Chatbots mit Persönlichkeit?

03.11.2026, 11:40 | Medien & Telekommunikation

Pressemitteilung von: *idw - Informationsdienst Wissenschaft*

Studie der Universität Mainz zeigt, wie KI-Sprachmodelle individuelle Merkmale in Texten ausdrücken und erkennen

Texte, die von sogenannten Large Language Models, also von KI-gestützten Chatbots stammen, weisen charakteristische „persönliche“ Merkmale auf. Außerdem können solche Chatbots individuelle Persönlichkeitseigenschaften anhand von Texten erkennen. Ob ein Text von einem Menschen oder einer KI stammt, erkennen Menschen aber zuverlässiger. Das haben Forschende des Psychologischen Instituts der Johannes Gutenberg-Universität Mainz (JGU) herausgefunden, wie sie jetzt in der Fachzeitschrift Scientific Reports berichten.

„Wir haben heute in immer mehr Lebensbereichen mit Chatbots zu tun – vom Kundenservice über digitale Beratungsangebote bis hin zu privat nutzbaren Chatbots, welche viele Menschen bereits alltäglich befragen und mit denen sie eigene Gefühle teilen“, sagt Sophie Körner, wissenschaftliche Mitarbeiterin in der Abteilung für Persönlichkeitspsychologie und Diagnostik der JGU und federführende Autorin der Studie. „Deshalb müssen wir verstehen, wie wir die Texte von Chatbots wahrnehmen – und ob wir sie überhaupt noch von menschlich geschriebenen Texten unterscheiden können.“

Gewissenhafter, verträglicher und emotional stabiler als Menschen

Für ihre Untersuchung verglich Körner zusammen mit Dr. Sarah Osterholz und Prof. Dr. Boris Egloff die Selbstbeschreibungen von sieben KI-Sprachmodellen – ChatGPT-4o, ChatGPT-3.5, Llama-3, Gemma 2, Gemini 2.5, Phi-3 und Orca 2 – mit denen von 70 Menschen. Dazu ließen sie die Chatbots und Menschen ihre jeweilige „Persönlichkeit“ mithilfe von psychologischen Fragebögen und offenen Fragestellungen auf Deutsch beschreiben. Beispielsweise sollten Aussagen zum eigenen Selbstwert oder den sogenannten Big Five – Eigenschaften wie Offenheit, Verträglichkeit oder Extraversion – bewertet werden. Bei den anschließenden offenen Fragen ging es beispielweise um die Interessen der Befragten, ihre Hoffnungen für die Zukunft und darum, wie sie sich typischerweise in Stresssituationen verhalten. Jedes Sprachmodell wurde zehnmal befragt, sodass die Forschenden insgesamt 140 Texte auswerteten.

Zunächst untersuchte das Team, wie ähnlich die Antworten der Sprachmodelle bei wiederholten Befragungen ausfielen. „Unsere Ergebnisse zeigen, dass die Sprachmodelle sich selbst durchaus recht zeitstabile Persönlichkeitsprofile zuschreiben“, sagt Ko-Autorin Osterholz. Dabei hätten sich die Sprachmodelle im Vergleich zu den menschlichen Teilnehmenden als gewissenhafter, verträglicher und emotional stabiler präsentiert.

Menschen erkennen KI-Texte zuverlässiger

Anschließend ließen die Forschenden zehn Menschen und sechs Chatbots alle Selbstbeschreibungen beurteilen. Anhand der Texte sollten sie Persönlichkeitsmerkmale wie Offenheit, Gewissenhaftigkeit und Selbstwertgefühl einschätzen. Sowohl den Menschen als auch den Chatbots gelang es dabei, Rückschlüsse auf die Persönlichkeitsmerkmale zu ziehen. Allerdings schätzten die Chatbots einige Merkmale der menschlichen Verfasser, darunter Offenheit und Selbstwertgefühl, tendenziell höher ein als die menschlichen Beurteilenden. Umgekehrt neigten Letztere dazu, den Sprachmodellen in allen abgefragten Merkmalen – allen voran in emotionaler Stabilität – höhere Werte zuzuschreiben als den menschlich geschriebenen Texten.

Außerdem sollten die Beurteilenden entscheiden, ob ein Mensch oder ein KI-Modell den jeweiligen Text geschrieben hatte. Dabei erreichten die menschlichen Beurteilenden eine Genauigkeit von 83 Prozent, die Chatbots dagegen nur 67 Prozent. Die beiden besten KI-Modelle bei dieser Aufgabe, Gemma und Gemini, konnten die Text-Urheberschaft etwa so zuverlässig bestimmen wie ein durchschnittlicher menschlicher Bewertender. Beide Gruppen neigten dazu, Texte

häufiger als von Menschen verfasst einzustufen; womöglich, weil eine Fehlattribution in umgekehrte Richtung vermieden werden sollte. Konkret heißt das, ein Chatbot-Text wurde eher fälschlicherweise als menschlich klassifiziert als es umgekehrt der Fall war. Jedoch ist nicht ausgeschlossen, dass KI-Modelle die Art des Verfassers zukünftig deutlich präziser werden einschätzen können, wenn ihre Trainingsdaten mehr künstlich generierte Texte enthalten.

Als Erkennungsmerkmale für KI-Texte nannten die Menschen vor allem deren Länge und eine gestelzte, besonders technische oder analytische Sprache. Auch eine computergestützte Sprachanalyse wies auf Unterschiede hin: Während KI-Texte mehr zeitbezogene Ausdrücke und Funktionswörter wie Artikel und Präpositionen enthielten, ging es in den menschlichen Selbstbeschreibungen öfter um Freizeit, Gefühle und soziale Beziehungen.

Möglichkeiten für die Diagnostik

Nach Einschätzung der Forschenden könnten KI-Sprachmodelle künftig auch in der psychologischen Diagnostik und im Personalwesen zum Einsatz kommen, etwa um ein erstes Bild von Persönlichkeitsmerkmalen zu gewinnen. Dies wird bereits in einigen Arbeitskontexten angestrebt. Wie zuverlässig solche Verfahren seien und unter welchen Bedingungen sie eingesetzt werden könnten, sei allerdings noch offen. „Sie bieten die Möglichkeit zu einer automatisierten, niedrigschwelligen Ersteinschätzung“, sagt Körner. Weil Sprachmodelle auf Grundlage menschlicher Texte trainiert werden, eignen sie sich ebenfalls sogenannte biases, das heißt Urteilsverzerrungen, an. Daher müsse erst noch sorgfältig und unter ethischen Gesichtspunkten geprüft werden, ob Persönlichkeitsbeurteilungen in sensiblen Bereichen Algorithmen überlassen werden sollten.

wissenschaftliche Ansprechpartner:
Sophie Körner
Abteilung für Persönlichkeitspsychologie und Diagnostik
Psychologisches Institut
Johannes Gutenberg-Universität Mainz
55099 Mainz
E-Mail: koerners@uni-mainz.de
www.ppd.psychologie.uni-mainz.de/

Originalpublikation:
S. T. Koerner et al., Personality expression and perception in large language models and humans, Scientific Reports, 3. November 2026,
DOI: 10.1038/s41598-026-75546-5,
https://doi.org/10.1038/s41598-026-75546-5

Johannes Gutenberg-Universität Mainz

JonasSiehoff (Mitarbeiter in der Presse- und Öffentlichkeitsarbeit)

01522 2524284

idw@uni-mainz.de

Link zur Pressemitteilung:

https://www.openpr.de/news/1329067/Chatbots-mit-Persoenlichkeit-idw.html