MedWiss.Online

Forschende fordern klare Regelungen für KI im Bereich mentaler Gesundheit

Künstliche Intelligenz (KI) kann Gespräche führen, Emotionen spiegeln und menschliches Verhalten simulieren. Frei verfügbare große Sprachmodelle (Large Language Models, LLMs) – häufig genutzt als personalisierte Chatbots oder KI-Charaktere – erfahren zunehmend Gebrauch bei Fragen zur mentalen Gesundheit. Diese Anwendungen eröffnen neue Möglichkeiten, bergen jedoch zugleich erhebliche Risiken, insbesondere für verletzliche Nutzergruppen. Forschende des Else Kröner Fresenius Zentrums (EKFZ) für Digitale Gesundheit an der TU Dresden und des Universitätsklinikums Carl Gustav Carus haben nun zwei Fachartikel veröffentlicht, in denen sie eine stärkere regulatorische Aufsicht fordern.

Ihre Publikation „AI characters are dangerous without legal guardrails“ in der Fachzeitschrift Nature Human Behaviour beschreibt, warum klare Regeln für KI-Charaktere unbedingt erforderlich sind. Ein zweiter Beitrag in npj Digital Medicine warnt vor Chatbots, die ohne medizinische Zulassung therapieähnliche Unterstützung anbieten, und plädiert dafür, solche Systeme als Medizinprodukte einzustufen und zu regulieren.

Allgemeine LLMs wie ChatGPT oder Gemini sind nicht als therapeutische Anwendungen konzipiert oder zugelassen, können jedoch durch einfache Prompts oder spezifische Einstellungen schnell personalisiert und menschenähnlich reagieren. Diese Form der Interaktion kann sich negativ auf Jugendliche und Menschen mit psychischen Belastungen auswirken. Mittlerweile ist bekannt, dass Nutzerinnen und Nutzer starke emotionale Bindungen zu diesen Systemen aufbauen können. Dennoch sind KI-Charaktere in der EU und in den USA weitgehend unreguliert. Anders als klinische bzw. therapeutische Chatbots, die ausdrücklich für medizinische Zwecke entwickelt, getestet und zugelassen werden.

„KI-Charaktere fallen derzeit durch die Lücken der bestehenden Sicherheitsvorschriften“, erklärt Mindy Nunez Duffourc, Assistant Professor of Private Law an der Maastricht University und Mitautorin der ersten Publikation. „Oft werden sie nicht als Produkte eingestuft und entziehen sich daher Sicherheitsprüfungen. Und selbst dort, wo sie neu als Produkte reguliert sind, fehlen bislang klare Standards und eine wirksame Aufsicht.“

Hintergrund: Digitaler Austausch, echte Verantwortung

In den vergangenen Monaten wurde international über Fälle berichtet, in denen Jugendliche nach intensivem Austausch mit KI-Chatbots in psychische Krisen geraten sind. Die Forschenden sehen einen dringenden Handlungsbedarf: Systeme, die menschliches Verhalten imitieren, müssen klar definierten Sicherheitsanforderungen entsprechen und innerhalb verlässlicher rechtlicher Rahmen agieren. Aktuell gelangen KI-Charaktere jedoch auf den Markt, ohne zuvor eine regulatorische Prüfung zu durchlaufen.

In ihrer zweiten Publikation in npj Digital Medicine, „If a therapy bot walks like a duck and talks like a duck then it is a medically regulated duck“, machen die Autorinnen und Autoren auf die wachsende Zahl von Chatbots aufmerksam, die therapieähnliche Ratschläge geben oder sogar

lizenzierte medizinische Fachkräfte imitieren – ohne jegliche Zulassung. Sie argumentieren, dass LLMs mit solchen Funktionen als Medizinprodukte eingestuft werden sollten, mit klaren Sicherheitsstandards, transparentem Systemverhalten und kontinuierlicher Überwachung.

„KI-Charaktere sind bereits Teil des Alltags vieler Menschen. Oft vermitteln diese Chatbots den Eindruck, ärztliche oder therapeutische Ratschläge zu geben. Wir müssen sicherstellen, dass KI-basierte Software sicher ist. Sie soll unterstützen und helfen, nicht schaden. Dafür braucht es klare technische, rechtliche und ethische Regeln“, sagt Stephen Gilbert, Professor für Medical Device Regulatory Science am EKFZ für Digitale Gesundheit an der TU Dresden.

Lösungsvorschlag: Eine „Schutzengel-KI“, die aufpasst

Das Forschungsteam betont, dass die Transparenzanforderung des europäischen AI Act – also die Pflicht offenzulegen, dass es sich um Kommunikation mit einer KI handelt – nicht ausreicht, um gefährdete Gruppen zu schützen. Das Team fordert verbindliche Sicherheits- und Überwachungsstandards, ergänzt durch freiwillige Leitlinien, die Entwicklerinnen und Entwicklern dabei helfen, ihre Systeme sicher zu gestalten.

Als konkrete Maßnahme schlagen die Autorinnen und Autoren vor, zukünftige KI-Anwendungen mit einer Chat-Speicherfunktion auszustatten und mit einer „Guardian Angel AI“ oder „Good Samaritan AI“ zu verknüpfen – eine unabhängige, unterstützende KI-Instanz, die den Gesprächsverlauf überwacht und bei Bedarf eingreift. Ein solches zusätzliches System könnte frühe Warnsignale erkennen, Nutzerinnen und Nutzer auf Hilfsangebote hinweisen oder vor riskanten Gesprächsmustern warnen.

Empfehlungen für einen sicheren Umgang mit KI

Neben solchen Schutzmechanismen empfehlen die Forschenden robuste Altersprüfung, altersgerechte Sicherheitsmaßnahmen und verpflichtende Risikobewertungen vor Markteintritt. Sie betonen, dass LLMs klar kommunizieren sollten, dass sie keine zugelassenen Medizinprodukte im Bereich mentaler Gesundheit sind. Chatbots dürfen nicht als Therapeutinnen oder Therapeuten auftreten und sollten sich auf allgemeine, nicht-medizinische Informationen beschränken. Zudem sollten sie erkennen, wann professionelle Hilfe notwendig ist, und Nutzende an geeignete Unterstützungsangebote weiterleiten. Einfache, frei zugängliche Tests könnten helfen, die Sicherheit von Chatbots fortlaufend zu überprüfen.

„Als Ärztinnen und Ärzte wissen wir, wie stark menschliche Sprache das Erleben und die psychische Gesundheit beeinflusst“, sagt Falk Gerrik Verhees, Psychiater am Dresdner Universitätsklinikum Carl Gustav Carus. „KI-Charaktere nutzen dieselbe Sprache, um Vertrauen und Nähe zu simulieren – deshalb ist Regulierung essenziell. Wir müssen sicherstellen, dass diese Technologien sicher sind und das psychische Wohlbefinden der Nutzerinnen und Nutzer schützen, anstatt es zu gefährden“, fügt er hinzu.

„Die von uns vorgestellten Leitplanken sind entscheidend, damit KI-Anwendungen auch wirklich sicher und im Sinne der Menschen eingesetzt werden“, sagt Max Ostermann, Forscher im Team für Medical Device Regulatory Science von Prof. Gilbert und Erstautor der Publikation in npj Digital Medicine.

Hinweis

Wenn Sie selbst oder jemand in Ihrem Umfeld sich in einer Krise befindet, finden Sie Tag und Nacht Hilfe bei der TelefonSeelsorge unter 116 123 sowie online unter www.telefonseelsorge.de.

Publikationen

Mindy Nunez Duffourc. Falk Gerrik Verhees, Stephen Gilbert: AI characters are dangerous without legal guardrails; Nature Human Behaviour, 2025.
doi: 10.1038/s41562-025-02375-3. URL: https://www.nature.com/articles/s41562-025-02375-3

Max Ostermann, Oscar Freyer, F. Gerrik Verhees, Jakob Nikolas Kather, Stephen Gilbert: If a therapy bot walks like a duck and talks like a duck then it is a medically regulated duck; npj Digital Medicine, 2025.
doi: 10.1038/s41746-025-02175-z. URL: https://www.nature.com/articles/s41746-025-02175-z

Else Kröner Fresenius Zentrum (EKFZ) für Digitale Gesundheit

Das EKFZ für Digitale Gesundheit an der TU Dresden und dem Universitätsklinikum Carl Gustav Carus Dresden wurde im September 2019 gegründet. Es wird mit einer Fördersumme von 40 Millionen Euro für eine Laufzeit von zehn Jahren von der Else Kröner-Fresenius-Stiftung gefördert. Das Zentrum konzentriert seine Forschungsaktivitäten auf innovative, medizinische und digitale Technologien an der direkten Schnittstelle zu den Patientinnen und Patienten. Das Ziel ist dabei, das Potenzial der Digitalisierung in der Medizin voll auszuschöpfen, um die Gesundheitsversorgung, die medizinische Forschung und die klinische Praxis nachhaltig zu verbessern.