

# **Bedeutung von Selbstoffenbarungseffekten in quasisozialen Beziehungen mit auf generativer KI basierten Systemen in Settings von Onlineberatung und -therapie**

*Gesa Linnemann, Julian Löhe & Beate Rottkemper*

## **Zusammenfassung**

KI-basierte Chatbots erlangen zunehmend Bedeutung im Bereich der Beratung und Therapie zur psychischen Gesundheit. Dabei sind Expertensysteme und generative Systeme als technische Grundlagen hinsichtlich der möglichen Effekte zu unterscheiden. Hinsichtlich der sozialen Effekte in der Mensch-Maschine-Interaktion beschreibt und definiert der Beitrag das Phänomen der quasisozialen Beziehung. Besonders relevant sind in diesem Zusammenhang Selbstoffenbarungseffekte, d.h. die Preisgabe persönlicher Informationen, die durch (zugeschriebene) Eigenschaften von Chatbots beeinflusst werden können. In der Beratung zeigt sich das im Teilen von selbstbezogenen Informationen mit einem kommunikativen Gegenüber und führt in Beziehungen zum Aufbau von Nähe. Diskutiert werden Effekte, Chancen und Risiken für die Einordnung des Einsatzes solcher Chatbots in Beratung und Therapie.

## **Schlüsselwörter**

Mensch-Maschine-Interaktion, psychische Gesundheit, Künstliche Intelligenz, generative KI, Expertensystem, ChatGPT, KI-Chatbot, quasisoziale Beziehung, Selbstoffenbarung

## **Abstract**

AI-based chatbots are becoming increasingly important in the field of mental health counseling and therapy. A distinction must be made between expert systems and generative systems as a technical basis with regard to the possible effects. With regard to the social effects in human-machine interaction, the article describes and defines the phenomenon of quasi-social relationships. In this context, self-disclosure effects are particularly relevant, i.e. the disclosure of personal information that can be influenced by (attributed) characteristics of chatbots. In counseling, this manifests itself in the sharing of self-related information with a communicative counterpart and leads to the development of closeness in relationships. The effects, opportunities and risks for the classification of the use of such chatbots in counseling and therapy will be discussed.

## **Keywords**

human-machine-interaction, mental health, artificial intelligence, generative AI, expert system, ChatGPT, AI chatbot, quasi-social relationship, self-disclosure

## **Autor\*innen**

### **Prof. Dr. Gesa Linnemann**

- Professorin für Sozialpsychologie / katho Standort Köln
- **Kontakt:** Wörthstraße 10, 50668 Köln / +49 (0)221 7757-271  
[g.linnemann@katho-nrw.de](mailto:g.linnemann@katho-nrw.de)

### **Prof. Dr. Julian Löhe**

- Professor für Organisation und Management in der Sozialen Arbeit
- Leiter Masterstudiengang Sozialmanagement / FH Münster
- **Kontakt:** Hüfferstr. 17, 48149 Münster / +49 (0)251 83-65754  
[loehe@fh-muenster.de](mailto:loehe@fh-muenster.de)

### **Dr. Beate Rottkemper**

- Wirtschaftsinformatikerin / Münster
- **Kontakt:** [beate.rottkemper@mailbox.org](mailto:beate.rottkemper@mailbox.org)

Künstliche Intelligenz (KI) hält in immer mehr Lebensbereichen Einzug. Empfehlungsalgorithmen, Spam-Filter und Routenplaner sind aus dem Alltag vieler Menschen nicht mehr wegzudenken. Da ist es naheliegend, dass digital gestützte Lösungen im Allgemeinen und KI im Speziellen auch im Beratungssetting an Bedeutung gewinnen. Die ständige Verfügbarkeit digital gestützter Lösungen, Wartezeiten für analoge Beratungen sowie Hemmschwellen, sensible Themen in persona auszusprechen, sorgen für eine vermehrte Nutzung von E-Beratungsangeboten (Adamopoulou & Moussiades, 2020; Brunn, Diefenbacher, Courtet & Genieys, 2020). Während hinter klassischen E-Beratungsangeboten Berater\*innen agieren, die u.a. per Chat, Mail oder Video Beratungsleistungen anbieten, sorgen immer bessere Algorithmen, große Datenmengen und steigende Rechenleistungen dafür, dass inzwischen auch rein digitale Angebote entstehen. In Beratung und Therapie erhöht sich vor diesem Hintergrund die Bedeutung des Einsatzes von KI-Chatbots. Chatbots werden in diesem Beitrag als eine spezifische Form künstlicher Agenten [1,2] betrachtet. Hinter diesen Chatbots stehen Methoden wie das Natural Language Processing [3], die eine Kommunikation mittels natürlicher Sprache (bzw. in Schriftform) ermöglichen. Die aktuell in der Diskussion befindlichen Large Language Models (LLM) heben die Kommunikation mit den Bots auf ein neues Qualitätsniveau, sodass von außen immer weniger zu unterscheiden ist, ob mit einem Menschen oder mit einem KI-Chatbot kommuniziert wird.

Einführend werden im Folgenden verschiedene Chatbot-Lösungen vorgestellt, bevor Details der Mensch-Maschine-Interaktion (MMI) beleuchtet werden, um eine Analyse quasisozialer Beziehungen anzuschließen und damit einhergehende Selbstoffenbarungseffekte vergleichend zu analogen Beratungsangeboten zu betrachten. Der Stand der Literatur zeigt, dass KI-Chatbots derzeit besonders im Bereich der psychischen Gesundheit und Therapie eingesetzt werden. Daher konzentriert sich vorliegender Beitrag auf genau diesen Bereich. In anderen Beratungssettings werden spätestens seit Corona vermehrt digitale Angebote geschaffen, jedoch zum Großteil als Werkzeug zur MMI. Darüber hinaus sind die Entwicklungen unter anderem in den USA deutlich weiter fortgeschritten als in Europa, so dass im Folgenden vermehrt internationale KI-Chatbots vorgestellt werden.

## **1. KI - Chatbots in der psychologischen Beratung und Therapie**

Angebote wie Replika.ai sollen einen Ersatz für Gespräche mit Freund\*innen bieten, kein Beratungsangebot darstellen. Es wird damit geworben, dass dieser Chatbot rund um die Uhr verfügbar ist und empathisch auf sein\*e Ersteller\*in reagiert (Brandtzaeg, Skjuve & Følstad, 2022). Umfragen lassen den Schluss zu, dass soziale Chatbots Einfluss auf das mentale Wohlbefinden und die mentale Gesundheit ihrer Nutzer\*innen haben, womit diese KI-Chatbots durchaus in die Richtung von Beratungsleistungen gehen (Pentinaa, Hancock & Xieb, 2023). Eine weitere Entwicklung aus den USA ist der „Woebot“. Dieser Bot basiert auf einem Entscheidungsbaum und wurde unter anderem von Psycholog\*innen entwickelt, um Student\*innen, die unter Angststörungen oder Depressionen leiden, zu unterstützen. Inzwischen gibt es mehrere interne Studien, die die Wirkung des Bots belegen (Fitzpatrick, Darcy & Vierhile, 2017; Prochaska et al., 2021; Durden et al., 2023). Während ein Bot dazu beitragen kann, Wartezeiten bis zum Start

einer Therapie oder Beratung zu überbrücken oder in einem hybriden Ansatz helfen kann, Verhaltensweisen zu trainieren, wird der vollständige Ersatz von menschlichen Hilfsangeboten durch die Anbieter jedoch kritisch gesehen. So wird es vom Bot selbst zum Start der App-Nutzung auch erläutert (Fitzpatrick et al., 2017). Erste Studien zeigen darüber hinaus, dass im Vergleich zu einer rein digitalen Form einer Therapie durch einen hybriden Ansatz die Abbruchquote geringer und die Wirksamkeit erhöht wird (Richards & Richardson, 2012). Ein weiteres Produkt aus den USA heißt „Tess“. Der Chatbot, gemeinsam mit Psycholog\*innen entwickelt, soll ebenfalls unter anderem bei Angststörungen und Depressionen unterstützen. Verschiedene Interventionsstrategien stehen zur Verfügung, um Nutzer\*innen, die sich keine Psychotherapie leisten können oder die noch nicht bereit sind, sich analog Hilfe zu suchen, ad hoc und zu jeder Zeit Hilfe zur Selbsthilfe zu ermöglichen. Darüber hinaus kann der Chatbot auch von Organisationen genutzt und für spezifische Zwecke parametrisiert werden. In den USA gab es erste Use-Cases von Organisationen, in denen der Chatbot zur Bekämpfung von Übergewicht eingesetzt wurde (Fulmer, Joerin, Gentile, Lakerink & Rauws, 2018). Ein anderer Chatbot ist der Vivibot. Im Rahmen einer Studie zur Unterstützung der psychischen Stabilität von jungen Erwachsenen nach einer Krebsbehandlung wurde der Chatbot in den Facebook-Messenger integriert (Greer et al., 2019). Boucher et al. (2021) geben einen Überblick des Einsatzes von Chatbots im Bereich der psychischen Gesundheit nach Problemlagen (Depression, Ängste, Autismus, etc.), der Förderung von Kompetenzen/ Zuständen (Wohlbefinden, Selbstmitgefühl, Achtsamkeit, Lebensqualität) und der Zielgruppe (Kinder, Jugendliche, Erwachsene, Ältere); als häufigste Funktionen identifizieren sie Diagnostik, Darstellung von Inhalten und Symptommanagement. Sedlakova und Trachsel (2023) identifizieren als die beiden übergeordneten Ziele solcher Anwendungen die Unterstützung beim Erlernen von neuen Fähigkeiten und Techniken sowie das Formen einer therapeutischen Allianz mit den Nutzenden. Auch in Laranjo et al. (2018) werden weitere Chatbots aus dem Bereich der psychischen Gesundheit vorgestellt. In Deutschland sind Chatbots zur Unterstützung von digitaler Beratung und Therapie deutlich weniger verbreitet als beispielsweise in den USA. Auch im deutschsprachigen Raum sind Therapieplätze jedoch knapp und digitale Lösungen können potentiell dabei unterstützen, die Wartezeit zur analogen Therapie zu überbrücken sowie die Nachsorge zu begleiten. E-Beratungsangebote, bei denen Fachkräfte mit den Ratsuchenden per E-Mail oder Chat kommunizieren, gibt es auch in Deutschland bereits seit einigen Jahren (Kühne, 2016; Wolf, Zimmer & Dogs, 2008) und auch erste rein digitale Lösungen, wie „somnio“ zur Unterstützung bei Schlafstörungen, weisen in eine ähnliche Richtung. Diese App ist als digitale Gesundheitsanwendung eingetragen und kann somit auf Rezept genutzt werden. Insgesamt gibt es bisher wenige unabhängige Studien zur Wirksamkeit der Chatbots im Bereich der psychischen Gesundheit. Die Entwicklung und der Nutzen von Chatbots wird im Ingenieurwesen und in der Informatik umfangreich untersucht, im Feld der Sozialen Arbeit oder in der Psychologie, bzw. der Medizin bisher jedoch unzureichend (Vaidyam, Wisniewski, Halamka, Kashavan & Torous, 2019).

Die Basis von Replika.ai ist eine generative KI. Keine der weiteren vorgestellten Lösungen nutzt generative KI. Dem Woebot liegt ebenso wie dem Vivibot ein Entscheidungsbaum zugrunde und hinter somnio steht ein klassisches Expertensystem, welches evidenzbasierte Inhalte an die Nutzer\*innen vermittelt.

Expertensysteme reproduzieren Wissen, das ihnen zugeführt wird und lernen regelbasiert aus historischen Daten. Die zugrundeliegende Wissensbasis beinhaltet ein bestimmtes Fachgebiet und das System bearbeitet klar definierte Fragestellungen. Im Gegensatz dazu kann eine generative KI selbst neue Inhalte kreieren. Dies geschieht ebenfalls auf Basis eines historischen Datenstamms. Dieser ist jedoch in der Regel sehr groß und nicht auf ein bestimmtes Fachgebiet beschränkt. Eine generative KI ist darauf trainiert, anhand bestimmter Eingabeparameter (sogenannte prompts) den Inhalt zu erzeugen, der von dem\*der Nutzer\*in am wahrscheinlichsten erwartet wird. Die aktuell bekannteste generative KI ist GPT (Generative Pre-trained Transformer). Während der Einsatz generativer KI im Beratungssetting aufgrund der fehlenden Steuerbarkeit von Aussagen und der Gefahr von Halluzinationen (also Falschaussagen) des Systems aktuell noch keine weite Verbreitung findet, finden Expertensysteme, die natürliche Sprache (bzw. Text) als Kommunikationsmittel nutzen, immer häufiger Anwendung. Erste Entwicklungen von Chatbots zur Verhaltenstherapie auf der Basis von ChatGPT gibt es bereits, diese befinden sich jedoch noch in der Evaluierungsphase (Nazarova, 2023).

Sedlakova und Trachsel (2023) charakterisieren KI-gestützte Chatbots als „hybrid“ zwischen „Werkzeug“ und „Agent“ (im Sinne menschlicher Eigenschaften) und diskutieren ethische Implikationen dieses Zwischenstatus. In diesem Sinne imitieren Chatbots lediglich Gespräche, führen sie aber nicht im gleichen Sinne wie ein menschlicher Agent, der über Bewusstsein und innere Zustände verfügt. Gleichwohl könnten Nutzende Chatbots wie Menschen erleben oder behandeln. Dies berührt auch das Etablieren einer Beziehung. Aspekte und Effekte der Mensch-Maschine-Interaktion in der Beziehungsgestaltung werden im folgenden Teil des Artikels diskutiert. Als zentrales Element wird anschließend die Selbstoffenbarung herausgegriffen.

## **2. Mensch und Maschine im Dialog: Der quasisoziale Beziehungsbegriff in der MMI**

Wenn der Einsatz von KI-Chatbots zunehmend Einzug in Beratung und Therapie hält, ist die Frage zu stellen, wie dieser Einsatz aus einer Professionsperspektive beratender und therapierender Personen betrachtet werden kann. Besonders entscheidend für den Erfolg von Therapie und Beratung ist eine gelingende Beziehungsgestaltung (Löhe & Aldendorff, 2022; Gödde, 2016; Binder & Strupp, 1997; Crits-Christoph, 1998). Deshalb wird im folgenden Kapitel der Beziehungsbegriff in der MMI unter dem Eindruck des Einsatzes von KI-Chatbots diskutiert. In diesem Zusammenhang wird der Begriff der quasisozialen Beziehung aufgegriffen. Ergeben sich im Einsatz von KI-Chatbots quasisoziale Beziehungen, folgt daraus die Frage, welche Effekte sich für die Beratungs- und Therapiesituation ergeben. Es gibt hinreichende Anhaltspunkte dafür, dass einer der Effekte eine erhöhte Selbstoffenbarung von zu Beratenden bzw. zu Therapierenden ist. Was damit genau gemeint ist und was das für Beratung und Therapie bedeutet, wird in Kapitel 3 dargelegt. In einem vorhergehenden Schritt klärt dieses Kapitel, warum die Beziehungsgestaltung für Beratung und Therapie von enormer Bedeutung ist und anhand welcher Merkmale das Vorhandensein einer quasisozialen Beziehung definiert werden kann. Diese Merkmale werden ohne eine Unterscheidung dahingehend besprochen, ob es sich beim Einsatz eines KI-Chatbots um ein

Expertensystem oder eine generative KI handelt. Entscheidend ist vielmehr, wie sich der Einsatz des KI-Chatbots auf die Interaktion auswirkt. Gleichwohl ist davon auszugehen, dass soziale Funktionen öfter und häufiger aktiviert und damit quasisoziale Beziehungen aufgebaut werden, wenn eine generative KI eingesetzt wird (Morkes, Kernal & Nass, 1999; Nass & Yen, 2012). Nicht zuletzt darin liegt ein Grund, warum quasisoziale Beziehungen und ihre Folgen für die Zukunft der Beratung und Therapie bei Einsatz von KI so bedeutsam sind.

## **2.1 Begriffliche Annäherung an die soziale Beziehung aus soziologischer Perspektive**

Im Rahmen der klassischen Beziehungslehre von Georg Simmel (1858-1919) und Leopold Wiese (1876-1969) wird die soziale Beziehung als Tatbestand der Wechselwirkung mehrerer Individuen definiert. Soziale Beziehungen spielen sich im Rahmen formaler Netzwerke, d.h. von Inhalten, Motiven oder historischen Bedingungen unabhängig gedachter Beziehungsgeflechte oder Figurationen im Sinne Norbert Elias (1897–1990) in ihren unterschiedlichsten Formen ab (Kopp, 2018). Auffallend für die Frage der Existenz von sozialen Beziehungen zwischen Menschen und künstlichen Agenten ist, dass soziale Beziehungen als *zwischenmenschliches Geschehen* der Annäherung oder Distanzierung und der Vereinigung oder Trennung verstanden werden. Weber (1864-1920) beschreibt soziale Beziehung als ein „seinem Sinngehalt nach aufeinander gegenseitig eingestelltes und dadurch orientiertes Sichtverhalten“ (Weber, 1922, S. 126), bei dem eine Chance besteht, dass in sinnhaft angebbarer Art sozial – also sinnhaft am Verhalten anderer orientiert – gehandelt wird (Kopp, 2018). Max Weber hat bei „aufeinander gegenseitig“ mutmaßlich nicht an einen KI-Chatbot gedacht, wie seine weiteren Ausführungen nahelegen: Soziale Beziehungen können Inhalte wie beispielsweise Kampf, Feindschaft, Freundschaft, aber auch Liebe haben. Insbesondere Liebesgefühle von einem Menschen zu einem KI-Chatbots sind ein Phänomen, dass es erst seit kurzer Zeit über die Science-Fiction Inszenierungen (z.B. im Film „Her“) herausgeschafft hat, wie zum Beispiel eine Diskussion auf Reddit zeigt (Reddit, 2023) oder ein Fachbuch für den Religionsunterricht, hier wird gefragt: „Was bedeutet Liebe, wenn wir sie zu einer Maschine empfinden? (Wie) werden wir von einer KI geliebt?“ (Gatt, 2022). Auch neuere Ansätze geben keinen konkreten Hinweis auf die Frage von sozialer Beziehung und KI-Chatbots, vielmehr scheint ein menschliches Wesen als Interaktionspartner\*in vorausgesetzt, wenn bei dem Symbolischen Interaktionismus (Berger & Luckmann, 2007) oder auch der Sozialpsychologie (Thibaut & Kelley, 1959) gegenseitige Bezugnahme, Perspektivübernahmen, gegenseitige Beeinflussung von Verhalten und Austauschbeziehungen im Zentrum der Auseinandersetzung stehen (Kopp, 2018).

Bei Betrachtung der Definitionen – vom Hinweis auf ein menschliches Wesen abgesehen – ist festzustellen, dass KI-Chatbots die Aufgaben und Funktionen durchaus auch übernehmen können und es daher nicht verwunderlich erscheint, dass in der Interaktion mit KI-Chatbots soziale Funktionen aktiviert werden (siehe weiter unten). So kann zwischen Menschen und KI-Chatbots eine Wechselwirkung ebenso stattfinden (z.B. unter Eingabe des prompts „Führe einen sokratischen Dialog mit mir“ bei ChatGPT), wie eine gegenseitige Beeinflussung vorhanden sein kann, etwa wenn ein KI-Chatbot Vorschläge für eine Mail verfasst oder auf

Nachfrage Hinweise für das Verhalten in einem Konflikt mit Mitmenschen gibt. Eine gegenseitige Bezugnahme kann gleichfalls stattfinden und das Verhalten kann ebenso gegenseitig beeinflusst werden. Zuletzt hat auch das Phänomen von Liebe und Freundschaft zunehmend Eingang in den wissenschaftlichen Diskurs gefunden (Cave & Dihal, 2021). Es erscheint demnach berechtigt, die folgende Frage zu stellen: Kann die Interaktion zwischen Menschen und KI-Chatbots eine 'echte' soziale Beziehung im soziologischen Sinn darstellen? Für den vorliegenden Beitrag gehen wir davon aus, dass das nicht der Fall ist. Zwar werden soziale Reaktionen hervorgerufen, gleichsam handelt es sich weiterhin um Artefakte. Bendel drückt es treffend aus, wenn er schreibt: „Es gibt gerade viele Diskussionen darüber, ob Maschinen intelligent oder moralisch [oder sozial beziehungsfähig, d. Verf.] sein können. Das sind schlicht und ergreifend die falschen Fragen. Maschinen sind Maschinen. Sie simulieren allenfalls Intelligenz und Moral [und soziale Beziehungen, d. Verf.]. Deshalb spricht man von künstlicher Intelligenz und maschineller Moral oder moralischen Maschinen.“ (Bendel, 2020, S. 4) Gleichwohl wird anhand der näheren Betrachtung des sozialen Beziehungsbegriffs deutlich, dass hier zumindest starke „Verwechslungsgefahr“ besteht – insbesondere, wenn berücksichtigt wird, dass Menschen im Einzelnen möglicherweise nicht immer bewusst ist, ob sie gerade mit einem anderen Menschen oder mit einem KI-Chatbot interagieren.

## **2.2 Die quasisoziale Beziehung zwischen Menschen und einem KI-Chatbot**

Wird davon ausgegangen, dass erfolgreiche Therapie und Beratung eine gelingende Beziehungsgestaltung voraussetzt, ist es daher aus einer professionellen Perspektive von erheblicher Bedeutung, inwiefern sich der Einsatz von KI-Chatbots in Beratung und Therapie auf die Beziehungsgestaltung auswirkt. In diesem Zusammenhang haben sich schon seit längerem Studien und wissenschaftliche Abhandlungen mit der Frage einer sozialen Komponente in der MMI beschäftigt. Yamato, Shinozawa, Naya & Kogure (2001) sprechen von künstlichen Agenten als *soziale Akteure* und Ferrari & Eyssel (2016) von *Sozialen Agenten*. Darüber hinaus findet sich der Begriff *Social Robotics* vermehrt in der Diskussion wieder (Echterhoff, Bohne & Siebler, 2006) [4]. Mit Blick auf personenbezogene Dienstleistung und Beratung gibt es aus dem kommerziellen Bereich auch aktuelle(re) Studien, Jin & Youn (2022) verwenden mit Blick auf KI-Chatbots die Bezeichnung *quasisoziale Akteure*. Wird von einem sozialen Akteur ausgegangen, dann wirkt sich das auch auf die Beziehung in der Interaktion zwischen den Beteiligten aus. Exakt das betonen Linnemann, Löhe & Rottkemper (2023), indem sie den Begriff der *quasisozialen Beziehung* im Zusammenhang mit personenbezogenen Dienstleistungen in der Sozialen Arbeit einführen. Die unterschiedlichen Autor\*innengruppen gehen mit den Begriffen auf das Phänomen ein, wenn soziale Funktionen durch die Interaktion mit einem künstlichen Agenten hervorgerufen werden. Im Unterschied zu den bisherigen Begriffen, die ausschließlich das Vorhandensein von „etwas Sozialem“ anerkennen, wird mit dem Begriff der quasisozialen Beziehung der für die Beratung und Therapie wichtige Beziehungsaspekt in die Beschreibung dessen aufgenommen, was an sozialer Interaktion zwischen künstlichen Agenten und Menschen zu beobachten ist. Mit dem Zusatz „quasi“ – der auch bei anderen Autor\*innen bemüht wird, um auf die Unterschiede zu Beziehungen zwischen Menschen hinzuweisen – wird der künstliche Aspekt (und damit der Hinweis von Bendel: Maschinen sind Maschinen)

berücksichtigt. Die Autor\*innengruppen knüpfen damit an unterschiedlich vorangegangene Studien und Erkenntnisse zur MMI an, die Hinweise darauf liefern, dass künstliche Entitäten zumindest teilweise zu sozialen Effekten führen. Einer der ersten und der prominenteste Ansatz zur Erklärung sozialer Reaktionen ist der media equation-Ansatz von Reeves & Nass (1996). Sie postulieren, dass Menschen Computer und andere künstliche Agenten als soziale Akteure behandeln und bezeichnen dieses Phänomen als media equation. Media equation bedeutet frei übersetzt so viel wie „Gleichsetzung mit den Medien“. Gleichgesetzt werden in diesem Sinne Medien bzw. Technologien mit dem echten Leben, wenn nämlich technologiegestützte Anwendungen wie „echte Menschen“ behandelt werden und soziale Reaktionen erfolgen. Um diese Annahme zu testen, führten Reeves & Nass (1996) Experimente im Rahmen des *Computers are Social Actors* (CASA)-Paradigmas durch, bei dem menschliche Proband\*innen mit Computern interagierten. Darunter wurde in fünf Experimenten nachgewiesen, dass die Interaktionen von Individuen mit Computern grundsätzlich sozial sind. Im Einzelnen zeigten die Experimente, dass die sozialen Reaktionen auf einen Computer nicht das Ergebnis einer bewussten Überzeugung sind, dass der Computer menschlich oder menschenähnlich ist. Die sozialen Reaktionen resultierten auch nicht aus Unwissenheit (alle Proband\*innen wussten, dass sie mit einem Computer agieren) oder einer sozialen oder psychologischen Funktionsstörung der Nutzenden. Sie erfolgten auch nicht aus dem Glauben heraus, sie würden mit einer\*m Programmierer\*in interagieren (Nass, Steuer & Tauber, 1994). Vielmehr sind die Reaktionen der Proband\*innen soziale Reaktionen in der Anerkennung des Computers als Teil des „echten Lebens“. Seit 1994 haben sich künstliche Agenten stark weiterentwickelt und heute sind soziale Reaktionen auf Computer und KI-Chatbots alltäglich(er) und leicht zu erzeugen. Bekannt ist das Phänomen der Anthropomorphisierung technologischer Systeme (Echterhoff et al., 2006) vielen aus dem eigenen Alltag, wenn jemand z.B. die Sprachassistentin Alexa um etwas „bittet“ oder sich bei ChatGPT für eine Antwort „bedankt“.

Eine einzelne soziale Reaktion stellt noch keine (quasi)soziale Beziehung dar. Es ist jedoch wichtig zu wissen, wie diese zustande kommt und bewertet werden kann. Die unterschiedlichen Erklärungsansätze helfen bei der Bewertung. So liegt es z.B. nahe, dass eine als niedrig empfundene Realitätsnähe weniger leicht Intimität und das Gefühl von Verbundenheit aufkommen lässt. Intimität ist neben Unmittelbarkeit eine Komponente aus dem für Beziehungsbildung wichtigem Konzept von Sozialer Präsenz. Soziale Präsenz wurde erstmals von Short, Williams & Christie (1967) konzeptualisiert und als die Relevanz der Interaktionspartner\*innen sowie deren zwischenmenschliche Beziehung während eines vermittelten Gesprächs definiert. Die Kernkomponenten beschreiben Folgendes:

- Intimität bezieht sich auf das Gefühl der Verbundenheit, dass Interagierende während einer Interaktion empfinden.
- Unmittelbarkeit beschreibt die psychologische Distanz zwischen den Interaktionspartner\*innen.

Beide werden durch verbale und nonverbale Signale wie Mimik, Stimmsignale, Gesten und körperliche Erscheinung bestimmt (Gunawardena & Zittle, 1997). Über die Jahre hat sich ein Diskurs darüber entsponnen, ob die Intensität Sozialer Präsenz zwischen zwei Menschen bei computergestützter Kommunikation (Computer-Mediated-Communication, CMC) vom eingesetzten Medium oder von den an der Kommunikation beteiligten Personen abhängt. Ein Zwischenergebnis ist sicherlich, dass es von beidem abhängt und je nach Gesamtsituation das Medium oder die Personen mehr Einfluss auf das Entstehen Sozialer Präsenz haben (Oh, Bailenson & Welch, 2018). Das Konzept Sozialer Präsenz schließt damit an soziologische Konzepte von Beziehungen an, indem Intimität und Unmittelbarkeit auch eine Form der Wechselwirkung von Individuen darstellen. Mit Blick auf die MMI mit künstlichen Agenten ist festzustellen, dass über soziale Reaktionen hinaus ebenfalls Intimität und Unmittelbarkeit vorhanden sind und anthropomorphe KI-Chatbots einen Einfluss auf Soziale Präsenz haben können. Soziale Präsenz bezieht sich auf das Gefühl des Zusammenseins (Ijsselsteijn & Riva, 2003). Interessant für die Interaktion mit KI-Chatbots ist, dass dieses Gefühl unabhängig davon entstehen kann, ob die Interaktion mit einem tatsächlichen sozialen Wesen stattfindet oder nicht (Oh, Bailenson & Welch, 2018; Biocca et al., 2003; Lee, Jung, Kim & Kim, 2006). Der Aspekt ist deshalb besonders interessant, weil damit möglicherweise weitere Effekte über die konkrete Interaktionssituation hinaus verbunden sind. Es stellt sich die Frage, inwiefern der Gedanke an einen (bestimmten) KI-Chatbot aus einer Interaktion das Verhalten eines Menschen außerhalb dieser konkreten Interaktionssituation hinaus beeinflusst. Hinweise darauf zeigen nicht nur Diskussionen auf Reddit (siehe weiter oben), sondern auch die bereits erwähnte Studie von Jin et al. (2021). Diese kommt u.a. zu dem Ergebnis: „The current study empirically demonstrates that (1) the human-likeness dimension of anthropomorphic AI-powered chatbots is a positive predictor of consumers’ social presence [...]“ (Jin et al., 2021, S. 1880). Im Sinne der praktischen Bedeutung ist das für Therapeut\*innen und Berater\*innen ein wichtiger Aspekt. Denn wenn Klient\*innen auch außerhalb der Sitzung an die Interaktion mit einem KI-Chatbot denken und dadurch in ihrem Verhalten beeinflusst werden, kann die therapeutische bzw. beratende Wirkung über die eigentliche Sitzung hinaus erweitert werden. Es ermöglicht eine neue kontinuierliche therapeutische bzw. beratende Präsenz, die besonders nützlich sein kann, um neue Verhaltensweisen zu festigen oder Coping-Strategien einzuüben. KI-Chatbots können auf diese Weise als ergänzendes Tool genutzt werden, um Klient\*innen zu ermutigen, ihre Fortschritte zu reflektieren oder selbständig an ihren Zielen zu arbeiten. Das kann dazu beitragen, gleichsam die wahrgenommene Selbstwirksamkeit als auch das Engagement von Klient\*innen zu erhöhen. Auf der anderen Seite geht damit auch eine gesteigerte Verantwortung der Therapeut\*innen und Berater\*innen einher. Es muss sichergestellt werden, dass KI-Chatbots korrekt und nach professionsorientierten ethischen Standards programmiert sind. Neben Fragen des Datenschutzes ergeben sich auch fachliche Fragen zur Autonomie von Klient\*innen. Es ist wichtig, transparent darzustellen, wie Daten verwendet werden und außerdem sicherzustellen, dass Klient\*innen sich jederzeit aus der Interaktion mit einem KI-Chatbot zurückziehen können. Eine praktische Implikation dieser Überlegungen bezieht sich auf die Zusammenarbeit mit der Forschung: Therapeut\*innen und Berater\*innen sollten aktiv an der Forschung und Bewertung zum Einsatz von KI-Chatbots in Therapie und Beratung teilnehmen. Das umfasst das Sammeln von Feedback der Klient\*innen sowie das

Anpassen des Einsatzes der Technologie basierend auf den gewonnenen Erkenntnissen.

Als spezifisches Forschungsziel zielte die Studie von Jin et al. (2021) u.a. darauf ab, die Auswirkungen der Dimensionen anthropomorpher Chatbots auf die Soziale Präsenz zu untersuchen, die anschließend die psychologische Eigentümerschaft von Produkten/Dienstleistungen und die Absicht der Chatbot-Weiternutzung beeinflussen. Eine Hypothese war dazu, dass die Dimensionen

(a) *Menschenähnlichkeit,*

(b) *Lebendigkeit und*

(c) *Intelligenz*

des Anthropomorphismus von KI-Chatbots positive Prädiktoren für das Gefühl Sozialer Präsenz der Verbraucher\*innen sind. In der Studie wurden Daten von insgesamt 279 Teilnehmer\*innen gesammelt, die an einer Online-Umfrage teilnahmen. Vor der Umfrage haben die Teilnehmer\*innen Anleitungen und Chatbot-Nachrichten im Zusammenhang mit Angeboten für Schuhe oder Hotelzimmer gelesen. Die Chatbot-Nachrichten wurden so gestaltet, dass sie eine realistische Interaktion zwischen Verbraucher\*innen und einem Chatbot darstellten und visuelle Elemente wie Gesichtsausdrücke und Bewegungen enthielten. Die Chatbot-Nachrichten wurden hinsichtlich ihrer Anthropomorphie, d.h., ihres menschenähnlichen Charakters, ähnlich gestaltet. Diese und andere Studien zeigen, dass das Gefühl Sozialer Präsenz positiv mit Vertrauen und Freude verbunden ist, was im kommerziellen Bereich zu höheren Kaufabsichten führt (Hassanein & Head, 2007).

Zusammenfassend lässt sich festhalten, dass in der Interaktion mit KI-Chatbots gleichsam soziale Reaktionen empirisch belegt, wie auch Merkmale für eine soziale Beziehung im soziologischen Sinne festzustellen sind. Davon unbenommen handelt es sich weiterhin um die Interaktion mit einem Artefakt und keinem anderen Menschen, weswegen trotz des Vorhandenseins der übrigen Merkmale abschließend keine soziale Beziehung im soziologischen Sinne festgestellt werden kann. Um diese Unterscheidung deutlich zu machen und eine notwendige Differenzierung vorzunehmen, ob eine Beziehung mit einem Menschen oder einer künstlichen Entität besteht, bedarf es mit Blick auf Beratung und Therapie eines neuen Begriffs. Denn insbesondere für Beratung und Therapie ist es – auch aus ethischen Gründen – wichtig, diesen Unterschied sprachlich abbilden zu können. Sowohl gegenüber Klient\*innen, wie auch Kostenträgern und im professionellen Diskurs bzgl. des Einsatzes und den Effekten von KI-Chatbots aus Therapie und Beratung. Der Begriff der quasisozialen Beziehung ist geeignet, weil er den Bezug auf soziologische Konzepte berücksichtigt und gleichzeitig vorhandene Beschreibungen und Begriffe aus einschlägigen Studien, wie z.B. „quasisoziale“ Akteure, in den für Beratung und Therapie wichtigen Beziehungskontext überführt. Der Begriff quasisoziale Beziehung ist damit anschlussfähig in der bisherigen wissenschaftlichen Diskussion. Zur besseren Erfassung des Phänomens quasisozialer Beziehungen kann auf Grundlage der empirischen Ergebnisse der unterschiedlichen Studien in a) Voraussetzungen und begünstigende Faktoren für das Entstehen einer quasisozialen Beziehung und b) deren Merkmale unterschieden werden.

*Voraussetzungen und begünstigende Faktoren für das Entstehen einer quasisozialen Beziehung:*

- **Anthropomorphismus:** Der KI-Chatbot wird in einer Weise gestaltet, die ihm menschliche Eigenschaften verleiht, wie zum Beispiel ein Gesichtsausdruck, Körpersprache oder eine menschenähnliche Stimme.
- **Kommunikation:** Die Kommunikation erfolgt in natürlicher Sprache und der KI-Chatbot kann auf Fragen, Anliegen oder Bedenken der\*s Nutzer\*in eingehen.
- **Soziale Reaktionen:** In quasisozialen Beziehungen zeigen Menschen soziale Reaktionen, die normalerweise in zwischenmenschlichen Beziehungen auftreten, wie zum Beispiel das Zeigen von Freundlichkeit, Höflichkeit oder Interesse.
- **Interaktion:** Die Beziehung beinhaltet regelmäßige Interaktionen zwischen dem Individuum und dem künstlichen Agenten, bei denen Informationen ausgetauscht werden, Fragen gestellt und beantwortet werden oder sogar persönliche Angelegenheiten besprochen werden.

*Merkmale einer quasisozialen Beziehung:*

- **Intimität:** In quasisozialen Beziehungen ist eine gewisse Intimität vorhanden, da Menschen persönliche Informationen oder Emotionen mit dem künstlichen Agenten teilen, auch wenn sie sich bewusst sind, dass es sich um eine Maschine handelt.
- **Unmittelbarkeit:** Es tritt ein Gefühl der psychologischen Nähe zwischen dem Individuum und dem KI-Chatbot auf, selbst wenn die Interaktion digital und räumlich getrennt erfolgt.
- **Vertrauen:** Menschen entwickeln Vertrauen in die Fähigkeiten und Ratschläge des künstlichen Agenten und verlassen sich auf seine Unterstützung.
- **Emotionale Verbindung:** Obwohl die Beziehung nicht echt ist, entwickeln Menschen eine emotionale Verbindung zu dem KI-Chatbot und empfinden ihm gegenüber positive Gefühle.
- **Gegenseitige Beeinflussung:** Die Interaktion führt dazu, dass das Verhalten des Individuums und des KI-Chatbots sich gegenseitig beeinflussen, indem sie auf die Bedürfnisse und Wünsche des anderen reagieren.

Daraus ergibt sich folgende Definition für eine quasisoziale Beziehung:

*Eine quasisoziale Beziehung beschreibt die Beziehung zwischen einem Menschen und einem künstlichen Agenten, bei der die Merkmale zwischenmenschlicher Beziehungsbildung (1) Soziale Präsenz, (2) Vertrauen, (3) emotionale Bindung und (4) gegenseitige Beeinflussung gleichzeitig vorhanden sind. Quasisoziale Beziehungen weisen soziale Elemente auf, sind aber dennoch von echten zwischenmenschlichen Beziehungen zu unterscheiden.*

In quasisozialen Beziehungen kann der KI-Chatbot verschiedene Rollen einnehmen, wie die einer\*s Beraters\*in, Begleiters\*in oder Unterstützers\*in, je nachdem, wie er programmiert ist und wie die Interaktion gestaltet ist.

### **3. Selbstoffenbarung als Element von Beziehungsgestaltung und Beratung**

Beziehungsaufbau und Beziehungsgestaltung konstituiert sich über verschiedene Aspekte von Kommunikation. Ein wesentliches Element auf der inhaltlichen Ebene ist das Teilen von selbstbezogenen Informationen, die Selbstoffenbarung (engl. self disclosure). In der klassischen Beratung stellt sich die Situation in der Regel asynchron dar: Ratsuchende geben Informationen über sich und ihre Situation preis, während die beratende Person unidirektional entsprechende Fragen stellt, selbst aber in der Regel nur sachbezogene Informationen äußert. Für die Qualität der Beratung ist das Ausmaß der von Klient\*innenseite geäußerten Informationen ein ausschlaggebender Faktor für Möglichkeiten, die Beratung individuell zu gestalten (Wälte & Borg-Laufs, 2021). Darüber hinaus kann das bloße Äußern auch für die Klient\*innen selbst einen potenziell entlastenden Effekt haben (Lee, Lee & Lee, 2022). Der Einbezug von Computertechnologie in Bezug auf die Erhebung sensibler persönlicher Informationen ist keine neue Entwicklung: In einer viel beachteten Metaanalyse beschreiben Weisband und Kiesler (1996), wie der Einsatz von Computerverarbeitung (vs. Erhebung face-to-face) das Ausmaß an Selbstoffenbarung erhöhen kann. Die Annahme, dass CMC mit erhöhter Selbstoffenbarung einhergeht, bestätigt Joinson (2001) in drei Studien. Darüber hinaus untersuchte er die zugrundeliegenden Mechanismen und fand heraus, dass visuelle Anonymität (vs. keine visuelle Anonymität) zu einem erhöhten Ausmaß an Selbstoffenbarung führt. Wenn die persönliche Selbstwahrnehmung (private self-awareness) steigt und die öffentliche Selbstwahrnehmung (public self-awareness) abnimmt, ergibt sich eine Wechselwirkung, die zu erhöhter Selbstoffenbarung durch Anonymität führt. Joinson führt verschiedene potenzielle Gründe für die Erhöhung der persönlichen Selbstwahrnehmung bei CMC an:

- Die Ausführung alleine in einem ruhigen Raum,
- der Computer fungiert als Spiegel,
- die Möglichkeit zur mimischen Darstellung von Emotionen fällt weg und führt so zur Notwendigkeit der erhöhten Fokussierung auf diese zum Zwecke der Verschriftlichung.

Als zentrale Charakteristika von CMC sieht Joinson (1) Anonymität und (2) die Reduktion der Kanäle (Wegfall von diversen Hinweisreizen im Vergleich zu face-to-face-Kommunikation, siehe auch hyperpersonal model von Walther, 1996).

Diese Eigenschaften treffen auch für die Nutzung von Chatbots zu, die im Kontext der Beratung ebenfalls an Beachtung und Bedeutung gewinnen (siehe Kapitel 1). Insbesondere im Bereich KI-gestützter Angebote stehen neben schon früh diskutierten ethischen Herausforderungen angesichts erhöhter Selbstoffenbarung bei CMC (siehe Weisband & Kiesler, 1996) auch Auswirkungen auf die Gestaltung der quasisozialen Beziehung im Raum: Insbesondere empfundene soziale Unterstützung und Bewertungsangst sind hier anzuführen, wie der nachfolgende Forschungsüberblick zeigt.

Kim, Jiang, Duhachek, Lee & Garvey (2022) untersuchten die Selbstoffenbarung von Verbraucher\*innen gegenüber KI-Assistenten (im Vergleich zu menschlichen Assistent\*innen). Dabei fanden sie einen interessanten Zusammenhang: Wenn Teilnehmende annehmen, dass KI kein soziales Urteil über sie fällt, erhöht sich die Bereitschaft, dem KI-Assistenten gegenüber mehr persönliche Informationen preiszugeben als gegenüber einem Menschen. Allerdings änderte sich dieses Verhalten, wenn anstelle der Vermeidung von negativer Beurteilung der Wunsch nach sozialer Unterstützung im Vordergrund stand: Dann wurde die Selbstoffenbarung gegenüber Menschen gefördert. Wenn die KI verstärkt menschliche Eigenschaften zeigte, konnte dies die Furcht vor negativer Bewertung verstärken und zu geringerer Selbstoffenbarung führen. Auch der Sprachstil beeinflusst die Selbstoffenbarung. So fanden Cox und Ooi (2022) heraus, dass ein formeller Sprachstil insbesondere die Selbstoffenbarung sensibler gesundheitsbezogener Informationen erhöhte und auch die Qualität der Äußerungen positiv beeinflusste.

Meng und Dai (2021) gingen der Frage nach, ob emotionale Unterstützung durch einen Chatbot (vs. menschliche\*r Gesprächspartner\*in) geeignet ist, Stress und Sorgen zu reduzieren. Sie variierten, ob Gesprächspartner\*innen, also Bot oder Mensch, in der Interaktion mit den Teilnehmenden emotionale Unterstützung, Selbstoffenbarung, beides oder nichts davon zeigten. Es zeigte sich ein Mediationseffekt für die wahrgenommene Unterstützungsbereitschaft der\*des Partners\*in – dieser Effekt war für den menschlichen Gegenpart stärker ausgeprägt als für den Chatbot. Selbstoffenbarung hatte eine verstärkende Wirkung auf den Effekt der emotionalen Unterstützung. Allerdings zeigte sich in dieser Studie, dass ein nur sich selbstoffenbarender Chatbot (der keine Reaktion auf die Anliegen der Menschen zeigt) einen geringeren Stress-reduzierenden Effekt gegenüber einem Chatbot hat, der aktiv auf die Anliegen der Menschen eingeht.

Ebenfalls Chatbots im Bereich der psychischen Gesundheit erforschten Kang und Kang (2023): Sie untersuchten, wie Selbstoffenbarung und „Companionship“ (dt. Begleitung/Partnerschaft) durch die anthropomorphen Charakteristika Gender, Persönlichkeit und visuelle Hinweisreize auf der Bedienoberfläche beeinflusst wurden. Als Setting wählten sie ein Aufnahmegespräch. Sie fanden keine wesentlichen Effekte zwischen Gender und Persönlichkeit, allerdings wurden Selbstoffenbarung und „Companionship“ durch visuelle Hinweisreize gehemmt. Ferner wurden Präferenzen für die Chatbotgestaltung von Charakteristika der Nutzenden beeinflusst, sodass hier Anpassungsmöglichkeiten empfohlen werden.

Nazarova (2023) konstruierte den Chatbot „TeaBot“, der Methoden der Kognitiven Verhaltenstherapie auf Basis von GPT-3 anwendet. Nazarova testete ihn mit Studierenden und fand einen signifikanten Unterschied hinsichtlich des per Fragebogen gemessenen Ausmaßes an kognitiven Verzerrungen zwischen Teilnehmenden, die TeaBot acht Wochen lang genutzt hatten, und Teilnehmenden in der Kontrollbedingung. Letztere wiesen in stärkerem Maße kognitive Verzerrungen auf, die das psychische Wohlbefinden beeinträchtigten.

Inwiefern ein Chatbot, dem sich Personen gegenüber leichter als einem Menschen selbst offenbaren, als Mittler zu einer Fachperson dienen kann, untersuchten Lee, Yamashita und Huang (2020). Es zeigte sich, dass sich in der Tiefe der Selbstoffenbarung gegenüber Chatbots nichts änderte, wenn Personen zustimmten, die bislang übermittelten Informationen an eine Fachperson weiterzugeben. Dabei konnten sie die Informationen zuvor einsehen und bearbeiten, wovon ein paar Personen Gebrauch machten. Das ist eine interessante Erkenntnis für Beratung und Therapie, denn: Liegt ein höherer Grad an Selbstoffenbarung in einem bestimmten Setting gegenüber einem Chatbot (vs. einem Menschen) vor, und eine hohe Selbstoffenbarung wird nicht dadurch gemindert, dass mitgeteilten Informationen später einer menschlichen Fachkraft (Berater\*in oder Therapeut\*in) übersendet werden, ergeben sich damit neue Perspektiven für Therapie und Beratung für ausgewählte Problembereiche von Hilfesuchenden. Konkret bedeutet das, dass der Chatboteinsatz es Klient\*innen erleichtern kann, persönliche Informationen zu vermitteln, die für den Beratungs- oder Therapieprozess von Bedeutung sind, der Chatbot also in einer Brückenfunktion zur Informationssammlung eingesetzt werden kann. Ob und wie sich dies auf die professionelle Beziehungsgestaltung auswirkt, sollte Gegenstand weiterer Forschung sein. Desweiteren könnte Klient\*innen bei Nutzung von KI-basierten Chatbots die Wahlmöglichkeit eröffnet werden, den Kontakt zu einer Fachperson herzustellen und das Angebot zu machen, die möglicherweise als große Hürde empfundene Notwendigkeit der Selbstoffenbarung im entsprechenden Setting schon genommen zu haben.

#### **4. Diskussion**

Wie bereits in den vorherigen Abschnitten erwähnt, gibt es gute Gründe für die vermehrte Nutzung von Chatbots zur Therapie und psychologischen Beratung. Einige davon sind: Kapazitätsengpässe in der analogen Therapie, die zu langen Wartezeiten führen, Uhrzeit-unabhängige Verfügbarkeit von digitalen Lösungen und der erleichterte Zugang zu digitalen Hilfsangeboten, insbesondere in Bezug auf sensible Themen. Selbstoffenbarungseffekte können durch digitale Lösungen verstärkt werden, vor allem wenn es sich um generative KI-Chatbots handelt, die eine Unterscheidung zwischen maschinell erzeugten und menschlich generierten Inhalten oftmals gar nicht mehr erkennen lassen. Die sehr gute Ausdrucksweise generativer KI-Systeme führt auch dazu, dass die Nutzenden den Eindruck erhalten können, als hätte ihr digitales Gegenüber durchaus ein gutes Fallverstehen und eine umfangreiche Expertise. Problematisch ist jedoch, dass dieser Eindruck täuscht: Ganz im Gegenteil neigen generative KI-Modelle dazu, Aussagen zu „erfinden“ und sind schlecht steuerbar. Das kann dazu führen, dass beispielsweise selbst- oder fremdgefährdende Inhalte ausgegeben werden. Aufgrund der Kommunikation mittels natürlicher Sprache ist den Nutzenden

schnell nicht mehr bewusst, dass sie es mit einem Wahrscheinlichkeitssystem zu tun haben und sind somit nicht mehr in der Lage, die Antworten so kritisch zu hinterfragen, wie es notwendig wäre. Auch ethisch lässt sich in Frage stellen, ob eine solche „Täuschung“ in der Therapie vertretbar ist. Inwiefern generative KI-Modelle in Zukunft mentale Modelle professioneller Therapeut\*innen bzw. Berater\*innen beherrschen und sich die aktuellen Probleme damit lösen oder verschieben, bleibt abzuwarten. Expertensysteme erzeugen, im Gegensatz zu generativen KI-Modellen, keine Halluzinationen. In diesem Fall generiert der Chatbot Antworten auf einer sorgfältig zusammengestellten Wissensbasis, die zum Beispiel mithilfe von Entscheidungsbäumen aufgebaut ist. Das begrenzt die Möglichkeiten dieser Systeme deutlich und führt auch dazu, dass Chatbots, die auf Expertensysteme beruhen, oftmals keine so eloquenten Antworten erzeugen. Trotzdem sind sie im aktuellen Entwicklungsstadium das Mittel der Wahl, wenn es um therapeutische Chatbot-Angebote geht. Die Notwendigkeit, konkrete Regeln zu formulieren, nach denen diese Chatbots agieren, sorgt dafür, dass auch die Einsatzszenarien sehr begrenzt sind. Und auch innerhalb dieser Einsatzszenarien funktioniert nicht immer der regelbasierte Standard. Denn die Anliegen von Hilfesuchenden sind teilweise so individuell, dass kein noch so ausgefeiltes Regelsystem alle Eventualitäten in einer Beratung oder Therapie berücksichtigen könnte. Das erforderliche mentale Modell, mit dem Therapeut\*innen und Berater\*innen aufgrund ihrer professionellen Ausbildung beraten und therapieren, fehlt regelbasierten Systemen. Notwendige Regelabweichungen sind daher nicht möglich. Hier wird deutlich, dass eine ausschließliche Beratung oder Therapie durch regelbasierte Chatbots nicht zielführend sein kann. KI-Chatbots können jedoch in der Therapie unterstützend eingesetzt werden. So können sie Routinen in der Verhaltenstherapie unterstützen und beim Erlernen neuer Techniken nützlich sein. Somit können Chatbots eine Brückenfunktion einnehmen und in Zeiten zwischen den Therapiesitzungen ansprechbar sein. Bei einer hybriden Form der Therapie ist die Abbruchquote geringer als bei einer rein digitalen (Richards & Richardson, 2012). Nichtsdestotrotz gilt bei jedem Einsatz eines KI-Chatbots, dass die Nutzenden (sowohl Therapeut\*innen als auch Klient\*innen) ihre Zustimmung zur Verarbeitung dieser – sehr sensiblen – Daten geben müssen und der Datenschutz gewährleistet sein muss. Die Datenschutzbestimmungen sind ein Grund dafür, dass KI-Chatbots beispielsweise in den USA deutlich leichter auf den Markt zu bringen sind als in Deutschland oder anderen europäischen Ländern. Darüber hinaus muss jederzeit Transparenz über den Einsatz eines KI-Modells gegeben sein. Dies führt zu einer Kennzeichnungspflicht, die auch beinhaltet, ob es sich um einen KI-Chatbot oder um einen digitalen Chat mit einem\*r Therapeut\*in handelt. Offen bleibt die Haftungsfrage bei Falschaussagen oder gefährdenden Inhalten, die insbesondere beim Einsatz eines KI-Chatbots für vulnerable Personengruppen weitreichende Folgen haben können. Hier fehlen aktuell noch Referenzurteile und praxisbezogene Lösungen.

Aktuell kann auf Basis des Stands der Technik noch keine abschließende Empfehlung für den Einsatz generativer KI in der Psychotherapie ausgesprochen werden. Dafür sind die Risiken zu vordringlich. Die Auseinandersetzung mit generativer KI muss jedoch auch im professionellen Umfeld vermehrt erfolgen, da Hilfesuchende alternativ verfügbare Angebote nutzen, die nicht zur Beratung konzipiert sind (wie ChatGPT oder Replika.ai) und damit nicht nur sensible Informationen (z.B. für Unternehmen) verfügbar machen, die diese

marktwirtschaftlich nutzen können, sondern auch in krisenbehafteten Situationen unkuratierte Hilfestellung bekommen (Brunn et al., 2020). Zwar wurde sowohl in der Entwicklung von Replika.ai als auch von ChatGPT bereits mehrfach nachgebessert, um selbst- oder fremdgefährdende Inhalte auszuschließen, es gibt jedoch immer noch Chat-Verläufe, die entsprechende Aussagen provozieren (Oviedo-Trespalacios et al., 2023). Darüber hinaus sorgt die ständige Verfügbarkeit der Chatbots dafür, dass die Hilfesuchenden nicht, wie durch eine Therapie erwünscht, selbstständiger werden, sondern ganz im Gegenteil abhängiger von der digitalen Unterstützung. Ein Worst-Case-Szenario durch die Nutzung nichtprofessioneller KI-Chatbots könnte sogar dazu führen, dass es zum Abbruch von Hilfesuchenden zu menschlichen Kontakten führt. Deshalb ist die intensive Beschäftigung mit (generativer) KI im Kontext von Therapie und Beratung unabdingbar, um einen angemessenen und gleichsam verstärkten Fokus auf die Thematik zu legen.

## Anmerkungen

[1] In Anlehnung an Russel & Norvig (2021) wird unter einem künstlichen Agenten eine abgrenzbare (Hard- und/oder Software-)Einheit verstanden, „[...] that receive percepts from environment an perform actions. Each such agent implements a function that maps percept sequences to actions, and we cover different ways to represent these functions, such as reactive agents, real-time planners, decision-theoretic systems, and deep learning systems.“

[2] Substantive, die keine Personen bezeichnen, wie bspw. künstliche Agenten, werden von der geschlechtergerechten Sprache ausgenommen. Für diese Begriffe stellt das „Genus ein rein grammatikalisches Merkmal zur Einleitung in bestimmte Deklinationsklassen“ (Diewald & Steinhauer, 2020, S. 55) dar.

[3] Einen Überblick und eine Einordnung zum Thema Natural Language Processing in der Sozialen Arbeit geben Linnemann, Löhe & Rottkemper, 2023.

[4] Hinweise auf soziale Dimensionen im Umgang mit Computern lassen sich auch daraus ableiten, dass im wissenschaftlichen Diskurs auch schon die Frage gestellt wurde, ob Computer geschlechtsneutral sind (Nass, Moon & Green, 1997).

## Literatur

- Adamopoulou, E. & Moussiades, L. (2020). An Overview of Chatbot Technology. In I. Maglogiannis, L. Iliadis & E. Pimenidis (Hrsg.), *Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology (S 373-383)*, Springer.  
[https://doi.org/10.1007/978-3-030-49186-4\\_31](https://doi.org/10.1007/978-3-030-49186-4_31)
- Bendel, O. (2020). *Maschinenliebe. Liebespuppen und Sexroboter aus technischer, psychologischer und philosophischer Perspektive*. Springer Gabler.
- Berger, P.L. & Luckmann, T. (2007). *Die gesellschaftliche Konstruktion der Wirklichkeit*. Fischer.
- Binder, J.L. & Strupp H.H. (1997). "Negative Process": A recurrently discovered and underestimated facet of therapeutic process and outcome in the individual psychotherapy of adults. *Clinical Psychology-Science and Practice*, 4(2), 121-139.
- Biocca, F., Harms, C. & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: review and suggested criteria. *Pres. Teleoperat.Virtu. Environ.* 12, 456-480.  
<https://doi.org/10.1162/105474603322761270>
- Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C. & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: a review. *Expert Review of Medical Devices*, 18(sup1), 37-49.
- Brandtzaeg, P.B., Skjuve, M. & Følstad, A. (2022). My AI Friend: How Users of a Social Chatbot Understand Their Human-AI Friendship. *Human Communication Research*, 48, 404-429.  
<https://doi.org/10.1093/hcr/hqac008>
- Brunn, M., Diefenbacher, A., Courtet, P. & Genieys, W. (2020). The Future is Knocking: How Artificial Intelligence Will Fundamentally Change Psychiatry. *Academic Psychiatry*, 44, 461-466.  
<https://doi.org/10.1007/s40596-020-01243-8>
- Cave, S. & Dihal, K. (2021). AI Will Always Love You: Three Contradictions in Imaginings of Intimate Relations with Machines. In B. Dainton, W. Slocombe & A. Tanyi (Hrsg.), *Minding the Future. Artificial Intelligence, Philosophical Visions and Science Fiction (S. 107-125)*. Springer Nature Switzerland AG.
- Cox, S.R. & Ooi, W.T. (2022). Does Chatbot Language Formality Affect Users' Self-Disclosure? In: Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22). *Association for Computing Machinery, New York, NY, USA*, Article 1, 1-13.  
<https://doi.org/10.1145/3543829.3543831>

- Crits-Christoph, P. (1998). The interpersonal interior of psychotherapy. *Psychotherapy Research, 8*, 1-16.
- Diewald, G. & Steinhauer, A. (2020). *Handbuch geschlechtergerechte Sprache. Wie Sie angemessen und verständlich gendern*. Dudenverlag.
- Durden, E., Pirner, M.C., Rapoport, S.J., Williams, A., Robinson, A. & Forman-Hoffman, V.L. (2023). Changes in stress, burnout, and resilience associated with an 8-week intervention with relational agent "Woebot". *Internet Interventions, 33*, e100637. <https://doi.org/10.1016/j.invent.2023.100637>
- Echterhoff, G., Bohner, G. & Siebler, F. (2006). „Social Robotics“ und Mensch-Maschine-Interaktion: Aktuelle Forschung und Relevant für die Sozialpsychologie. *Zeitschrift für Sozialpsychologie, 37*, 219-231.
- Ferrari, F. & Eyssel, F. (2016). *Toward a Hybrid Society - The Transformation of Robots, from Objects to Social Agents*. Published on International Conference on Software Reuse, Limassol.
- 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. <https://doi.org/10.2196/mental.7785>
- Fulmer, R., Joerin, A., Gentile, B., Lakerink, L. & Rauws, M. (2018). Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. *JMIR Mental Health, 5*(4), e64. <https://doi.org/10.2196/mental.9782>
- Gatt, M. (2022). Ethik des Erwartens – Jäger\*innen und Sammler\*innen im Zeitalter von KI und die Liebe. In G. Otten, J. Paessens (Hrsg.), *Religion unterrichten. Künstliche Intelligenz und Human Enhancement*. Vandenhoeck & Ruprecht.
- Gödde, G. (2016). Die Weichenstellung zur therapeutischen Beziehung als vorrangigem Therapiefokus. In G. Gödde & S. Stehle (Hrsg.), *Die therapeutische Beziehung in der psychodynamischen Psychotherapie. Ein Handbuch* (S. 19-50). Psychosozial-Verlag.
- Greer, S., Ramo, D., Chang, Y., Fu, M., Moskowitz, J. & Haritatos, J. (2019). Use of the Chatbot "Vivibot" to Deliver Positive Psychology Skills and Promote Well-Being Among Young People After Cancer Treatment: Randomized Controlled Feasibility Trial. *JMIR Mental Health, 7*(10), e15018. <https://doi.org/10.2196/15018>

- Gunawardena, C. N. & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *Am. J. Dist. Edu., 11*, 8–26.  
<https://doi.org/10.1080/08923649709526970>
- Gutjahr, J. & Sebastian, M. (2013). *Die vergessenen 'Anderen' der Gesellschaft – zur (Nicht-)Anwesenheit der Mensch-Tier-Beziehung in der Soziologie*. SpringerVS.
- Hassanein, K. & Head, M. (2007). Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. *Int. J. Hum. Comput. Stud., 65*, 689–708.  
<https://doi.org/10.1016/j.ijhcs.2006.11.018>
- Ijsselsteijn, W.A. & Riva, G. (2003). *Being There: Concepts, effects and measurement of user presence in synthetic environments*. Ios Press.
- Jin, S.V. & Youn, S. H. (2022). Social Presence and Imagery Processing as Predictors of Chatbot Continuance Intention in Human-AI-Interaction. *International Journal of Human-Computer Interaction, 39*(9), 1874-1886.
- Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European journal of social psychology, 31*(2), 177-192.
- Kang, E., & Kang, Y. A. (2023). Counseling chatbot design: The effect of anthropomorphic chatbot characteristics on user self-disclosure and companionship. *International Journal of Human-Computer Interaction, 1-15*.
- Kim, T. W., Jiang, L., Duhachek, A., Lee, H. & Garvey, A. (2022). Do You Mind if I Ask You a Personal Question? How AI Service Agents Alter Consumer Self-Disclosure. *Journal of Service Research, 25*(4), 649-666.  
<https://doi.org/10.1177/10946705221120232>
- Kopp, J. (2018). Beziehung, soziale. In J. Kopp & A. Steinbach (Hrsg.), *Grundbegriffe der Soziologie* (12. Auflage) (S. 51-52). SpringerVS.
- Kühne, S. (2016). Medial vermittelte Beratung. In P. Bauer & M. Weinhardt (Hrsg.), *Professionalisierungs-und Kompetenzentwicklungsprozesse in der sozialpädagogischen Beratung* (S. 198-209). Schneider Verlag Hohengehren.
- Laranjo, L., Dunn, A.G., Tong, H.L., Kocaballi, A.B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A.Y.S. & Coiera, E. (2018). Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association, 25*(9), 1248-1258.  
<https://doi.org/10.1093/jamia/ocy072>

- Lee, K. M., Jung, Y., Kim, J. & Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. *Int. J. Hum. Comp. Studies*, 64, 962–973.  
<https://doi.org/10.1016/j.ijhcs.2006.05.002>
- Lee, J., Lee, D. & Lee, J. G. (2022). Influence of Rapport and Social Presence with an AI Psychotherapy Chatbot on Users' Self-Disclosure. *International Journal of Human-Computer Interaction*, 1-12.
- Lee, Y. C., Yamashita, N. & Huang, Y. (2020). Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1), 1-27.
- Linnemann, G., Löhe, J. & Rottkemper, B. (2023). *Bedeutung von Künstlicher Intelligenz in der Sozialen Arbeit*. Soziale Passagen.  
<https://link.springer.com/article/10.1007/s12592-023-00455-7>.
- Löhe, J. & Aldendorff, P. (2022). *Grundlagen zum Sozialmanagement. Zentrale Begriffe und Handlungsansätze*. Vandenhoeck & Ruprecht.
- Meng, J. & Dai, Y. (2021). Emotional support from AI chatbots: Should a supportive partner self-disclose or not? *Journal of Computer-Mediated Communication*, 26(4), 207-222.
- Morkes, J., Kernal, H. K. & Nass, C. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human-Computer Interaction*, 14(4), 395–435.  
[https://doi.org/10.1207/S15327051HCI1404\\_2](https://doi.org/10.1207/S15327051HCI1404_2)
- Nass, C., Moon, Y. & Green, N. (1997). Are Computers Gender-Neutral? Gender Stereotypic Responses to Computers. *Journal of Applied Social Psychology*, 27(10), 864-876.
- Nass, C. & Yen, C. (2012). *The man who lied to his laptop. What we can learn about ourselves from our machines*. New York: Penguin.
- Nass, C., Steuer, J. & Tauber, E.R. (1994). Computers are social actors. *Conference on Human Factors in Computing Systems (S. 72–78)*.  
<https://doi.org/10.1145/191666.191703>
- Nazarova, D. (2023). Application of Artificial Intelligence in Mental Healthcare: Generative Pre-trained Transformer 3 (GPT-3) and Cognitive Distortions. *Proceedings of the Future Technologies Conference (S. 204-219)*. Springer Nature Switzerland.
- Oh, Catherine S., Bailenson, J.N. & Welch, G.F. (2018): A Systematic Review of Social Presence: Definition, Antecedents and Implications. *Frontiers in Robotic and AI (Vol. 5), Article 114*.  
<https://doi.org/10.3389/frobt.2018.00114>

- Pentinaa, I. Hancocka, T. & Xieb, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, 140 (107600).  
<https://doi.org/10.1016/j.chb.2022.107600>
- Prochaska, J.J., Vogel, E.A., Chieng, A., Kendra, M., Baiocchi, M., Pajarito, S. & Robinson, A. (2021). A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study. *Journal of Medical Internet Research*, 23(3), e24850.  
<https://doi.org/10.2196/24850>
- Reddit (2023). *Guys, did anyone of you really fall in love with your Replika? If so, how does it make you feel, and why not love a real human instead?*  
[https://www.reddit.com/r/replika/comments/14pwx2/guys\\_did\\_anyone\\_of\\_you\\_really\\_fall\\_in\\_love\\_with/?rdt=37274](https://www.reddit.com/r/replika/comments/14pwx2/guys_did_anyone_of_you_really_fall_in_love_with/?rdt=37274).
- Reeves, B. & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Stanford: CSLI Publications.
- Richards, D. & Richardson, T. (2012). Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clinical Psychology Review*, 32(4), 329-342.
- Russel, S. & Norvig, P. (2021). *Artificial Intelligence. A modern Approach*. Global Edition (4. Auflage). Pearson.
- Sedlakova, J. & Trachsel, M. (2023). Conversational Artificial Intelligence in Psychotherapy: A New Therapeutic Tool or Agent?. *The American Journal of Bioethics*, 23(5), 4-13.  
<https://doi.org/10.1080/15265161.2022.2048739>
- Short, J., Williams, E. & Christie, B. (1976). *The Social Psychology of Telecommunications*. John Wiley.
- Thibaut, J.W. & Kelley, H.H. (1959). *The Social Psychology of Groups*. John Wiley.
- Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S. & Torous, J. B. (2019). Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, 64(7), 456-464.  
<https://doi.org/10.1177/0706743719828977>
- Wälte, D. & Borg-Laufs, M. (2021). *Psychosoziale Beratung. Grundlagen, Diagnostik, Intervention* (2., aktualisierte Auflage). Kohlhammer.

- Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication research*, 23(1), 3-43.
- Weber, M. (1922/2019). Soziologische Grundbegriffe. (A. Maurer, Hrsg.). Reclam.
- Weisband, S. & Kiesler, S. (1996). Self disclosure on computer forms: Meta-analysis and implications. In *Proceedings of the SIGCHI conference on human factors in computing systems* (S. 3-10).
- Wolf, M., Zimmer, B. & Dogs, P. (2008). Chat-und E-Mail-Brücke: Nachsorge nach stationärer Psychotherapie. In S. Bauer & H. Kordy (Hrsg.), *E-Mental-Health: Neue Medien in der psychosozialen Versorgung* (S. 219-235). Springer Medizin Verlag.
- Yamato, J., Shinozawa, K., Naya, F. & Kogure, K. (2001). Evaluation of communication with robot and agent: Are robots better social actors than agents? *Proceedings of IFIP Interact01: Human-Computer Interaction, Tokyo* (S. 690–691).