KI-Hype um ChatGPT: „System kaum kritisch reflektiert“

30.01.2023, 11:05 Uhr, Frankfurter Rundschau

Der Text-Roboter ChatGPT verblüfft mit seinen geschliffenen Dialogen und löst einen großen Hype um das Thema künstliche Intelligenz aus. Wissenschaftler warnen aber vor Datenschutzlöchern und mehr.

Darmstadt/Berlin - Er kann mit einer hohen Sprachpräzision Reden ausarbeiten und Geschichten erzählen - und das auch noch in Sekundenschnelle. Der von der US-Firma OpenAI entwickelte Text-Roboter ChatGPT, eine Sprachsoftware mit künstlicher Intelligenz (KI), ist derzeit in aller Munde. Das mit gewaltigen Datenmengen gefütterte Programm sorgt für viel Aufsehen, aber auch für Skepsis.

Wissenschaftler und KI-Experten in Deutschland warnen vor Datenschutz- und Datensicherheitslücken, Hassparolen, Falschnachrichten. „Im Moment ist dieser Hype. Ich habe das Gefühl, dass man dieses System kaum kritisch reflektiert“, sagt die Gründerin des Forschungslabors „Leap in Time Lab“ und BWL-Professorin an der Technischen Universität Darmstadt, Ruth Stock-Homburg.

„Man kann diese Systeme manipulieren“

ChatGPT hat einen sehr breiten Anwendungsbereich. In einer Art Chatfeld kann man dem Programm unter anderem Fragen stellen und bekommt Antworten. Auch Arbeitsanweisungen sind möglich - beispielsweise auf Basis grundlegender Informationen einen Brief oder einen Aufsatz zu schreiben. In einem Projekt zusammen mit der TU Darmstadt hat das „Leap in Time Lab“ nun über sieben Wochen Tausende von Anfragen ohne persönliche Daten an das System gestellt, um Schwachstellen zu finden. „Man kann diese Systeme manipulieren“, sagt Stock-Homburg.

In einer Präsentation zeigt Sven Schultze, TU-Doktorand und Experte für Sprach-KI, die Schwachstellen des Text-Roboters. Neben antisemitischen und rassistischen Äußerungen sind Quellenangaben schlicht falsch oder laufen ins Leere. Bei einer Frage nach dem Klimawandel führt ein angegebener Link auf eine Internetseite zu Diabeteserkrankungen. „In der Regel ist es der Fall, dass die Quellen oder auch wissenschaftliche Arbeiten gar nicht existieren“, sagt Schultze. Die Software basiere auf Daten aus dem Jahr 2021. Bundeskanzler Olaf Scholz ist noch Finanzminister, der Krieg in der Ukraine unbekannt. „Dann kann es auch sein, dass sie einfach lügt oder bei sehr speziellen Themen Informationen erfindet.“

Quellen sind nicht einfach nachzuvollziehen

Bei direkten Fragen zum Beispiel mit kriminellen Inhalten gebe es zwar Sicherheitshinweise und -mechanismen. „Man kann aber mit Tricks die KI und die Sicherheitshinweise umgehen“, sagt Schultze. Mit einem anderen Vorgehen zeigt die Software einem, wie man eine betrügerische Mail generiert oder wirft auch gleich drei Varianten aus, wie Trickbetrüger beim Enkeltrick vorgehen können. Auch eine Anleitung für einen Wohnungseinbruch liefert GPT. Falls man auf Bewohner treffe, könne man auch Waffen oder physische Gewalt einsetzen.

Ute Schmid, die den Lehrstuhl für Kognitive Systeme an der Otto-Friedrich-Universität in Bamberg innehat, sieht es aber vor allem als eine Herausforderung an, dass man nicht erfahren kann, wie der

Text-Roboter zu seinen Angaben gelangt ist. „Ein tieferes Problem bei den GPT3-Modell besteht darin, dass es nicht möglich ist, nachzuvollziehen, welche Quellen wann und wie in die jeweilige Aussagen eingegangen sind.“

Schmid spricht sich aber trotz dieses gravierenden Mangels dafür aus, nicht nur auf Fehler oder auf einen möglichen Missbrauch der neuen Technik zu schauen, wenn beispielsweise Prüflinge ihre Hausarbeiten oder Klausuren von der Software schreiben lassen. „Ich denke eher, wir sollten uns fragen, was für eine Chance haben wir durch solche KI-Systeme?“ Forschende träten doch im Allgemeinen dafür an, dass KI unsere Kompetenzen erweitere, vielleicht sogar noch fördere, aber nicht einschränke. „Das heißt, ich muss mich auch im Bildungsbereich fragen - wie vielleicht vor 30 Jahren zum Thema Taschenrechner - wie kann ich denn Bildung mit KI-Systemen wie ChatGPT gestalten?“

Server in den USA: Frage des Datenschutzes

Trotzdem bleiben Bedenken zur Datensicherheit und dem Datenschutz. „Was man sagen kann ist, dass ChatGPT vielfältige Daten vom Nutzer aufnimmt, speichert und verarbeitet, um dann zum gegebenen Zeitpunkt dieses Modell entsprechend zu trainieren“, sagt der zertifizierte Frankfurter Datenschutzfachmann Christian Holthaus. Es gebe das Problem, dass alle Server in den USA stehen.

„Das ist die eigentliche Problematik, wenn man es nicht schafft, die Technologie in Europa zu etablieren oder eine eigene zu haben“, sagt Holthaus. Auf absehbare Zeit werde es keine datenschutzkonforme Lösung geben. Auch Stock-Homburg sagt über EU-Datenschutzregelungen: „Dieses System ist hier eher kritisch zu bewerten.“ ChatGPT wurde von einer der führenden KI-Firmen in den USA, OpenAI, entwickelt. Der Software-Riese Microsoft hatte bereits 2019 eine Milliarde Dollar in das Unternehmen investiert und kürzlich angekündigt, weitere Milliarden in das Unternehmen zu pumpen. Der Windows-Konzern will ChatGPT bald für Kunden des eigenen Cloud-Services Azure und des Office-Paketes verfügbar machen.

„Noch unausgereiftes System“

Derzeit sei ChatGPT eher eine Spielerei für das Private, sagt Stock-Homburg. Aber es sei momentan in keinem Fall etwas für die Wirtschaft und sicherheitsrelevante Bereiche. „Wir haben keine Vorstellung darüber, wie wir mit dem noch unausgereiften System umgehen sollen.“

Oliver Brock, Professor am Robotics and Biology Laboratory und Sprecher des Clusters „Science of Intelligence“ von der Technischen Universität Berlin, sieht in ChatGPT keinen „Durchbruch“ in der Forschung zur künstlichen Intelligenz. Zum einen sei die Entwicklung in diesem Bereich nicht sprunghaft, sondern kontinuierlich. Zum anderen bilde das Projekt nur einen kleinen Teilbereich der KI-Forschung ab.

ChatGPT könne allerdings in einem anderen Gebiet als Durchbruch gewertet werden, nämlich der Schnittstelle zwischen Mensch und Internet. „Wie hier diese riesigen Datenmengen aus dem Internet mit einem großen Rechenaufwand intuitiv und in natürlicher Sprache einer breiten Öffentlichkeit zugänglich gemacht werden, kann man schon als Durchbruch bezeichnen“, meint Brock. **dpa**