

p

Psychotherapiewissenschaft  
in Forschung, Profession und Kultur

Band 48

Paolo Raile, Peter Geißler

# KI in der Psychotherapie (-wissenschaft)



WAXMANN

# Psychotherapiewissenschaft in Forschung, Profession und Kultur

Schriftenreihe der  
Sigmund-Freud-Privatuniversität Wien

Herausgegeben von Bernd Rieken

Band 48

Die Sigmund-Freud-Privatuniversität in Wien ist die erste akademische Lehrstätte, an der die Ausbildung zum Psychotherapeuten integraler Bestandteil eines eigenen wissenschaftlichen Studiums ist. Durch das Studium der Psychotherapiewissenschaft (PTW) wird dem Umstand Rechnung getragen, dass Psychotherapie eine hoch professionelle Tätigkeit ist, die – wie andere hoch professionelle Tätigkeiten auch – neben einer praktischen Ausbildung eines eigenen akademischen Studiums bedarf. Das hat zur Konsequenz, dass die wissenschaftliche Beschäftigung mit ihr nicht mehr ausschließlich den Nachbardisziplinen Psychiatrie und Klinische Psychologie mit ihrer nomologischen Orientierung obliegt, sodass die PTW als eigene Disziplin an Konturen gewinnen kann.

Vor diesem Hintergrund wird die Titelwahl der wissenschaftlichen Reihe transparent: Es soll nicht nur die Kluft, welche zwischen Psychotherapieforschung und Profession besteht, verringert, sondern auch berücksichtigt werden, dass man der Komplexität des Gegenstands am ehesten dann gerecht wird, wenn neben den üblichen Zugängen der Human- und Naturwissenschaften auch Methoden und/oder Fragestellungen aus dem Bereich der Kultur-, Sozial- und Geisteswissenschaften Berücksichtigung finden.

Paolo Raile, Peter Geißler

# KI in der Psychotherapie(-wissenschaft)

Erkenntnisse aus Forschung und Praxis



Waxmann 2025  
Münster • New York

**Bibliografische Informationen der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

**Psychotherapiewissenschaft in Forschung, Profession und Kultur, Band 48**

ISSN 2192-2233

Print-ISBN 978-3-8188-0019-2

E-Book-ISBN 978-3-8188-5019-7

© Waxmann Verlag GmbH, Münster 2025

Steinfurter Straße 555, 48159 Münster

[www.waxmann.com](http://www.waxmann.com)

[info@waxmann.com](mailto:info@waxmann.com)

Umschlaggestaltung: Anne Breitenbach, Münster

Umschlagabbildung: © Thitisan/Shutterstock.com

Satz: MTS. Satz & Layout, Münster

Alle Rechte vorbehalten. Nachdruck, auch auszugsweise, verboten.  
Kein Teil dieses Werkes darf ohne schriftliche Genehmigung des Verlags in irgendeiner Form reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

# Vorwort

Im November 2024 führte uns die Vermittlung einer Kollegin zusammen: zwei Autoren mit einer gemeinsamen Leidenschaft. Ich verspürte damals den dringenden Wunsch, jemanden an meiner Seite zu wissen, mit dem ich mich fundiert über die Rolle der KI im Berufsfeld der Psychotherapeut\*innen austauschen konnte. Die Möglichkeiten, die sich durch KI ergeben, faszinierten mich sehr. Doch innerhalb meines bisherigen beruflichen Netzwerks stieß ich überwiegend auf Skepsis.

Umso größer war mein Glück, als Paolo Raile unmittelbar auf meine E-Mail reagierte und mir erfreut mitteilte, dass auch er sich – als Psychotherapeut und Wissenschaftler – intensiv mit KI und insbesondere ChatGPT beschäftigt. Er hat dazu sogar schon Veröffentlichungen vorzuweisen.

Den ersten Impuls für meine Beschäftigung mit KI verdanke ich jedoch meinem Freund Andi, der mich frühzeitig auf ChatGPT aufmerksam machte und mit dem ich viele spannende Experimente durchführte. Diese gemeinsamen Erkundungen halfen mir sehr, die Möglichkeiten und Einsatzbereiche besser zu verstehen.

In unserem ersten Gespräch erzählte ich Paolo von meiner Idee, ein Buch zu diesem spannenden Thema zu schreiben – und wie zwei Verlage meine Vorschläge zuvor abgelehnt hatten. Schon nach wenigen Tagen wurde uns klar, dass wir dieses Vorhaben gemeinsam verwirklichen wollten – und genau das nahmen wir dann in Angriff.

Uns war wichtig, das Projekt zügig voranzubringen. Die Entwicklungen im Bereich der KI schreiten so rasant voran, dass eine lange Planungsphase schnell zu einem veralteten Werk führen könnte. Deshalb bin ich Paolo Raile dankbar, dass er bereit war, dieses Projekt mit mir gemeinsam zu realisieren und mit großem Engagement zu begleiten.

Mit diesem Buch möchten wir unseren Leser\*innen ein Gefühl dafür vermitteln, welche Möglichkeiten KI in Zukunft für unser Berufsfeld bereithält. Unser Leitsatz dabei: KI als Werkzeug begreifen und lernen, effektiv mit ihr zu kooperieren.

Peter Geißler im Januar 2025



# Inhalt

1	Einführung – KI in der Psychotherapie(-wissenschaft) .....	9
2	Künstliche Intelligenz und Chatbots .....	12
2.1	Funktionsweise von Künstlicher Intelligenz .....	12
2.2	Die Entwicklung der Künstlichen Intelligenz seit den 1950er Jahren .....	13
2.3	ChatGPT – Die KI für Jedermann .....	15
2.4	Das Training von ChatGPT .....	16
2.5	Wie ChatGPT Texte erstellt und ausgibt .....	18
2.6	Fähigkeiten und Begrenzungen von ChatGPT .....	19
2.7	Ethische Überlegungen und Herausforderungen .....	21
2.8	KI in der Psychotherapie .....	23
2.9	Ideen und Tools zur Nutzung von KI in der Psychotherapie .....	26
2.10	KI und die DSGVO – Wichtige Hinweise für Europäer .....	28
3	KI in der Psychotherapiewissenschaft und -ausbildung .....	30
3.1	In der Experimentellen Psychotherapiewissenschaft .....	30
3.2	Im Bereich Psychotherapie-Integration .....	34
3.3	Als Hilfe zur Seminarvorbereitung .....	43
3.4	In der Psychotherapieausbildung – Ein Patient*innensimulator .....	54
4	KI in der Psychotherapieforschung .....	63
4.1	In der Planung von Forschungsprojekten .....	64
4.2	Zur Unterstützung bei Drittmittelanträgen .....	70
4.3	Im Projekt- und Qualitätsmanagement .....	74
4.4	Zur Unterstützung bei der Recherche .....	77
4.5	Zur Unterstützung bei der Durchführung von Forschungsprojekten .....	82
4.6	Zur Aufbereitung und Analyse von Daten .....	87
4.7	Beim Schreiben von Publikationen und ethische Bedenken .....	90
5	KI in der Psychotherapiepraxis .....	96
5.1	Meine Reise mit der KI: Von der Skepsis zur alltäglichen Nutzung .....	96
5.2	Der Schlüssel in der Anwendung von KI: das „Prompten“ .....	100
5.3	KI als Orientierungshilfe in einer laufenden Therapie .....	106
5.4	KI als Hilfe zur Prozessreflexion über eine lange Therapie .....	114
5.5	KI als Hilfe bei einer strukturellen Einschätzung .....	119
5.6	KI als Wiedererinnerungshilfe .....	124
5.7	KI als Hilfe zur Vorbereitung einer Intervention .....	128
5.8	KI als Hilfe zur Beleuchtung spezifischer Themenfelder .....	137
5.9	KI als Reflexionshilfe beim Nachdenken über bisher Ungedachtes .....	166
5.10	KI als Unterstützungstool bei der Diagnostik .....	198
5.11	KI als Unterstützungstool bei der Psychoedukation .....	204
5.12	KI als Unterstützung bei der Traumdeutung .....	206

6	KI in der Verwaltung einer Psychotherapiepraxis .....	217
6.1	Zur Unterstützung von Kommunikation und Terminmanagement .....	217
6.2	Zur Erstellung von Webseiten und Informationsplattformen .....	219
6.3	Als Unterstützung bei der Dokumentation und Verrechnung .....	222
7	Zusammenfassung .....	226
8	Literatur .....	229

# 1 Einführung – KI in der Psychotherapie(-wissenschaft)

Die rapide Weiterentwicklung der Künstlichen Intelligenz (KI) hat in den letzten Jahren nicht nur technologische, sondern auch gesellschaftliche, ethische und berufspraktische Dimensionen beeinflusst. Im Bereich der Psychotherapie, einer Disziplin, die sich traditionell stark auf menschliche Interaktion und Empathie stützt, wird die Integration von KI oft mit einer Mischung aus Faszination und Skepsis betrachtet. Dieses Buch widmet sich der vielschichtigen Frage, wie KI die psychotherapeutische Praxis, Wissenschaft und Ausbildung bereichern, unterstützen und transformieren kann. Dabei steht stets die Frage im Raum, wie diese Technologie als Werkzeug eingesetzt werden kann, ohne den Kern der Psychotherapie – die zwischenmenschliche Beziehung und das Vertrauen – zu kompromittieren.

Die Digitalisierung des Gesundheitswesens, insbesondere der Psychotherapie, schreitet unaufhaltsam voran. KI-basierte Systeme haben das Potenzial, eine Vielzahl von Prozessen effizienter zu gestalten, von der administrativen Unterstützung über diagnostische Anwendungen bis hin zur Entwicklung neuer therapeutischer Ansätze. Gleichzeitig stellen sie Psychotherapeut\*innen und Forschende vor Herausforderungen, die von technischen Limitationen über ethische Dilemmata bis hin zur Frage der Menschlichkeit in der Therapie reichen. Dieses Buch ist ein Versuch, diese Herausforderungen zu beleuchten und eine fundierte Grundlage für den kritischen und zugleich konstruktiven Umgang mit KI in der Psychotherapie(-wissenschaft) zu schaffen. Dabei wird nicht nur der praktische Nutzen von KI untersucht, sondern auch deren potenzielle Risiken und Grenzen kritisch reflektiert. Es geht darum, einen bewussten Umgang mit dieser Technologie zu fördern, der sowohl die Möglichkeiten als auch die Verantwortung, die mit ihrem Einsatz einhergehen, berücksichtigt.

Ein zentrales Anliegen dieses Buches ist es, die verschiedenen Möglichkeiten und Grenzen der KI aufzuzeigen. Dabei wird die Perspektive von Behandelnden, Lehrenden, Lernenden und Forschenden gleichermaßen berücksichtigt. KI soll hier nicht als Bedrohung, sondern als Werkzeug verstanden werden, das durch kompetente und reflektierte Anwendung erhebliche Vorteile bieten kann. Es geht darum, wie KI in der Praxis und Forschung nicht nur Arbeitsprozesse optimieren, sondern auch innovative Lösungen für komplexe Fragestellungen bieten kann. Die Diskussion über den Einsatz von KI ist dabei keineswegs neu, hat jedoch in den letzten Jahren durch die Entwicklung leistungsfähiger Sprachmodelle wie ChatGPT eine neue Dringlichkeit und Aktualität gewonnen. Dieses Buch versucht, sowohl eine Orientierungshilfe als auch eine Inspirationsquelle zu sein und den Leser\*innen konkrete Werkzeuge an die Hand zu geben, mit denen sie KI in ihren beruflichen Alltag integrieren können.

Die Idee zu diesem Buch entstand aus der Beobachtung heraus, dass die Diskussion über die Rolle von KI in der Psychotherapie oft polarisiert geführt wird. Auf der

einen Seite gibt es Überzeugte, die in der KI die Antwort auf nahezu alle Probleme des Gesundheitswesens sehen. Auf der anderen Seite stehen Skeptiker\*innen, die jede technologische Neuerung als potenzielle Gefahr für die therapeutische Beziehung und die Integrität der Psychotherapie betrachten. Dieses Buch versucht, eine Brücke zwischen beiden Positionen zu schlagen und eine ausgewogene Sichtweise zu präsentieren. Dabei wird die Diskussion stets von der Frage geleitet, wie KI sinnvoll eingesetzt werden kann, ohne dabei die zentrale Rolle der menschlichen Expertise und Empathie zu untergraben. Es geht darum, die Chancen und Risiken gleichermaßen zu erkennen und einen Weg zu finden, der sowohl die Effizienz steigert als auch die ethischen Grundprinzipien wahrt.

Die Zielsetzung ist dabei zweifach: Zum einen soll die Leserschaft ein Verständnis dafür entwickeln, wie KI funktioniert und welche Potenziale sie bietet. Zum anderen werden konkrete Anwendungen und Szenarien vorgestellt, die aufzeigen, wie KI verantwortungsvoll und effektiv eingesetzt werden kann. Dabei wird auch auf Grenzen, Risiken und ethische Implikationen eingegangen, um einen ganzheitlichen Zugang zu dem Thema zu ermöglichen. Gleichzeitig wird der Blick auf die Praxis gerichtet, um zu zeigen, wie KI bereits heute in der Psychotherapie genutzt wird und welche Erfahrungen dabei bisher gemacht wurden. Es geht darum, sowohl theoretische Grundlagen zu schaffen als auch praxisnahe Impulse zu geben, die den Leser\*innen helfen, eigene Ansätze und Anwendungen zu entwickeln.

Die Struktur dieses Buches orientiert sich an den verschiedenen Einsatzbereichen von KI in der Psychotherapie(-wissenschaft). Es beginnt mit einer grundlegenden Betrachtung der technischen Hintergründe und möglichen Anwendungsfelder von KI. Hierbei werden zentrale Technologien wie maschinelles Lernen, Sprachmodelle und KI-gestützte Systeme erklärt, um ein solides Verständnis für die folgenden Kapitel zu schaffen. Besonderes Augenmerk liegt dabei auf der Funktionsweise von Sprachmodellen wie ChatGPT, die als ein Paradebeispiel für die Leistungsfähigkeit moderner KI-Technologien dienen.

Im darauffolgenden Kapitel wird die Rolle von KI in der Psychotherapiewissenschaft untersucht. Dieses Kapitel widmet sich sowohl der theoretischen Fundierung als auch der praktischen Anwendung in der Lehre und Ausbildung. Es werden Möglichkeiten aufgezeigt, wie KI in der universitären Psychotherapieausbildung eingesetzt werden kann, beispielsweise durch Patient\*innensimulatoren oder adaptive Lernplattformen. Gleichzeitig wird erörtert, wie KI dazu beitragen kann, schulübergreifende Perspektiven und neue theoretische Ansätze zu entwickeln.

Das vierte Kapitel beschäftigt sich mit der Psychotherapieforschung. Hier wird detailliert beschrieben, wie KI als Werkzeug zur Datenanalyse und zur Strukturierung von Forschungsprozessen eingesetzt werden kann. Beispiele umfassen den Einsatz von Machine-Learning-Algorithmen zur Mustererkennung in großen Datensätzen, die automatisierte Literaturrecherche und die Erstellung wissenschaftlicher Manuskripte. Dieses Kapitel hebt hervor, wie KI nicht nur bestehende Forschungsmethoden effizienter gestalten, sondern auch neue Forschungsfragen generieren kann.

Der praktische Einsatz von KI in der therapeutischen Arbeit steht im Mittelpunkt des fünften Kapitels. Es werden sowohl inhaltliche als auch organisatorische Aspekte beleuchtet, darunter die Nutzung von KI zur Prozessreflexion, zur Vorbereitung von Interventionen, zur Diagnostik und zur psychoedukativen Arbeit. Praktische Beispiele und Fallstudien illustrieren, wie KI in der alltäglichen Praxis eingesetzt werden kann, um die Arbeit von Psychotherapeut\*innen zu erleichtern und die Versorgung der Patient\*innen zu verbessern.

Abschließend widmet sich das Buch den organisatorischen Fragen der Psychotherapiepraxis, die mit dem Einsatz von KI effizienter beantwortet werden können. Es wird diskutiert, wie KI-Tools wie ChatGPT Webseiten für Psychotherapeut\*innen erstellen, bei der Dokumentation unterstützen oder die Kommunikationstasks effizienter abarbeiten können.

Dieses Buch versteht sich als Einladung, die Rolle von KI in der Psychotherapie nicht nur zu beobachten, sondern aktiv mitzugestalten. Es geht darum, sowohl die Möglichkeiten als auch die Herausforderungen zu erkennen und sich damit auseinanderzusetzen, wie diese Technologie das Berufsfeld der Psychotherapie verändern kann und wird. Dabei wird immer wieder betont, dass KI kein Ersatz für die menschliche Expertise ist, sondern ein Werkzeug, das – richtig eingesetzt – einen echten Mehrwert bieten kann. Die Leser\*innen werden ermutigt, die vorgestellten Technologien auszuprobieren und ihre eigenen Erfahrungen damit zu machen. Gleichzeitig soll dieses Buch helfen, eine kritische Haltung zu entwickeln und die ethischen Implikationen zu berücksichtigen. Denn nur durch einen reflektierten Umgang mit KI kann sichergestellt werden, dass diese Technologie zum Wohl der Patient\*innen und zur Weiterentwicklung der Psychotherapie genutzt wird.

Die Reflexion über den Einsatz von KI in der Psychotherapie eröffnet eine neue Dimension der Auseinandersetzung mit der Digitalisierung und Automatisierung im Gesundheitswesen. Dabei wird deutlich, dass es nicht nur um technische Herausforderungen geht, sondern auch um die grundsätzliche Frage, wie wir als Gesellschaft mit neuen Technologien umgehen wollen. Dieses Buch bietet einen Rahmen für diese Diskussion und möchte dazu beitragen, dass KI nicht als Bedrohung, sondern als Chance wahrgenommen wird. Es fordert dazu auf, die Möglichkeiten der KI mutig zu erkunden, dabei jedoch stets die ethischen und professionellen Prinzipien der Psychotherapie zu bewahren. Die kommenden Kapitel laden dazu ein, die Welt der KI in der Psychotherapie zu entdecken, ihre Chancen zu erkennen und sich mit ihren Herausforderungen auseinanderzusetzen. Lassen Sie uns gemeinsam erkunden, wie wir mit KI die psychotherapeutische Praxis, Wissenschaft und Lehre in eine neue Ära führen können.

## 2 Künstliche Intelligenz und Chatbots

### 2.1 Funktionsweise von Künstlicher Intelligenz

Maschinelles Lernen bildet das Herzstück moderner Künstlicher Intelligenz (KI). Es ermöglicht Maschinen, Muster in Daten zu erkennen und basierend darauf Entscheidungen zu treffen oder Vorhersagen zu machen. Die Grundprinzipien des maschinellen Lernens lassen sich in drei Kategorien einteilen: überwachtes Lernen, unüberwachtes Lernen und bestärkendes Lernen (Roumeliotis & Tselikas, 2023). Überwachtes Lernen ist der am weitesten verbreitete Ansatz. Hierbei wird ein Algorithmus mit einem Datensatz trainiert, der sowohl Eingabedaten als auch die dazugehörigen Ausgaben enthält. Ziel ist es, die Beziehung zwischen diesen Variablen zu lernen, um zukünftige Eingaben korrekt vorherzusagen (Wu et al., 2023). Ein typisches Beispiel ist die Bildklassifikation, bei der ein Modell darauf trainiert wird, ein Bild einer bestimmten Kategorie zuzuordnen, etwa „Hund“ oder „Katze“ (Hariri, 2024).

Im Gegensatz dazu steht das unüberwachte Lernen, bei dem der Algorithmus nur mit unmarkierten Daten arbeitet. Ziel ist es, versteckte Muster oder Strukturen in den Daten zu erkennen. Dies wird häufig bei Clusteranalysen eingesetzt, beispielsweise zur Segmentierung von Kunden in der Marktforschung (Liu & Wang, 2024). Bestärkendes Lernen wiederum basiert auf einem Belohnungssystem. Hier lernt ein Agent, eine Sequenz von Aktionen zu wählen, die zu einem optimalen Ergebnis führen. Dies wird häufig in der Robotik und in Computerspielen eingesetzt, wo der Algorithmus durch Versuch und Irrtum lernt, Aufgaben zu optimieren (Briganti, 2024).

Die Qualität und Menge der Daten, die ein KI-System nutzt, sind entscheidend für seine Leistungsfähigkeit. Große Datensätze sind notwendig, um Muster zuverlässig zu erkennen und generalisierbare Modelle zu entwickeln. Sprachmodelle wie GPT-3 oder GPT-4, die Millionen von Parametern haben, wurden auf Textmengen trainiert, die hunderte Milliarden Wörter umfassen (Wu et al., 2023). Ohne diese riesigen Datensätze könnten solche Modelle nicht die sprachliche Vielfalt und Genauigkeit erreichen, die sie auszeichnet. Gleichzeitig ist die Qualität der Daten entscheidend. Verzerrungen in den Trainingsdaten können zu systematischen Fehlern führen, was die Bedeutung der Kuratierung und Aufbereitung von Daten unterstreicht (Roumeliotis & Tselikas, 2023).

Eine der eindrucksvollsten Anwendungen von maschinellem Lernen ist die maschinelle Übersetzung. Systeme wie Google Translate basieren auf neuronalen Netzen, die auf großen mehrsprachigen Korpora trainiert werden. Sie nutzen selbstaufmerksame Mechanismen innerhalb der Transformer-Architektur, um die Bezie-

hungen zwischen Wörtern und Sätzen zu verstehen (Liu & Wang, 2024). Dadurch können sie nicht nur Wörter übersetzen, sondern auch idiomatische und grammatische Feinheiten zwischen Sprachen berücksichtigen.

Ein weiteres Beispiel sind personalisierte Empfehlungssysteme, die Algorithmen verwenden, um individuelle Vorlieben zu analysieren und relevante Inhalte vorzuschlagen. Diese Systeme finden Anwendung in Streaming-Diensten wie Netflix oder E-Commerce-Plattformen wie Amazon. Sie lernen aus dem Verhalten der Nutzer – etwa durch Klicks oder Kaufhistorien – und verwenden diese Daten, um präzise und relevante Vorschläge zu machen (Briganti, 2024). Solche Modelle verwenden häufig unüberwachte Lerntechniken, um Cluster ähnlicher Nutzer zu identifizieren, oder hybride Ansätze, die sowohl überwachtes als auch unüberwachtes Lernen kombinieren (Hariri, 2024).

Sprachmodelle wie ChatGPT sind ein Beispiel für die Weiterentwicklung dieser Technologien. Sie kombinieren unterschiedliche maschinelle Lernansätze, um natürliche Sprache zu verstehen und zu generieren. Während des Trainingsprozesses werden riesige Mengen an Textdaten verwendet, um semantische Beziehungen zwischen Wörtern und Sätzen zu modellieren (Wu et al., 2023). Dadurch können sie menschenähnliche Antworten in einer Vielzahl von Kontexten erzeugen, von technischen Anfragen bis hin zu kreativen Aufgaben wie dem Schreiben von Gedichten oder Geschichten (Roumeliotis & Tselikas, 2023).

Die Fortschritte im maschinellen Lernen haben dazu geführt, dass KI-Systeme heute in einer Vielzahl von Anwendungen eingesetzt werden können. Übersetzungsdienste, personalisierte Empfehlungen und sprachbasierte Modelle sind nur einige Beispiele dafür, wie KI unseren Alltag bereichert und vereinfacht. Trotz dieser Fortschritte bleibt die Weiterentwicklung von Trainingsmethoden und Datenaufbereitung eine zentrale Herausforderung, um die Leistungsfähigkeit und Verlässlichkeit dieser Systeme weiter zu verbessern.

## 2.2 Die Entwicklung der Künstlichen Intelligenz seit den 1950er Jahren

Künstliche Intelligenz (KI) hat in den letzten Jahrzehnten eine bemerkenswerte Entwicklung durchlaufen und sich von einer theoretischen Disziplin zu einer Schlüsseltechnologie entwickelt, die unser Leben und unsere Arbeit auf vielfältige Weise beeinflusst. Im Kern beschreibt KI die Fähigkeit von Maschinen, menschliche kognitive Prozesse wie Lernen, Problemlösen und Entscheidungsfindung zu imitieren (Wolfram, 2023). Das Ziel der KI besteht darin, komplexe Aufgaben zu lösen, die traditionell menschliche Intelligenz erfordern. Die frühe Forschung im Bereich KI reicht bis in die 1950er Jahre zurück, als Wissenschaftler wie John McCarthy den Begriff prägten und die ersten regelbasierten Systeme entwickelten. Ein bekanntes Beispiel ist „Eliza“, ein simples textbasiertes System, das als virtueller Therapeut fungierte. Es konnte einfache Dialoge führen und wird oft als Vorläufer moderner Sprachmodelle betrachtet (Liu & Wang, 2024).

Die Einführung von neuronalen Netzwerken in den 1980er Jahren markierte einen entscheidenden Wendepunkt, da diese Technologie Maschinen in die Lage versetzte, Muster aus Daten zu erkennen und zu lernen (Wolfram, 2023). In jüngerer Zeit hat die Transformer-Architektur, die 2017 von Vaswani und Kollegen eingeführt wurde, die Entwicklung von KI-Systemen revolutioniert. Diese Architektur bildet die Grundlage für sogenannte große Sprachmodelle wie GPT (Generative Pretrained Transformer), die in Systemen wie ChatGPT Anwendung finden (Wu et al., 2023).

Der Einsatz von KI in der Psychotherapie eröffnet neue Möglichkeiten, um sowohl Patient\*innen als auch Therapeut\*innen zu unterstützen. Beispielsweise können KI-Systeme Frühwarnsysteme für psychische Krisen entwickeln, Therapieprozesse überwachen und personalisierte Behandlungspläne erstellen. ChatGPT, als eines der fortschrittlichsten Beispiele für KI im Bereich natürlicher Sprache, wird bereits zur Erstellung psychoedukativer Inhalte und zur Unterstützung einfacher therapeutischer Interventionen eingesetzt (Rogan et al., 2024; Woodnutt et al., 2024). Solche Anwendungen tragen dazu bei, die Zugänglichkeit und Effizienz von psychotherapeutischen Angeboten zu verbessern, indem sie Barrieren wie lange Wartezeiten oder geografische Entfernungen überwinden (Li et al., 2024).

Die Leistungsfähigkeit von KI beruht auf ihrer Fähigkeit, aus großen Mengen an Daten zu lernen. Sprachmodelle wie ChatGPT nutzen eine Kombination aus unüberwachtem Lernen, um grundlegende Sprachstrukturen zu erfassen, und überwachtem Lernen, um spezifische Aufgaben zu verfeinern. Dieser zweistufige Ansatz ermöglicht es den Modellen, sinnvolle und kontextgerechte Antworten auf natürliche Spracheingaben zu geben (Wolfram, 2023; Liu & Wang, 2024). Der Erfolg solcher Systeme in der Psychotherapie hängt jedoch davon ab, dass sie ethisch verantwortungsvoll eingesetzt werden. Besonders in der Interaktion mit Patient\*innen müssen Algorithmen so gestaltet werden, dass sie keinen Schaden verursachen und kulturelle sowie individuelle Unterschiede berücksichtigen (Mensah & Addy, 2023).

Trotz der beeindruckenden Fortschritte gibt es bedeutende Herausforderungen und Begrenzungen bei der Nutzung von KI. Eines der Hauptprobleme liegt in der Transparenz solcher Modelle, da die Mechanismen, die zu spezifischen Vorhersagen oder Entscheidungen führen, oft nur schwer nachvollziehbar sind. Dies erschweren die Akzeptanz und das Vertrauen sowohl bei Patient\*innen als auch bei Therapeut\*innen (Li et al., 2024). Ein weiteres Problem ist die potenzielle Verstärkung von Vorurteilen, die in den Trainingsdaten vorhanden sein könnten. Eine unkritische Verwendung von KI könnte bestehende Ungleichheiten und Stigmata in der psychischen Gesundheitsversorgung verschärfen. Daher ist es entscheidend, dass KI-Systeme sorgfältig geprüft und überwacht werden, um diese Risiken zu minimieren (Rogan et al., 2024; Woodnutt et al., 2024).

Die Zukunft der KI in der Psychotherapie könnte durch multimodale Systeme geprägt sein, die Text-, Bild- und Sprachdaten kombinieren, um noch tiefere Einblicke in die psychische Gesundheit zu ermöglichen. Solche Systeme könnten nicht nur präzisere Diagnosen stellen, sondern auch individuelle Therapieansätze entwickeln, die speziell auf die Bedürfnisse aller Patient\*innen zugeschnitten sind (Wu et al.,

2023; Olawade et al., 2024). Dennoch bleibt die Frage, wie weit KI gehen kann, ohne die menschliche Komponente in der Therapie zu ersetzen. Der zwischenmenschliche Aspekt der Psychotherapie ist von zentraler Bedeutung, und KI sollte eher als Werkzeug zur Unterstützung, denn als Ersatz für menschliche Therapeut\*innen betrachtet werden (Mensah & Addy, 2023).

Die Integration von KI in die Psychotherapie ist eine spannende, aber auch herausfordernde Entwicklung. Sie bietet die Möglichkeit, die psychische Gesundheitsversorgung zu revolutionieren, indem sie die Effizienz steigert und die Reichweite erweitert. Gleichzeitig erfordert sie jedoch einen verantwortungsvollen Umgang, um sicherzustellen, dass ethische Standards eingehalten und die Bedürfnisse der Patient\*innen berücksichtigt werden. Mit der richtigen Balance zwischen technologischem Fortschritt und menschlicher Expertise könnte KI ein wertvolles Instrument zur Verbesserung der psychischen Gesundheit werden.

### 2.3 ChatGPT – Die KI für Jedermann

ChatGPT, ein Generative Pretrained Transformer-Modell (GPT), ist ein bedeutendes Beispiel für die jüngsten Fortschritte in der künstlichen Intelligenz (KI). Es handelt sich um ein Sprachmodell, das von OpenAI entwickelt wurde und auf der Transformer-Architektur basiert, die erstmals von Vaswani et al. im Jahr 2017 vorgestellt wurde. Diese Architektur nutzt Self-Attention-Mechanismen, um lange Abfolgen von Textdaten effizient zu analysieren und zu generieren (Liu & Wang, 2024). Die Hauptfunktion von ChatGPT ist die Generierung von Texten in natürlicher Sprache, die sowohl kohärent als auch kontextbezogen sind, was es besonders für dialogbasierte Anwendungen prädestiniert (Wu et al., 2024).

Die technische Grundlage von ChatGPT basiert auf einer mehrstufigen Entwicklung, beginnend mit der Vortrainierungsphase. In dieser Phase wird das Modell auf großen Textmengen trainiert, um sprachliche Muster und Strukturen zu erfassen. Hierbei wird ein selbstüberwachtes Lernverfahren eingesetzt, das es dem Modell ermöglicht, ohne manuelle Kennzeichnung der Daten zu lernen. Das Ziel dieser Vortrainierung ist es, ein universelles Sprachverständnis zu schaffen, das anschließend für spezifische Aufgaben optimiert werden kann (Li et al., 2024).

Eine entscheidende Komponente der Transformer-Architektur ist der Mechanismus der Multi-Head Attention. Dieser ermöglicht es, verschiedene Teile eines Textes gleichzeitig zu analysieren und dabei auch langfristige Abhängigkeiten zwischen Wörtern zu berücksichtigen. Diese Fähigkeit ist zentral für die Effizienz und Genauigkeit von Modellen wie ChatGPT, insbesondere bei Aufgaben, die eine detaillierte Kontextmodellierung erfordern (Liu & Wang, 2024).

Die Feinabstimmung von ChatGPT erfolgt durch Reinforcement Learning with Human Feedback (RLHF). Dieses Verfahren integriert menschliches Feedback, um die Qualität und Kohärenz der vom Modell erzeugten Antworten zu verbessern. Es sorgt dafür, dass ChatGPT nicht nur präzise, sondern auch nutzerzentrierte Er-

gebnisse liefert, was seine Anwendungsmöglichkeiten erheblich erweitert (Briganti, 2024).

Ein bemerkenswertes Merkmal von ChatGPT ist seine Vielseitigkeit. Neben der Generierung von Texten wird es für Aufgaben wie Übersetzungen, Textzusammenfassungen und kreative Textgenerierungen eingesetzt. In kreativen Prozessen, beispielsweise beim Verfassen von Geschichten oder Gedichten, zeigt das Modell ein hohes Maß an Flexibilität und Anpassungsfähigkeit (Mensah & Addy, 2023). Diese Vielseitigkeit wird durch die kontinuierliche Optimierung des Modells und den Zugang zu riesigen Datenmengen ermöglicht (Wu et al., 2024).

ChatGPT wird in verschiedenen Branchen genutzt, darunter im Kundenservice, in der Medizin und dem Bildungswesen. In der Medizin wird es beispielsweise zur Unterstützung von Ärzten in der Entscheidungsfindung und zur Erstellung medizinischer Berichte eingesetzt. Diese Anwendungen verdeutlichen, wie KI-Technologien zunehmend in spezialisierte Berufsfelder integriert werden (Woodnutt et al., 2024).

Die Weiterentwicklung des Modells hat auch zur Einführung multimodaler Fähigkeiten geführt. So kann die neueste Version von ChatGPT sowohl Texte als auch Bilder verarbeiten, was seine Anwendbarkeit in Bereichen wie Bildunterschriftenerstellung und komplexer Datenvisualisierung erweitert (Wu et al., 2024). Diese Fähigkeit hebt ChatGPT von früheren Sprachmodellen ab und zeigt, wie KI zunehmend anspruchsvollere Aufgaben bewältigen kann (Liu & Wang, 2024).

Trotz seiner beeindruckenden Fähigkeiten bleibt ChatGPT ein Werkzeug, das von Nutzern verantwortungsbewusst eingesetzt werden muss. Die Bedeutung des Modells liegt nicht nur in seinen technischen Fortschritten, sondern auch in der Art und Weise, wie es das Verständnis und die Interaktion mit KI in verschiedenen Anwendungsbereichen verändert hat (Olawade et al., 2024). Es ist ein Paradebeispiel für die Möglichkeiten und Herausforderungen, die KI in einer zunehmend digitalisierten Welt bietet.

## 2.4 Das Training von ChatGPT

Das Training von ChatGPT erfolgt in mehreren Schritten, die darauf abzielen, das Modell so zu entwickeln, dass es natürliche Sprache präzise verstehen und sinnvoll generieren kann. Ein zentraler Bestandteil dieses Prozesses ist die sogenannte Vortrainierung. Dabei wird das Modell mit einer riesigen Menge an Texten trainiert, die aus Büchern, Artikeln, Webseiten und anderen Quellen stammen. Ziel ist es, grundlegende Muster und Strukturen der Sprache zu erkennen. Während der Vortrainierung lernt das Modell beispielsweise, dass Wörter in einem Satz eine bestimmte Reihenfolge und grammatikalische Beziehung haben (Radford et al., 2019; Vaswani et al., 2017).

Eine Schlüsselinnovation bei ChatGPT ist die Verwendung der sogenannten Transformer-Architektur. Diese Technologie ermöglicht es, große Textmengen effizient zu analysieren. Anders als ältere Modelle, die Wörter einzeln oder nachein-

ander verarbeiten, betrachtet der Transformer alle Wörter eines Textes gleichzeitig. Dadurch können auch Verbindungen zwischen weit auseinanderliegenden Wörtern oder Sätzen erkannt werden. Diese Fähigkeit macht die Architektur besonders geeignet für Aufgaben wie Textzusammenfassungen oder Konversationen (Vaswani et al., 2017).

Innerhalb der Transformer-Architektur wird der Mechanismus der Self-Attention verwendet. Dieser Ansatz hilft dem Modell, wichtige Wörter in einem Satz hervorzuheben. Ein Beispiel: In dem Satz „Der Hund, der die Katze jagt, ist schnell“, erkennt Self-Attention, dass „der Hund“ und „ist schnell“ zusammengehören, auch wenn die Worte nicht nebeneinanderstehen. Dadurch kann das Modell den Zusammenhang besser verstehen und natürliche Antworten generieren (Radford et al., 2019).

Nach der Vortrainierung wird das Modell durch ein Verfahren namens Reinforcement Learning with Human Feedback (RLHF) weiter verbessert. Hier bewerten menschliche Experten die von ChatGPT generierten Texte und geben Feedback, ob die Antworten korrekt und hilfreich sind. Das Modell nutzt diese Rückmeldungen, um seine Parameter so anzupassen, dass die Qualität seiner Antworten steigt. Zum Beispiel könnten Nutzer bewerten, ob eine Antwort sachlich korrekt ist oder ob sie unklar formuliert wurde. Dies macht ChatGPT in der Praxis vielseitiger einsetzbar (Wu et al., 2023).

Eine der größten Herausforderungen bei diesem Prozess ist die enorme Rechenleistung, die benötigt wird. Das Training von ChatGPT erfordert spezialisierte Hardware wie GPUs oder TPUs, die große Datenmengen parallel verarbeiten können. Dies macht das Training nicht nur aufwendig, sondern auch teuer, was die Weiterentwicklung solcher Modelle erschwert (van Kooten et al., 2024).

Zusätzlich spielt die Qualität der Daten eine zentrale Rolle. Verzerrungen in den Daten, wie etwa gesellschaftliche Vorurteile oder einseitige Perspektiven, können dazu führen, dass das Modell entsprechende Fehler übernimmt. Daher ist es wichtig, die Trainingsdaten sorgfältig auszuwählen und zu überprüfen, um solche Verzerrungen zu minimieren (Ferrara, 2024).

Ein weiterer Aspekt ist die feingranulare Optimierung des Modells für spezielle Anwendungsbereiche. Dies erfolgt durch sogenanntes Feintuning. Dabei wird ChatGPT auf kleinere, spezifischere Datensätze trainiert, um es beispielsweise für medizinische oder technische Anwendungen zu spezialisieren. Diese zusätzliche Trainingsphase ermöglicht es dem Modell, in Fachbereichen gezielte und präzisere Antworten zu liefern (Radford et al., 2019).

Trotz der technologischen Fortschritte stellt der Trainingsprozess weiterhin ethische Fragen. Beispielsweise könnte die Verwendung personenbezogener Daten oder sensibler Informationen Risiken bergen. Entwickler müssen sicherstellen, dass alle Daten im Einklang mit Datenschutzrichtlinien verarbeitet werden, und Transparenz bei der Nutzung solcher Systeme gewährleisten (Wu et al., 2023).

Die Fortschritte, die durch diesen rigorosen Prozess erzielt werden, machen ChatGPT zu einem leistungsstarken Werkzeug für unterschiedlichste Anwendungen. Es kann nicht nur einfache Konversationen führen, sondern auch komplexe Texte

generieren und kreative Aufgaben übernehmen. Diese Vielseitigkeit ist ein direkter Beweis für die Effektivität moderner KI-Trainingsmethoden und ihres Potenzials, die Art und Weise, wie wir Technologie nutzen, grundlegend zu verändern (Vaswani et al., 2017).

## 2.5 Wie ChatGPT Texte erstellt und ausgibt

ChatGPT entscheidet über die Generierung von Texten anhand von Wahrscheinlichkeiten, die es während seines Trainings gelernt hat. Das Modell analysiert dabei, welches Wort im gegebenen Kontext mit der höchsten Wahrscheinlichkeit als nächstes sinnvoll ist. Diese Technik wird als probabilistische Textgenerierung bezeichnet und bildet die Grundlage der Arbeitsweise von Sprachmodellen. Die Wahrscheinlichkeit basiert auf den Mustern und Strukturen, die das Modell aus umfangreichen Textkorpora gelernt hat. Beispielsweise könnte in einem Satz, der mit „Die Sonne scheint am“ beginnt, das Wort „Himmel“ eine höhere Wahrscheinlichkeit haben als „Regenschirm“ (Brown et al., 2020).

Der Mechanismus, der hinter dieser Entscheidungsfindung steht, wird durch die Transformer-Architektur ermöglicht. Diese Architektur verwendet Self-Attention-Mechanismen, um Beziehungen zwischen Wörtern zu gewichten und sicherzustellen, dass wichtige Informationen im Text erhalten bleiben. Selbst wenn Wörter oder Sätze weit voneinander entfernt sind, kann das Modell Verbindungen herstellen und relevante Zusammenhänge analysieren. So wird sichergestellt, dass die generierten Texte nicht nur lokal, sondern auch im größeren Kontext sinnvoll bleiben (Vaswani et al., 2017).

Ein entscheidender Parameter bei der Steuerung der Generierung ist die sogenannte „Temperatur“. Dieser Wert beeinflusst, wie stark die Wahrscheinlichkeiten bei der Auswahl der Wörter berücksichtigt werden. Eine niedrige Temperatur führt dazu, dass das Modell konservativer arbeitet und stets die wahrscheinlichste Fortsetzung wählt. Dies erzeugt stabile, aber vorhersehbare Ergebnisse. Eine höhere Temperatur hingegen erhöht die Wahrscheinlichkeit, dass auch weniger vorhersehbare Wörter gewählt werden, was zu kreativeren, aber möglicherweise weniger konsistenten Antworten führt (Radford et al., 2019).

Ein Beispiel verdeutlicht diesen Mechanismus: Soll das Modell eine Geschichte über einen „mutigen Ritter“ generieren, könnte bei einer niedrigen Temperatur eine klassische Abenteuergeschichte entstehen, die sich stark an bekannten Mustern orientiert. Mit einer höheren Temperatur könnten jedoch ungewöhnlichere Elemente wie ein philosophischer Drache oder ein Ritter, der lieber kocht als kämpft, eingefügt werden. Dies zeigt, wie die Temperatur die Balance zwischen Kreativität und Konsistenz beeinflusst (Brown et al., 2020).

Die Entscheidungsfindung von ChatGPT basiert jedoch nicht nur auf der Wahl einzelner Wörter. Der gesamte Kontext eines Gesprächs wird berücksichtigt, um sicherzustellen, dass die Antworten kohärent und inhaltlich relevant sind. Dieser Ansatz wird durch die Self-Attention-Mechanismen der Transformer-Architektur

unterstützt, die dafür sorgen, dass das Modell sowohl auf die unmittelbare als auch auf die weiter entfernte Umgebung eines Textes reagieren kann (Vaswani et al., 2017).

Ein weiteres Beispiel für die Anwendung probabilistischer Textgenerierung ist die Textzusammenfassung. Wenn ein Benutzer einen langen Text eingibt und eine Zusammenfassung anfordert, analysiert das Modell den Inhalt und generiert eine kompakte Darstellung der wichtigsten Punkte. Hierbei wird die Wahrscheinlichkeitsberechnung angepasst, um den Fokus auf Schlüsselinformationen zu legen und unnötige Details auszulassen (Rawas, 2024).

Obwohl ChatGPT leistungsstark ist, bleibt die probabilistische Natur seiner Entscheidungsfindung eine Herausforderung. Da das Modell die wahrscheinlichste Fortsetzung eines Textes auswählt, können Antworten entstehen, die zwar plausibel klingen, aber faktisch falsch sind. Dieses Problem, bekannt als „Halluzination“, ist ein bekanntes Phänomen bei Sprachmodellen und erfordert fortlaufende Forschung und Optimierung, um solche Fälle zu minimieren (Radford et al., 2019).

Die Temperatur, kombiniert mit der Wahrscheinlichkeitsberechnung, gibt Entwicklern und Nutzern die Möglichkeit, die Generierung von Texten flexibel zu steuern. Während niedrige Werte für verlässliche und vorhersehbare Ergebnisse sorgen, fördern höhere Werte Kreativität und Innovation. Diese Balance ist entscheidend für Anwendungen, die sowohl präzise als auch originelle Texte erfordern, wie z. B. Dialogsysteme, kreative Schreibprozesse oder automatische Zusammenfassungen (Brown et al., 2020).

## 2.6 Fähigkeiten und Begrenzungen von ChatGPT

ChatGPT hat durch seine fortschrittlichen technischen Grundlagen beeindruckende Fähigkeiten im Umgang mit Sprache entwickelt. Eine seiner größten Stärken ist das kontextuelle Sprachverständnis, das es dem Modell ermöglicht, auch bei komplexen Anfragen relevante und präzise Antworten zu generieren. Dies wird durch die Transformer-Architektur ermöglicht, die es erlaubt, große Textmengen effizient zu analysieren und dabei den Kontext von Wörtern und Sätzen zu berücksichtigen. Anders als ältere Modelle wie rekurrente neuronale Netze, die nur den unmittelbaren Zusammenhang zwischen Wörtern betrachten konnten, kann ChatGPT durch Self-Attention-Mechanismen auch weiter entfernte Verbindungen herstellen. Diese Fähigkeit ist besonders nützlich in Szenarien, in denen längere Konversationen oder komplexe Fragestellungen analysiert werden müssen (Vaswani et al., 2017).

Ein weiteres bemerkenswertes Merkmal von ChatGPT ist seine Kreativität. Das Modell kann originelle Texte generieren, die sowohl informativ als auch unterhaltsam sind. Dies umfasst das Verfassen von Gedichten, Geschichten oder kreativen Ideen für Projekte. Die zugrunde liegende Wahrscheinlichkeitsberechnung ermöglicht es dem Modell, zwischen vielfältigen möglichen Textausgaben zu wählen, was zu innovativen und manchmal unerwarteten Ergebnissen führt. Studien zeigen, dass ChatGPT in kreativen Prozessen als unterstützendes Werkzeug eingesetzt werden

kann, indem es neue Perspektiven oder Ideen einbringt, die von menschlichen Nutzern weiterentwickelt werden können (Liu et al., 2023).

Eine weitere Stärke von ChatGPT ist seine Flexibilität. Das Modell kann in einer Vielzahl von Domänen eingesetzt werden, darunter in der Generierung technischer Dokumentationen, medizinischer Berichte oder im Schreiben von Code. Diese Vielseitigkeit macht es zu einem wichtigen Werkzeug für Fachleute in unterschiedlichen Branchen. Durch die Feinabstimmung auf spezifische Datensätze kann das Modell außerdem an spezielle Anforderungen angepasst werden, was seine Anwendbarkeit in hochspezialisierten Bereichen wie der Medizin oder der Forschung erhöht (Aljamaan et al., 2024).

Trotz dieser beeindruckenden Fähigkeiten weist ChatGPT auch Schwächen auf, die seine Anwendung einschränken können. Eine der bedeutendsten ist die sogenannte Halluzination. Hierbei erzeugt das Modell Informationen, die zwar plausibel erscheinen, aber faktisch falsch oder irreführend sind. Diese Schwäche ist auf die probabilistische Natur der Textgenerierung zurückzuführen, bei der das Modell die wahrscheinlichste Fortsetzung eines Textes wählt, ohne den Inhalt auf Richtigkeit zu überprüfen. Solche Halluzinationen können besonders problematisch sein, wenn ChatGPT in sensiblen Bereichen wie der Medizin oder der Wissenschaft eingesetzt wird, wo genaue und verlässliche Informationen entscheidend sind (Chelli et al., 2024).

Ein weiteres Problem ist der Bias in den Trainingsdaten. Da ChatGPT auf öffentlich zugänglichen Daten trainiert wurde, spiegelt es oft die Vorurteile und Verzerrungen wider, die in diesen Daten enthalten sind. Beispielsweise können geschlechtsspezifische, kulturelle oder soziale Stereotype unbeabsichtigt verstärkt werden, wenn das Modell Inhalte generiert. Diese Verzerrungen stellen eine ethische Herausforderung dar, da sie in bestimmten Kontexten diskriminierend wirken oder bestehende Ungleichheiten verstärken können. Die Entwicklung neuer Ansätze zur Reduzierung solcher Biases bleibt ein zentrales Forschungsgebiet (Metze et al., 2024).

Eine weitere Begrenzung von ChatGPT ist das Fehlen eines tiefen Verständnisses. Obwohl das Modell in der Lage ist, kontextuell kohärente und sprachlich korrekte Antworten zu generieren, basiert seine Leistung auf statistischen Mustern und nicht auf echtem Wissen oder logischem Denken. Das Modell hat kein tatsächliches Verständnis für die Bedeutung der generierten Inhalte, was dazu führen kann, dass es inhaltlich unlogische oder sachlich falsche Aussagen macht. Dies ist ein grundlegendes Problem probabilistischer Sprachmodelle, das ihre Anwendung in bestimmten Bereichen einschränkt (Aljamaan et al., 2024).

Die potenziellen Gefahren bei der Nutzung von ChatGPT gehen über technische Schwächen hinaus. Eine der größten Herausforderungen ist der mögliche Missbrauch des Modells. Beispielsweise kann ChatGPT verwendet werden, um täuschend echte Texte für Desinformationskampagnen oder Social-Media-Bots zu erstellen, die die öffentliche Meinung manipulieren. Solche Szenarien verdeutlichen die Notwendigkeit, Mechanismen zu entwickeln, die den Missbrauch solcher Modelle begrenzen und ihre verantwortungsvolle Nutzung fördern (Chelli et al., 2024).

Auch der Datenschutz ist ein kritischer Punkt. Da ChatGPT auf riesigen Textmengen trainiert wird, besteht das Risiko, dass personenbezogene oder sensible Daten in die Trainingsdaten einfließen. Dies könnte zu Verstößen gegen Datenschutzgesetze führen oder sensible Informationen unabsichtlich reproduzieren. Die Implementierung strenger Datenschutzrichtlinien und die sorgfältige Überprüfung der verwendeten Daten sind entscheidend, um solche Risiken zu minimieren (Siontis et al., 2023).

Ethische Überlegungen spielen ebenfalls eine zentrale Rolle bei der Weiterentwicklung von ChatGPT. Besonders in der Medizin ist es entscheidend, dass die generierten Inhalte genau und vertrauenswürdig sind. Studien zeigen, dass ungenaue Antworten in klinischen Kontexten schwerwiegende Folgen für Patient\*innen haben können. Es ist daher unerlässlich, dass Modelle wie ChatGPT durch kontinuierliche Überwachung und Feinabstimmung verbessert werden, um die Sicherheit und Zuverlässigkeit ihrer Anwendungen zu gewährleisten (Metze et al., 2024).

Trotz dieser Herausforderungen bietet ChatGPT enormes Potenzial. Durch die kontinuierliche Forschung und Entwicklung können die Schwächen des Modells minimiert und seine Stärken weiter ausgebaut werden. Wenn es verantwortungsvoll eingesetzt wird, hat ChatGPT das Potenzial, Arbeitsprozesse zu optimieren, kreatives Denken zu fördern und in zahlreichen Bereichen einen positiven Einfluss zu haben (Aljamaan et al., 2024).

## 2.7 Ethische Überlegungen und Herausforderungen

Die ethische Nutzung von ChatGPT in der Therapie und anderen Bereichen ist eine der zentralen Herausforderungen moderner KI-Systeme. Eine verantwortungsvolle Anwendung erfordert die Berücksichtigung vielfältiger Faktoren, einschließlich des Schutzes der Privatsphäre, der Vermeidung von Diskriminierung und der Förderung von Transparenz. Generative KI wie ChatGPT hat die Fähigkeit, riesige Mengen an Textdaten zu analysieren und menschenähnliche Antworten zu generieren. Dies birgt sowohl große Chancen als auch erhebliche Risiken (Hua et al., 2024).

Ein wichtiges ethisches Anliegen ist der Datenschutz. ChatGPT verarbeitet Daten aus einer Vielzahl von Quellen, was das Risiko von Datenlecks oder die unbeabsichtigte Offenlegung personenbezogener Informationen erhöht. Obwohl OpenAI Maßnahmen wie die Blockierung von persönlichen Daten ergriffen hat, bestehen weiterhin Bedenken hinsichtlich der Transparenz der Datennutzung und möglicher Angriffe auf die Sicherheitsmechanismen (Wu et al., 2024).

Darüber hinaus werfen die Verzerrungen in den Trainingsdaten erhebliche ethische Fragen auf. ChatGPT lernt aus umfangreichen Datensätzen, die oft Vorurteile und Stereotypen enthalten. Diese können in den generierten Antworten widergespiegelt werden, was insbesondere in sensiblen Bereichen wie der Gesundheitsversorgung oder der Therapie problematisch ist. Verzerrungen könnten dazu führen, dass diskriminierende oder fehlerhafte Informationen generiert werden, was die Notwendigkeit von robusten Korrekturmaßnahmen unterstreicht (Zhou et al., 2024).

Die Integration multimodaler Daten, wie Text, Bild und Audio, stellt eine vielversprechende Zukunftsperspektive dar. In der Therapie könnte dies bedeuten, dass ChatGPT nicht nur auf Texteingaben reagiert, sondern auch visuelle und auditive Signale interpretiert, um beispielsweise emotionale Zustände besser zu verstehen. Dies könnte die Effizienz und Präzision bei der Diagnose und Behandlung erheblich verbessern, wirft jedoch zusätzliche ethische und regulatorische Fragen auf, insbesondere in Bezug auf den Schutz der Privatsphäre und die Einwilligung der Nutzer (Adhikari et al., 2023).

Ein weiteres Problem ist die potenzielle Abhängigkeit von KI-Systemen. Es besteht die Gefahr, dass menschliche Fachkräfte zunehmend auf KI vertrauen und ihre eigenen Fähigkeiten vernachlässigen. Diese Entwicklung könnte negative Auswirkungen auf die Qualität der Therapie haben, insbesondere in Situationen, die ein hohes Maß an menschlichem Einfühlungsvermögen und Urteilsvermögen erfordern (Hua et al., 2024).

Regulatorische Rahmenbedingungen spielen eine entscheidende Rolle bei der Sicherstellung der verantwortungsvollen Nutzung von ChatGPT. Es bedarf klarer Vorschriften, die den Einsatz solcher Technologien regeln, insbesondere in sensiblen Bereichen wie der Medizin oder der Therapie. Diese Vorschriften sollten Aspekte wie die Haftung bei fehlerhaften Antworten, die Transparenz der Datenquellen und den Schutz der Privatsphäre umfassen (Stahl & Eke, 2024).

Die Missbrauchsmöglichkeiten von ChatGPT sind ein weiteres ethisches Anliegen. Es besteht die Gefahr, dass das Modell für die Verbreitung von Desinformationen oder für betrügerische Aktivitäten genutzt wird. Dies unterstreicht die Notwendigkeit von Mechanismen, die den Missbrauch solcher Technologien erschweren und ihre Nutzung auf legitime Zwecke beschränken (Wu et al., 2024).

Trotz dieser Herausforderungen bietet ChatGPT auch bedeutende Vorteile. Die Fähigkeit, Informationen schnell zu analysieren und zu generieren, kann dazu beitragen, Arbeitsabläufe zu optimieren und die Effizienz in der Therapie zu erhöhen. Bei richtiger Anwendung kann ChatGPT die Qualität der Versorgung verbessern und den Zugang zu therapeutischen Ressourcen erweitern, insbesondere in unterversorgten Gebieten (Zhou et al., 2024).

Die Entwicklung einer klaren Ethikrichtlinie ist unerlässlich, um sicherzustellen, dass die Nutzung von ChatGPT mit menschlichen Werten und sozialen Normen übereinstimmt. Solche Richtlinien sollten auf einer breiten Konsultation mit Experten aus verschiedenen Bereichen basieren und die Bedürfnisse der Nutzer in den Mittelpunkt stellen (Hua et al., 2024).

Die Nutzung von ChatGPT in der Therapie bringt erhebliche ethische und regulatorische Herausforderungen mit sich. Gleichzeitig bietet es enorme Chancen, die durch eine sorgfältige und verantwortungsvolle Implementierung genutzt werden können, um die Therapie und andere Bereiche zu revolutionieren (Adhikari et al., 2023).

Im Abschnitt 4.7 wird genauer auf die ethischen Herausforderungen bei der Nutzung von KI im wissenschaftlichen Arbeiten eingegangen. In vielen anderen Kapiteln

werden ethische Gedanken gestreift. Die starke Präsenz des Themas weist jedenfalls darauf hin, dass hier noch viel regulatorische Arbeit geleistet werden muss, bevor KI sicher und ethisch bedenkenlos eingesetzt werden kann. Bis dahin gilt der Grundsatz: Handle nach bestem Wissen und Gewissen!

## 2.8 KI in der Psychotherapie

Die Integration künstlicher Intelligenz (KI) in die mentale Gesundheitsversorgung hat das Potenzial, bestehende Prozesse zu transformieren und neue Möglichkeiten zu schaffen. Verschiedene KI-gestützte Tools haben spezifische Anwendungsbereiche im Bereich Mental Health gefunden, von der Frühintervention bis hin zur Unterstützung komplexer therapeutischer Prozesse (Schneider & Eichenberg, 2025).

Eines der prominentesten Beispiele ist Woebot, ein chatbot-basierter virtueller Therapeut, der Elemente der kognitiven Verhaltenstherapie verwendet. Woebot wurde entwickelt, um niedrigschwellige emotionale Unterstützung zu bieten und Stimmungen zu analysieren. Studien zeigen, dass Woebot insbesondere bei der Selbsthilfe und in der Prävention effektiv ist, indem es Nutzer durch regelmäßige Check-ins unterstützt und evidenzbasierte psychologische Techniken bereitstellt (Fitzpatrick et al., 2017).

Ein weiteres Tool, das in der mentalen Gesundheitsversorgung breite Anwendung findet, ist Wya. Dieses KI-gestützte System kombiniert geführte Meditationen, personalisierte Selbsthilfeprogramme und psychologische Reflexionen. Besonders in der Frühintervention für Menschen mit milden bis moderaten Beschwerden hat Wya positive Ergebnisse erzielt (Inkster et al., 2018).

Replika bietet eine andere Herangehensweise. Ursprünglich entwickelt, um soziale Interaktionen zu fördern, wird Replika zunehmend in der mentalen Gesundheitsversorgung eingesetzt. Es ermöglicht personalisierte Gespräche, die das Gefühl von Einsamkeit reduzieren können. Jedoch weist Replika Einschränkungen auf, insbesondere bei der Behandlung von schweren psychischen Erkrankungen, da es ihm an tieferer emotionaler Intelligenz und klinischem Fachwissen fehlt (Jiang et al., 2022).

Neben chatbot-basierten Ansätzen existieren auch datengetriebene Plattformen wie Mindstrong, die Verhaltensdaten von Smartphones nutzen, um Frühwarnzeichen psychischer Erkrankungen zu erkennen. Mindstrong analysiert Tippmuster und andere Interaktionsmuster, um Veränderungen in der kognitiven Funktion zu identifizieren. Solche datengetriebenen Ansätze bieten neue Möglichkeiten in der Früherkennung, werfen jedoch ethische Fragen hinsichtlich des Datenschutzes und der Transparenz auf (Onnela & Rauch, 2016).

Ellie, ein KI-gestütztes virtuelles Agentensystem, ist ein weiteres Beispiel für den Einsatz von KI in der Diagnostik. Ellie nutzt maschinelles Lernen und Sprachverarbeitung, um psychologische Merkmale aus Sprache und Gesichtsausdrücken abzuleiten. Diese Technologie wird in der Forschung verwendet, um Symptome von Depression und posttraumatischer Belastungsstörung zu identifizieren (DeVault et al., 2014).

Mit der Einführung generativer Sprachmodelle wie ChatGPT ergeben sich zahlreiche neue Anwendungen in der Psychotherapie. Eine der zentralen Stärken von ChatGPT liegt in der Erstellung psychoedukativer Materialien. Es kann dazu verwendet werden, wissenschaftlich fundierte Inhalte in eine leicht verständliche Sprache zu übersetzen, die den Bedürfnissen von Patient\*innen mit unterschiedlichen Bildungsniveaus gerecht wird. Beispielsweise kann ChatGPT detaillierte Informationen über Angststörungen bereitstellen, einschließlich der zugrunde liegenden biologischen Mechanismen und bewährter Bewältigungsstrategien. Diese Materialien könnten nicht nur schriftlich, sondern auch in interaktiven Formaten angeboten werden, indem ChatGPT als virtueller Gesprächspartner auf spezifische Fragen der Patient\*innen eingeht. In der Patient\*innenaufklärung könnte das Modell Ärzt\*innen und Therapeut\*innen helfen, Erklärungen zu standardisieren, um sicherzustellen, dass alle Patient\*innen konsistente und qualitativ hochwertige Informationen erhalten. Darüber hinaus könnte es genutzt werden, um personalisierte Informationspakete zu erstellen, die auf die individuellen Bedürfnisse und Fragen eines\*einer Patient\*in zugeschnitten sind, beispielsweise in der Onkopsychologie oder der Suchttherapie (Brown et al., 2020).

Neben der Patient\*innenaufklärung zeigt ChatGPT großes Potenzial in der Forschung. Besonders in der qualitativen Forschung kann das Modell dazu beitragen, große Mengen an Interviewdaten effizient zu analysieren. Es kann Muster und Themen identifizieren, die sonst zeitaufwendige manuelle Kodierungsprozesse erfordern würden. So könnten beispielsweise Daten aus Interviews mit Patient\*innen, die an Depressionen leiden, schneller analysiert werden, um zentrale Belastungsfaktoren oder hilfreiche Interventionen zu identifizieren. Darüber hinaus bietet ChatGPT Werkzeuge zur Durchführung literaturbasierter Analysen, indem es schnell relevante Studien zusammenfasst, zentrale Ergebnisse extrahiert und Querverweise zwischen Forschungsartikeln herstellt. Dies könnte in systematischen Übersichtsarbeiten oder Metaanalysen genutzt werden, um den Forschungsprozess zu beschleunigen und gleichzeitig die Genauigkeit zu erhöhen (Liu et al., 2023).

Trotz dieser Vorteile gibt es deutliche Grenzen für den Einsatz von ChatGPT in der direkten Patient\*inneninteraktion. Obwohl das Modell empathisch klingende Antworten generieren kann, basiert es auf statistischer Wahrscheinlichkeitsberechnung und nicht auf einem echten Verständnis von Emotionen oder Kontext. Das Fehlen einer tiefen emotionalen Intelligenz bedeutet, dass ChatGPT zwar Sätze wie „Das klingt wirklich schwierig für Sie“ formulieren kann, jedoch nicht in der Lage ist, die tatsächliche emotionale Lage eines\*einer Patient\*in zu erkennen oder darauf einzugehen. Dies ist besonders problematisch bei komplexen psychischen Zuständen wie einer schweren Depression oder posttraumatischen Belastungsstörung, bei denen nicht nur fachliches Wissen, sondern auch ein hohes Maß an emotionaler Präsenz erforderlich ist. Darüber hinaus könnte die standardisierte Natur der Antworten zu Missverständnissen führen, wenn spezifische Nuancen in den Anliegen eines\*einer Patient\*in übersehen werden (Bender et al., 2021).