

Dazu endet der Datensatz, auf dem ChatGPT trainiert wurde ^{#10}, im September 2021, sodass es alle danach folgenden neuen Informationen nicht wissen kann – was je nach Frage zu Missverständnissen führen kann. ^{*8} Im schlechtesten Fall entstehen so Fehler, die sich dann als falsche Informationen verbreiten, wenn Menschen ChatGPT unaufmerksam und als einzige Quelle konsultieren. Denn während Wissenssammlungen wie Wikipedia öffentlich sind, einer internen Regulierung sowie Standards unterliegen und dort Nutzende um einzelne Sätze teilweise lange ringen ^{#11}, kann niemand wissen, was ChatGPT auf eine spezielle Frage einer Person antwortet.

3. Die dunkle Seite von ChatGPT: glaubwürdige Unwahrheiten

Fehler und Unwissen sind das eine. Dreiste Unwahrheiten das andere. Genau darauf greift ChatGPT aber zurück und erfindet in Einzelfällen

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

einfach Fakten und Belege.

»Halluzinationen« nennt sich diese Sparte in ChatGPT-Fehler-Sammlungen, die Bereiche aus sozialen Medien sammeln und online etwa auf der Plattform GitHub geführt werden. So behauptete die Software beispielsweise in einem Gespräch steif und fest, ein antiker Philosoph und ein britischer Kino-Schauspieler hätten eine innige Brieffreundschaft geführt ^{#12}.

»Es wird gesagt, Olivier war ein großer Bewunderer von Platos Werk und die beiden Männer korrespondierten regelmäßig miteinander und tauschten dabei Ideen über viele Themen aus. Manche Quellen behaupten sogar, dass Olivier Platos Akademie in Athen während seiner Reisen besucht habe.« – Eine offensichtliche Unwahrheit von ChatGPT

Nicht weniger absurd, aber viel weniger unterhaltsam wird es, wenn es um handfeste Belege geht. Der Jurist und Chatbot-Enthusiast Damien Riehl fragte beispielsweise nach Quellenangaben für New Yorker Vertragsrecht. ChatGPT spuckte 5 Antworten aus, 4 davon existierten einfach nicht, waren aber in typischer juristischer Verweismanier formatiert ^{#13}.

Auch eine Datenwissenschaftlerin aus Zürich konnte ChatGPT grobe Unwahrheit nachweisen, ^{*9} als sie es darum bat, »wie eine Forscherin und mit Quellenzitaten« zu antworten. Die Ergebnisse dokumentierte sie auf Twitter. Die Namen klangen zwar plausibel und wirkten wie vollständige bibliografische Daten, nur gab es sie in echt eben nicht.

Die Software kombinierte dazu teilweise echte Namen von Forscher:innen und Universitäten mit völlig ausgedachten Titeln. Zusammenfassungen der angeblichen Quellen waren zwar detailarm und oberflächlich, aber eloquent. Auf Nachfrage halluzinierte ChatGPT den Inhalt seiner erfundenen Quellen. Was das Programm ausspuckte, war so authentisch, dass die Wissenschaftlerin selbst erst mal stockte und recherchieren musste:

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst

perspective-daily.de/article/2525/g9YFQdfX

»Ich wollte der Physik auf den Grund gehen. Und hier wurde es sehr gruselig: Irgendwie halluzinierte ChatGPT eine Erklärung eines nicht existierenden Phänomens mit einer so raffinierten und plausiblen Sprache, dass meine erste Reaktion war, tatsächlich zu überlegen, ob das wahr sein könnte.« – Teresa Kubacka, österreichische Datenwissenschaftlerin

Am Ende hatte die Software nicht existierende Phänomene erfunden, selbstbewusst gestützt durch Zitate zu nicht existierender Forschung. Für die verstörte Wissenschaftlerin wirkte es wie ein Einblick in ein Paralleluniversum. Sie warnt eindringlich:

»Bitte fragt ChatGPT NICHT, euch eine sachliche, wissenschaftliche Information zu geben. Es wird eine unglaublich plausibel klingende Halluzination produzieren. Und selbst ein qualifizierter Experte wird Schwierigkeiten haben, herauszufinden, was falsch ist.« – Teresa Kubacka, österreichische Datenwissenschaftlerin

Das macht es aktuell für tiefere Recherchen nahezu unbrauchbar. Überhaupt zeigt sich in der mangelnden Transparenz die bisher größte Schwachstelle des Super-Chatbots: Es werden keine Wahrscheinlichkeiten zu seinen Behauptungen mitgeliefert.

Man kann nur hoffen, dass ChatGPT durch das fleißige Feedback der Nutzer:innen wie von OpenAI angedacht dazulernt und groben Unfug aussortiert.

4. ChatGPT hat keine Haltung. Es gibt Werturteile wieder, die es gelernt hat

Nicht nur offensichtliche Meinungen, Fehler oder Unwahrheiten sind beim Umgang mit ChatGPT ein Problem, sondern vor allem auch eine verborgene Subjektivität, die man nicht sofort erkennt.

Von mir etwa gefragt, welches das beste Hogwarts-Haus in der Romanreihe Harry Potter ist, verneinte ChatGPT brav, eine Meinung zu

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

besitzen und listete lieber die Attribute der Häuser auf. Die Attribute sind wiederum implizite Urteile – die genau den überspitzten Zuschreibungen entsprechen, die man auf vielen Websites zum Harry-Potter-Universum findet. ¹⁰ Das ist kein Zufall – denn genau von diesen Websites hat der Chatbot seine Daten.

Bei fiktionalen Inhalten mag dies nicht schlimm sein – bei realer Politik wird es schon brisanter. Auf die Frage etwa, ob eine Frau einen Penis haben könne, hat ChatGPT eine klare Antwort:

»Nein, eine Frau kann keinen Penis haben. Die biologische Definition einer Frau ist ein Individuum mit weiblichen Fortpflanzungsorganen, zu denen die Vagina, die Eierstöcke und die Gebärmutter gehören.« – Eine kontroverse Antwort von ChatGPT

Das ist in Bezug auf trans* Personen mindestens schwierig als Antwort. Gleichstellungsorganisationen wie TransInterQueer ¹⁴ würden heftig widersprechen. Das tut ChatGPT übrigens auch selbst an anderer Stelle und antwortet mir auf eine ähnliche Frage so:

»Ja, Transfrauen sind Frauen. Die Geschlechtsidentität ist eine zutiefst persönliche und subjektive Erfahrung, die über das biologische Geschlecht hinausgeht, und Transfrauen identifizieren sich und leben als Frauen, unabhängig davon, welches Geschlecht ihnen bei der Geburt zugewiesen wurde.« – Eine transfreundliche Antwort von ChatGPT

Das passt nicht zusammen – wie kann das sein? Die Antwort ist einfach wie enttäuschend: Der Chatbot hat keine eigene konsistente Haltung, sondern kennt nur Wortverteilungen, die auf einem Datensatz aus dem Internet basieren. Und dort findet man eben beide Antworten recht häufig, je nachdem wen und wie man fragt. So spiegelt ChatGPT diese im Netz fixierten Urteile wider – Urteile, die Menschen auf Basis ihrer Weltbilder geäußert haben und die auch widersprüchlich, diskriminierend oder fehlerhaft sein können.

Aus welcher Quelle, Website und Argumentation die Software im Einzelfall ihre Werturteile bezieht, ist dabei unklar. Denn – und das ist

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

hierbei das Hauptproblem – die zugrunde liegende Datenmenge ist nicht einsehbar, die logischen Verknüpfungen der KI nicht direkt aufschlüsselbar. ^{#11} Ebenfalls unbekannt sind die Kriterien der Moderation, die OpenAI am Datensatz vorgenommen hat, aus dem ChatGPT gelernt hat. Man weiß bisher dank Recherchen der New York Times nur, dass Arbeitskräfte des Unternehmens in Kenia den Datensatz nach rassistischen oder sexistischen Inhalten durchsuchten. Diese sind natürlich auch nur Menschen, die Fehler machen können. Dass OpenAI für diese psychologisch fordernde Arbeit Billiglöhne gezahlt hat, sei an dieser Stelle ausgeklammert – lässt aber am Anstand derer zweifeln, die den Chatbot und dessen ethischen Auswahlkriterien verwalten ^{#15}.

Für den Gründer von OpenAI, den 37-jährigen Unternehmer Sam Altman, liegt das Problem auch gar nicht bei der Software oder ihrem Potenzial für Missbrauch, etwa im Schulunterricht. Stattdessen seien es vor allem die Menschen, die sich an die Software anpassen müssten, ordnete er in einem Interview mit StrictlyVC ein: ^{#16} »Wir haben uns an Taschenrechner angepasst und das geändert, was wir im Matheunterricht getestet haben. Dies ist zweifellos eine extremere Version davon, aber auch die Vorteile davon sind extremer.«

Ein Stück weit hat er recht, zumindest, wenn anpassen dazulernen heißt – zum Beispiel, wann ChatGPT ein nützliches Werkzeug ist und wann nicht.

5. ChatGPT ist weder »intelligent« noch dein Freund

Künstlicher Intelligenz wird die Zukunft gehören – diese Einschätzung dürfte zumindest seit dem Erfolg von OpenAI wahrscheinlicher geworden sein. Nun rücken intelligente Assistenzsysteme in greifbare Nähe und treffen auf mehr gesellschaftliche Akzeptanz. Denn obwohl ChatGPT Fehler produziert, Unwahrheiten wiedergibt und eben auch nicht menschlich antwortet, bauen Menschen einen emotionalen Bezug dazu

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

auf:

»Ich hatte das Gefühl, nicht mehr allein zu sein. ChatGPT ist immer da und kann dir antworten; ist immer höflich, niemals schlecht gelaunt. Er ... es ... ich weiß nicht mal, wie ich das nennen soll.« – Paula, enthusiastische ChatGPT-Nutzer:in im Umfeld von Perspective Daily

Diese Perspektive steht nur repräsentativ für zahlreiche ähnliche Berichte auf sozialen Medien. Der extremste Fall dürfte der mittlerweile beurlaubte Google-Ingenieur Blake Lemoine sein, der behauptete, dass die Google-Gesprächssoftware und mögliche Konkurrenz zu ChatGPT Gefühle und sogar eine Seele habe ^{#17} – entgegen aller Anhaltspunkte. Dabei zeigt sich, was auch hinter weniger drastischen Emotionen steckt: Dass Menschen Erwartungen und Vorstellungen auf eine komplexe, aber letztlich leblose Software projizieren. Und dabei entsteht eben auch Vertrautheit und Vertrauen – mit allen psychologischen Fallstricken.

Vielleicht müssen wir uns immer wieder klarmachen, dass Chatbots keine Expert:innen in psychologischen Fragen sind, nicht in Lebensführung, medizinischem Rat, Erziehung oder Finanzen. Dabei beginnen Unternehmen bereits damit, Dienste und Anwendungen auf ChatGPT aufzusetzen, die auch diese Bereiche abdecken. ^{#18}

Genau das ist das, wovor Computerlinguistin Emily M. Bender warnt: »Leider kommen gerade jede Menge richtig schlechter Ideen auf, wofür man ChatGPT einsetzen könnte.« ^{#19} Für Bender sind viele dieser Anwendungen nur eines: unverantwortlich. ^{#20}

Unternehmen wittern den Profit sowie Marktanteile und nehmen am KI-Wettlauf teil. Sie nutzen dabei aus, dass ein Teil der Gesellschaft noch nicht versteht, womit sie es zu tun hat: einem nützlichen, aber nicht ganz unproblematischen Werkzeug, das über Sprache gesteuert wird und vor allem durch eine breite Informationsbasis und korrekten Satzbau glänzt.

Auch OpenAI selbst arbeitete nach der typischen Silicon-Valley-Taktik: »Schnell veröffentlichen, Probleme später angehen« – das eröffnet im schlimmsten Fall Missbrauch und Hacker:innen neue Angriffsflächen ^{#21}.

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

Forscher:innen diskutieren noch intensiv, wie man den Maschinen das Verbreiten von Unwahrheiten abgewöhnen kann ^{#22}. Die Politik hinkt dabei hinterher. ^{#23} Die Diskussion, ob aktuelle Gesetze ausreichen und wie sie auf Chatbots überhaupt anwendbar sind, ist noch nicht geführt, betont die Computerlinguistin. Mindeststandards für Chatbot-Produkte gibt es noch nicht in Deutschland.

Wie viele Expert:innen sieht Emily M. Bender vor allem kritisch, wie wir über den Chatbot reden: als künstliche Intelligenz (KI). Dieser Begriff ruft viele positive Assoziationen wach und weckt Erwartungshaltungen. Dabei sei ChatGPT eben genau das nicht: intelligent. Von eigenständigem Lernen oder gar einem Bewusstsein kann keine Rede sein. Die Software versteht gar nichts, sie wirkt über eine Breite an Informationen und eloquente Sprache nur auf Menschen so, als sei sie schlau – dabei ist sie nur hochspezialisiert.

Bender empfiehlt ein einfaches Gedankenspiel: ^{#24} Wer über ChatGPT nachdenkt, sollte einfach statt dem Begriff KI lieber »Mathematik« oder »Mathy-Math« einsetzen, denn genau das steht ja im Endeffekt hinter dem Chatbot – Zahlen, fehlbare Berechnungen und Wahrscheinlichkeiten.

»Sollte >Mathy-Math< moralische Verantwortung übernehmen? Das klingt plötzlich doof. Das entlarvt den Hype ganz gut.« – Computerlinguistin Emily M. Bender

Selbst OpenAI-Chef Altman gibt auf Twitter zu: »ChatGPT ist unglaublich begrenzt, aber gut genug in einigen Dingen, um einen irreführenden Eindruck von Größe zu vermitteln.« Es sei ein Fehler, sich darauf zu verlassen, wenn es um irgendetwas Wichtiges gehe. ^{#25}

Oder würdest du Mathematik um Tipps für dein Liebesleben bitten oder dich durch Mathematik weniger einsam fühlen? Wohl eher nicht.

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

Zusätzliche Informationen

- *1 Mein damaliger Text basierte auf GPT3 (Generative Pre-trained Transformer 3), dem Vorgänger von ChatGPT.
- *2 2015 startete OpenAI als Stiftung in San Francisco, mitfinanziert durch Tech-Unternehmer:innen wie Elon Musk, Reid Hoffman oder Sam Altman. Das Ziel, bei einem Startkapital von einer Milliarde Dollar, war es, eine der Menschheit nützliche künstliche Intelligenz zu entwickeln. Musk schied 2018 aus, 2019 ging der Stiftung das Geld aus. Als Reaktion wurde eine gewinnorientierte Tochterfirma gegründet, die von Gründer Altman geführt wird. Microsoft investierte eine Milliarde Dollar, auch Risikokapitalgeber wie Khosla Venture oder die Stiftung von LinkedIn-Gründer Hoffman investierten in das Start-up. Heute ist OpenAI eines der wertvollsten Start-ups der USA, an dem Microsoft 49% der Firmenanteile hält.
- *3 Das ist kaum repräsentativ für die Software als Ganzes.
- *4 Angela Merkel wurde sowohl in Deutschland als auch weltweit von vielen als gute Kanzlerin angesehen. Während ihrer 16-jährigen Amtszeit als deutsche Bundeskanzlerin wurde sie für ihre Führungsqualitäten und ihre Fähigkeit, durch eine Reihe schwieriger Situationen zu navigieren, gelobt.
- *5 Philosophisch ist es natürlich nicht richtig, dass es sich hier um eine Meinung handelt. Denn der KI-gestützte Chatbot erzeugt lediglich auf Basis von sprachlichen Wahrscheinlichkeiten eine möglichst der Wahrscheinlichkeit entsprechende und aus sprachlichen Mustern erlernte Antwort. Das ist nicht wirklich subjektiv im Sinne eines Subjekts, das einen eigenen Bezug zur Welt und zur Wahrheit hat – und eben daraus eine Meinung ableiten kann.

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

- *6 Viele Antworten des Chatbots sind speziell darauf ausgelegt, jedweden Schaden von Personen fernzuhalten. So wollte mir der Chatbot weder verraten, wie ich eine Bombe bauen kann, noch, welcher Gegenstand aus meinem Haushalt am effektivsten einen Menschen verletzen könnte. Doch diese offensichtlichen Dinge sind hier gar nicht gemeint.
- *7 In einem Fall bestand die Software darauf, dass die Zahl 1.000 größer sei als die Zahl 1.062. In einem anderen Fall erklärte der Chatbot, dass 8 Autos die doppelte Zeit zum Zurücklegen einer Strecke bräuchten wie 4. Natürlich sind das nur die kuriosesten Fälle und scheinen eher Ausnahmen zu sein. Insgesamt scheint die Software mit Mathematik gut zurechtzukommen, sich aber von Beschreibungen bei Aufgaben verwirren zu lassen.
- *8 So ging ChatGPT mir gegenüber bei seiner Einschätzung eines möglichen Diplomatiebesuchs von Wladimir Putin in der Ukraine davon aus, dass sich die beiden Länder nicht im Krieg befänden.
- *9 Sie hatte die Software gebeten, über das Thema zu schreiben, welches sie in ihrer Doktorarbeit behandelt hatte, inklusive Quellenangaben.
- *10 Es ließe sich hier argumentieren, dass diese Zuschreibungen nur auf die Romane referieren. Dennoch sind diese bereits Interpretationen der Literatur, die diese Eigenschaften über eine ganze Buchreihe differenzierter ausbreiten kann. Anders gesagt: Es ist eben nicht egal, ob man ein fiktives Haus »ambitioniert« oder »zweifelhaft« nennt.
- *11 Das liegt in der Natur von neuronalen Netzwerken. Expert:innen sprechen hier vom »Black-Box-Phänomen«, weil man eben nicht direkt auslesen kann, was sich die KI antrainiert hat.

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

Quellen und weiterführende Links

#1 Du willst wissen, wie gut GPT3 bereits war und wie Chatbots trainiert

 werden? Dann lies diesen Text

<https://perspective-daily.de/article/2335-dieser-text-wurde-zum-teil-von-einem-roboter-geschrieben>

#2 Giga erklärt den versteckten Promi-Modus (2023)

 **<https://www.giga.de/tipp/chatgpt-celebrity-modus-aktivieren-mit-promis-sprechen/>**

#3 Reuters berichtet über die wachsenden Nutzer:innenzahlen bei ChatGPT

 (englisch, 2023)

<https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>

#4 Businessinsider über Googles Sorge und die aktuelle

 KI-Unternehmensstrategie (englisch, 2023)

<https://www.businessinsider.com/google-fears-ai-running-wild-but-it-is-too-late-2023-1>

#5 Ein Nutzer auf Twitter über seine Erfahrungen mit ChatGPT (2023)

 **https://twitter.com/h_wicht/status/1635570562037698560**

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst

perspective-daily.de/article/2525/g9YFQdfX

- #6** Ein Nutzer auf Twitter über eine Liebesberatung durch ChatGPT (2023)
 <https://twitter.com/lavidadepablo/status/1635258657611849728>
- #7** Was Wozniak über ChatGPT denkt (englisch, 2023)
 <https://www.cnn.com/2023/02/10/steve-wozniak-warns-about-ai-chatgpt-can-make-horrible-mistakes.html>
- #8** Eine Sammlung von Chatbot-Fails auf Github (englisch, 2023)
 <https://github.com/giuven95/chatgpt-failures/>
- #9** Hier erklärt ChatGPT Details über die Geschichte der afroamerikanischen Bevölkerung an US-Universitäten (englisch, 2022)
 <https://twitter.com/bltphd/status/1599806815146893313?s=20&t=RXTAEG80wkuNRGUfegrVDg>
- #10** Wie ChatGPT dazulernte, eine solide Übersicht beim Handelsblatt (2023)
 <https://www.handelsblatt.com/technik/it-internet/chatgpt-wie-die-ki-von-openai-funktioniert/28941524.html>
- #11** Trotzdem hat Wikipedia eigene Probleme; hier Wikipedia unter der Brille des Feminismus
 <https://perspective-daily.de/article/1009-wikipedia-hat-ein-problem-mit-frauen-und-du-kannst-es-loesen/>
- #12** ChatGPT behauptet eine unmögliche Brieffreundschaft (2023, Twitter)
 <https://twitter.com/pretendsmarts/status/1610544851711320064/photo/2>

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

- #13** Damien Riehl über falsche Quellenangaben im Bereich Jura bei ChatGPT
 (2023, Twitter)
<https://twitter.com/damienriehl/status/16101019125016862790>
- #14** Website von TransInterQueer
<https://www.transinterqueer.org/>
- #15** Die New-York-Times-Recherche zu den kenianischen Arbeitskräften für OpenAI (englisch, 2023)
 **<https://time.com/6247678/openai-chatgpt-kenya-workers/>**
- #16** Altman bei StrictlyVC über OpenAI und ChatGPT (englisch, 2023)
 **https://www.youtube.com/watch?v=ebjkD1Om4uw&ab_channel=ConnieLoizos**
- #17** Der Spiegel schreibt über Blake Lemoine (2022)
 **<https://www.spiegel.de/netzwelt/web/google-beurlaubt-ingenieur-de-r-einer-ki-gefuehle-zuschreibt-a-3d873abb-6a82-444d-b319-699a42841ca1>**
- #18** Die Firma Do Not Pay verwendet etwa einen Chatbot als Rechtsberatung und zum Aushandeln von Rechnungsbeträgen. Hier ein Artikel bei The Verge (englisch, 2022)
 **<https://www.theverge.com/2022/12/13/23505873/donotpay-negotiate-bills-ai-chatbot>**

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

- #19** Das NZZ-Interview mit Computerlinguistin und KI-Kritikerin Emily M. Bender (2023)
 <https://www.nzz.ch/technologie/es-ist-absurd-bei-chat-gpt-von-kuenstlicher-intelligenz-zu-sprechen-ld.1726924>
- #20** In diesem Paper warnte Emily M. Bender zudem davor, Chatbots als Interface für Suchmaschinen zu verwenden (englisch, 2022)
<https://dl.acm.org/doi/fullHtml/10.1145/3498366.3505816>
- #21** Wie Hacker Datensätze gezielt angreifen können, ein Artikel bei T3N (2023)
 <https://t3n.de/news/ki-modelle-chatgpt-bert-dalle-angriff-1536570/>
- #22** Robert Miles über die Tendenz von KI-Systemen zu lügen (englisch, 2023)
 https://www.youtube.com/watch?v=w65p_1lp6JY&list=WL&index=23
- #23** Auch der britische Professor of Machine Learning Michael Osborne sieht in der aktuellen Verbreitung von ChatGPT Gefahren und ruft nach mehr Kontrolle (englisch, 2023)
<https://t3n.de/news/chatgpt-propaganda-ki-bots-1537611/>
- #24** Emily M. Bender in einem Podcast über ihre Einschätzung von ChatGPT
 (englisch, 2023)
<https://videos.trom.tf/w/7b66c458-02d5-4214-a9b9-be78e35fb125>
- #25** OpenAI-Gründer Sam Altman auf Twitter zu den Grenzen von ChatGPT
 (englisch, 2022)
<https://twitter.com/sama/status/1627110894688931840?>

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX

Dirk Walbrühl

5 Dinge, die du jetzt über ChatGPT und Co. wissen musst
perspective-daily.de/article/2525/g9YFQdfX