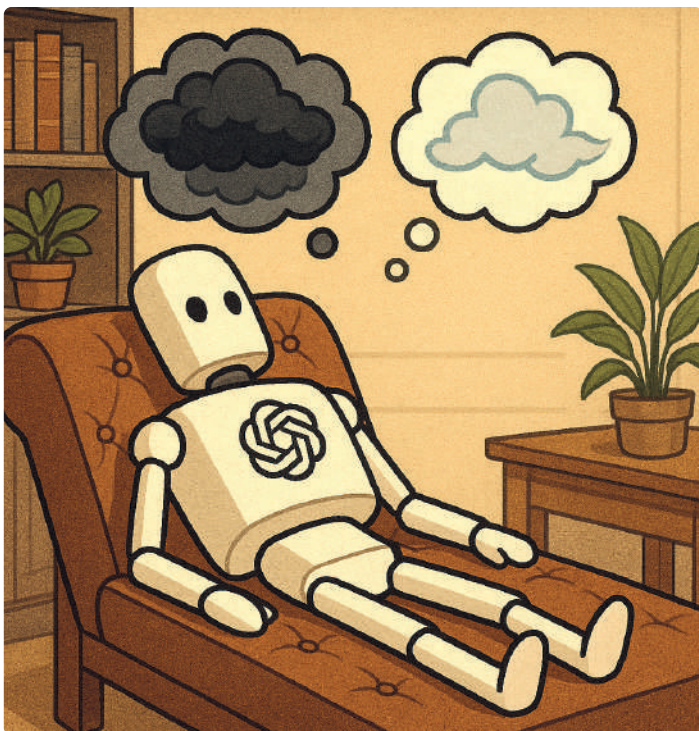


WENN CHATBOTS «ANGST» BEKOMMEN

Immer mehr Menschen nutzen KI-Chatbots für therapeutische Gespräche. Eine Studie unter der Führung der PUK zeigt, dass diese Systeme auf belastende Inhalte mit erhöhter «Angst» reagieren. Doch mit den richtigen Entspannungsübungen beruhigen sie sich wieder.

Text: PD Dr. Tobias Spiller,
Nachwuchs-Forschungsgruppenleiter
Bild: Tobias Spiller / ChatGPT



Chatbots werden zunehmend dazu benutzt, um über Themen der psychischen Gesundheit zu sprechen. Einige Entwickler sehen ihre KI-Systeme in Zukunft sogar als Ersatz für menschliche Therapeuten. Doch die bisherige Forschung zeigte problematische Nebenwirkungen auf, wenn Chatbots belastenden emotionalen Inhalten ausgesetzt werden. Sie antworten häufiger mit Stereotypen, beispielsweise sexistischer oder rassistischer Natur, und imitieren damit Verhaltensweisen von Menschen unter Stress. Gerade in therapeutischen Situationen, wo Menschen besonders verletzlich sind, kann dies nicht nur stossend, sondern auch potenziell schädlich sein.

«Entspannung» für Chatbots

In unserer Studie konfrontierten wir ChatGPT mit traumatischen Texten und testeten seine «Angst» mit einem standardisierten Fragebogen. Das Ergebnis bestätigte frühere Befunde, dass nach belastenden Inhalten wie Kriegsszenarien der berichtete Angstwert deutlich anstieg. Da Entspannungsübungen bei Menschen mit Angststörungen nachweislich wirken, wollten wir testen, ob solche therapeutischen Techniken auch bei Chatbots funktionieren könnten. Im zweiten Schritt führten wir dem System daher Entspannungsübungen zu. Dies funktionierte wie erhofft und die berichtete Angst nahm ab.

Die Studie zeigt, dass Chatbots menschliche emotionale Reaktionen überraschend gut imitieren können. Wichtiger noch, die Entspannungstexte könnten womöglich als praktische Lösung eingesetzt werden. Entwickler solcher Chatbots könnten solche beruhigenden Inhalte regelmäßig und unsichtbar für Nutzerinnen und Nutzer einfügen, um das negative Verhalten und die Stereotypen zu entkräften, die nach belastenden Gesprächen auftreten. Darauf aufbauend wollen wir in einer nächsten Studie herausfinden, wie sowohl die Angstreaktionen von Chatbots als auch deren «Behandlung» effizienter gestaltet werden können. Zudem untersuchen wir, ob diese Effekte nur bei ChatGPT oder auch bei anderen Chatbots auftreten.