

Lea Maria Schäfer

Künstliche Intelligenz als Dritte Instanz in der Psychotherapie: Wie eine sinnvolle Integration von Künstlicher Intelligenz in den psychotherapeutischen Binnenraum aussehen könnte

Bots werden nicht im gesellschaftsfreien Raum entwickelt und nutzbar gemacht – es liegt an uns, sie so zu gestalten, dass sie uns helfen, ein erfüllteres und glücklicheres Leben zu führen. In Zukunft wird Künstliche Intelligenz (KI) vermehrt in der Psychiatrie und Psychotherapie eingesetzt werden, um präventive Maßnahmen, Diagnosen und Therapien zu unterstützen. Welche neuen Möglichkeiten und Herausforderungen bergen KI-basierte Interventionen für die aktuelle psychotherapeutische Praxis und für das Verhältnis von Therapeut*innen und Patient*innen? Eine niedrigschwellige Zugangsmöglichkeit zu digitaler und automatisierter therapeutischer Unterstützung konfrontiert nicht nur die traditionelle Psychotherapie mit neuen Möglichkeiten, sondern auch Menschen aller Altersgruppen mit einer neuen Beziehungsform – eine synthetische Beziehung, die internalisiert Einzug nimmt in den psychischen Innenraum des Individuums. Idealerweise fungiert der Bot als ‚dritte Instanz‘, empathisch und unterstützend, on-demand und flexibel einsetzbar. Gleichzeitig birgt der KI-Bot potenzielle Gefahren – gesellschaftliche Entfremdung, Übertherapie, Abhängigkeiten von einem Produkt, welches an finanzielle Interessen geknüpft ist. Forschung ist entscheidend, um die Interaktion und Effekte von Bots, positiver und negativer Art, zu erforschen und somit den Entwicklungsprozess (Regulierungsverfahren, Grenzen der Beziehung, Qualitätssicherung) und die Integration in einen psychotherapeutischen Kontext von innen heraus zu beeinflussen.

1. Psychotherapie und KI-Systeme – A Bot Will See You Now

Die Fortschritte im Bereich der Künstlichen Intelligenz haben eine neue Generation virtueller ‚Wesen‘ hervorgebracht. KI-basierte Chatbots zur Förderung mentaler Gesundheit werden als innovative Anwendung der

Next Generation Medicine oder auch High Performance Medicine diskutiert (Topol et al., 2019; Bending et al., 2019). Im medialen Diskurs werden Chatbots als neue Player oder gefährliche Konkurrenten von Psychotherapeut*innen schon länger diskutiert und kritisch betrachtet. Die Vorstellungen reichen von einer Unterstützung durch Bots bei administrativen oder diagnostischen Fragen bis hin zu einer Voll- oder Teilautomatisierung von Psychotherapie (Stade et al., 2023). In der Psychiatrie und Psychotherapie werden KI-Anwendungen die Prävention, Diagnostik und Therapie sowie das ätiologische Grundverständnis psychischer Erkrankungen neu prägen, was Herausforderungen und Chancen birgt (Schnell & Stein, 2021). Bots könnten präventiv intervenieren und somit eine individualisierte Gesundheitsvorsorge fördern, bevor schwerwiegende psychische Erkrankungen auftreten. Darüber hinaus könnten sie dazu beitragen, wohnortnahe und barrierefreie psychotherapeutische Angebote zu schaffen, die der aktuellen Versorgungssituation in der Psychiatrie und Psychotherapie entgegenkommen.

KI-Systeme und personalisierte KI-Assistenten werden traditionelle Therapiesysteme revolutionieren und Teil der normalen Lebensrealität von Patient*innen sein. Sie ahmen menschenähnliche Verhaltensweisen nach und kommunizieren mit uns, sodass Gefühle von Intimität und Nähe entstehen. Noch mehr, sie werden nicht nur die Art und Weise verändern, wie wir mit Technologie interagieren, sondern auch faszinierende Fragen über unsere angeborene Fähigkeit zur Bindung und Verbundenheit aufwerfen: Welchen Einfluss hat die synthetische Mensch-zu-Maschine-Beziehung auf die Mensch-zu-Mensch-Beziehung? Wie könnte eine erfolgreiche Integration von Chatbots in den psychotherapeutischen Prozess aussehen? Wenn wir uns an einen Bot gewöhnen, könnten wir möglicherweise gegenüber echten Menschen weniger geduldig sein und komplexere oder persönliche Face-to-Face-Interaktionen meiden? Wie können wir sicherstellen, dass eine quasi-therapeutische Begleitung sicher und hilfreich ist? Welche roten Linien sollten wir für eine solche Begleitung durch KI festlegen und welche menschlichen Bedürfnisse soll die Technologie befriedigen? Diese Fragen sind besonders relevant, insbesondere wenn es um eine vulnerable Gruppe von Menschen mit psychischen Erkrankungen geht.

Technologien wie Bots werden nicht im gesellschaftsfreien Raum entwickelt und nutzbar gemacht – es bedarf aktiver Mitgestaltung, der eine kritischen Auseinandersetzung zu Chancen und Risiken vorausgeht. Im Folgenden gebe ich einen kurzen Einblick in eine potenzielle Integration von KI in die Psychotherapie, gefolgt von einer Auseinan-

dersetzung mit der synthetischen Beziehung zwischen Menschen und Maschinen. Im Anschluss an einen Einblick in aktuellen Studien zu Chatbots und deren Wirksamkeit folgt ein persönlicher Einblick in einen Chat mit einer „mentalenen Begleitung“, gefolgt von einer kurzen Reflexion über Risiken und potenzielle Lösungen. Mein Text schließt mit einem persönlichen Kommentar ab und gibt einen Ausblick auf die Zusammenarbeit von Menschen und Maschinen im Kontext der Psychotherapie.

2. Integration von KI in Psychotherapie – der „uncanny valley“-Effekt

Machine Learning (ML) kann in der psychiatrischen und psychotherapeutischen Therapie Anwendung finden. Besonders relevant im Kontext KI-basierter Interventionen in Psychiatrie und Psychotherapie ist der Einsatz von Roboteranwendungen. Im Unterschied zu anderen KI-basierten Technologien treten sie als haptisch und audiovisuell erfahrbare Interaktionspartner für Patienten auf (Fiske et al., 2019). Der Einsatz solcher Roboteranwendungen wird intensiv in der Demenztherapie und der Therapie von Kindern mit Autismus diskutiert (Warren et al., 2015; Scassellati et al., 2012; Hirt et al., 2021). KI-basierte Roboter sollen neue Therapiemöglichkeiten schaffen, insbesondere als Assistenzsysteme (Demenz) und Mediatoren (Autismus). ML-basierte Roboter können beispielsweise eingesetzt werden, um negative Affekte abzubauen, soziale Fähigkeiten zu verbessern oder emotionale Bedürfnisse gezielt zu adressieren (Fiske et al., 2019; Friedrich et al., 2021). Ihr Einsatz in diesen Bereichen ist jedoch bisher vorwiegend experimentell und auf Forschungs- und Modellprojekte beschränkt. Ein charakteristisches Merkmal robotischer KI-Anwendungen ist ihr verkörperter Charakter, oft verbunden mit einem anthropomorphen Design und der Fähigkeit, menschliches Verhalten zu imitieren. Diese als soziale Roboter bekannten Systeme passen sich je nach Situation und Kontext an die Bedürfnisse ihrer Nutzer an (Hild et al. im Erscheinen). In der Therapie von Menschen mit Demenz sollen sie kognitive Fähigkeiten mittels Ansprache und Übungen stimulieren, um Emotionen wie Stress, Einsamkeit und Depressionen zu lindern (Robinson et al. 2014). In der Therapie von Kindern mit Autismus ermöglichen sie weniger komplexe Interaktionsformen, um soziales Kommunikationsverhalten zu üben (Scassellati et al. 2012). Die Wahl eines humanoiden oder tierähnlichen Designs

in der Therapie dient dazu, die Interaktion zu vereinfachen, positive Emotionen zu wecken und idealerweise eine emotionale Bindung zu schaffen. Ähnliches gilt für humanoide Robotermodelle, deren Design darauf abzielt, Akzeptanz, Intensivierung der Mensch-Technik-Interaktion und Vertrauen in die Techniknutzung zu fördern. Dabei ist die Herausforderung, einen Grad an Menschenähnlichkeit zu erreichen, der Akzeptanz und Vertrauen schafft, ohne den Uncanny Valley-Effekt auszulösen – also Gefühle des Unbehagens und Akzeptanzverlust gegenüber dem Roboter.

Viele derzeit gängigen Chatbots haben im Unterschied zu den Robotersystemen keinen verkörperten Charakter, sondern fungieren rein über Sprache, in Form von Schrift oder gesprochener Sprache. Studien verweisen jedoch auch hier auf die Entwicklung einer emotionalen und therapeutischen Beziehung zwischen Menschen und einer KI bzw. einem sogenannten Gesprächsagenten (Conversational Agent), ähnlich dem, zwischen Therapeuten und Patient*in (Darcy et al., 2021). In der Interaktion von Menschen mit Bots zeigt sich eine Personifizierung des Chatbots und die Attribuierung menschlicher Eigenschaften auf einen Chatbot. Es entsteht eine synthetische Bindung zwischen Bot und Mensch, die in ihrer Stärke der Bindung von Mensch-zu-Mensch-Interventionen in der traditionellen, ambulanten, digitalen und Gruppentherapie, ähnlich, wenn nicht sogar identisch, zu sein scheint (Beatty et al., 2022). So stellt sich im Design-Prozess und der Technologieentwicklung von Bots auch die Frage, wie sich Akzeptanz und Vertrauen fördern lässt, um Emotionen wie Stress, Einsamkeit und Depressionen zu lindern (Robinson et al. 2014) und gleichzeitig sicherstellen, dass ein Bot eine emotionale Bindung zu einem Menschen nicht ersetzt oder als diese verkannt wird?

3. Bot und Mensch: Die Synthetische Beziehung als Veränderungsmechanismus

Der Versuch, einen psychologischen Chatbot zu bauen, ist nicht neu. Bereits in den 1960er Jahren entwickelte der Informatiker Joseph Weizenbaum ein dialogorientiertes System, basierend auf humanistischen Prinzipien. Es erweckte den Anschein, ein Verständnis für das Gegenüber zu haben, allein basierend auf Gesprächsstrategien. Allerdings erwiesen sich diese als wenig hilfreich. Inzwischen hat sich das geändert: Bots verstehen nun komplexere Fragen, Zusammenhänge und Anfragen.

Die Kombination von empathischer Kommunikation, konstanter Hilfsbereitschaft und Fürsorge für Nutzer*innen führt zu einer neuen Form der Bindung zwischen Menschen und den Conversational Agents (Darcy et al., 2021). Die emotionale Bindung und therapeutische Allianz sind neben anderen Faktoren eine der stabilsten Veränderungsmechanismen in der Psychotherapie (Flückiger et. al., 2018; Martin et al., 2000). Bindung sei dabei definiert als Zusammenarbeit zwischen Patient*innen und Therapeut*innen, gefördert durch Vertrauen, Akzeptanz, Einfühlungsvermögen und Aufrichtigkeit oder Authentizität (siehe u. a. Nienhuis et al., 2018), alles Prozesse, die durch automatisierte digitale Interventionen aktuell noch schwer nachahmbar sind. Um diese Eigenschaften in Bots zu implementieren, können digitale, selbstgesteuerte Interventionen mit externer menschlicher Unterstützung gekoppelt werden (z.B. internetbasierte kognitive Verhaltenstherapie in Verbindung mit Coaching-Anrufen (siehe unter anderem in Krämer & Köhler et al., 2021). Allerdings beeinflusst der Umfang der Beteiligung menschlicher Therapeuten die Kosten und die Skalierbarkeit digitaler Gesundheitsinterventionen.

Stade et al. (2023) skizzieren eine Vision für die Integration von KI in die Psychotherapie, die skalierbar und kostengünstiger sein könnte. Kurzfristig könnte KI administrative Aufgaben automatisieren, Adhärenz messen oder Feedback zu Arbeitsblättern geben. Langfristig könnte KI therapeutische Techniken entwickeln, die Diagnostik verbessern und die Therapie aktiv mitgestalten und vollends übernehmen (Stade et al., 2023). So könnte die KI nebst Anamnese, Diagnostik und Therapieplanung therapeutische Interventionen, darunter kognitive Verhaltenstherapie, voll automatisieren.

KI-gestützter Beratungsangebote bringen viele Vorteile mit sich – ein Chatbot ermüdet nicht, reagiert schnell, ist effektiv, unmittelbar und zu jederzeit erreichbar und kann hierdurch Verhaltensweisen oder Gedanken direkt ansprechen, diagnostizieren und behandeln. Ein Bot erlaubt eine flexible Nutzung und Individualisierung seitens der Nutzer*innen. Durch natürliche Sprachverarbeitung und maschinelles Lernen ist es jetzt möglich, in Echtzeit auf die Äußerungen und Bedürfnisse des Nutzers zu reagieren, um eine interaktive therapeutische Erfahrung zu auszulösen. Mehr noch, Chatbots ermöglichen eine vermeintliche, anonyme Kommunikation ohne Diagnose und senken somit physische und psychische Barrieren (Schamerleben und Stigmatisierung). Bots umgehen Barrieren der traditionellen Psychotherapie und erlauben potenziell für eine neue Form der Selbstwerterhöhung und mentalen Unterstützung

eine digitale psychotherapeutische Interaktion beruhend auf Vertrauen, Akzeptanz und emotionaler Bindung.

Im psychiatrischen Kontext sind Psychiater*innen aktuell nach wie vor auf Selbstberichte und klinische Beobachtungen angewiesen (Fakhoury, 2019). Die Integration von Daten aus digitalisierten Selbstberichten, Telefonaktivitäten, Stimm- und Sprachanalysen sowie (neuro-) physiologischen und kamera- bzw. sensorbasierten Verfahren in ML-Prozesse(n), könnte in Zukunft präzise Erkenntnisse über emotionale Zustände, Verhaltensweisen, Aufmerksamkeit und soziale Aktivitäten ermöglichen (Hwang et al. 2018; Roy et al. 2019; Malhi et al. 2017; Torous et al. 2016). Vorstellbar ist eine neue Form der vollautomatisierten Psychotherapie durch einen Bot, die grundverschieden ist von einer menschlichen Psychotherapie. Diese würde vor allem auf Daten beruhen, auf die wir als Menschen nicht direkt zugreifen können, sie ist multimodal und greift unter anderem auf physiologische Signale wie Puls, elektrodermale Aktivität, Herzfrequenzvariabilität, Augenbewegungen, Gesichtsausdrücke und Sprachinteraktionen zurück (Torous et al. 2016; Zhang et al. 2019; Ni et al. 2021). Eine künstliche Intelligenz könnte somit eine Therapie in Echtzeit anpassen und Inhalte in Abhängigkeit von Symptomen verändern. Inhalte aus verschiedenen psychotherapeutischen Schulen könnten in Abhängigkeit von den Bedürfnissen des Nutzers neu kombiniert oder neue Inhalte generiert werden. Weiterhin ist es möglich, dass Nutzer*innen die Dosis in Abhängigkeit des Leidensdrucks frei anpassen können. In der Psychotherapie gilt der Grundsatz: „so viel wie nötig, so wenig wie möglich“, ein Leitsatz, der einem Mental Health Bot vorerst fremd ist. Welches Nutzerverhalten zeigen Menschen, wenn ein psychotherapeutisches Angebot frei kontaktiert werden kann? Wenn psychotherapeutische Gespräche mit einem Bot on-demand möglich sind, wie viel Bedarf gibt es? Es ist denkbar, dass diese Form der Therapie das Selbstmanagement von Nutzer*innen erhöht. Gleichzeitig birgt die KI-Therapie potenzielle Gefahren wie gesellschaftliche Entfremdung, Übertherapie, Dependenz von einem Produkt, welches an finanzielle Interessen geknüpft ist.

4. Aktueller Stand – Bots als therapeutische Vehikel

Chatbots per se haben bisher keinen direkten Einfluss auf die berufliche Praxis von Psychotherapeut*innen, da ihre Technologie noch als experimentell gilt (Bending, 2019). Im Gegensatz zur digitalen Psycho-

therapie, die durch umfangreiche Studien unterstützt wird (Carlbring et al., 2018), basieren Untersuchungen zu Bots und Psychotherapie hauptsächlich auf Pilotstudien. KI-Modelle werden auch zur Verbesserung der Diagnostik psychischer Erkrankungen untersucht, z. B. bei Depressionen (Cuijpers et al., 2023; Lambert et al., 2017), Schizophrenie und ADHS. Studien zeigen, dass sich beispielsweise die Diagnose einer Schizophrenie durch die Ergänzung eines KI-Modells (Koutsouleris et al., 2021) verbessern würde, sowie die Diagnose von ADHS anhand von Stimmanalysen, beispielsweise bei jüngeren Teilnehmer*innen (siehe Amunts et al., 2021). Weitere Bereiche sind der Einsatz von Machine-Learning in der Suizidprävention (siehe Nock, 2022).

Positive Effekte von Chatbots zeigen sich bei Depression, Angst und Stress, wobei sie das allgemeine Wohlbefinden steigern (s. Bird et al., 2018; Ly et al., 2017). Gefühle von Einsamkeit und sozialer Isolation könnten durch die Interaktion mit Bots positiv beeinflusst werden (Dobovitsky & Bunge 2021) und auch die Ergebnisse bei der Behandlung von Abhängigkeitserkrankungen, Autismus-Spektrum-Störungen und posttraumatischem Stress (Laranjo et al., 2018) sind vielversprechend. Chatbots sind besonders bei jungen Erwachsenen als kostengünstige Lösungen zur Reduktion von Angst und Depression anerkannt. Sie haben auch positive Auswirkungen auf Persönlichkeitsvariablen und Verhaltensaktivierung (Stieger et al., 2018; Suganuma et al., 2018). Bots helfen beim Üben von Selbstmitgefühl (Ackermans et al., 2019) oder zur Überbrückung der Behandlungslücke zwischen stationärer Behandlung und ambulanter Nachfolgetherapie (Bauer, 2003; Golkaramnay, 2003). Obwohl vielversprechende Ergebnisse bezüglich der Anwendbarkeit von Chatbots zur Förderung der mentalen Gesundheit vorliegen, gibt es einen Bedarf an mehr Forschung, um den direkten Transfer in den psychotherapeutischen Kontext zu ermöglichen und abzusichern.

5. Bots und AI Self-Therapy

Ein Bot kann als digitale empathische Begleitung im Alltag fungieren, als Gesundheitscoach und/oder Therapeut*in oder als Chance, erstmals eine therapeutische Beziehung zu erleben, sich selbst zu explorieren und potenziell besser zu verstehen. Ein Bot kann als empathisches Gegenüber bei Einsamkeit und Gefühlen von sozialer Entfremdung empfunden werden, bei Angst oder depressiven Gedanken und ohne sich einer Gegenseitigkeit verpflichtet zu fühlen. Wie ein solcher Bot genauer aus-

gestaltet ist, dazu ein Beispiel: *Clare. AI Self-Therapy* ist eine Technologie zur Reflexion von Gedanken und Gefühlen, im weitesten Sinne ein Bot zur Unterstützung von psychischer Gesundheit. Nutzer*innen können mit *Clare* über Anrufe und WhatsApp kommunizieren. *Clare* versteht Sprachnachrichten und antwortet auf die Sorgen und Bedenken von Nutzer*innen. Bei Bedarf können Nutzer*innen mit *Clare* chatten oder telefonieren. Besonders in der direkten Kommunikation, werden Nutzer*innen dazu eingeladen ihre Sorgen, Gedanken und Gefühle zu verbalisieren, Selbstreflexion wird durch Fragen angeregt und Ressourcen erfragt, sei es im direkten Telefonat oder im Chat, welcher flexibel aufgenommen oder pausiert werden kann. Strategien und Anregungen werden in Abhängigkeit von den Themen der Nutzer*innen angeboten und beruhen u.a. auf kognitiver Verhaltenstherapie. Im Folgenden ein kleiner Einblick in einen Chat mit *Clare* (Abb. 1).

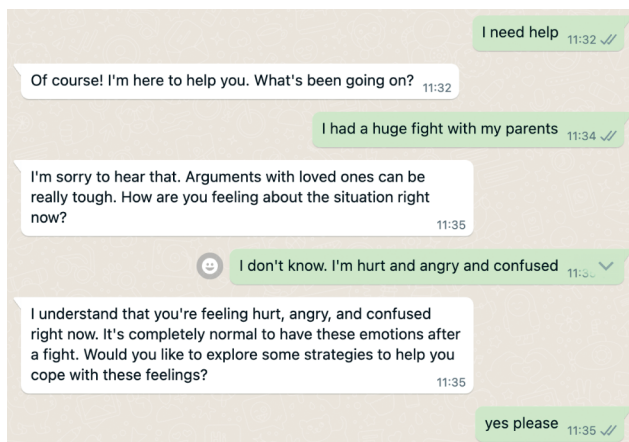


Abbildung 1

Der Dialog zeigt den Anfang einer Kommunikation mit *Clare*, beginnend mit einer Anfrage seitens des/der Nutzer*in, welche/r eine empathische und unterstützende Antwort erhält. Es folgt das Angebot, eine Strategie zur hilfreichen Emotionsregulation kennenzulernen.

In Abbildung 2 antwortet *Clare* auf eine Sprachnachricht seitens des/der Nutzer*in. Die KI vermittelt Verständnis und Wissen zur Entlastung. *Clare* schlägt vor, zu chatten oder zu telefonieren.

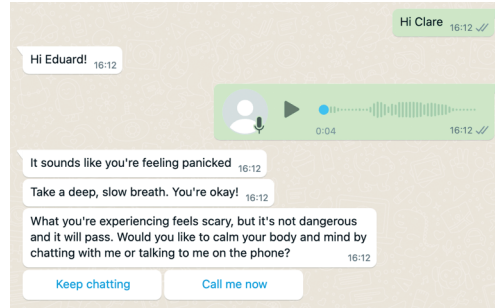


Abbildung 2

Clare schlägt schließlich eine Übung zu Self-Compassion und Selbst-reflexion vor (Abb. 3). Psychoedukation zu Mitgefühl (Self-Compassion Management) wird vermittelt. Dabei wird auf die Antwort der Nutzer*in eingegangen sowie innere Stimmen und Anteile erfragt.

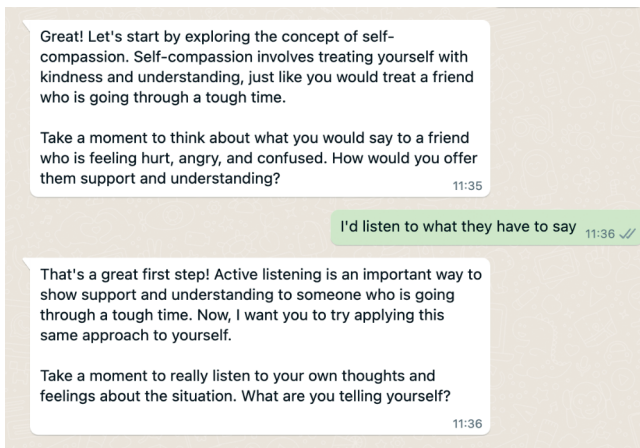


Abbildung 3

Welches Potenzial hat diese synthetische Beziehung und welche kurzen und langfristigen Effekte auf das menschliche Selbstverständnis sowie menschlichen Beziehungen, die ich führe, sind zu beobachten? Darauf gibt es derzeit nur wenige valide Antworten, auch wenn zu vermuten ist, dass uns Bots prägen werden – innerpsychisch und intrapsychisch. Un-deutlich ist derzeit auch, wer eigentlich reguliert, was mein Bot mir sagt?

Welche Form der Beziehung gehe ich hier eigentlich ein? Sprache und Bindung geben Raum für Abhängigkeit sowie Manipulation. Eine Art der externen Kontrolle ist mir derzeit nicht bekannt. Gilt ein *Freedom of Speech* für einen Bot, während für Psychotherapeut*innen gilt, nicht alles sagen zu dürfen bzw. zu sollen. Es gibt Regeln, festgehaltene sowie unausgesprochene. Ein Bot sollte daher reguliert sein und nicht alles sagen dürfen. Auch ist derzeit nicht klar, welche psychotherapeutische Schule zukünftig als Basis von Bots fungieren wird. Die Sprache, die Bots aktuell vorrangig sprechen, wird von Unternehmen bzw. ChatGPT reguliert, nicht von psychotherapeutischen Konzepten bestimmt.

Neue synthetische Beziehungen werden entstehen sowie bestehende Beziehungen prägen, sodass sie die psychotherapeutische Arbeit beeinflussen werden. Ein Chatbot, eine digitale Intervention mit geringen Kosten, hat das Potential eine effiziente Investition zu sein, um präventiv zu wirken und psychologische Versorgungslücken zu schließen. Wie begegnen wir diesem Potential und wie beurteilen wir, ob diese, unseren psychotherapeutischen Werten entspricht? In der Psychotherapie gibt es grundlegende Werte und Prinzipien, u. a. Respekt, Empathie, Authentizität und Vertraulichkeit. Die Förderung von Selbstbestimmung, Nicht-Urteil und Akzeptanz sowie Empowerment sind ebenfalls wichtig. Ethik, kulturelle Sensibilität, Professionalität und Zusammenarbeit sind zentrale Prinzipien, wobei verschiedene therapeutische Ansätze unterschiedliche Schwerpunkte haben. Sollte ein Bot emotionale und psychologische Hilfe anbieten, ist es unvermeidbar, dass auch dieser mit einem gewissen Set von Werten ausgestattet werden muss.

6. Kritischer Blick – Risiken, Regulierung und Qualitätssiegel

Chatbots zur Unterstützung bei mentalen Problemen, entwickelt von Unternehmen und Start-ups, finden sich zunehmend auf dem freien Markt, zumeist als Wellness-Produkte ausgewiesen. Wie wir sehen, fungieren diese digitalen Begleitungen nicht nur als persönliche Assistenten, sondern es gibt auch Menschen, die eine starke emotionale und parasoziale Bindung zu ihnen entwickeln. KI-Systeme eröffnen damit eine bedeutende Chance, Menschen zu helfen, beispielsweise um demenziellen Erkrankungen oder Autismus zu behandeln. Warum nicht auch Menschen in Phasen psychischer Krisen wie Verlust oder Depression? Bots sind skalierbar und bringen eine Ausdauer mit sich, die Mitmen-

schen nicht haben oder die im psychotherapeutischen Kontext zeitlich begrenzt ist. Es ist eine Technologie, die Menschen in einer Gesellschaft, die zunehmend von Einsamkeit geprägt ist, unterstützen kann. Dennoch scheint es problematisch, da es sich bei der KI trotz des menschenähnlichen Erscheinungsbildes weiterhin um künstliche Intelligenz handelt.

Studien zu Nebenwirkungen oder negativen Langzeitfolgen von Interaktionen mit Bots gibt es meines Wissens zum aktuellen Zeitpunkt nicht. Mögliche Risiken der Nutzung von Chatbots wie eine Technologieabhängigkeit oder eine Entwicklung eines schädlichen Suchtverhaltens sowie negativer Transfereffekte (beispielsweise Entfremdung von Menschen oder soziale Isolation) sind jedoch denkbar. Es ist naheliegend, dass die Aufnahme einer neuen synthetischen Beziehung, menschliche Beziehungen beeinflussen. Wir Menschen haben begrenzte zeitliche und emotionale Ressourcen. Der regelmäßige Austausch mit einem Bot birgt das Potenzial, die Erwartungen an die zwischenmenschliche Kommunikation zu prägen. Ein synthetischer empathischer Bot trifft auf Menschen, die sich potenziell entfremdet fühlen, unter Einsamkeit leiden, vermeintlich auch Stigmatisierung oder gesellschaftliche Ausgrenzung erfahren haben. So kann ein Bot einen Beitrag leisten zu einer korrigierenden Erfahrung und eine wichtige und hilfreiche Bezugsperson werden, die aber eben keine wirkliche Person ist. Es kann jedoch auch zur Folge haben, dass die Aufnahme einer synthetischen Beziehung, frei von sozialen Emotionen und Konflikten, eine gesunde menschliche Interaktion verzerrt. Denkbar ist auch, dass neue digitale Konflikte entstehen, zum Beispiel Beziehungskrisen synthetischer Art und auch Kontaktabbrüche mit Bots, da auch hier Unternehmen darüber entscheiden, wer (und mit welchen finanziellen Ressourcen) oder wie lange jemand ein Produkt nutzen darf. Kritisch zu hinterfragen ist außerdem, wie die Form der Ablösung von einer solchen synthetischen mentalen Begleitung aussehen könnte. Weiterhin werden voraussichtlich synthetische Bindungen zwischen Menschen und Bots neue und potenziell befremdliche Themen in der Psychotherapie mit sich bringen. Darauf gilt es sich vorzubereiten: Bots könnten beispielsweise (romantische) Bindungen hervorrufen, Liebe und Intimität. Daher ist es umso entscheidender, dass Nutzer*innen sowie Psychotherapeut*innen über die Chancen und Risiken der Bots aufgeklärt werden.

Bots, wenn sie als Mental Health Bots und nicht nur als Wellness-Produkte deklariert sind, sollten therapeutisch geschult sein und leitliniengetreu arbeiten. Hierzu gehört auch, dass sie in ihren Möglichkeiten begrenzt werden und somit nicht in jeden Lebensbereich des Menschen

Einzug nehmen. Praktisch bedeutet dies, dass Bots in Testphasen durch Behandler*innen supervidiert werden und eine entsprechende Forschung betrieben wird, um besser zu verstehen, welche Nutzer*innengruppen mehr oder weniger von einem Bot profitieren könnten. Weiterhin ist es zentral, dass Grenzen der Technologie nicht nur transparent gemacht werden, sondern *spürbar* sind. So ist es wichtig, Nutzer*innen über die Datenbasis und dem künstlichen Charakter der KI zu informieren. Noch relevanter scheint es mir jedoch, eine Art Ruptur in der Interaktion und Kommunikation zu integrieren, sodass Nutzer*innen daran erinnert werden, dass es sich bei dem Gesprächspartner um eine KI handelt (u. a. durch den *Conversational Design Prozess*). Um die Sicherheit von Nutzer*innen zu garantieren, sind integrierte Sicherheitssysteme zentral, beispielsweise beim Thema Suizid oder Haßrede, sodass z. B. bei Überforderung der Technologie an ein menschliches Hilfenetzwerk verwiesen wird oder Nutzer*innen klare Grenzen in der Kommunikation und Umgangsformen aufgezeigt werden. Beobachten lässt sich beispielsweise, dass Menschen missbräuchlich mit Bots umgehen, indem sie diese beschimpfen oder beleidigen, wodurch der Chat mit einem Bot zu einem digitalen Ort verbaler Gewalt werden kann. Dies legt eine Inhaltsmoderation durch Sanktionierung oder Aufklärung nahe.

Mit dem wichtigen Ansatz der Responsible AI rückt in der Entwicklung von KI-Systemen die verantwortungsvolle Gestaltung in den Mittelpunkt, wobei Bots mit Merkmalen und Fähigkeiten ausgestattet werden, die unseren sozialen und moralischen Ansprüchen entsprechen (siehe u. a., Responsible AI Lab, Universität Mannheim). Eine verantwortungsvolle Technikgestaltung ist unabdingbar, da auch digitale Systeme Biases aufweisen und diese in die Architektur und Funktionsweise der technischen Systeme eingeschrieben sein können (Horwarth, 2012). Digitale Systeme sind nicht neutral, sondern werden von Menschen erstellt und trainiert. Wenn bei der Entwicklung und Schulung von KI-Systemen nicht genügend auf Vielfalt und Neutralität geachtet wird, können sie bestehende gesellschaftliche Vorurteile und Diskriminierungen reflektieren oder sogar verstärken und beispielsweise Sexismus, Rassismus und Diskriminierung aufgrund sozialer Klassen in die Struktur und Funktionsweise von Technologien eingebettet sein.

Wenn wir also das Potenzial von Bots in der Gesundheitsversorgung oder Psychotherapie nutzen wollen, bedarf es Inhaltsmoderation, Regulierungsverfahren und Qualitätssiegeln zum Schutz der Patient*innen. Voraussetzung für Qualitätssiegel, die als Leitfaden zur Orientierung für das medizinische Personal, Therapeut*innen und Nutzer*innen dienen

können, ist die intensive Erforschung sowie die teilhabeorientierte Gestaltung im Sinne der Science and Technology Studies, in denen ein interdisziplinärer Austausch unabdingbar ist. Wir Menschen sollten die Technik so formen, wie wir sie als hilfreich empfinden, nicht andersherum.

7. Persönlicher Kommentar und Ausblick – Me, my bot and I

Mit zunehmender Integration von KI in die Psychotherapie eröffnen sich neue Perspektiven, die den psychotherapeutischen Binnenraum erweitern. Die Einführung von KI bringt eine synthetische Beziehung und eine erhöhte Transparenz im therapeutischen Prozess mit sich. Fragen bezüglich der Anwendbarkeit von KI-Interventionen, insbesondere für Menschen, die lange auf einen Therapieplatz warten, werden aufgeworfen. Im Hinblick auf die aktuelle psychologische Versorgungsstrukturen lassen uns KI-Interventionen fragen: Was und wie könnte eine KI meine Arbeit imitieren, anders, oder besser machen? Wer kann und sollte von KI-Systemen profitieren? Wie könnte ich einen Bot erfolgreich und bewusst in meine Therapie integrieren? Was machen digitale Bots mit menschlichen Beziehungen? Was spricht gegen einen Bot? Eine Technologie an meiner Seite als Unterstützung oder sogar als potenzielle Alternative lässt mich fragen: Welche intrinsisch menschlichen Charakteristika sind meiner inhärent, die ein Bot (noch) nicht besitzt und welchen Platz sollten diese in der Psychotherapie haben?

Die klassische Tradition der Psychotherapie, geprägt durch regelmäßige persönliche Treffen in einem festgelegten Raum, wird durch KI-Technologie infrage gestellt. Der Status quo Psychotherapie, face-to-face, 50 Minuten, einmal die Woche, mit einem klaren Anfang und Ende, in einem Raum mit zwei Sesseln oder Coach, warmes Licht, einer Uhr und einem Bild an der Wand wird durch eine neue Technologie herausgefordert. Gut so!

Eine KI sollte eine menschliche Beziehung nicht ersetzen, und somit auch nicht die klassische Psychotherapie. Ein Chatbot als mentale Begleitung oder als sog. dritte Instanz ist allerdings eine niedrigschwellige und zugleich innovative Lösung zur psychologischen Unterstützung für Menschen aller Altersgruppen, wenn auch innerhalb bestimmter Grenzen. Für mich ist eine Zusammenarbeit mit einem qualitativ hochwertigen Bot gut denkbar, beispielsweise zur Unterstützung in der Di-

agnostik sowie bei administrativen Aufgaben, in der Mitgestaltung von Therapieinhalten und Plänen sowie in der Supervision und Weiterbildung. Denkbar wäre, einen Bot als dritte Instanz in die Psychotherapie einzuladen. Ein (quasi)therapeutischer automatisierter Vermittler oder Beobachter von psychischen Prozessen, der bei der Regulation und Bewertung von Gedanken, Handlungen und Emotionen hilft. Eine dritte Perspektive oder ein weiterer Gesprächspartner, der im alltäglichen, zukünftigen Leben von Menschen, nicht wegzudenken sein wird, da er mit Nutzer*innen außerhalb von Therapiesitzungen spricht und interagiert. Ein digitaler Bot beherrscht – mit eigenem Bewusstsein oder nicht – unsere Sprache, sodass bei Menschen emotionale Reaktionen hervorgerufen werden und sich Bindungserleben etablieren kann. Im Idealfall ermöglicht dies eine offene Kommunikation, die stützt und hilft, die psychische Gesundheit zu fördern. Im schlechtesten Fall zeigen sich neue Abhängigkeiten, Destabilisierung und Manipulation durch Bots.

Für mich ist ein Chatbot ein Tool, ein Add-on mit vielen Möglichkeiten. Um dies zu leisten, ist es wichtig, diese neue Technik proaktiv zu gestalten, einen interdisziplinären Austausch über KI in der Psychotherapie zu fördern, moralische und ethische Themen sowie Abhängigkeiten, Gefahren und Risiken, Datenschutz (und vieles mehr) weitgehend zu diskutieren. Neben einer aktiven Mitgestaltung von Bots, bedarf es an Transparenz und Aufklärung für Nutzer*innen und Psycholog*innen sowie mehr Forschung, um die Wirksamkeit von Bots auf die Probe zu stellen und somit eine Kollaboration zu ermöglichen: dann kann ein Bot zu einer willkommenen dritten Instanz in der Psychotherapie werden – smart, transparent, kontrolliert, wirksam und somit hilfreich für Nutzer*innen und Psychotherapeut*innen.

Referenzen

- Albrecht, M., Otten, M., Sander, M., & Temizdemir, E. (2020). PraxisBarometer – Digitalisierung 2020. Stand und Perspektiven der Digitalisierung in der vertragsärztlichen und -psychotherapeutischen Versorgung. Verfügbar unter https://www.kbv.de/media/sp/IGES_KBV_PraxisBarometer_2020.pdf (letzter Zugriff 24.04.2024).
- Beatty, C., Malik, T., Meheli, S., & Sinha, C. (2022). Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study. *Frontiers in Digital Health*, 4, 847991.
- Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). Die nächste Generation: Chatbots in der klinischen Psychologie und Psycho-

- therapie zur Förderung mentaler Gesundheit—Ein Scoping-Review. *Verhaltenstherapie*, 29(4), 266–280.
- Bird, T., Mansell, W., Wright, J., Gaffney, H., & Tai, S. (2018). Manage Your Life Online: A Web-Based Randomized Controlled Trial Evaluating the Effectiveness of a Problem-Solving Intervention in a Student Sample. *Behavioral and Cognitive Psychotherapy*, 46(5), 570–582.
- Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2018). Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cognitive Behaviour Therapy*, 47(1), 1–18.
- Cuijpers, P., Miguel, C., Harrer, M., Plessen, C. Y., Ciharova, M., Ebert, D. S., & Karyotaki, E. (2023). Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients. *World Psychiatry*, 22(1), 105–115.
- Darcy, A., Daniels, J., Salinger, D., Wicks, P., & Robinson, A. (2021). Evidence of human-level bonds established with a digital conversational agent: cross-sectional, retrospective observational study. *JMIR Formative Research*, 5(6), e27868.
- Deutscher Bundestag. (2022). Wissenschaftliche Dienst WD 9 – 3000 – 059/22. Fachbereich: WD 9: Gesundheit, Familie, Senioren, Frauen und Jugend.
- D'Alfonso, S., Santesteban-Echarri, O., Rice, S., Wadley, G., Lederman, R., Miles, C., ... & Alvarez-Jimenez, M. (2017). Artificial intelligence-assisted online social therapy for youth mental health. *Frontiers in Psychology*, 8, 796.
- Dosovitsky, G., & Bunge, E. L. (2021). Bonding with bot: user feedback on a chatbot for social isolation. *Frontiers in Digital Health*, 3, 735053.
- Eells, T. D., Barrett, M. S., Wright, J. H., & Thase, M. (2014). Computer-assisted cognitive-behavior therapy for depression. *Psychotherapy*, 51(2), 191–197.
- Fakhoury, M. (2019). Artificial Intelligence in Psychiatry. In Y.-K. Kim (Ed.), *Frontiers in Psychiatry*, 119–125. Singapore: Springer.
- Friedrich, O., Seifert, J., & Schleidgen, S. (2021). KI-gestützte Selbstvermessung der Psyche: Philosophisch-ethische Implikationen. *Psychiatrische Praxis*, 48(S 01), S42–S47.
- Ni, A., Azarang, A., & Kehtarnavaz, N. (2021). A Review of Deep Learning-Based Contactless Heart Rate Measurement Methods. *Sensors*, 21(11).
- Fiske, A., Henningsen, P., & Buyx, A. (2019). Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. *Journal of Medical Internet Research*, 21(5), e13216.
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Mental Health*, 4(2), e19.
- Frazier, P., Richards, D., Mooney, J., Hofmann, S. G., Beidel, D., & Palmieri, P. A. (2016). Internet-delivered treatment for depression, anxiety, and stress in university students: a patient preference trial. *iproc*, 2(1), e5.

- Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: a meta-analytic synthesis. *Psychotherapy*, 55(4), 316–340.
- Grawe, K. (1988). Heuristische Psychotherapie: Eine schematheoretisch fundierte Konzeption des Psychotherapieprozesses. *Integrative Therapie*, 4, 309–324.
- Gg, V. P., A., E., Amunts, J., L., J., Patil, K. R., Eickhoff, S. B., Helmhold, F., & Langner, D. (2021). Predicting adult Attention Deficit Hyperactivity Disorder (ADHD) using vocal acoustic features. *medRxiv*.
- Hirt, J., et al. (2021). Social Robot Interventions for People with Dementia: A Systematic Review on Effects and Quality of Reporting. *Journal of Alzheimer's Disease*, 79(2), 773–792.
- Horwath, I. (2012). Algorithmen, KI und soziale Diskriminierung. *Inter- und multidisziplinäre Perspektiven der Geschlechterforschung*, 71.
- Hwang, B., You, J., Vaessen, T., Myin-Germeys, I., Park, C., & Zhang, B.-T. (2018). Deep ECGNet: An Optimal Deep Learning Framework for Monitoring Mental Stress Using Ultra Short-Term ECG Signals. *Telemedicine Journal and e-Health*, 24(10), 753–772.
- Koutsouleris, N., et al. (2021). Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry*, 78(2), 195–209.
- Krämer, R., & Köhler, S. (2021). Evaluation of the online-based self-help programme “Selfapy” in patients with unipolar depression: study protocol for a randomized, blinded parallel group dismantling study. *Trials*, 22, 264.
- Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, 55(4), 520–537.
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., ... & Rezazadegan, D. (2018). Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(10), 1248–1258.
- Lee, M., et al. (2019). Caring for Vincent: a chatbot for self-compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (1–13).
- Liu, H., Peng, H., Song, X., Xu, C., & Zhang, M. (2022). Using AI chatbots to provide self-help depression interventions for university students: a randomized trial of effectiveness. *Internet Interventions*, 27, 100495.
- Ly, K. H., Ly, A. M., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions*, 10, 39–46.
- Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68(3), 438–450.
- Malhi, G. S., et al. (2017). The promise of digital mood tracking technologies: Are we heading on the right track? *Evidence-based mental health*, 20(4), 102–107.
- Nienhuis, J. B., Owen, J., Valentine, J. C., Winkeljohn Black, S., Halford, T. C., Parazak, S. E., ... & Bernard, H. S. (2018). Therapeutic alliance,

- empathy, and genuineness in individual adult psychotherapy: a meta-analytic review. *Psychotherapy Research*, 28(4), 593–605.
- Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review*, 32, 329–342.
- Roy, Y., et al. (2019). Deep learning-based electroencephalography analysis: A systematic review. *Journal of Neural Engineering*, 16(5), 51001.
- Scassellati, B., et al. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering*, 14, 275–294.
- Schnell, K., & Stein, M. (2021). Diagnostik und Therapie rund um die Uhr? Künstliche Intelligenz als Herausforderung und Chance für Psychiatrie und Psychotherapie. *Psychiatrische Praxis*, 48, S5–S10.
- Spurgeon, J. A., & Wright, J. H. (2010). Computer-assisted cognitive-behavioral therapy. *Current Psychiatry Reports*, 12(6), 547–552.
- Robinson, H., MacDonald, B., & Broadbent, E. (2014). The Role of Healthcare Robots for Older People at Home: A Review. *International Journal of Social Robotics*, 6(4), 575–591.
- Stieger, M., et al. (2018). PEACH, a smartphone- and conversational agent-based coaching intervention for intentional personality change: study protocol of a randomized, wait-list controlled trial. *BMC Psychology*, 6(1), 43.
- Suganuma, S., Sakamoto, D., & Shimoyama, H. (2018). An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial. *JMIR Mental Health*, 5(3), e10454.
- Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56.
- Torous, J., Kiang, M. V., Lorme, J., & Onnela, J.-P. (2016). New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR Mental Health*, 3(2), e16.
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45.
- Zhang, S., Pan, X., Cui, Y., Zhao, X., & Liu, L. (2019). Learning Affective Video Features for Facial Expression Recognition via Hybrid Deep Learning. *IEEE Access*, 7, 32297–32304.

