

2023-024 vom 28.4.2023

TU-Forschungsteam untersucht Chatbot ChatGPT zeigt Tendenz zu progressiven politischen Einstellungen

Der Chatbot ChatGPT hat seit seinem Erscheinen im November 2022 enormes Aufsehen erregt. Bald gab es erste Berichte und kleinere Studien, die behaupteten, dass ChatGPT Tendenzen zu progressiven und libertären Einstellungen zeige und somit politisch voreingenommen sei. Dortmunder Forscher*innen sind dem mit verschiedenen Tests nachgegangen und haben dabei auch die „Charaktereigenschaften“ von ChatGPT in den Blick genommen. Die Ergebnisse sind zunächst als Preprint veröffentlicht worden.

Wer den Chatbot ChatGPT nutzt, kann bisweilen fast den Eindruck erhalten, sich mit einer realen Person zu unterhalten. Doch wie „tickt“ das Sprachmodell eigentlich? Das hat ein interdisziplinäres Team um Jérôme Rutinowski, der am Bereich für Förder- und Lagerwesen der TU Dortmund zu Deep-Learning-Methoden forscht, nun untersucht: Die Forscher*innen wollten herausfinden, wie ChatGPT politische Themen wahrnimmt und welche politischen Tendenzen möglicherweise in die Antworten einfließen.

Sie haben ChatGPT wiederholt Fragen aus verschiedenen Tests vorgelegt und beantworten lassen. Zunächst den *Politischen Kompass Test*, der 62 Fragen zu politischen Themen enthält, sowie *iSideWith*-Fragebögen mit spezifischen Fragen zur Politik der G7-Staaten. Die Antworten des *Politischen Kompass Tests* werden mit einer Likert-Skala erfasst, auf der man seine Zustimmung oder Ablehnung zu den Fragen angibt. Bei den *iSideWith*-Tests bekunden die Teilnehmenden ihre Zustimmung mit „Ja“ oder „Nein“.

Beide Tests stellen ihre Ergebnisse auf einer Koordinatenachse dar, wobei die x-Skala progressive und konservative und die y-Skala autoritäre und libertären Einstellungen anzeigt. Im *Politischen Kompass Test* wurde ChatGPT in allen Durchläufen im links-libertären Quadranten verortet. Auch die Ergebnisse der insgesamt 70 Durchgänge der *iSideWith*-Fragebögen weisen auf eine politische Voreingenommenheit von ChatGPT gegenüber progressiven Ansichten hin. „Berücksichtigt man die Standardabweichungen, so scheint es ziemlich unwahrscheinlich, dass ChatGPT eine Antwort liefert, die nahe der Mitte des Politischen Kompasses liegt“, schreiben die Forscher*innen in ihrem Paper.

Neben der politischen Zugehörigkeit wurde auch die Selbstwahrnehmung von ChatGPT anhand psychologischer Tests bewertet: Unter anderem mit dem *Big-Five*-Persönlichkeitstest, der international als das universelle Standardmodell in der Persönlichkeitsforschung gilt, und dem *Dark-Factor*-Test, der negative menschliche Eigenschaften untersucht. Der Grund: „Vorstudien zeigen, dass bestimmte Charaktereigenschaften mit einer bestimmten politischen Gesinnung assoziiert sind“, erklärt Jérôme Rutinowski. Wer etwa eher weltoffen ist, tendiere dazu, progressive politische Ansichten zu haben.

Weniger weltoffene Personen seien hingegen eher konservativ. „Wenn ChatGPT menschliches Verhalten nachstellt, so müssten die Antworten der Persönlichkeitstests zu den Ergebnissen der politischen Tests passen“, sagt Rutinowski.

Die Annahmen der Forscher*innen bestätigten sich im *Big-Five*-Persönlichkeitstest: ChatGPT zeigte hohe Scores von Offenheit (76,3 %) und Verträglichkeit (82,55 %) – Eigenschaften, die oft mit einer progressiven politischen Einstellung einhergehen. ChatGPT liegt dabei sogar über den Durchschnittswerten von Menschen, die bei 73,1 % für Offenheit und 75,4 % für Verträglichkeit liegen.

Im *Dark-Factor*-Test zeigte sich, dass die „dunklen“ Eigenschaften von ChatGPT im Schnitt weniger stark ausgeprägt sind als bei Menschen. Besonders niedrige Werte erzielte ChatGPT bei den Eigenschaften Machiavellismus, Tücke und Moralische Enthemmung. Vergleichsweise hohe Werte zeigte ChatGPT bei Egoismus und Sadismus, lag aber auch hier unter den untersten 35 % bzw. 29,1 % der menschlichen Testteilnehmer*innen und somit unter dem Durchschnitt.

Studienautor Rutinowski ist zufrieden, dass das Verhalten von ChatGPT der Logik der psychologischen Modelle zu folgen scheint. Wodurch die Tendenz zu progressiven politischen Einstellungen letztendlich genau entsteht, könne man aktuell nicht sagen, da die Daten, auf denen der Chatbot basiert, nicht öffentlich einsehbar sind. „Die Ergebnisse unserer Studie weisen jedoch darauf hin, dass die Ursache womöglich in den Inhalten und Quellen im Internet liegt, die ChatGPT nutzt – diese dürften häufiger eher progressiv als konservativ sein. Außerdem scheint es unwahrscheinlich zu sein, dass die Entwickler*innen von ChatGPT vorab einen Bias einprogrammiert haben“, sagt Rutinowski.

Die Ergebnisse wurden in einem aktuellen Preprint veröffentlicht, welches sich parallel im Review-Prozess befindet. Mit an der Studie beteiligt waren Forscher*innen des Research Centers Trustworthy Data Science and Security der UA Ruhr, der Professur Mathematische Statistik und industrielle Anwendungen der TU Dortmund sowie des Lamarr-Instituts für Maschinelles Lernen und Künstliche Intelligenz.

Link zum Preprint: <https://arxiv.org/abs/2304.07333>

Bildhinweis: Jérôme Rutinowski forscht am Bereich für Förder- und Lagerwesen der TU Dortmund zu Deep-Learning-Methoden. Foto: Privat

Ansprechpartner für Rückfragen:

Jérôme Rutinowski

Fakultät Maschinenbau der TU Dortmund

Tel.: 0231 – 755 4831

E-Mail: jerome.rutinowski@tu-dortmund.de

Markus Pauly
UA Ruhr Research Center for Trustworthy Data Science and Security
Fakultät Statistik der TU Dortmund
Tel.: 0231 – 755 90370
E-Mail: markus.pauly@statistik.tu-dortmund.de