dgppn

Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde e.V.

Positionspapier
03.11.2026

Einsatz von KI in Psychiatrie und Psychotherapie

Künstliche Intelligenz (KI) bietet vielfältige Potenziale, die psychiatrische und psychotherapeutische Versorgung in den Bereichen Diagnostik, Therapie und Administration sinnvoll zu ergänzen und zu unterstützen. Das vorliegende Positionspapier gibt einen Überblick über Einsatzmöglichkeiten, ordnet den gegenwärtigen Stand der wissenschaftlichen Evidenz ein und formuliert zentrale Anforderungen an eine verantwortungsvolle Nutzung.

Hintergrund

Vor dem Hintergrund des zunehmenden Bedarfs an Unterstützungsangeboten im Bereich psychischer Gesundheit sowie gleichzeitig begrenzter personeller Ressourcen im Versorgungssystem gewinnt die Diskussion über die potenzielle Rolle digitaler Technologien und KI auch in der psychiatrischen und psychotherapeutischen Versorgung in wachsendem Maße an Bedeutung.

KI-basierte Anwendungen werden zunehmend sichtbar und finden Anwendung in unterschiedlichen Kontexten, z. B. nutzen mittlerweile Menschen bei psychischen Belastungen digitale Hilfsangebote wie *Large Language Models* (LLM)-basierte *Chatbots* oder allgemeine digitale Gesundheitsinterventionen. Diese Entwicklung zeigt somit den Bedarf an niedrigschwellig erreichbaren Unterstützungsangeboten, macht gleichzeitig aber deutlich, dass der Einsatz dieser Technologien im Bereich psychischer Gesundheit besondere Sorgfalt und verantwortungsvolle Gestaltung erfordert.

Gleichermaßen werden KI-basierte Anwendungen auch zunehmend von Patientinnen und Patienten mit manifesten psychischen Erkrankungen in Anspruch genommen. Die Nutzung erfolgt hier vor allem zur Informationssuche, zur emotionalen Unterstützung und Stabilisierung, aber auch als leicht zugängliche Alternative oder Ergänzung zu etablierten

Behandlungsansätzen [1]. Zum Einsatz kommen dabei jedoch vielfach keine dezidiert für die therapeutische Anwendung entwickelten Angebote, sondern allgemein verfügbare KI-Systeme, wie z. B. *ChatGPT* und andere LLM, deren Entwicklung primär nicht auf klinische Kontexte ausgerichtet ist.

Zudem wachsen das öffentliche Interesse an möglichen Einsatzfeldern sowie die Aufmerksamkeit bezüglich möglicher Chancen und Risiken dieser Technologien im Bereich der psychischen Gesundheit [2].

Wenngleich aktuell bereits – begrenzte – Evidenz für einen gewissen Nutzen KI-basierter Anwendungen in Diagnostik und Therapie psychischer Erkrankungen vorliegt [3, 4], bestehen substanzielle Evidenzlücken hinsichtlich Wirksamkeit, Sicherheit, Datenschutz, ethischer Verantwortung, möglicher Fehlinformationen sowie der Gefahr einer unzureichenden oder inadäquaten Versorgung vulnerabler Patientengruppen [5].

Potenzial von KI für die psychiatrische und psychotherapeutische Versorgung

Die Einsatzmöglichkeiten von KI in Psychiatrie und Psychotherapie sind vielfältig und umfassen neben Diagnostik, Monitoring und Therapie auch administrative Aspekte [6].

Diagnostik: Im Bereich Diagnostik ermöglicht KI die Analyse großer, heterogener und multimodaler Datensätze, wodurch komplexe Zusammenhänge zwischen biologischen, psychologischen und verhaltensbezogenen Faktoren ermittelt werden könnten [7, 8, 9]. Dies eröffnet neue Möglichkeiten der Identifikation bislang schwer erkennbarer Krankheitsmuster über die bestehenden Diagnosekriterien hinaus [7, 10, 11]. KI könnte somit helfen, die bekannten Unschärfen und Grenzen etablierter Diagnosesysteme (*Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* [DSM-5], *International Statistical Classification of Diseases and Related Health Problems, 11th Revision* [ICD-11]) zu überwinden. In der individuellen Diagnostik scheint der Einsatz derzeit noch nicht sinnvoll.

Monitoring: Parallel erweisen sich digitale Technologien wie Smartphones oder *Wearables* als wertvolle Datenquellen für Diagnostik, Therapie- und Verlaufsmonitoring. Diese ermöglichen nicht nur die alltagsnahe Symptomerfassung, sondern auch eine frühzeitige Rückfallprävention durch unmittelbar (in Echtzeit stattfindende) erfolgende psychotherapeutische Interventionen (*Just-in-Time Adaptive Interventions,* JITAI). Allerdings sind Nutzen und Validität passiver Datenerfassung mittels Smartphones und Wearables zur Sammlung und Analyse relevanter Verhaltensindikatoren für die routinemäßige Versorgung bislang noch nicht ausreichend belegt [12].

Therapie: KI kann als ergänzende Maßnahme sowohl in der Face-to-Face-Interaktion als auch im Hybridkontakt eingesetzt werden [13]. Dabei könnten KI-basierte *Mental-Health-Chatbots* oder KI-basierte Digitale Gesundheitsanwendungen (DiGA) neue Perspektiven in der Behandlung eröffnen, die Prädiktion von Krankheitsverläufen und Therapieresponse verbessern und damit eine individualisiertere Behandlungsplanung ermöglichen. Vorstellbar ist auch die Einbeziehung biologischer und klinischer Variablen, um *Machine-Learning*-basiert Therapiewirksamkeit und Nebenwirkungen vorherzusagen oder Auswahl und Dosierung therapeutischer Interventionen zu optimieren.

Administration: Schließlich kann KI Fachkräfte entlasten, indem sie administrative Aufgaben automatisiert, und so zeitliche Ressourcen für die eigentliche Behandlung schafft [6]. Potenzielle Anwendungsbereiche umfassen dabei u. a. die automatisierte Transkription von Dokumentationen, effizienteres Wartezeitmanagement, automatisierte Triagefunktionen oder digitale Lotsensysteme.

Evidenz zum Einsatz von KI in der psychotherapeutischen Behandlung

Angesichts der strukturellen Ressourcenknappheit in der psychotherapeutischen Versorgung steht insbesondere hier der Einsatz KI-basierter Anwendungen im Fokus des Interesses. So bieten vor allem KI-basierte Chatbots aus Sicht der Nutzenden Vorteile, die in den aktuellen Versorgungsstrukturen nicht oder zumindest nicht ausreichend abgebildet werden. Hierbei spielen Aspekte wie die Ad-hoc-Verfügbarkeit sowie die Möglichkeit einer anonymen und

wertungsfreien Kommunikation eine wichtige Rolle. Erste Studien zeigen, dass solche Systeme insbesondere bei leichten bis moderaten psychischen Belastungen zur Förderung des Selbstmanagements beitragen und den Zugang zu weiterführenden Versorgungsangeboten erleichtern können [3, 4]. Perspektivisch könnten KI-basierte Chatbots als ergänzende Instrumente die bestehenden psychotherapeutischen und psychiatrischen Versorgungsstrukturen unterstützen und deren Reichweite erhöhen.

Im Hinblick auf die Wirksamkeit dieser Anwendungen liegen inzwischen einige wissenschaftliche Studien vor. So zeigt eine umfassende Meta-Analyse von Lau et al. [3] positive Effekte KI-basierter psychologischer Interventionen. Insgesamt wurden 30 randomisierte kontrollierte Studien mit 6.100 Teilnehmenden aus neun Ländern ausgewertet. Die untersuchten Interventionen basierten überwiegend auf KI-basierten-Chatbot-Oberflächen, die Ansätze aus der kognitiven Verhaltenstherapie nutzten. Die Ergebnisse zeigen, dass KI-gestützte Interventionen depressive Symptome signifikant reduzieren konnten. Moderate Effekte waren auch in Follow-Up-Untersuchungen nach sechs bis zwölf Monaten nachweisbar. Das Autorenteam betont, dass solche Systeme insbesondere angesichts des weltweiten Mangels an psychotherapeutischen Angeboten eine sinnvolle Ergänzung zur regulären Versorgung darstellen könnten. Mögliche Vorteile einer KI-basierten Therapie lägen u. a. in der besseren Zugänglichkeit, geringeren Kosten sowie der Reduktion von Stigmatisierung psychischer Erkrankungen. Gleichzeitig wird jedoch auf methodische Einschränkungen hingewiesen, darunter hohe Heterogenität der Studien sowie teilweise geringe Evidenzqualität. Zudem fehlten ausreichende Langzeitdaten und standardisierte Interventionskonzepte [3].

In einer weiteren systematischen Übersichtsarbeit und Metaanalyse von Li et al. [4] wurde die Wirksamkeit KI-basierter Gesprächsagenten zur Förderung der psychischen Gesundheit untersucht. Hier zeigt sich, dass KI-gestützte Chatbots psychische Belastungen, insbesondere depressive Symptome und Stress, signifikant reduzieren können. Dagegen konnten keine belastbaren Nachweise für eine nachhaltige Verbesserung des allgemeinen psychischen Wohlbefindens erbracht werden. Einschränkend wird auf Herausforderungen hinsichtlich Kommunikationsfehlern, Personalisierung, Sicherheit und langfristiger Wirksamkeit hingewiesen [4].

Einer Übersichtsarbeit von Paulus [14] zufolge können strukturierte Chatbots kleine, aber signifikante Verbesserungen bei depressiver Symptomatik und Angstsymptomatik erzielen, wobei der Nutzen vor allem in der Förderung von Therapieadhärenz und psychoedukativer Unterstützung liegt [14]. Gleichzeitig bestehen erhebliche Sicherheitsbedenken, insbesondere hinsichtlich fehlerhafter Kriseninterventionen, unzureichender Risikoeinschätzungen bei Jugendlichen sowie der Verstärkung von Zwangs- und Angststörungen durch Rückversicherungsverhalten [14].

Bei der Interpretation der Studienlage zu LLM-basierten Chatbots ist allerdings zu berücksichtigen, dass sich die genannten systematischen Übersichtsarbeiten überwiegend auf KI-basierte Chatbots aus der Zeit vor der Einführung von ChatGPT beziehen. Angesichts des mittlerweile erheblichen technologischen Fortschritts und der deutlich gestiegenen Leistungsfähigkeit heutiger Systeme ist davon auszugehen, dass die berichteten Effekte das Potenzial derzeitiger LLM-basierter Anwendungen nur unzureichend abbilden. Perspektivisch sind somit weitere Studien notwendig, um valide Aussagen zur langfristigen Wirksamkeit im klinischen Alltag mit jeweils aktuell zur Verfügung stehenden Anwendungen gewinnen zu können.

Übergreifend betont die Literatur zudem die Notwendigkeit eines *Human-in-the-loop*-Prinzips, bei dem KI ausschließlich unterstützend eingesetzt und durch klare Eskalations- und Kontrollmechanismen begrenzt wird [13, 14].

Therapeutische Beziehung

Für den Einsatz von KI in der Psychotherapie kommt dem Aspekt der therapeutischen Beziehung eine besondere Bedeutung zu. So gelten beispielsweise Vertrauen und Empathie als zentrale Wirkfaktoren der Psychotherapie. Diese sind nach bisheriger Auffassung bislang nicht vollständig automatisierbar, da sie auf wechselseitiger zwischenmenschlicher Interaktion beruhen und sich nicht allein durch algorithmische Prozesse abbilden lassen. Allerdings gibt es erste Hinweise, dass Patientinnen und Patienten auch gegenüber digitalen Interventionen eine gewisse Form der therapeutischen Beziehung bzw. Allianz entwickeln können [15, 16, 17, 18]. Der Beziehungsaspekt spricht somit nicht grundsätzlich gegen den Einsatz von KI,

erfordert jedoch eine differenzierte Betrachtung der Möglichkeiten und Grenzen der durch KI-basierte Interventionen ergänzten Psychotherapie.

Risiken, Grenzen und ethische Verantwortung

Der Einsatz von KI im Bereich psychischer Gesundheit ist auch mit Risiken, Grenzen und besonderen ethischen Herausforderungen verbunden [14, 19, 20].

Grenzen allgemeiner KI-Systeme (z. B. ChatGPT)

Viele der allgemeinen KI-basierten-Chatbot-Systeme wie z. B. ChatGPT verfügen nicht über spezifisch entwickelte klinische, psychiatrische oder psychotherapeutische Funktionen oder über klar definierte Algorithmen für den Umgang mit psychischen Krisensituationen [21]. Daraus können Risiken entstehen, insbesondere wenn Menschen mit schweren psychischen Erkrankungen, beispielsweise psychotischen Erkrankungen [22], oder in akuten Krisen digitale Systeme die nicht eigens für diesen Zweck entwickelt wurden, als primäre Unterstützungsquelle nutzen. Allerdings könnten, im Gegensatz zu den allgemeinen KI-basierten-Chatbot-Systemen, speziell für die therapeutische Anwendung entwickelte, leitlinienorientierte KI-basierte DiGA deutlich geringere Risiken bergen. So werden DiGA entsprechend als Medizinprodukt regulatorisch klassifiziert und gekennzeichnet, um verbindliche Anforderungen insbesondere an Sicherheit, Qualität und Verantwortlichkeit zu gewährleisten.

Risiken in der Interaktion

Kritisch diskutiert wird mittlerweile auch die Gefahr einer „Pseudo-Allianz“, die letztlich auch durch die intendierte Eigenschaft von KI-basierten Chatbots zustande kommt, Antworten an die Erwartungen der Nutzenden anzupassen und auf diese Weise die Interaktion möglichst lange aufrechtzuerhalten (*Sycophancy*).

Aus solchen Interaktionen können dysfunktionale emotionale Abhängigkeiten mit daraus resultierenden Gefahren entstehen, für die beispielsweise Patientinnen und Patienten mit Persönlichkeitsstörungen und hoher interpersoneller Dysfunktionalität besonders anfällig sind.

Auch besteht die Gefahr inadäquater Reaktionen auf Krisensituationen, wenn beispielsweise Suizidalität nicht erkannt oder paranoides Erleben verstärkt wird [23].

Mangelnder Datenschutz und Transparenz

Der Einsatz von KI im Gesundheitsbereich wirft auch grundlegende Fragen zu Datenschutz, Transparenz und Verantwortung auf [13]. (Psychische) Gesundheitsdaten gehören zu den besonders sensiblen personenbezogenen Informationen und erfordern daher besondere Sorgfalt im Umgang. Digitale Anwendungen müssen höchsten Standards der Datensicherheit entsprechen. Transparenz ist erforderlich hinsichtlich Datenerhebung, Speicherung und Verwendung. In Ländern außerhalb der EU betriebene, frei zugängliche LLM erfüllen diese Anforderungen nicht und sollen entsprechend auch nicht zur Nutzung empfohlen werden.

Ein weiterer zentraler Aspekt betrifft die Transparenz gegenüber Nutzenden. Patientinnen und Patienten müssen jederzeit erkennen können, dass sie mit einem KI-System interagieren. Ebenso sind Möglichkeiten und Grenzen der Systeme für Nutzende nicht immer ersichtlich, weshalb eine entsprechende Aufklärung unbedingt erforderlich ist.

Obwohl der *EU-AI-Act* erste Schutzmaßnahmen vorsieht, z. B. gegen Manipulation, erscheinen die Regelungen insgesamt noch unzureichend. Erforderlich ist daher ein erweiterter ethischer und regulatorischer Rahmen, der mögliche Risiken des KI-Einsatzes bei psychischen Erkrankungen systematisch berücksichtigt und so einen verantwortungsvolleren Einsatz von KI in der Gesundheitsversorgung ermöglicht [24]. Gleichzeitig bedarf es neben der Regulierung zum Schutz der Privatsphäre auf europäischer Ebene einer gezielten Innovationsförderung und offenheit. Die intensive Nutzung frei verfügbarer LLM zeigt, dass Menschen unter Leidensdruck niedrigschwellige Hilfe häufig höher gewichten als Datenschutz. Fehlen ausreichend zugängliche Alternativen, werden Patientinnen und Patienten weiterhin auf unregulierte, aber verfügbare und als hilfreich wahrgenommene Angebote zurückgreifen.

Wirksamkeitsnachweis

Hinsichtlich der Beurteilung der Wirksamkeit ist die bislang begrenzte wissenschaftliche Evidenz zu berücksichtigen. Der therapeutische Einsatz KI-basierter Anwendungen setzt einen unabhängigen wissenschaftlichen Nachweis von Wirksamkeit und Sicherheit voraus. Dabei müssen dieselben medizinischen und wissenschaftlichen Standards wie für etablierte Behandlungsverfahren gelten; eine entsprechende Konformitätsbewertung und Kennzeichnung als Medizinprodukt sind daher zu empfehlen.

Fazit und Anforderungen

Die DGPPN sieht klare Chancen, bestehende Versorgungsstrukturen durch den Einsatz von KI sinnvoll zu ergänzen. Zugleich müssen diese Entwicklungen im Einklang mit den zentralen Werten und Prinzipien der klinischen Praxis stehen: wissenschaftliche Evidenz, ethische Verantwortung, Datenschutz, die Sicherheit von Patientinnen und Patienten sowie die sorgfältige Bewertung möglicher Risiken und Nebenwirkungen.

Vor dem Hintergrund bestehender Ressourcenknappheit im Bereich der Psychotherapie können verantwortungsvoll entwickelte KI-basierte Anwendungen perspektivisch dazu beitragen, den Zugang zu Unterstützungsangeboten zu erleichtern und Versorgungslücken zu verringern [25]. Insbesondere bei leichten bis moderaten psychischen Erkrankungen könnten sie Fachpersonal entlasten und dadurch Versorgungskapazitäten für Patientinnen und Patienten mit schweren und komplexen Erkrankungen schaffen.

Schließlich könnten KI-basierte Anwendungen aufgrund ihrer diskreten und niedrigschwelligen Zugänglichkeit für manche Patientinnen und Patienten eine bevorzugte Form der psychischen Unterstützung darstellen.

Aus den bislang vorliegenden Erkenntnissen lassen sich aus Sicht der DGPPN somit folgende Voraussetzungen für den verantwortungsvollen Einsatz von KI in der Psychiatrie und Psychotherapie formulieren:

1. **Wissenschaftliche Evidenz als Grundvoraussetzung**

   KI-Lösungen sollen nur dann eingesetzt werden, wenn Wirksamkeit, Sicherheit und Nutzen nach wissenschaftlichen Standards nachgewiesen wurden.

2. **Schutz vulnerabler Patientengruppen**

   Der Einsatz von KI-basierten Anwendungen bei schweren psychischen Erkrankungen ist kritisch zu prüfen. Insbesondere der Einsatz bei Patientinnen und Patienten mit schweren depressiven Störungen, Suizidalität, psychotischen Symptomen, Erkrankungen aus dem schizophrenen Formenkreis, bipolaren Störungen und Persönlichkeitsstörungen könnte derzeit Risiken bergen. Kinder, Jugendliche sowie ältere Patientinnen und Patienten sind zudem Gruppen mit besonderen Schutzbedürfnissen.

3. **KI-Nutzung als Unterstützung und nicht als Ersatz menschlicher Expertise**

   KI-Systeme können Fachkräfte bei Diagnostik, Therapie und Administration unterstützen, sollten jedoch ohne entsprechenden Evidenznachweis zur Wirksamkeit und Nebenwirkungsarmut die persönliche therapeutische Beziehung und klinische Verantwortung nicht vollständig ersetzen. Die Einbindung ärztlicher oder psychologischer Therapeutinnen und Therapeuten sollte gemäß dem *Human-in-the-Loop*-Prinzip Grundvoraussetzung bleiben.

4. **Datenschutz und Datensicherheit als unbedingte Voraussetzungen**

   Der Schutz sensibler Gesundheitsdaten stellt eine zentrale Forderung an KI-Produkte dar. Nutzende müssen darauf vertrauen können, dass ihre Daten sicher verarbeitet und gespeichert werden.

5. **Transparenz über Möglichkeiten und Grenzen von KI**

   Patientinnen und Patienten müssen jederzeit erkennen können, dass und wann sie mit KI interagieren, und über deren Fähigkeiten sowie Grenzen aufgeklärt werden.

6. **Gezielter KI-Einsatz zur Verbesserung von Zugang und Versorgung**

   Insbesondere empfohlen wird der Einsatz zur integrierten oder abgestimmten Nutzung zur Überbrückung von Wartezeiten, zur Psychoedukation, zum Monitoring von

Symptomen und zur Entlastung administrativer Prozesse, z. B. Terminvereinbarungen, im Sinne von *Stepped-Care*-Modellen.

7. **Schaffung verbindlicher regulatorischer Rahmenbedingungen**

   Für den verantwortungsvollen Einsatz von KI sind klare Qualitätsstandards und Verantwortlichkeiten, Anforderungen an den Wirksamkeits- und Sicherheitsnachweis und kontinuierliche Kontrolle von Bedeutung. Demzufolge erscheint die Entwicklung solcher Standards für die Anwendung von KI im Kontext psychiatrisch-psychotherapeutischer Behandlung unter Berücksichtigung des EU-AI-Acts bzw. in Anlehnung an etablierte Verfahren und Prinzipien der Medizinprodukteregulierung als nächster erforderlicher Schritt.

AG KI in Psychiatrie und Psychotherapie

Autorinnen und Autoren:

Peter Zwanzger

Philipp Klein

Andreas Reif

Harald Baumeister

Steffen Moritz

Sabine Herpertz

Glossar

Künstliche Intelligenz (KI) bzw. *Artificial Intelligence (AI)* bezeichnet allgemein computergestützte Systeme, die Aufgaben ausführen können, für die üblicherweise menschliche kognitive Fähigkeiten erforderlich sind. Dazu zählt beispielsweise das Erkennen von Sprache und Bildern, das „Verstehen" und Generieren von Texten, Entscheidungsunterstützung oder das Erkennen von Mustern in großen Datenmengen. KI umfasst dabei sehr unterschiedliche Technologien und Verfahren.

Maschinelles Lernen bzw. *Machine Learning* ist ein zentrales Element von KI. Dabei werden Computersysteme nicht ausschließlich durch fest programmierte Regeln gesteuert, sondern lernen statistische Zusammenhänge aus großen Datenmengen. Ziel ist es, Vorhersagen oder Entscheidungen auf Basis zuvor erkannter Muster zu treffen. Anwendungen des maschinellen Lernens finden sich beispielsweise in der medizinischen Bildanalyse oder in Vorhersagemodellen für Krankheitsverläufe.

Deep Learning ist eine Unterkategorie des maschinellen Lernens, bei dem künstliche neuronale Netzwerke mit vielen Verarbeitungsebenen eingesetzt werden. Diese Verfahren ermöglichen besonders leistungsfähige Anwendungen, etwa in der Sprach- und Bilderkennung. Viele moderne KI-Systeme beruhen auf *Deep-Learning*-Methoden.

Generative KI bezeichnet KI-Systeme, die auf Grundlage erlernter Muster neue Inhalte wie Texte, Bilder, Audio- oder Videoinhalte erzeugen können. *Large Language Models (LLM,* siehe unten) stellen eine spezifische Form generativer KI dar, die auf die Verarbeitung und Generierung von Sprache ausgerichtet ist. Die Fähigkeit zur Generierung neuer Inhalte eröffnet vielfältige Anwendungsmöglichkeiten, birgt jedoch auch Risiken, beispielsweise durch die absichtliche (*deep fake*) oder unabsichtliche (Halluzination, siehe unten) Erzeugung fehlerhafter oder irreführender Informationen.

Large Language Models (LLM) sind eine Unterkategorie des maschinellen Lernens. Dabei handelt es sich um große Sprachmodelle, die auf sehr umfangreiche Textmengen trainiert wurden. Sie können Sprache analysieren und natürlich klingende Texte generieren. Bekannte Beispiele sind Systeme wie *Generative Pre-trained Transformer* (GPT)-Modelle. LLM können Fragen beantworten, Texte zusammenfassen, Gespräche simulieren oder Vorschläge formulieren. Sie „verstehen" Inhalte jedoch nicht im menschlichen Sinne, sondern berechnen statistisch sprachlich passende Antworten auf Grundlage ihrer Trainingsdaten.

Chatbots sind konkrete Anwendungen, die eine dialogartige Kommunikation mit Nutzenden ermöglichen. Chatbots können regel- oder datenbasiert sein, wobei LLM zur zweiten Gruppe gehören. Während Chatbots früherer Generationen nur vorgegebene Antworten lieferten, ermöglichen neuere KI-basierte Chatbots dynamische, personalisierbare und komplexe Gesprächsverläufe; am bekanntesten dürften die US-Systeme wie *ChatGPT*, *Gemini* und *Claude* sowie chinesische Modelle wie *DeepSeek* sein, aber auch das EU-basierte LLM von *Mistral* sei an dieser Stelle genannt. In der psychischen Gesundheitsversorgung finden solche generischen LLM-Systeme, sofern frei zugänglich, zunehmend Anwendung zur Psychoedukation, emotionalen Unterstützung oder zur niedrigschwelligen Begleitung als unregulierte Selbsthilfe.

Halluzinationen, manchmal auch als *Konfabulationen* bezeichnet, beschreiben im Zusammenhang mit generativer KI die von Nutzenden nicht intendierte Erzeugung inhaltlich falscher, frei erfundener oder irreführender Aussagen durch ein Sprachmodell. Obwohl solche Antworten sprachlich überzeugend wirken können, sind sie sachlich nicht richtig. Dies kann u. a. im psychiatrisch-psychotherapeutischen Bereich erhebliche Risiken verursachen.

Sycophancy ist eine Tendenz von KI-Modellen, Aussagen, Erwartungen oder Überzeugungen der Nutzenden übermäßig zu bestätigen, auch wenn dies zulasten der sachlichen Richtigkeit oder ethischer Grundsätze geht. Dies kann sich etwa darin zeigen, dass falsche Annahmen nicht korrigiert, erwartungskonforme Fehlinformationen ausgegeben oder problematische Ratschläge erteilt werden.

Digitale Gesundheitsanwendungen (DiGA) sind digitale medizinische Anwendungen, die diagnostische oder therapeutische Zwecke unterstützen und in Deutschland unter bestimmten Voraussetzungen erstattungsfähig sein können. Aktuell basieren die verfügbaren DiGA noch eher selten auf KI, sondern sind regelbasiert und daher auch nicht Gegenstand dieses Positionspapiers.

Literatur

1. Diaz L, Huguet A (2026) Spontaneous use of ChatGPT for mental health support: an exploratory study. J Public Health. https://doi.org/10.1007/s10389-025-02668-x
2. Powell A (2026) Is a chatbot therapist better than nothing? https://news.harvard.edu/gazette/story/2026/01/is-a-chatbot-therapist-better-than-nothing/ (Zugegriffen: 17.08.2026)
3. Lau Y, Ang WHD, Ang WW et al. (2025) Artificial intelligence-based psychotherapeutic intervention on psychological outcomes: a meta-analysis and meta-regression. Depress Anxiety 2025:8930012. https://doi.org/10.1155/da/8930012
4. Li H, Zhang R, Lee YC et al. (2023) Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. NPJ Digit Med 6:236. https://doi.org/10.1038/s41746-023-00979-5
5. Frances A (2026) Warning: AI chatbots will soon dominate psychotherapy. Br J Psychiatry 228:474–478. https://doi.org/10.1192/bjp.2025.10380
6. Sun J, Lu T, Shao X et al. (2025) Practical AI application in psychiatry: historical review and future directions. Mol Psychiatry 30:4399-4408. https://doi.org/10.1038/s41380-025-03072-3
7. Sharma SK, Alutaibi AI, Khan AR et al. (2025) Early detection of mental health disorders using machine learning models using behavioral and voice data analysis. Sci Rep 15:16518. https://doi.org/10.1038/s41598-025-00386-8
8. Terhorst Y, Sander LB, Ebert DD et al. (2023) Optimizing the predictive power of depression screenings using machine learning. Digit Health 9. https://doi.org/10.1177/20552076231194939
9. Lenz E, Naamanka J, Trabert W et al. (2026) Benchmarking large language models against practicing clinicians on psychopathological assessment. NPJ Digit Med 9:518. https://doi.org/10.1038/s41746-026-02852-7
10. Moshe I, Terhorst Y, Opoku Asare K et al. (2021) Predicting Symptoms of Depression and Anxiety Using Smartphone and Wearable Data. Front Psychiatry 12:625247. https://doi.org/10.3389/fpsyt.2021.625247
11. Terhorst Y, Knauer J, Baumeister H (2023) Smart Sensing Enhanced Diagnostic Expert Systems. In: Montag C, Baumeister, H. (Hrsg) Digital Phenotyping and Mobile Sensing. New Developments in Psychoinformatics. Springer
12. Simon L, Steubl LS, Terhorst Y et al. (2026) Digital Phenotyping – Lessons Learned, Challenges, and Future Perspectives for Psychodiagnostics. Diagnostica 72:29–40. https://doi.org/10.1026/0012-1924/a000363
13. Schwartz B, Alpers GW, Baumeister H et al. (2026) Potenziale und Grenzen Künstlicher Intelligenz in der klinisch-psychologischen Forschung, Lehre und Praxis. Psychologische Rundschau 77:136–142. https://doi.org/10.1026/0033-3042/a000764
14. Paulus MP (2026) Current themes of AI in mental health: Actionable evidence and guardrails for mood and anxiety care. J Mood Anxiety Disord 14:100181. https://doi.org/10.1016/j.xjmad.2026.100181
15. Malouin-Lachance A, Capolupo J, Laplante C et al. (2025) Does the Digital Therapeutic Alliance Exist? Integrative Review. JMIR Ment Health 12:e69294. https://doi.org/10.2196/69294
16. Shoshani A, Gurfinkel B, Kor A et al. (2026) Efficacy of a Conversational AI Agent for Psychiatric Symptoms and Digital Therapeutic Alliance. JAMA Netw Open 9:e266713. https://doi.org/10.1001/jamanetworkopen.2026.6713
17. Steinhoff M, Steubl L, Baumeister H (2020) Social Distancing oder Emotional Closeness? Familiendynamik 45:318–327. https://doi.org/10.21706/fd-45-4-318
18. Wehmann E, Köhnen M, Härter M et al. (2020) Therapeutic Alliance in Technology-Based Interventions for the Treatment of Depression: Systematic Review. J Med Internet Res 22:e17195. https://doi.org/10.2196/17195
19. Blease C, Rodman A (2025) Generative artificial intelligence in mental healthcare: an ethical evaluation. Curr Treat Options Psychiatry 12:5. https://doi.org/10.1007/s40501-024-00340-x
20. Torous J, Blease C (2024) Generative artificial intelligence in mental health care: potential benefits and current challenges. World Psychiatry 23:1–2. https://doi.org/10.1002/wps.21148

21. Rolvien L, Kruse L, Rings S et al. (2026) Intelligent virtual agents in psychotherapy: a safety evaluation across high-risk mental health scenarios. Sci Rep 16:13411. https://doi.org/10.1038/s41598-026-49764-w
22. Augustin M, Pollak TA, Morrin H (2026) Characterizing the spiral: potential mechanisms in AI-associated delusions. NPP Digit Psychiatry Neurosci 4:14. https://doi.org/10.1038/s44277-026-00065-0
23. Chen S, Gao M, Sasse K et al. (2025) When helpfulness backfires: LLMs and the risk of false medical information due to sycophantic behavior. NPJ Digit Med 8:605. https://doi.org/10.1038/s41746-025-02008-z
24. Tavory T (2024) Regulating AI in Mental Health: Ethics of Care Perspective. JMIR Ment Health 11:e58493. https://doi.org/10.2196/58493
25. Hillebrand M, Baumeister H (2025) Chance oder Risiko? KI-basierte Tools in der Psychotherapie. Psychotherapeutenjournal 24:158–160. https://doi.org/10.61062/ptj202502.006