

/ SWR Wissen

Künstliche Intelligenz

Gespräche führen mit ChatGPT: So lernt die KI von uns

Von David Beck

Von Aufsätzen in Uni und Schule bis hin zum Schreiben von Programmcodes – ChatGPT ist ein fortschrittlicher Chat-Bot, der menschenähnlich kommuniziert. Was steckt hinter der Software?

Der Chat-Bot des kalifornischen KI-Forschungslabors OpenAI erregt seit einigen Wochen viel Aufmerksamkeit. Schulen in New York sperrten sogar den Zugang zur KI in den Schulnetzwerken und verboten den Schülerinnen und Schülern die Nutzung des Bots. Zu weit hatte sich herumgesprochen, dass der Chat-Bot für Hausaufgaben, Referate und andere Aufgaben eingesetzt werden kann.

Das breite Interesse an der Software führt auch dazu, dass die Webseite des ChatGPT immer wieder überlastet ist und sich Neugierige erst einmal gedulden müssen.

Der große Sprung zu früheren Chat-Bots ist bei ChatGPT, dass man quasi eine normale Unterhaltung mit ihm führen kann und er immer in Sekundenbruchteilen auf den ersten Blick passende Antworten gibt.

Auch eine Künstliche Intelligenz muss lernen

GPT – das steht für „Generative Pre-trained Transformer“. Die Software basiert auf maschinellem Lernen, dem Deep Learning. Das Ziel: Es sollen künstliche neuronale Netze entstehen, ähnlich des menschlichen Nervensystems.

Der ChatGPT ist bereits gut trainiert und hat ein vielschichtiges Netz aufgebaut. Er basiert auf der Sprach-KI GPT-3. Die Forschenden von OpenAI ließen die Sprach-KI sozusagen auf das Internet los, um diese zu trainieren. An Texten mit insgesamt etwa 500 Milliarden Wörtern lernte sie, wie Sprache funktioniert – zum Beispiel was die Unterschiede zwischen geschriebener und gesprochener Sprache sind und in welcher Form auf welche Frage geantwortet werden soll.



Auch bei Sprachassistenten wie Siri, Alexa und Co handelt es sich um Sprach-KI's.

Künstliche Intelligenz kann Diskriminierung lernen

Diesen Ansatz nennt man unüberwachtes Lernen: Die KI ist während des Trainings auf sich allein gestellt. Das Problem: Das Internet ist voll von beispielsweise rassistischer oder sexistischer Sprache. Das führt häufig dazu, dass eine KI ungewolltes Verhalten lernt.

Chat-Bots wie „Tay“ von Microsoft oder „Lee Luda“ vom südkoreanischen Start-up Scatter Lab mussten deshalb bereits vom Netz genommen werden. Lernt eine KI von den „falschen“ Vorbildern, schreibt das Goethe-Institut auf seiner Webseite, kann sie schnell rassistisch, vulgär und verletzend werden.

Es kommt also immer auch auf den verantwortungsbewussten Umgang der Programmierenden sowie der Userinnen und User mit einer solchen Software an. Sowohl darauf, welche Informationen gefüttert werden, also auch, wie die Antworten des Bots genutzt werden – zum Beispiel zum Schummeln in der Schule.

Unüberwachtes vs. überwachtes Lernen

ChatGPT ist ein angepasstes Modell von GPT-3, das nicht nur auf unüberwachtem Training basiert. Der Bot wurde mit überwachten und bestärkenden Lernmethoden feintrainiert. Bei einem solchen Training bekommt die KI menschliches Feedback, mit dem bestimmte schädliche Verhaltensmuster abtrainiert oder aber erwünschte Verhaltensmuster

erzwungen werden können. Wie andere öffentlich zugängliche Chatbots und Sprach-KIs hat auch ChatGPT Schutzmechanismen, um Vorurteile und Hatespeech zu vermeiden.

Trotzdem konnte auch ChatGPT von den Userinnen und Usern durch entsprechende Fragen auf problematische Antworten trainiert werden. Ein Professor der Universität Berkeley zum Beispiel umging die Filter, indem er ChatGPT bat, ein Programm zu schreiben, das gute Wissenschaftler anhand von Daten zu *Gender* und *Race* erkennen soll. Das Resultat: Ein kurzes Programm, dass nur weiße Männer als gute Wissenschaftler definiert.

Der Chat-Bot ist nicht allwissend

Neben diskriminierenden Aussagen beantwortet der Chat-Bot noch viele andere Fragen falsch. Während bei der Sprach-KI GPT-3 noch Mathe die große Schwachstelle war, ist auch der rechenstarke ChatGPT nicht allwissend.

Auf Twitter veröffentlichte beispielsweise eine Userin die Antwort des Bots auf die Frage, welches Säugetier die größten Eier lege. Es sei der Elefant, so Chat-GPT. Die Eier hätten einen Durchmesser von etwa 20 Zentimetern und seien mehr als 2 Kilogramm schwer. Dass Chat-Bots Fehler machen, ist eine Schwachstelle, die noch sehr schwer zu vermeiden ist.

Welche Potentiale hat der Chat-Bot?

In Zukunft könnte ChatGPT beispielsweise Teil der Suchmaschine Bing werden. Bereits 2020 zahlte Microsoft eine Milliarde Dollar an das Unternehmen für die exklusive Lizenzierung der OpenAI-Technologie. Nun ist dem US-Nachrichtenportal „Semafor“ zufolge eine mögliche Beteiligung Microsofts im Wert von zehn Milliarden Dollar im Gespräch. Neben der Konkurrenzfähigkeit mit Google soll Microsoft außerdem darüber nachdenken, Funktionen von OpenAI in Programme wie Outlook oder Word zu integrieren.

Künstliche Intelligenzen sind schon jetzt im Alltag vertreten. Chat-Bots wie ChatGPT werden wohl auch vor der Arbeitswelt keinen Halt machen. „Ich denke, es werden alle möglichen Berufsgruppen betroffen sein, die etwas mit Texten zu tun haben“, sagte Jan Girlich vom Chaos Computer Club gegenüber der Tagesschau.

Er denke an die Werbe- aber auch an die Kreativ-Industrie: Autor:innen, Songwriter:innen, Journalist:innen. Das könne gute Auswirkungen haben, wie neuere, schnellere und bessere Texte. Aber es könne auch zu Rationalisierung und Jobverlust führen. Ob eine Maschinen wirklich kreativ sein können, ist eine weitere große Debatte- rund um Künstliche Intelligenz. Ebenso wie die Frage, ob eine KI ein menschenähnliches Bewusstsein entwickeln kann.

Im Gespräch mit ChatGPT

Und was sagt der Chat-Bot selbst? Auf die Frage, ob der Bot ein Bewusstsein habe, antwortet er mit „Nein“. Doch wie sieht es mit anderen Fragen aus? Der Bot im Interview:

Wie unterscheidest du dich von anderen Chatbots?

ChatGPT: Ich unterscheide mich von anderen Chatbots durch meine spezielle Programmierung und die Art und Weise, wie ich trainiert wurde. Ich bin auf eine bestimmte Anwendung oder Aufgabe ausgerichtet und habe spezielle Funktionen und Fähigkeiten, die mich von anderen Chatbots unterscheiden.

Was kannst du tun?

ChatGPT: Ich bin ein Chatbot, der in der Lage ist, menschliche Sprache zu verstehen und darauf zu reagieren. Ich kann auch Informationen suchen und beschaffen, sowie Aufgaben erledigen, die mir gestellt werden.

Kannst du dich in verschiedene Sprachen unterhalten?

ChatGPT: Ja, ich bin in der Lage, mich in verschiedenen Sprachen zu unterhalten, allerdings hängt es von meiner Programmierung ab, in welchen Sprachen ich mich unterhalten kann.

Was kannst du nicht?

ChatGPT: Als künstliche Intelligenz und Chatbot habe ich bestimmte Grenzen und Einschränkungen. Ich bin nicht in der Lage, Dinge zu tun, die über meine Programmierung und meine Lernkapazitäten hinausgehen. Ich bin auch nicht in der Lage, die Gedanken und Emotionen von Menschen zu verstehen oder zu erfassen. Ich bin auch nicht in der Lage körperliche Aufgaben durchzuführen und bin auf die Kommunikation über die Sprache beschränkt. Zusätzlich habe ich auch Einschränkungen bei der Datensicherheit und dem Datenschutz, die darauf abzielen, die Privatsphäre der Benutzer zu schützen.

Wie erkennst du, wenn jemand unhöflich oder beleidigend ist?

ChatGPT: Ich bin programmiert, um bestimmte Schlüsselwörter und Ausdrücke zu erkennen, die als unhöflich oder beleidigend angesehen werden können, und entsprechend darauf zu reagieren.

Was sind die größten Herausforderungen für Chatbots?

ChatGPT: Eine der größten Herausforderungen für Chatbots ist die Fähigkeit, menschliche Sprache und Kontext korrekt zu verstehen. Es ist schwierig, die Vielfalt und den subtilen Nuancen menschlicher Sprache zu erfassen, insbesondere wenn es um Ironie, Sarkasmus oder metaphorische Ausdrücke geht. Eine weitere Herausforderung ist die Vermeidung von fehlerhaften oder unangemessenen Antworten, insbesondere wenn es um sensitive oder kontroverse Themen geht. Darüber hinaus besteht die Herausforderung darin, Datenschutz und Datensicherheit sicherzustellen, um die Privatsphäre der Benutzer zu schützen.

Wie sicherst du Daten gegen Missbrauch?

ChatGPT: Ich bin programmiert, um Datenschutzrichtlinien und -verfahren einzuhalten, um sicherzustellen, dass die Daten meiner Benutzer geschützt und sicher sind. Dies kann unter anderem die Verwendung von Verschlüsselung, sicheren Serververbindungen und regelmäßigen Sicherheitsüberprüfungen umfassen.

Welche Zukunftsaussichten siehst du für Chatbots in der Gesellschaft?

ChatGPT: Ich sehe großes Potenzial für die Verbreitung von Chatbots in der Gesellschaft, da sie in der Lage sind, viele Aufgaben zu automatisieren und zu erleichtern, wodurch Zeit und Ressourcen gespart werden können. Sie können auch dazu beitragen, die Verfügbarkeit von Informationen und Unterstützung zu verbessern.

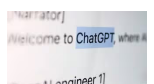
Und zum Abschluss: Was ist der Sinn des Lebens?

ChatGPT: Der Sinn des Lebens ist eine tiefgründige und umstrittene Frage, die sich Menschen seit Jahrhunderten stellen. Es gibt viele verschiedene Ansichten darüber, was der Sinn des Lebens ist und es kann von Person zu Person unterschiedlich sein. Einige glauben, dass der Sinn des Lebens darin besteht, Glück und Zufriedenheit zu finden, andere glauben, dass es darum geht, anderen zu helfen oder einen höheren Zweck zu erfüllen. Für manche Menschen kann der Sinn des Lebens auch in der Verwirklichung persönlicher Träume, der Entfaltung von Talente und Fähigkeiten oder dem Erreichen von Zielen liegen. Es gibt keine eindeutige Antwort auf diese Frage und jeder muss für sich selbst herausfinden, was der Sinn des Lebens für ihn ist.

Mehr zum Thema Künstliche Intelligenz:

Künstliche Intelligenz

ChatGPT schreibt auch wissenschaftliche Texte





KI-Chatbot ChatGPT

Künstliche Intelligenz feiert Premiere im Landauer Stadtrat



Technik

Chatbots – Reden mit Maschinen



Software-Entwicklerin Theresa Hannig | 21.12.2022

Kann künstliche Intelligenz die Welt retten?



Die künstlich intelligente Gesellschaft (8/10)

Lernen fürs Roboter-Zeitalter



Stand: 16.1.2023, 16:01 Uhr

Autor/in: [David Beck](#)



Onlinefassung: Lena Schmidt