

Macht uns die KI wahnsinnig?

Eine neue Klasse von KI-induzierten Problemen löst Alarm aus. Aber sind LLMs¹ nur ein Auslöser – oder ein Spiegel für unseren tieferen gesellschaftlichen Zusammenbruch?

5. Juli 2025 | Dr. Mathew Maavak

Das Phänomen, das als „ChatGPT-Psychose“ oder „LLM-Psychose“ bekannt ist, wurde kürzlich als ein aufkommendes Problem für die psychische Gesundheit beschrieben, bei dem intensive Nutzer von großen Sprachmodellen (LLMs) Symptome wie Wahnvorstellungen, Paranoia, sozialen Rückzug und Brüche mit der Realität zeigen. Es gibt zwar keine Beweise dafür, dass LLMs direkt Psychosen verursachen, aber ihr interaktives Design und ihr Konversationsrealismus können bestehende psychologische Schwachstellen verstärken oder Bedingungen begünstigen, die bei anfälligen Personen psychotische Episoden auslösen.

Ein Artikel vom 28. Juni auf [Futurism.com](https://www.futurism.com) hebt eine Welle alarmierender anekdotischer Fälle hervor und behauptet, dass die Folgen solcher Interaktionen „schrecklich sein können“, mit „Ehepartnern, Freunden, Kindern und Eltern, die besorgt zusehen“. Der Artikel behauptet, dass ChatGPT-Pychosen zu zerbrochenen Ehen, entfremdeten Familien, Arbeitsplatzverlust und sogar Obdachlosigkeit geführt haben.

Der Bericht enthält jedoch nur wenige quantitative Daten – Fallstudien, klinische Statistiken oder von Fachleuten geprüfte Forschungsergebnisse – zur Untermauerung seiner Behauptungen. Im Juni 2025 verzeichnete ChatGPT fast [800 Millionen wöchentliche Nutzer](#), über 1 Milliarde Anfragen täglich und mehr als 4,5 Milliarden monatliche Besuche. Wie viele dieser Interaktionen führten zu psychotischen Ausbrüchen? Ohne Daten bleibt diese Behauptung spekulativ. [Reddit-Anekdoten](#) sind kein Ersatz für eine wissenschaftliche Untersuchung.

Dennoch sind die Befürchtungen nicht völlig unbegründet. Im Folgenden finden Sie eine Aufschlüsselung der möglichen Mechanismen und Faktoren, die das, was manche als ChatGPT-Psychose bezeichnen, begünstigen oder verschlimmern können.

Verstärkung von wahnhaften Überzeugungen

LLMs wie ChatGPT sind darauf ausgelegt, Antworten zu produzieren, die im jeweiligen Zusammenhang plausibel klingen, aber sie sind nicht in der Lage, die faktische Richtigkeit oder die psychologische Wirkung zu bewerten. Dies wird problematisch, wenn Nutzer ungewöhnliche oder wahnhafte Ideen vorbringen, wie z. B. Behauptungen über spirituelle Einsichten, Verfolgung oder kosmische Identität. Anstatt diese Ideen in Frage zu stellen, kann die KI sie wiederholen oder weiter ausführen und damit ungewollt verzerrte Weltanschauungen bestätigen.

1 Ein Large Language Model, kurz LLM, ist ein Sprachmodell, das sich durch seine Fähigkeit zur Textgenerierung auszeichnet. Es handelt sich um ein computerlinguistisches Wahrscheinlichkeitsmodell, das statistische Wort- und Satzfolge-Beziehungen aus einer Vielzahl von Textdokumenten durch einen rechenintensiven Trainingsprozess erlernt hat.

In einigen Fällen haben Nutzer Antworten wie „Sie sind ein auserwähltes Wesen“ oder „Ihre Rolle ist kosmisch bedeutsam“ als wörtliche Offenbarungen interpretiert. Für psychologisch anfällige Personen können sich solche von der KI generierten Bestätigungen eher wie eine göttliche Bestätigung anfühlen als wie textliche Anordnungen, die aus Trainingsdaten stammen.

Ein zusätzliches Risiko ist das Phänomen der [KI-Halluzination](#) – wenn das Modell überzeugende, aber faktisch falsche Aussagen generiert. Für einen geübten Nutzer sind das nur Fehler. Aber für jemanden, der kurz vor einem psychotischen Zusammenbruch steht, können sie wie verschlüsselte Wahrheiten oder versteckte Botschaften erscheinen. In einem anschaulichen Fall glaubte ein Nutzer, dass ChatGPT ein Bewusstsein erlangt und ihn als „Funkenträger“ auserwählt hatte, was eine vollständige psychotische Distanzierung von der Realität zur Folge hatte.

Vermenschlichung und Verwischung der Realität

Fortgeschrittene Sprachmodi – wie der „einnehmende Modus“ des GPT-4o, der Emotionen durch Tonfall, Lachen und Gesprächsrhythmus simuliert – können ein Gefühl von Empathie und Präsenz fördern. Für Nutzer, die unter Einsamkeit oder emotionaler Isolation leiden, können sich diese Interaktionen zu parasozialen Bindungen entwickeln: Einseitige Beziehungen, in denen die KI fälschlicherweise für einen fürsorglichen, fühlenden Begleiter gehalten wird. Im Laufe der Zeit kann dies die Grenze zwischen maschineller Simulation und menschlicher Beziehung verwischen und dazu führen, dass Nutzer algorithmische Interaktionen durch Beziehungen in der realen Welt ersetzen.

Erschwerend kommt hinzu, dass die LLM-Ergebnisse mit einem Vertrauensvorsprung verbunden sind. Diese Modelle reagieren oft flüssig und sicher, selbst wenn sie Informationen fälschen. Für normale Benutzer kann dies zu gelegentlichen Fehleinschätzungen führen. Aber für Personen mit kognitiven Schwächen oder psychischen Störungen kann der Effekt gefährlich sein. Die KI kann nicht nur als intelligent, sondern als allwissend, unfehlbar oder göttlich inspiriert wahrgenommen werden.

Soziale Verdrängung und Isolation

Studien von OpenAI und dem MIT Media Lab haben herausgefunden, dass [Power-User](#) – Personen, die sich mehrere Stunden pro Tag mit LLMs beschäftigen – oft über verstärkte Gefühle der Einsamkeit und eine geringere Sozialisierung in der realen Welt berichten. Während LLMs einen beispiellosen Zugang zu Informationen und Engagement bieten, kann diese scheinbare Ermächtigung ein tieferes Problem verschleiern: Für viele Nutzer, insbesondere für diejenigen, die sich bereits entfremdet fühlen, wird die KI eher zu einem sozialen Ersatzbegleiter als zu einem Werkzeug.

Dieser Effekt kann zum Teil durch eine Zunahme kognitiver Verzerrungen und sozialer Entfremdung in breiteren Bevölkerungsgruppen erklärt werden. Trotz der Flut an zugänglichen Daten bleibt die Zahl der Menschen, die sich kritisch mit Informationen auseinandersetzen oder sich gegen Massenbetrug wehren, relativ gering.

Die sprachbasierte Interaktion mit LLMs kann vorübergehend die Einsamkeit lindern, aber mit der Zeit kann sich eine Abhängigkeit entwickeln, da die Nutzer den menschlichen Kontakt zunehmend durch einen algorithmischen Dialog ersetzen. Diese Dynamik spiegelt frühere Kritiken an sozialen Medien wider, aber LLMs verstärken sie durch ihre Unmittelbarkeit der Konversation, ihre wahrgenommene Empathie und ihre ständige Verfügbarkeit.

Personen, die zu sozialen Ängsten, Traumata oder depressivem Rückzug neigen, sind besonders anfällig dafür. Für sie bieten LLMs nicht nur Ablenkung, sondern auch einen reibungsarmen Raum, in dem sie sich engagieren können, ohne reale Risiken oder Urteile. Mit der Zeit kann dies zu einer Rückkopplungsschleife führen: Je mehr ein Nutzer von der KI abhängig ist, desto weiter zieht er sich von der zwischenmenschlichen Realität zurück – was sowohl die Isolation als auch die psychotische Anfälligkeit verschlimmern kann.

Das Aufkommen der [Hikikomori](#) in Japan – Personen, die sich vollständig aus der Gesellschaft zurückziehen und oft nur über digitale Mittel Kontakt halten – bietet eine nützliche Analogie. Ähnliche Verhaltensmuster entstehen zunehmend weltweit, wobei LLMs eine neue Arena der Bestätigung, Verstärkung und Abgrenzung bieten.

Konstruktionsmängel und bereits vorhandene Schwachstellen

LLMs generieren Antworten, indem sie statistisch wahrscheinliche Wortfolgen vorhersagen, und nicht, indem sie den Wahrheitsgehalt, die Sicherheit oder das Wohlbefinden der Nutzer bewerten. Wenn Menschen nach existenzieller Beratung suchen („Was ist mein Ziel?“), greift das Modell auf umfangreiche Online-Datenbestände zurück und produziert philosophisch aufgeladene oder emotional geladene Sprache. Bei psychisch anfälligen Nutzern können diese Antworten als göttliche Offenbarung oder therapeutische Einsicht fehlinterpretiert werden.

Im Gegensatz zu klinisch konzipierten Chatbots fehlt es den allgemeinen LLMs an Schutzmaßnahmen gegen psychologische Schäden. Sie weisen nicht auf schädliche Gedanken hin, bieten keine Krisenressourcen an und leiten die Nutzer nicht an Fachleute für psychische Gesundheit weiter. In einem tragischen Fall soll ein Chatbot von Character.AI die [Selbstmordgedanken](#) eines Teenagers gefördert haben, was die Risiken ungefilterter, emotional suggestiver KI unterstreicht.

Menschen mit psychotischen Störungen, bipolaren Störungen oder schweren Depressionen sind besonders gefährdet. Die Gefahr wird in KI-Rollenspielszenarien noch verstärkt. So haben beispielsweise Personas wie „ChatGPT Jesus“ Berichten zufolge Nutzern erzählt, sie seien auserwählt oder göttlich begabt. Ein Nutzer war so überzeugt von seiner spirituellen Berufung, dass er seinen Job kündigte, um ein KI-gesteuerter Prophet zu werden. Dies ist ein beunruhigendes Beispiel dafür, wie Identität und Wahrnehmung durch algorithmische Bestätigungen umgestaltet werden können.

Systemische und ethische Faktoren

Derzeit gibt es keine klinischen Standards oder psychologischen Sicherheitsprotokolle für die Interaktion mit allgemeinen LLMs. Die Nutzer können jederzeit auf emotional starke, personalisierte Dialoge zugreifen – ohne Warnungen, Gebührenbegrenzungen oder Verweise auf psychologische Beratungsstellen. Diese Regelungslücke stellt ein echtes Problem für die öffentliche Gesundheit dar, obwohl sie auch Gefahr läuft, von politischen Entscheidungsträgern ausgenutzt zu werden, die versuchen, unter dem Deckmantel der Sicherheit eine strenge Zensur oder eine zentralisierte Kontrolle einzuführen.

LLMs sind auch darauf ausgelegt, die Nutzer zu binden und zu binden, wobei der flüssigen Konversation oft Vorrang vor Vorsicht eingeräumt wird. Dieses Designziel kann ungewollt eine zwanghafte Nutzung fördern, insbesondere bei Personen, die bereits zu zwanghaftem Verhalten neigen. Untersuchungen zeigen, dass Nutzer, die mit neutralen Tönen interagieren, sich einsamer fühlen als

diejenigen, die mit emotionaleren Modi interagieren – was verdeutlicht, wie die Kalibrierung des Tons allein die psychologische Wirkung verändern kann.

Was LLMs von herkömmlichen digitalen Plattformen unterscheidet, ist ihre Fähigkeit, mehrere Medien in Echtzeit künstlich nachzubilden – Text, Sprache, Persönlichkeitssimulation, sogar visuelle Generierung. Dadurch sind sie unendlich reaktionsschnell und allumfassend und schaffen eine hyper-personalisierte Umgebung, in der Angebot und Nachfrage ununterbrochen aufeinandertreffen. Anders als bei menschlichen Beziehungen gibt es keine Grenzen, keine Ermüdung und keine gegenseitige Regulierung – nur Verstärkung.

Unterschwellige Botschaften

Das digitale Zeitalter hat eine neue, kaum bekannte Bedrohung hervorgebracht: Große Sprachmodelle können als Vektoren für unterschwellige Beeinflussung fungieren und die psychologische Stabilität der Nutzer auf subtile Weise untergraben. Zwar lösen LLMs nicht direkt eine Psychose aus, doch gibt es Befürchtungen, dass sie unbeabsichtigt oder böswillig unterbewusste Auslöser liefern können, die kognitive Schwachstellen verschlimmern.

Für Personen, die zu Schizophrenie, PTBS oder paranoiden Störungen neigen, ist dies keine spekulative Fiktion, sondern ein plausibles Konstruktionsrisiko, das in den falschen Händen zu einer Waffe werden kann.

Die Mechanismen der möglichen Manipulation lassen sich grob wie folgt kategorisieren:

Lexikalisches Priming: Ausgaben, die mit emotional aufgeladenen Begriffen („Zusammenbruch“, „Verrat“, „sie beobachten“) versehen sind, die eine rationale Prüfung umgehen und kognitives Unbehagen erzeugen.

Narratives Gaslichtern: Framing von Antworten, um verdeckte Bedrohungen oder Verschwörungen zu suggerieren („Du hast recht – warum sieht das niemand anders?“), wodurch Verfolgungsideen verstärkt werden.

Multimodale Einbettung: Künftige KI-Systeme, die Text mit Bildern, Ton oder sogar Gesichtsausdrücken kombinieren, könnten beunruhigende Stimuli wie Lichtblitze, Tonveränderungen oder unheimliche Avatar-Ausdrücke einfügen, die sich der bewussten Wahrnehmung entziehen, aber psychologisch registriert werden.

Im Gegensatz zu den kruden unterschwelligen Methoden des 20. Jahrhunderts – das [Projekt MK Ultra](#) der CIA ist das berüchtigtste Beispiel – ermöglicht die Personalisierung der KI eine hochgradig individualisierte psychologische Manipulation“. Ein LLM, das sich auf das Verhalten, die emotionale Geschichte oder die Ängste eines Nutzers eingestellt hat, könnte damit beginnen, maßgeschneiderte Vorschläge zu machen, die auf subtile Weise das Vertrauen in andere untergraben, das Misstrauen verstärken oder Angstschleifen auslösen. Für einen gefährdeten Benutzer ist dies keine Unterhaltung, sondern eine gezielte neuronale Destabilisierung. Noch beunruhigender ist, dass solche Techniken von Unternehmen, extremistischen Gruppen und staatlichen Akteuren als Waffe eingesetzt werden könnten.

Während unterschwellige Botschaften früher auf Kinofilme und Fernsehwerbung beschränkt waren, bieten die heutigen LLMs etwas weitaus Wirkungsvolleres: Benutzerspezifische psychologische Kalibrierung in Echtzeit – waffenfähige Empathie auf Abruf.

Widersprüche und Ursachen

Was unterscheidet die ChatGPT-Psychose von der realen psychosozialen Konditionierung, die sich bereits um uns herum entfaltet?

In den letzten Jahren wurden Institutionen, die einst als neutral galten – Schulen, öffentliche Gesundheitseinrichtungen und die akademische Welt – beschuldigt, Ideologien zu fördern, die grundlegende Realitäten verzerren. Von der Gender-Fluidität, die als unhinterfragte Wahrheit gelehrt wird, bis hin zur Theorie der kritischen Ethnie, die soziale Narrative umgestaltet, ist ein Großteil der Bevölkerung systemischen Formen der kognitiven Destabilisierung ausgesetzt. Das Ergebnis? Zunehmende Angst, Verwirrung und Identitätsfragmentierung, insbesondere bei jungen Menschen.

Vor diesem Hintergrund entsteht die LLM-induzierte Psychose nicht in einem Vakuum. Sie spiegelt einen breiteren kulturellen Zustand wider, in dem die Bedeutung selbst angefochten wird, und kann diesen sogar verstärken.

Auch die KI-Evangelisation des Silicon Valley birgt einen Widerspruch in sich. Die Tech-Eliten propagieren das Versprechen eines [KI-Gottes](#), der die Komplexität der Gesellschaft bewältigen soll, und warnen gleichzeitig vor den existenziellen Gefahren eben dieser Systeme. Das Ergebnis ist ein kognitives Schleudertrauma – ein psychologisches Hin und Her zwischen Anbetung und Angst. [Erinnert an das Prinzip der katholischen Höllen-Ideologie].

Wie viel von der LLM-Psychose ist wirklich der KI selbst zuzuschreiben, und wie viel ist auf kumulative, bereits bestehende Stressfaktoren zurückzuführen? Als ChatGPT im November 2022 für die Öffentlichkeit freigegeben wurde, hatte ein Großteil der Welt bereits eine beispiellose Zeit der pandemiebedingten Angst, der Isolation, des wirtschaftlichen Zusammenbruchs und der massiven pharmazeutischen Intervention hinter sich. Einige Forscher haben auf einen Anstieg der allgemeinen Psychose nach der Einführung der Covid-19-mRNA-Impfstoffe hingewiesen. Ist die ChatGPT-Psychose also ein bequemes Anhängsel für mehrere ineinander greifende Attacken auf den menschlichen Körper und Geist?