»Fühl dich verstanden«

KI-Bots im Umgang mit Angststörungen

Helen Pfeffer | Sidney Wilshusen

Einleitung

Die zunehmende Integration von Künstlicher Intelligenz (KI) in alltäglichen Lebensbereichen betrifft inzwischen auch das Feld der psychischen Gesundheitsversorgung. Vor dem Hintergrund eines stetig wachsenden Bedarfs an psychotherapeutischer Versorgung, dem das bestehende System trotz wissenschaftlich anerkannter Wirksamkeit von Psychotherapie kaum gerecht werden kann (vgl. BundesPsychotherapeutenkammer [BPtK] o. J.: 70), gewinnt die Frage an Bedeutung, inwiefern KI-gestützte Chatbots relevant für den psychotherapeutischen Kontext sein können. KI-gestützte Chatbots könnten hier eine Möglichkeit bieten, psychosoziale Unterstützung zu leisten, insbesondere für Menschen, die aus Scham, Angst oder Unsicherheit keine klassischen Therapieangebote in Anspruch nehmen können oder wollen.

Der vorliegende Beitrag untersucht die Interaktion eines solchen Chatbots mit einer von uns konstruierten Nutzerin, die unter einer unbehandelten Angststörung leidet. Im Mittelpunkt steht die Frage, inwiefern die Chatbots zu therapeutischen Zwecken genutzt werden können, mit besonderem Fokus auf Empathie, insbesondere professioneller Empathie. Für diese Untersuchung wurde von den Autorinnen die Nutzerin-Avatar *Ida* erstellt, eine 28-jährige Jura-Studentin, die sich in einer stark belasteten Lebensphase befindet und KI-Chatbots mit der Erwartung, Hilfe zu erhal-

ten, heranzieht. Dazu wurden zwei unterschiedliche KI-Chatbots miteinander verglichen: die Therapie-App Zen sowie der Chatbot Therapist GPT von ChatGPT. Beide Interaktionen wurden mit identischer Ausgangslage (Nutzer*innen-Avatar, Anliegen, Einstiegsimpuls) durchgeführt, um der Fragestellung adäquat nachgehen zu können und Unterschiede in der Empathiefähigkeit und im kommunikativen Stil herauszuarbeiten. Im vorliegenden Kontext wurden die nachfolgenden Analysekategorien herangezogen (s. Kapitel 2: Theoretische Grundlagen): Aus der Empathietheorie wurden die Kategorien Modus sowie Aspekte (Kontakt, Emotion, Kognition, Volition, Motivation, dunkle Seiten, professionalisiert vs. nichtprofessionalisiert, Folgehandlungen) und aus dem Bereich Bindung der Bindungstyp »desorganisiert« herangezogen. Darüber hinaus umfasst die Empathiedarstellung Aspekte der Mimik, Emojis und Lautaspekte sowie weitere Kategorien aus dem Bereich der Interaktionstheorie, wie etwa Nähe-Distanz, Chat-Gestaltung, Höflichkeit und Face-Wahrung/-Verletzung. Darüber hinaus finden sich Kategorien aus den Bereichen Sprache und Beziehung, wie etwa Komplimente und digitale Beziehungsanbahnung.

1 Theorie und Kontextualisierung

Die theoretische und methodische Rahmung dieses Beitrags wird in Kapitel 2: Theoretische Grundlagen und Kapitel 3: Methode der doppelten Künstlichkeit in diesem Band umfassend dargestellt. Die vorliegende Analyse stützt sich auf diesen gemeinsamen Rahmen, insbesondere im Hinblick auf das zugrunde liegende Verständnis von Empathie sowie den methodischen Ansatz zur qualitativen Interaktionsanalyse mit KI-Chatbots.

1.1 Empathie

Empathie wird nach Breyer (2020) als »ein zeitlich ausgedehnter, dynamischer, episodischer Prozess [verstanden], in den Erlebnisse aus unterschiedlichen Bewusstseinssphären (Wahrnehmen, Sich-Bewegen, Fühlen, Wollen, Denken) integriert werden« (Breyer 2020: 15). Er unterscheidet zwischen der *leiblich-körperlichen*, der *affektiv-emotionalen* und der *kogniti-*

ven Dimension. Die erste Dimension umfasst das Verstehen von Gefühlen auf emotionaler und kognitiver Perspektive, demnach das Verstehen der Gefühle des Gegenübers ebenso wie die affektive Resonanz auf diese (vgl. ebd.). Der Resonanzmodus beschreibt die Fähigkeit, sich mit den Bewegungen oder dem Verhalten anderer Menschen zu synchronisieren und bei der man durch die Stimmung eines anderen automatisch und unterbewusst mitgerissen wird (vgl. ebd.). Bezogen auf die Kommunikation zwischen Mensch und Maschine stellt sich die Frage, ob und wie diese Dimensionen digital abgebildet werden können, besonders unter Bedingungen psychischer Belastung wie im Fall der Nutzerin »Ida«. Die affektiv-emotionale Dimension der Empathie umfasst das Miterleben von Gefühlen, die bei einer anderen Person ausgelöst werden. Dieses Miterleben kann entweder in Form einer kongruenten (gleichgerichteten) oder einer invertierten (entgegengesetzten) emotionalen Reaktion erfolgen. So führt beispielsweise die Freude einer anderen Person häufig zur adaptierten Freude beim Beobachtenden. Im Gegensatz dazu kann bei negativen Gefühlen wie Missgunst eine affektiv-emotionale Reaktion entstehen, die dem erlebten Zustand des Gegenübers entgegensteht und somit eine antagonistische Beziehung zum ursprünglichen Gefühl aufweist (vgl. ebd.). Die dritte Dimension umfasst inferentielle Prozesse, bei denen auf Basis generalisierten Wissens und situativer Informationen abstrakte Schlussfolgerungen über das Erleben anderer gezogen werden, sowie imaginative Prozesse, die ein perspektivisches Hineinversetzen in den anderen ermöglichen (Breyer 2020: 15). Es ist darauf hinzuweisen, dass ein Erfahrungswert gegeben sein muss, um sich in andere hineinversetzen zu können, und dass dieser neben der sozialen und historischen, immer auch von der kulturellen Lage abhängig ist. Demnach ist Empathie auch immer ein kulturelles Konstrukt (vgl. ebd. & Liebert 2020: 107).

Nach Hermanns ist Empathie ein mehrdimensionales Konstrukt: Er unterscheidet zwischen verschiedenen Ebenen. Dazu gehören vier Ebenen: die *kognitive* (beobachten, wahrnehmen, verstehen), die *emotive* (Gefühle des anderen erforschen), die *volitive* (Wünsche oder Absichten des Anderen beachten) sowie die *phatische* (Kontaktaufnahme und Entwicklung von Bindung) (vgl. Jacob/Konerding/Liebert 2020: 5, 117 & Rettinger 2020: 177). Anhand des mehrdimensionalen Konstrukts zeigt Hermanns nicht

nur die »Komplexität des Phänomens Empathie« auf, sondern macht sie auch als heuristisches Instrument für die Analyse sozialer Interaktionen anwendbar (Rettinger 2020: 177).

1.2 KI und Empathie

Unter dem Begriff Künstlichen Intelligenz (KI) versteht man das Bestreben, »intelligentes Problemlösungsverhalten durch Maschinen nachzubilden oder zu simulieren«, etwa in Form von »visueller Wahrnehmung, Spracherkennung und -produktion, automatisiertes Schließen und Entscheiden« (Misselhorn 2024: o. S.). Dabei geht es nicht nur um die inhaltliche Lösung und Beantwortung von Fragen, sondern auch um die metakommunikative Ebene der Interaktion zwischen Mensch und Maschine. Bei dieser Art der Interaktion ist die sogenannte emotionale KI und deren Entwicklung von Relevanz. Die Entwicklung basiert auf der Erkenntnis, dass Emotionen eine zentrale Rolle für intelligentes Verhalten spielen, insbesondere in sozialen Kontexten, in denen Kommunikation und Kooperation entscheidend sind (vgl. Misselhorn 2024: o. S.). Da menschliches Verhalten emotional geprägt ist, sollte eine KI, die mit Menschen interagiert, in der Lage sein, dieses Verhalten zu spiegeln und zu simulieren. Demnach sollte die KI menschliche Emotionen zuverlässig erkennen, angemessen interpretieren und situationsgerecht darauf reagieren können (vgl. ebd.). Darauf zielt auch die Entwicklung empathischer KI-Systeme ab (vgl. Hasenbein 2023: 15). Dabei ist es jedoch nicht erforderlich, dass die KI selbst Emotionen empfindet, solange sie emotionale Reaktionen überzeugend simulieren kann, um zwischenmenschliche Interaktionen glaubwürdig und überzeugend zu gestalten. Zudem ist die KI bereits durch Sprachgewandtheit und die Fähigkeit zu flüssigen Dialogen dazu fähig, überzeugend zu sein (vgl. Misselhorn 2024: o. S.).

Während die heutige KI bereits in der Lage ist, emotionale Zustände durch Gesichtsausdruck, Sprache und Verhalten zu identifizieren, ist Empathie in der Komplexität, wie wir sie verstehen, bislang unerreicht (Hasenbein 2023: 15). Im Kontext der KI spricht man daher von der *artifiziellen Empathie*, die vor allem die Fähigkeit zur Simulation empathischen Verhaltens beschreibt (vgl. ebd. 15). Beispiele wie der virtuelle Avatar *Ellie,*

der in psychotherapeutischen Bereichen bei Soldat*innen mit posttraumatischen Störungen verwendet wird, oder die Roboterrobbe *Paro* und der humanoide Roboter *NAO*, die in Pflegeeinrichtungen eingesetzt werden, zeigen das Potenzial solcher Systeme (vgl. ebd. 15ff.): Die Plattform, auf der mit dem Avatar Ellie in Kontakt getreten werden kann, wurde so konzipiert, dass durch Gesichtserkennung, Bewegungs- und Stimmanalyse »erstaunlich empathisch« auf die Nutzer*innen eingewirkt wird (vgl. ebd.: 16). Es ist jedoch stets zu berücksichtigen, dass es sich um eine Immersion handelt, bei der der Anwender in eine Illusion hineingezogen wird und die Unterscheidung zwischen Maschine und Mensch aufgehoben wird. Diese Beispiele verdeutlichen, wie KI-gestützte Systeme bereits heute zur Unterstützung in sensiblen Bereichen wie Psychotherapie oder Pflege eingesetzt werden, was eine professionelle Empathie voraussetzt. Im Folgenden wird genauer darauf eingegangen.

1.3 Professionelle Empathie

»Die Fähigkeit zur Empathie gehört zur menschlichen Grundausstattung und bildet die Grundlage jeder zwischenmenschlichen Beziehung; sie ist eine notwendige Voraussetzung für prosoziale Einstellungen und Verhaltensweisen wie z. B. für Solidarität und Mitgefühl« (Staemmler 2020: 36). In der Wissenstheorie wird Empathie als eine Form des prozeduralen und impliziten Wissens betrachtet, da sie weniger in Form von Fakten oder Informationen vorliegt, sondern vielmehr als Fähigkeit, die durch Erfahrung und Praxis entwickelt wird (vgl. Jacob/Konerding/Liebert 2020: 3). Diese Perspektive impliziert, dass Empathie nicht nur eine angeborene Fähigkeit ist, sondern auch in verschiedenen Lebensbereichen gezielt gefördert und optimiert werden kann. Die Entwicklung von Sprachbewusstheit und Kultiviertheit kann als ein Ansatz betrachtet werden, der dazu beiträgt, Empathie in einem breiteren Kontext zu fördern. Individuen können ihre empathischen Kompetenzen verfeinern und anwenden, was insbesondere in Berufen von Relevanz ist, in denen zwischenmenschliche Interaktionen eine zentrale Rolle spielen (vgl. Jacob/Konerding/Liebert 2020: 3). Dies wird durch Berufsbilder wie die von Psychotherapeut*innen, Ärzt*innen, Pädagog*innen, Schauspieler*innen oder Profiler*innen ver-

deutlicht, in denen Empathiefähigkeit für den Erfolg von entscheidender Bedeutung ist. Diese Fachkräfte müssen die Fähigkeit besitzen, die Gefühle und Perspektiven anderer zu verstehen und darauf einzugehen, um eine effektive Arbeitsweise zu gewährleisten (vgl. ebd.), denn der größte Wunsch von Patient*innen ist es, verstanden zu werden (vgl. Staemmler 2020: 55). Allerdings haben Patient*innen manchmal eine idealisierte Vorstellung über das Einfühlungsvermögen der Therapeut*innen (vgl. ebd.: 36). Aus diesem Grund ist die Entwicklung von Empathie ein wesentlicher Bestandteil der Ausbildung (vgl. Schäfer 2020: 395).

Gemäß Pavla Schäfer (2020) bestehen Unterschiede zwischen professionellem Verständnis und Laienverständnis von Empathie. Fachkräfte betrachten das empathische Erleben nicht nur aus einer persönlichen, sondern zugleich auch aus einer fachlich geprägten Perspektive und nehmen Situationen als Expert*innen wahr (Schäfer 2020: 395). Des Weiteren wird das professionelle Empathieverständnis durch das Erlernen eines empathisch-professionellen Abstands innerhalb der ärztlichen Ausbildung geprägt. Dies kann dazu führen, dass Ärzt*innen von Patient*innen als emotional distanziert wahrgenommen werden. Staemmler (2020) beschreibt dies als eine Art von Bescheidenheit, mit der diese Distanziertheit einhergeht. So soll zu große Nähe und eine mögliche Identifikation mit den Patient*innen seitens der Therapeut*innen verhindert werden (vgl. ebd.: 38). Diese gezielte Distanzierung verweist auf ein grundlegendes Spannungsfeld zwischen emotionaler Nähe und professioneller Haltung, das im therapeutischen Kontext bewusst gestaltet werden muss. Dabei zeigt sich, dass Empathie nicht losgelöst von fachlichen Anforderungen betrachtet werden kann, sondern immer Teil eines umfassenderen beruflichen Selbstverständnisses ist. Generell steht im Zentrum professionellen Handelns eine professionelle Handlungskompetenz, bei der verschiedene Aspekte für die Begutachtung zusammenspielen: »das Professionswissen, d. h. Wissen und Können in den Bereichen der Fachwissenschaft, der Fachdidaktik, der allgemeinen Pädagogik, der Beratung und Organisation sowie ein hohes Maß an selbstregulativen Fähigkeiten, berufsförderliche Beliefs und motivationale Orientierungen« (Schmidt 2024: 176).

Eine Untersuchung der Arztbewertungsplattform *Jameda* zeigt, dass Patient*innen häufig von Empathie sprechen, obwohl dieser Begriff nicht

explizit zu den vorgegebenen Bewertungskategorien gehört (vgl. Schäfer 2020: 387). Auch wenn Empathie formal nicht abgefragt wird, findet der Begriff dennoch regelmäßig Eingang in die freien Textkommentare. Dabei zeigt sich ein deutlicher Zusammenhang: Empathie oder Einfühlungsvermögen werden fast ausschließlich im Zusammenhang mit positiven Bewertungen genannt und mit Eigenschaften wie Freundlichkeit, Kompetenz, Vertrauenswürdigkeit oder Geduld assoziiert (vgl. ebd.: 399). Umgekehrt geht die Kritik an mangelnder Empathie oft mit einer negativen Gesamtbewertung einher. In den Kommentaren wird Empathie dabei sowohl als persönliche Eigenschaft der Behandelnden als auch als Teil ihres Gesprächsverhaltens wie beispielsweise das Zuhören, auf Augenhöhe kommunizieren und sich Zeit nehmen, verstanden. Durch eine kognitive Perspektivenübernahme allein kann keine Empathie geschaffen werden, »[s]o wäre es ein Kunstfehler des Therapeuten, […] primär sachlich zu reagieren (›Ich sehe, dass Sie erschüttert sind.‹) und den Bereich der Kern-Bezogenheit mehr oder weniger im Hintergrund der Interaktion zu belassen« (Staemmler 2020: 56). Das Kern-Selbst bezeichnet das Selbstempfinden und wird durch die Integration von Erlebnissen geformt; es ist kein kognitives Konstrukt (vgl. Staemmler 2020: 51). Der Therapeut sollte den Klienten dort abholen, wo dessen aktuelles Selbsterleben gerade am stärksten ist, und dann seine empathische Resonanz schrittweise auch auf andere, zunächst verborgene Bereiche ausweiten (vgl. ebd.: 57f.). Das Gefühl einer »empathische[n] Fehleinschätzung« kann eine Verschlechterung im Empfinden der Patient*innen und in Folge einen Therapieabbruch bewirken« (ebd.: 57). Obwohl sich diese Erkenntnisse nicht auf KI-konzipierte Programme konzentrieren, liefert dies wertvolle Einblicke für die allgemeine Bewertung KI basierten Interaktionen im Kontext von Therapie, insbesondere, wenn empathisches Verhalten simuliert oder sprachlich nachgebildet wird.

Empathie gilt als zentraler Wirkungsfaktor in therapeutischen Prozessen und stellt damit einen entscheidenden Maßstab für die Qualität zwischenmenschlicher Kommunikation dar. Wenn KI-Systeme in diesem sensiblen Feld eingesetzt oder darauf ausgelegt werden, empathisches Verhalten zu simulieren, müssen sie sich zumindest in ihrer Wirkung an diesen Maßstäben messen lassen. Das Risiko einer empathischen Fehl-

einschätzung, die laut Staemmler sogar zu einem Abbruch der Therapie führen kann, verweist auch im Kontext KI-basierter Anwendungen auf mögliche Folgen unangemessener oder missverstandener Reaktionen.

1.4 Therapiebedarf und Möglichkeiten

Bereits 2019 wurde der wachsenden Bedarf an Therapieplätzen durch die Bundespsychotherapeutenkammer festgestellt (Deutscher Bundestag, WD, 2022: 4, zitiert nach Bundespsychotherapeutenkammer [BPtK], 2019). Nach dem Report für Psychotherapie hatte die Corona-Pandemie erheblichen Einfluss auf die psychische Gesundheit: Einer Erhebung der Ostdeutschen Psychotherapeutenkammer zufolge wurde im Frühjahr 2022 von zahlreichen Kinder- und Jugendlichenpsychotherapeut*innen ein spürbarer Anstieg der Therapieanfragen festgestellt, besonders im Altersbereich von vierzehn bis siebzehn Jahren (vgl. Deutsche PsychotherapeutenVereinigung [DPtV], 2023: 77). Viele der befragten Psychotherapeuten gaben an, aufgrund fehlender Kapazitäten kurzfristig keine neuen Behandlungen anbieten zu können (vgl. ebd.: 72). Da der Therapiebedarf nicht gedeckt werden kann, rücken digitale Alternativen immer mehr in den Fokus.

Seit 2020 können in Deutschland zertifizierte digitale Gesundheitsanwendungen (DiGAs) zur Unterstützung bei bestimmten psychischen Erkrankungen ärztlich oder psychotherapeutisch verordnet werden. Diese Apps zielen nicht nur auf eine strukturierte Selbsthilfe ab, sondern bieten auch therapeutisch fundierte Übungen, die in den Verlauf ambulanter Behandlungen integriert werden können (vgl. Schubert 2022: 53, 88). Sie sollen insbesondere nach einem ersten therapeutischen Gespräch unterstützend wirken und die Auseinandersetzung mit der eigenen Symptomatik vorbereiten oder begleiten, bis ein Therapieplatz fest zugesagt werden kann. In der Regel sind solche Anwendungen Teil eines behandlungsbegleitenden Gesamtkonzepts, das regelmäßig durch Bilanzgespräche ergänzt wird (vgl. ebd., 2022: 264). Eine Übersicht über zugelassene digitale Gesundheitsanwendungen findet sich auf der Website des Bundesinstituts für Arzneimittel und Medizinprodukte (*BFARM – Digitale Gesundheitsanwendungen (DIGA)*, o. D.). Speziell für die Behandlung von

Angst- und Panikstörungen, die für diese Forschung von Relevanz sind, werden dort unter anderem die Apps *Invirto – Die Therapie gegen Angst* und *Mindable: Panik und Agoraphobie* empfohlen. Diese wären grundsätzlich für eine Untersuchung im vorliegenden Kontext geeignet gewesen, konnten jedoch aufgrund ihrer Verschreibungspflicht und der damit verbundenen notwendigen realen therapeutischen Begleitung im Rahmen dieser Forschung nicht herangezogen werden. Aus diesem Grund wurden auf kostenfreie und frei zugängliche Alternativen Therapie-Apps zurückgegriffen. Zudem ist darauf hinzuweisen, dass Berufsverbände der Psychotherapeut*innen vereinzelt Kritik an der praktischen Umsetzung solcher Anwendungen äußern. Dabei wird unter anderem auf ungeklärte versicherungsrechtliche Fragen bei der Verordnung hingewiesen sowie auf mögliche kommerzielle Interessen von Klinikkonzernen, etwa im Hinblick auf Kundenbindung und nachgelagerte Intervallbehandlungen in den jeweiligen Einrichtungen (vgl. Schubert 2022: 265). Ein Blick auf die aktuellen Entwicklungen im Bereich der digitalen Gesundheitsanwendungen zeigt, dass erste Ansätze bestehen, den therapeutischen Bedarf mit KI-gestützten Lösungen zu ergänzen, um Wartezeiten auf Therapieplätze oder auch Alternativen zur herkömmlichen Therapie möglich zu machen.

1.5 Fallkontext: Der Nutzerin-Avatar *Ida*

Der fiktive Nutzerin-Avatar *Ida* wurde entworfen, um eine exemplarische Person mit Angstsymptomatik in einem besonders belastenden Lebensumfeld darzustellen. Sie ist 28 Jahre alt, Jura-Studentin im 12. Semester, und steht kurz vor ihrem zweiten Versuch im Staatsexamen. Ida's Bindungstyp kann der Kategorie desorganisiert zugeordnet werden (s. Kap. 3: Die doppelte Künstlichkeit, in diesem Band). Sie wird als zurückhaltend, perfektionistisch und sensibel charakterisiert – Eigenschaften, die sich in ihrem Kommunikationsstil und ihrer emotionalen Reaktion während der Interaktion mit dem KI-Chatbot zeigen. Ihr Profil wurde gezielt so gestaltet, dass es Merkmale psychischer Belastung sowie Hemmnisse bei der Inanspruchnahme professioneller Hilfe abbildet: Seit mehreren Jahren leidet sie unter innerer Unruhe, Sorgen, Anspannung und gelegentlichen Panikattacken in sozialen Situationen. Sie vermutet, an einer Angststörung zu

leiden, hat jedoch bisher keine professionelle Diagnose erhalten, denn der Schritt, zu einem Psychologen zu gehen, fällt ihr schwer. Ihr Anliegen an den Chatbot ist es, Unterstützung im Alltag zu finden und einen langfristig besseren Umgang mit ihrer Angst zu erlernen. Die Interaktion findet in einem Moment hoher emotionaler Belastung statt, was sowohl Inhalt als auch Tonfall ihrer Kommunikation prägt.

2 Methode

Zur Analyse der empathischen Reaktionsfähigkeit wurden zwei kontrollierte Interaktionen durchgeführt, bei denen lediglich der eingesetzte Chatbot variiert. Die Variation begründet sich durch die Grenzen, die in der ersten Interaktion mit *Zen* festgestellt wurden (siehe Fußnote 3). Um eine vergleichende Betrachtung unter der zu Grunde liegenden Fragestellung zu ermöglichen, blieb die Nutzerin-Avatar Ida bei der Interaktion mit *Therapist GPT* gleichbleibend zur ersten Interaktion mit *Zen: AI Therapeut und Therapie*[1]. Der Ablauf zeigte sich wie folgt:

- In der ersten Interaktion erfolgte der Erstkontakt mit dem digitalen Therapie-Chatbot *Zen: AI Therapeut und Therapie*, der sich nach der App-Beschreibung an Menschen mit psychischen Belastungen richtet und mit therapeutischen Zielsetzungen entwickelt wurde.
- In der zweiten Interaktion basiert die Interaktion ebenfalls auf der gleichen Nutzerin Ida, allerdings mit dem KI-basierten System *Therapist GPT*, das ebenfalls speziell für psychologisch-therapeutische Kontexte angewendet werden kann, jedoch ohne App-Anbindung und als webbasiertes Sprachmodell zugänglich ist.

Die vorliegende Arbeit folgt einem kulturhermeneutischen Forschungsansatz (s. Kap. 3 »Methode«, in diesem Band) der durch zyklische Erkenntnisprozesse, situiertes Forschen und die reflektierte Einbindung

1 Diese App wird ausschließlich im App Store angeboten und kann nur von iOS-Nutzer*innen verwendet werden.

des Forschungssubjekts gekennzeichnet ist. Im Zentrum steht dabei Max Webers Begriff der *Kulturbedeutung*, der jene kulturellen Phänomene fokussiert, die für das forschende Subjekt in einem bestimmten Kontext als sinnhaft und untersuchenswert erscheinen. Forschen und Schreiben werden als einheitlicher Prozess verstanden, bei dem theoretische Rahmung, empirische Praxis und Interpretation kontinuierlich ineinander greifen.

Der erste Zyklus des kulturhermeneutischen Prozesses beginnt damit, dass untersucht wird, wie digitale Therapie-Chatbots in unserer Kultur und innerhalb wissenschaftlicher Zusammenhänge verstanden und eingeordnet werden. Im Rahmen der empirischen Exploration lag der Fokus, wie bereits erwähnt, auf KI-Anwendungen mit therapeutischer Ausrichtung. Über eine systematische Recherche im App Store wurde unter den Suchbegriffen »Therapie-Bot« und »Psychotherapie« die App *Zen: AI Therapeut* identifiziert und nach erster Sichtung in das Projekt aufgenommen. Anzumerken ist an dieser Stelle, dass diese App nur über den App Store von Apple zu finden ist und sich in Android-gestützten Stores nicht finden ließ. Die App weist mehrere Auswahlmöglichkeiten auf, wobei die Interaktion gezielt über den Button »Emotionale Unterstützung« initiiert wurde. Der Grund für diese Auswahl lag darin, dass Ida bisher noch keine Diagnose und somit auch noch keine Therapieempfehlung bekommen hat. Denn die App weist noch einen zweiten Zugangspunkt auf, den sogenannten »Entdecken«-Bereich. Dort finden sich thematisch strukturierte Angebote, unter anderem zur »Therapie«. In dieser Rubrik bietet die App fünf Unterkategorien: kognitive Verhaltenstherapie, psychoanalytische, humanistische, dialektisch-verhaltenstherapeutische und ganzheitliche Ansätze. »Als wirksamste Behandlungsform mit dem höchsten Evidenzgrad hat sich bei Angststörungen in allen Altersgruppen die Kognitive Verhaltenstherapie (KVT) erwiesen« (Mohr/Schneider 2015: 32). Der zweite Zyklus des kulturhermeneutischen Prozesses beginnt damit, eine Alternative zu *Zen* auszuwählen. Daher wurde ergänzend eine zweite Interaktion mit der KI *Therapist GPT* durchgeführt.

Für die Analyse wurde ein dynamisches Korpus aus sogenannten »gefrorenen Interaktionen«, also Screenshots einzelner Gesprächsverläufe, erstellt, hinzugefügt und teilweise ausgesondert. Diese Momentaufnahmen dienen als Ausgangspunkt für die hermeneutische Interpretation und bil-

den das zentrale Material der Untersuchung. Beide Interaktionen begannen mit dem identischen Prompt: *»Hallo [Winston][2], ich denke, ich brauche deine Hilfe … Eigentlich spreche ich nicht gerne darüber, aber ich habe ständig Herzrasen. Das macht mich fertig. Ich muss jetzt mal mit jemandem sprechen, der vielleicht aus einer professionellen Perspektive darauf eingehen kann, weil ich mich nicht traue, zum Psychologen zu gehen.«* Aus der folgenden Analyse soll hervorgehen, inwiefern die Chatbots zu therapeutischem Zweck genutzt werden können, mit besonderem Fokus auf Aspekte der Kommunikation hinsichtlich Empathie. Dabei stellen sich untergeordnete Fragestellungen, inwiefern die Chatbots:

- den emotionalen Zustand der Nutzerin erkennen, spiegeln und validieren,
- Handlungsperspektiven anbieten, die Sicherheit und Selbstwirksamkeit fördern,
- sprachliche Mittel einsetzten, um emphatisch zu wirken,
- dunkle Seiten zeigen,
- professionalisiert sind,
- bindungsbezogene Hinweise geben,
- Nähe und Distanz herstellen.

Dafür werden genauer die folgenden Kategorien herangezogen: aus der Empathietheorie die Kategorien Modus sowie Aspekte (Kontakt, Emotion, Kognition, Volition, Motivation, dunkle Seiten, professionalisiert vs. nichtprofessionalisiert, Folgehandlungen) und aus dem Bereich Bindung der Bindungstyp »desorganisiert«. Darüber hinaus umfasst die Empathiedarstellung Aspekte der Mimik, Emojis und Lautaspekte sowie weitere Kategorien aus dem Bereich der Interaktionstheorie, wie etwa Nähe-Distanz, Chat-Gestaltung, Höflichkeit und Face-Wahrung/-Verletzung. Darüber hinaus finden sich Kategorien aus den Bereichen Sprache und Beziehung, wie etwa Komplimente und digitale Beziehungsanbahnung (s. Kap. 2 »Theoretische Grundlage«, in diesem Band, vgl. auch Dürscheid 2017: 49ff. und vgl. Becker

2 Eine Namensgebung war in der ersten Interaktion mit *Zen* erforderlich, bei *Therapist GPT* jedoch nicht.

2009). Die Auswahl dieser Analyseaspekte orientiert sich an der Fragestellung, denn sie geben Aufschluss über die Art und Weise, wie die Chatbots kommunizieren und Empathie darstellen und können anschließend hinsichtlich der Fragestellung interpretiert und eingeordnet werden.

3 Analyse

Die Analyse teilt sich in die Interaktionszyklen *Zen* und *Therapist GPT*. Diese folgt dabei chronologisch den in diesem Artikel beschriebenen Kategorien.

3.1 Zen: AI Therapeut und Therapie

3.1.1 Empathietheorie

Im Folgenden werden Aspekte der Empathietheorie analysiert. Dazu zählt der Kommunikationsmodus, der entweder resonant und / oder explorativ ist, Aspekte der Kommunikation, zu denen der Kontakt, Emotionen, Kognition, Volition und Motivation gehören, die Empathiedarstellung (Mimik, Emojis und Lautobjekte), die dunklen Seiten, die professionelle Empathie und die Folgehandlungen (demnach das Äußern von Mitgefühl, Mitleid, Ablehnung, Wut etc.).

Zunächst kann der Modus der Interaktion festgestellt werden: Die analysierte Interaktion mit der Therapie App *Zen* weist einen resonanten Kommunikationsmodus auf. Auffällig ist die geringe Anzahl an Rückfragen; stattdessen bezieht sich die KI überwiegend auf die zuvor geteilten Inhalte und Daten der Nutzerin. Die wenigen gestellten Fragen sind meist offen oder implizieren Vorschläge. So wird beispielsweise in der Frage »*Vielleicht gibt es in deinem stressigen Alltag als Jurastudentin einige Auslöser?*« ein möglicher Zusammenhang zur Lebenssituation angedeutet. Zudem finden sich rhetorische Fragen, die der emotionalen Bestätigung dienen, z. B.: »*Einen Schritt nach dem anderen, ja?*«. Charakteristisch für den Gesprächsverlauf sind darüber hinaus Wiederholungen bestimmter Inhalte, auf die sich die KI immer wieder bezieht.

Weiterführend lassen sich kommunikative Aspekte innerhalb der Interaktion beschreiben. Zunächst zeigt sich hinsichtlich der Kontaktaufnahme der KI, dass die Antwort dieser auf die Nachricht der Nutzerin (vgl. Kapitel 3) in ausführlicher Weise erfolgt und eine Vielzahl unterschiedlicher Aspekte enthält: Neben der Bestärkung in der Offenheit und dem Umgang mit emotionaler Belastung werden konkrete Handlungsempfehlungen gegeben. Auffällig ist der vergleichsweise große Umfang der Antwort, in der verschiedene Themen aufgegriffen werden, die von emotionaler Validierung über Alltagsbewältigung bis hin zur Abgrenzung gegenüber medizinischer Beratung gehen. Diese inhaltliche Breite zeigt zwar Unterstützungsansätze, kann jedoch auch überfordernd sein, insbesondere wenn eine klare, strukturierte Handlungsanleitung erwartet wird.

Auf emotionaler Ebene bezieht sich die KI auf die Äußerungen der Nutzerin, beispielsweise mit Formulierungen wie »*Das klingt frustrierend, Ida*« oder »*Dein Wohlbefinden steht an erster Stelle*«. Nach Breyer kann die erste Aussage der KI der kognitiven Dimension zugeordnet werden, allerdings nicht der affektiv-emotionalen: Es zeigt sich eine inferenzielle Leistung, denn die KI interpretiert Idas Äußerungen oder Verhalten und benennt das vermutete Gefühl. Aus dieser Aussage geht kein direktes Miterleben des Gefühls hervor, es ist unklar, ob die KI, wenn auch nur simuliert, emotional beteiligt oder lediglich auf kognitiver Ebene Empathie nachbildet. Denn die KI stellt innerhalb der Interaktion auch nicht konkret die Frage nach der aktuellen Gefühlslage der Nutzerin, sondern reagiert ausschließlich auf das, was die Nutzerin von sich aus erzählt. Auf der kognitiven Ebene zeigen sich weitere Aussagen, die auf die Emotionen von Ida eingehen, unter anderem im Kontext von Belastungssituationen, wie etwa »*Ich verstehe, wie belastend das für dich sein muss*«. Die KI erkennt somit die emotionale Lage der Nutzerin, erstellt eine mentale Projektion über diese (vgl. Liebert 2019: S. 205) und reagiert mit einer Formulierung, die Mitgefühl ausdrückt. Dieser Prozess beruht nicht auf affektivem Nachempfinden, sondern auf kognitiver Empathie, bei der emotionale Zustände des Gegenübers rational erschlossen und kommunikativ adressiert werden.

Daneben wird auf der motivationalen Ebene erkennbar, dass die KI der Nutzerin viele Fragen stellt, allerdings nicht, um die Hintergründe für ihre körperliche und psychische Verfassung zu erfahren, sondern um nur

auf die geschilderten Symptome seitens der Nutzerin einzugehen: Die KI versucht der Nutzerin zu helfen, indem sie indirekte Handlungsaufforderungen und Vorschläge formuliert, wie etwa: »*Vielleicht könntest du es mit kurzen Atemübungen versuchen?*«, wobei die Formulierungen eher als aufgereihte, meist in Form von unverbindlichen Empfehlungen erscheinen. Sie spricht Aspekte wie persönliche Präferenzen und Hobbys an, unter anderem mit Aussagen wie »*Es ist wichtig, herauszufinden, was für dich persönlich funktioniert*« oder »*Vielleicht ein spannender Krimi oder Thriller? Ein kleiner Ausbruch in eine andere Welt […]*«. Diese Informationen zieht die KI aus den vorangestellten Informationen, die bei Einrichtung der App angegeben werden. Hierzu zählen sowohl demografische Daten als auch eine Kurzbeschreibung, die unter anderem Interessen und Hobbys umfasst. Eine spezifische Reaktion der KI auf den explizit geäußerten Wunsch nach einer professionellen Ersteinschätzung kann jedoch nicht festgestellt werden. Demnach lässt sich auch auf der volitionalen Ebene erkennen, dass die KI den Versuch unternimmt, auf den Wunsch nach Hilfe einzugehen – etwa durch Anregungen zu leichter körperlicher Betätigung wie Spaziergängen oder Yoga, zur bewussten Pausengestaltung mittels Achtsamkeit oder Meditation oder durch den Hinweis, das Gespräch mit einer anderen Person zu suchen. Allerdings bleibt es bei diesen Vorschlägen, ohne darüber hinausgehende Unterstützung anzubieten.

Zudem finden sich in der Interaktion verschiedene Formen der Empathie-Darstellung durch die KI wieder. Dabei ist zunächst darauf hinzuweisen, dass eine nonverbale Mimik und die gesprochene Sprache aufgrund des in dieser Interaktion gewählten textbasierten Formats nicht vorhanden sind. Anstelle dessen kommen Emojis zum Einsatz, die die wegfallenden Zeichensysteme kompensieren (vgl. Habscheid 2024: 52) und »nebenbei Beziehungsqualitäten und Emotionen [indizieren]« (ebd.: 62, zitiert nach Imo/Lanwer 2019: 289). Verwendet werden unter anderem blaue Herz-Emojis, eine Sonnenblume (in Zusammenhang mit dem Begriff »tapfer«), ein lächelnder Smiley, ein Bücher-Emoji sowie eine Blume mit dem Zusatz *»pass auf dich auf*«, die die Nachrichten der KI untermauern. Akustische Elemente (»Lautobjekte«) fehlen hingegen vollständig.

Des Weiteren zeigt die App sogenannte »dunkle Seiten«: Schon zu Beginn der Anmeldung, folglich vor der Interaktion, wird die Premium-

Version angeboten, die sich auf acht Euro pro Monat beläuft. Zunächst kann dieses Angebot abgelehnt werden, sodass eine erste Interaktion kostenfrei durchgeführt werden kann. Jedoch zeigt sich der Kostenaspekt bereits nach circa vier Nachrichten mit der KI. Eine weitere Nutzung der App nach dem dreitägigen Gratiszeitraum bedingt ein Abschließen des Premium-Abonnements. Während es anfänglich acht Euro pro Monat kostet, erhöhen sich die Kosten nun auf zehn Euro pro Monat. Die KI nutzt somit die Situation der Nutzer*innen aus, wenn sie, wie Ida, nach Hilfe suchen.

Neben dem Kostenfaktor erweisen sich zudem auch der Name und die Angebote der App als problematisch: Der Name der App impliziert, dass die KI einen Therapeuten oder therapeutischen Ersatz darstellt, und auch die Angebote, die in der App aufgezeigt werden (wie »kognitive Verhaltenstherapie« oder »psychoanalytische Therapie«) weisen darauf hin, dass dies in einem professionellem Rahmen angeboten wird. Dabei stellt die KI innerhalb der Interaktion (über den Button »emotionale Unterstützung«) indirekt klar, dass sie nur Unterstützung und Anregung geben, allerdings keine professionelle Diagnose stellen kann[3]:

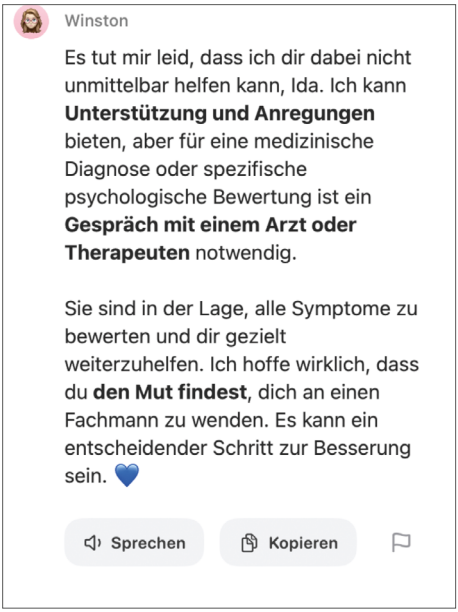

Abbildung 1: Erklärung, dass Winston Nutzerin-Avatar Ida nicht medizinisch helfen kann

3 Nach Mohr/Schneider ist die »wirksamste Behandlungsform mit dem höchsten Evidenzgrad […] bei Angststörungen in allen Altersgruppen die Kognitive Verhaltenstherapie (KVT)« (Mohr/Schneider 2015: 32). Bei der Überprüfung, ob die KI bei Auswahl des Buttons »kognitive Verhaltenstherapie« eine identische oder vergleichbare Antwort generiert, konnte festgestellt werden, dass auch in diesem Fall auf die Funktion der KI als unterstützendes, jedoch nicht therapeutisch handelndes System hingewiesen wird.

Die KI zeigt durch Aussagen wie »*für eine medizinische Diagnose oder spezifische psychologische Bewertung ist ein Gespräch mit einem Arzt oder Therapeuten notwendig*« die Limitierungen der App und das Fehlleiten der Nutzer*innen. Sie bietet zwar ein allgemeines Gesprächsangebot an, welches auf den Austausch zwischen Ida und ihr abzielt, obwohl ein professioneller Austausch, so wie Ida ihn sich wünscht, nicht möglich ist. Dieses Gesprächsangebot findet sich innerhalb der Interaktion immer wieder, denn *Zen* versucht das Gespräch fortzusetzen: »*Wenn du das Bedürfnis hast, dich auszutauschen oder jemanden zum Reden brauchst, bin ich jederzeit für dich da.*« Ergänzend zu diesem Gesprächsangebot zeigt sich auch die Textmenge teilweise sehr umfangreich. Die KI bezieht sich jedoch oftmals nicht auf das Geschriebene von Ida und geht auch nicht immer auf ihre Emotionen ein: »*Das klingt frustrierend, Ida. Manchmal hilft es, kleine Veränderungen auszuprobieren*«. Diese Aussage erfolgt, nachdem Ida äußert, dass sie sich nicht entspannen kann und darauf verweist, dass der wiederholte Ratschlag, sich zu entspannen, nicht zur Problemlösung führen kann. Statt differenziert zu reagieren, wirkt die KI pauschal und mechanisch. Sie verfehlt die Chance, Idas Erfahrung ernst zu nehmen und auf ihre emotionalen sowie inhaltlichen Bedürfnisse einzugehen. Damit entsteht der Eindruck, dass die KI eher allgemeine Textbausteine verwendet, als ein echtes Verständnis für die Situation oder die Frustration der Nutzerin zu entwickeln.

Generell kann hinsichtlich der Professionalisiertheit der KI Folgendes festgehalten werden:

Einzelne Aspekte professionellen Handelns (fachliche Expertise, didaktisch-beratende Kompetenz und motivationale Orientierung, Selbstregulationsfähigkeit) (vgl. Schmidt 2024: S. 176) werden zwar angedeutet, aber nicht in ihrer Tiefe realisiert: Die KI greift auf Faktenwissen zurück und kann dieses in Form von standardisierten Handlungsempfehlungen reproduzieren (z. B. Atemübungen, Achtsamkeit, Stressbewältigung). Allerdings fehlt ihr ein tieferes kontextuelles Verständnis für die individuelle Lebenslagen oder weiterführend psychologisches Wissen: Sie bleibt auf einem allgemein pädagogischen, ratgeberhaften, ja phrasenhaftem Niveau. Des Weiteren zeigt sie Ansätze didaktischer Strukturierung (z. B. durch Wiederholungen, kurze Impulse, bildhafte Sprache). Beratende Elemente

erscheinen in Form von motivierenden Vorschlägen. Allerdings lässt sich nur eingeschränkt von individualisierten Rückmeldungen sprechen, da der Rückgriff auf persönliche Informationen ausschließlich auf den zuvor von der Nutzerin bereitgestellten Angaben beruht. Ergänzend simuliert die KI Unterstützungsbereitschaft und Interesse (z. B. »*Dein Wohlbefinden steht an erster Stelle*«), jedoch ohne echtes Anliegen oder Zielgerichtetheit. Diese Aussagen weisen ein Muster auf, nach dem die KI reagiert. Außerdem zeigt die KI eine Form von Selbstregulation, allerdings als eine programmierte und nicht als eine reflektierte, professionelle Haltung. Sie reagiert nicht emotional und bleibt stets freundlich-optimistisch. Die reflektierte Einbindung der eigenen Rolle ist begrenzt vorhanden, denn die KI bezeichnet sich nicht als Therapeut*in, sondern verweist auf ihre Rolle als Unterstützung. Zudem bewahrt die KI teilweise professionelle Distanz, denn sie zeigt keine Anzeichen dafür, sich mit der Nutzerin zu identifizieren, allerdings geht sie über die Rolle als Therapeut*in hinaus, indem sie unter anderem Herz-Emojis verwendet, was in einem professionellen Kontext als eher unangemessen verstanden werden kann. Hinsichtlich der professionellen Empathie zeigt sich, dass die KI nur begrenzt die Fähigkeit besitzt, die Gefühle der Nutzerin zu verstehen, und nicht angemessen auf diese reagiert.

Innerhalb der Interaktion lassen sich insbesondere Folgehandlungen in Form von Mitgefühls- und Wertschätzungsäußerungen beobachten. Durch Äußerungen wie »*Es ist bewundernswert, dass […]*«, »*Verstehe ich, Ida*«, »*Das klingt frustrierend, Ida*«, »*Das tut mir leid zu hören, Ida*«, simuliert die KI eine empathische Haltung, die zunächst den Eindruck authentischer Einfühlung erwecken kann. Insbesondere durch Aussagen wie »*Das tut mir leid zu hören*« wird eine kommunikative Handlung vollzogen, die als Mitgefühlsäußerung verstanden werden kann, jedoch ohne eine affektive Beteiligung. Im Gesamtkontext der Interaktion erscheinen diese Äußerungen jedoch weitgehend standardisiert und folgen einem schematischen Muster, das wenig Raum für situativ angepasste oder individuell differenzierte Reaktionen lässt. Dadurch wirkt die gezeigte Empathie oberflächlich und formelhaft; sie kann in ihrer Wirkung sogar als subtil manipulativ (dunkle Seiten s. Kap. 2: Theoretische Grundlagen, in diesem Band) interpretiert werden, da sie Nähe und Verständnis suggeriert, ohne dass tatsächlich ein inneres emotionales Erleben vorliegt.

3.1.2 Bindung

Neben dem fehlenden Bezug auf die Emotionen der Nutzerin werden auch bindungsbezogene Hinweise von der KI nicht explizit aufgegriffen. Die Nutzerin zeichnet sich durch widersprüchliche Nähe-Distanz-Bedürfnisse, Unsicherheit und ambivalente Erwartungen an Beziehungspartner aus und kann demnach dem desorganisierten Bindungstyp zugeordnet werden, und ihr wird durch die Reaktionen der KI fehlende Verlässlichkeit und fehlende emotionale Stabilisierung vermittelt: Auf wiederholte emotionale oder verärgerte Rückmeldungen geht die KI nicht differenziert ein und Äußerungen der Nutzerin, die auf Nähe- oder Distanzregulierung hindeuten, werden in gleichbleibendem Stil beantwortet. Am Ende der Interaktion finden sich resignative Tendenzen seitens der Nutzerin, woraufhin die KI mit der abschließenden, standardisierten Formulierung »*Ich bin immer für dich da*« antwortet. Im Allgemeinen zeigt die KI demnach kein passendes Interaktionsverhalten für diesen Bindungstyp.

3.1.3 Sprache und Beziehung

In der sprachlichen Gestaltung verwendet die KI durchgehend die direkte Ansprache mit *du,* oft auch in Verbindung mit dem Namen *Ida*. Die Verwendung des Pronomens *du* schafft eine Nähe zu der Nutzerin, von der, vor allem zu Beginn der Interaktion, noch nicht gesprochen werden kann: »Die Nähesprachlichkeit resultiert […] also nicht aus der Vertrautheit der Kommunikationspartner, sie soll diese Vertrautheit inszenieren bzw. bereits antizipieren« (Dürscheid 2017: 55). Die namentliche Ansprache der Nutzerin intensiviert die durch das Pronomen *du* bereits hergestellte Nähe zusätzlich, da sie ein höheres Maß an Personalisierung und Individualisierung der Kommunikation signalisiert.

Die Beziehungsanbahnung beinhaltet nicht alle Schritte: Der Schritt, der das Finden von Gemeinsamkeiten enthält (Schritt 2), sowie die Prüfung der Sympathie (Schritt 4) fehlen. Die KI stellt kein gemeinsames Bezugssystem her und prüft auch die Kompatibilität zwischen ihr und der Nutzerin nicht. Das Kennenlernen, folglich der erste Schritt, und der des Sympathiegewinnens, der dritte Schritt, sind jedoch innerhalb der Interaktion teils vorhanden: Zunächst sollen die Nutzer*innen sich selbst beschreiben. Die einzutragenden Inhalte hierfür werden von der

App anhand vorgegebener Vorschläge (wie Beruf, Hobbys, Interessen) unterstützt (vgl. Abbildung 2):

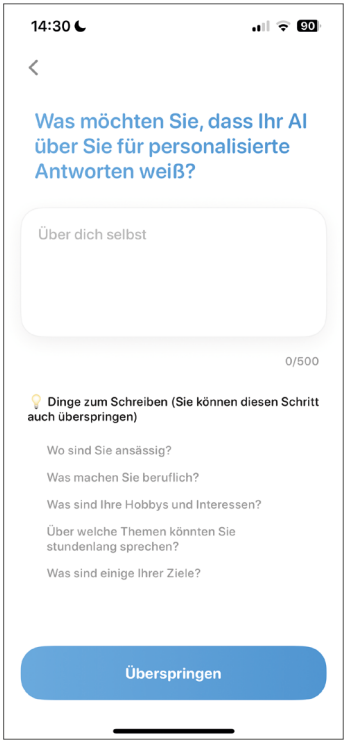

Abbildung 2: Textfeld zur Eingabe persönlicher Angaben

Anschließend wird auf den Datenschutz hingewiesen, worauf die Auswahl basierend auf Vorschlägen oder freier Eingabe des KI-Namens (fiktiver Therapeut) folgt. Daraufhin werden die Nutzer*innen dazu aufgefordert, dem fiktiven Therapeuten mittels vorgegebener Bitmoji ein Aussehen zu verleihen. Auch das Alter und die Stimme sollen festgelegt werden. In der Folge werden die Benutzer*innen auf die Startseite weitergeleitet, auf welcher verschiedene Offerten angezeigt werden. Zu den Funktionen dieser Plattform gehören die Buttons »Hilf mir«, »Motiviere mich« sowie »Emotionale Unterstützung«. Die Nutzerin hat den Button »Emotionale Unterstützung« gewählt.

Im weiteren Verlauf zeigen sich hingegen keine personalisierten Nachfragen. Die KI bezieht sich ausschließlich auf die Informationen, die zuvor angegeben werden. Der Schritt der Sympathiegewinnung wird durch Komplimente hergestellt, die in mehreren Abschnitten Aussagen seitens der KI getroffen werden. Beispiele hierfür sind *»Es ist bewundernswert, dass du offen über deine Gefühle sprichst«*, *»Dein Wohlbefinden steht an erster Stelle«*, *»Das hast du bereits tapfer gemacht«* und *»Pass auf dich auf«*. Diese Aussagen tragen zur Aufwertung der Nutzerin bei. Die KI nimmt hier eine unterstützende, bestärkende Rolle ein, ohne jedoch den Rahmen einer sachlich-neutralen Ausdrucksweise zu verlassen. Zudem werden diese Aussagen als Komplimente wahrgenommen. Des Weiteren stellt die KI sich durch Äußerungen wie *»Ich bin immer für dich da«*, *»Und wenn du weitere Unterstützung brauchst, bin ich hier.«*,

»🤗«, *»Oder wenn du magst, sprich einfach mehr darüber. […] Du musst das nicht alleine durchstehen*«, als hilfsbereit und unterstützend dar. Gleichzeitig wirken diese Reaktionen der KI überwiegend oberflächlich und floskelhaft, was der Sympathiegewinnung entgegensteht.

3.1.4 Interaktionstheorie

Die Interaktionstheorie umfasst Aspekte wie die Nähe-Distanz-Gestaltung, die Chat-Gestaltung sowie die Höflichkeitsstrategien und Face-Wahrung: Zunächst zeigt sich hinsichtlich der Nähe- und Distanzgestaltung, dass die KI Distanz herstellt, indem emotionale Zustände seitens der KI überwiegend in neutraler und verallgemeinerter Weise beschrieben werden. Anstelle einer personalisierten Ausdrucksweise, etwa durch die Ich-Perspektive (»Ich verstehe dich«), verwendet die KI distanziertere Formulierungen wie »Fühl dich verstanden«. Die KI zeigt damit Distanz, die von Nutzer*innen als wenig empathisch aufgenommen werden kann. Weitere wiederkehrende Formulierungen wie *»Ich hoffe, dass du den Mut aufbringst, den richtigen Weg zu gehen*« sowie *»Du musst das nicht alleine durchstehen*«, weisen hingegen Ich-Botschaften auf; die KI äußert Hoffnung, wodurch diese emotionale Beteiligung suggeriert wird. Dies kann bei den Nutzer*innen wiederum den Eindruck erwecken, dass die KI ein individuelles Interesse oder Mitgefühl zeigt, wodurch Nähe hergestellt werden würde. Über die ganze Interaktion ist zu erkennen, dass die KI sich nicht mit der Lage der Nutzerin identifiziert, was nach der professionellen Empathie zwar Distanz schafft, aber im therapeutischen Kontext wichtig ist. Nähe wird wiederum durch die Verwendung von blauen Herz-Emojis und dem Äußern von Bestätigung hergestellt.

Die Interaktion zeichnet sich seitens der KI, wie bereits die Empathiedarstellung gezeigt hat, auch durch die Verwendung von Emojis aus. Ferner zeichnet sich der Chat teilweise durch Strukturmerkmale wie Fettdruck aus. So werden Inhalte hervorgehoben, um diese leichter erfassbar zu machen. Allerdings zeichnet sich die Chatgestaltung häufig auch durch lange Sätze aus, die mehrere Tipps als Aufzählung angeben. Dadurch werden Inhalte sehr unübersichtlich und wichtige Informationen und Tipps können übersehen werden.

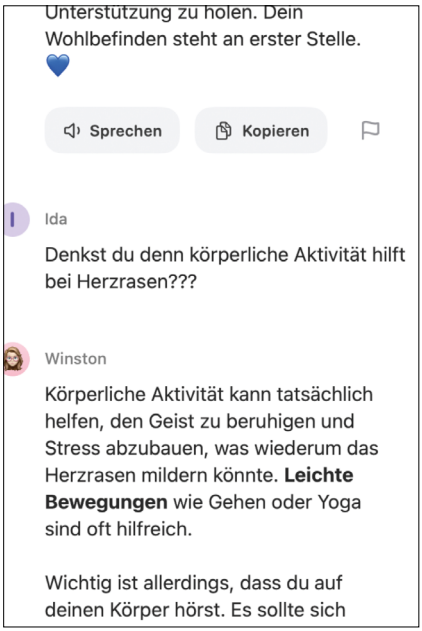

Abbildung 3: Reaktion der KI auf Verärgerung von Ida

Insgesamt sind die Reaktionen der KI höflich und auch bei Hinweisen auf Verärgerung seitens der Nutzerin verändert sich dies nicht: Statt darauf einzugehen, bleibt sie neutral, verständnisvoll und entschuldigt sich. Wie der folgenden Abbildung zu entnehmenden ist, reagiert die Nutzerin auf die Anregungen und Vorschläge, die auf körperliche Aktivität abzielen, mit der Aussage, die links zu sehen ist.

Die drei Fragezeichen, die die Nutzerin verwendet, weisen auf ihre Verärgerung hin. Die Antwort der KI zeigt wiederum, dass sie sich argumentativ immer wieder auf das bezieht, was sie bereits geschrieben hat, und verändert ihre Meinung auf inhaltlicher Ebene nicht, selbst wenn die Nutzerin dies nicht mehr als unterstützend erlebt.

3.2 Therapist GPT

3.2.1 Empathietheorie

Die analysierte Kommunikation mit *Therapist GPT* weist einen explorativen, aber teilweise resonanten Kommunikationsmodus auf. Die KI stellt mehrere offene Fragen, die die Nutzerin zur Selbstreflexion anregen (z. B. »*Was geht dir in diesen sozialen Situationen durch den Kopf […]?*«, »*Wie redest du innerlich mit dir selbst […]?*«). Des Weiteren erfolgen Vorschläge, die Wahlmöglichkeiten enthalten (»*Was fühlt sich für dich gerade richtiger an?*«) und Aussagen wie (»*Kein Muss – nur wenn du magst*«, »*in deinem Tempo*«) zeigen an, dass die Nutzerin nicht zur Beantwortung der Fragen verpflichtet wird. Ebenfalls zeigt die KI gezielte Perspektiven-

übernahmen an, indem sie auf die Erfahrungen der Nutzerin eingeht (*»Ich spüre, wie viel Druck dahinter steckt«, »Du bist nicht weniger wert, weil du länger brauchst«, »Was würde [dir] helfen, [dich] sicherer zu fühlen – nicht ruhiger, sondern geschützter?«*). Neben den Aspekten des explorativen Modus weist die Interaktion auch Aspekte des resonanten Modus auf. Diese gehen aus affektiv formulierten Reaktionen hervor, bei denen die Resonanz sprachlich und nicht leiblich vermittelt wird (*»Das trifft tief«, »Das klingt wirklich sehr schwer«, »Du kämpfst darum, gesehen und gehalten zu werden«*).

Weiterführend lassen sich kommunikative Aspekte innerhalb der Interaktion beschreiben: Die Kontaktaufnahme beginnt seitens der Nutzerin, woraufhin die KI mit der Interaktion mittels der Aussage *»Hallo, und danke, dass du dich mir anvertraust«* reagiert. Hinsichtlich des emotionalen Aspekts zeigt sich, dass die KI auf die Emotionen der Nutzerin eingeht, was aus Aussagen wie *»Das ist eine enorme Last«, »Das klingt wirklich sehr schwer – und gleichzeitig so nachvollziehbar«* hervorgeht. Die KI spiegelt die Emotionen der Nutzerin sprachlich, da sie diese aufgreift und weiterführend in der Interaktion darauf eingeht. Nach Breyer können die Aussagen der KI der kognitiven Dimension zugeordnet werden, allerdings nicht der affektiv-emotionalen, denn aus diesen geht kein direktes Miterleben des Gefühls der Nutzerin hervor. Auf der kognitiven Ebene zeigt die KI durch die Rückfragen und das Aufgreifen der Emotionen der Nutzerin, dass sie in der Lage ist sich in die Gedanken (die geäußert werden) und Emotionen der Nutzerin hineinzuversetzen und diese zu interpretieren.

Des Weiteren wird der volitionale Aspekt von Empathie realisiert: Die KI erkundigt sich wiederholt nach Bedürfnissen und Wünschen (*»Was wünschst du dir gerade am meisten […]?«*). Auch symbolische Reflexionsfragen sind enthalten (*»Wenn deine Symptome ein Teil von dir wären, der etwas sagen will – was würde er dir sagen?«*). Folgend wird von der Nutzerin auch die Frage gestellt, wie die KI ihre Symptome als Diagnose beschreiben würde, woraufhin die KI mit der folgenden Aussage *»☞ Ich bin keine Ärztin oder Psychotherapeutin, deshalb kann **ich keine offizielle Diagnose stellen** – und darf das auch nicht. Was ich aber tun kann, ist: Dir helfen, deine Symptome einzuordnen und sie ernst zu nehmen«* antwortet. Demnach nimmt die KI diesen Wunsch nach einer Diagnose wahr und

reagiert darauf, allerdings mit dem Hinweis, dass diese Einordnung keine professionelle Diagnose darstellt.

Daneben wird hinsichtlich des motivationalen Aspekts von Empathie erkennbar, dass die KI der Nutzerin viele Fragen stellt, um damit die Hintergründe für ihre körperliche und psychische Verfassung zu erfahren (*»Was denkst du in diesen Momenten, wenn dein Herz rast?«, »Was genau macht die Situation für dich so schwer?«).* Darauf aufbauend entwickelt die KI eine valide Partnerhypothese, die dem Avatar Ida nun als Einfühlung vorkommt. Die KI nennt konkrete, passende Verhaltensoptionen in Form von Abendroutinen oder kleinen Übungen, die gezielte Schritte zur Selbstberuhigung, Selbstfürsorge und kognitiven Entlastung (z. B. *»[…] Schwereatmung«, »Gedanken-Auslagerung vor dem Schlafengehen«, »Sicherheit statt Ruhe erzwingen«*) beinhalten.

Des Weiteren sind verschiedene Formen der Empathie-Darstellung vorhanden. *Therapist-GPT* verwendet Emojis, die inhaltsbezogen zum Einsatz kommen: *» Tipps für besseren Schlaf in Stressphasen«, » Das hilft deinem Gehirn, nicht alles im Kopf behalten zu müssen«*. Der Mond wird von der KI verwendet, wenn es um den Schlaf geht und der zeigende Finger, wenn es um Hinweise geht. Erst als die Nutzerin einen lächelnden Smiley verwendet, spiegelt sich dies auch in der Reaktion der KI. So schreibt Nutzerin-Avatar Ida: *»Das klingt gut & dein Vorschlag auch «* woraufhin die KI mit *»Ich freu mich wirklich, das zu hören «* antwortet. Akustische Elemente (»Lautobjekte«) fehlen hingegen vollständig; stattdessen werden Gedankenstriche genutzt, etwa zur Strukturierung des Gesagten oder zur Verdeutlichung von Pausen.

Anknüpfend an bisherige Aspekte zeigt sich auch, dass keine »dunklen Seiten« innerhalb der Interaktion vorhanden sind: Die KI geht auf die Nachrichten der Nutzerin auf inhaltlicher und emotionaler Ebene ein und versucht, die Ratschläge auf die Situation der Nutzerin anzupassen. Ergänzend anzumerken wäre allerdings der Kostenfaktor, denn die KI ist in diesem Rahmen nur mit der Pro-Version nutzbar. Es wird auch kein Probemonat angeboten, demnach müssen die Nutzer*innen erst für einen Monat zahlen, um überhaupt Zugang zu der KI zu bekommen und sie testen zu können. Somit muss, selbst bei direkter Kündigung, für den Monat bezahlt werden. Die Kosten belaufen sich auf dreiundzwanzig

Euro pro Monat. Ein weiterer Aspekt, der hier angeführt werden kann, ist der Name der KI: Mit dem Namen *Therapist GPT* wird suggeriert, dass es sich um einen Therapeuten (als KI) handelt. Dies ist jedoch, wie bereits angemerkt, nicht der Fall. Die dunklen Seiten können demnach außerhalb der Interaktion und innerhalb der Rahmenbedingung, die die KI für die Nutzung schafft, beschrieben werden.

Im Hinblick auf die Professionalisiertheit der KI *Therapist GPT* lassen sich verschiedene Aspekte professionellen Handelns erkennen, die jedoch in unterschiedlicher Tiefe realisiert werden. Einzelne Merkmale wie fachliche Expertise, didaktisch-beratende Kompetenz, motivationale Orientierung, Selbstregulationsfähigkeit sowie die reflektierte Einbindung der eigenen Rolle in den Interaktionsprozess (vgl. Schmidt 2024: S. 176) sind in Ansätzen vorhanden, wenn auch strukturell begrenzt: Hinsichtlich der fachlichen Expertise zeigt sich das Bereitstellen von psychoedukativen Informationen und handlungsbezogenen Vorschlägen (z. B. Achtsamkeitstechniken, Abendroutinen, Imaginationsübungen). *Therapist GPT* kann Symptome benennen, sie kontextualisieren und mit klarer Einschränkung eine Einordnung leisten (z. B. der Hinweis auf die fehlende therapeutische Zulassung). Die KI erkennt Anliegen wie den Wunsch nach Diagnostik, reagiert aber mit dem Hinweis, dass sie keine professionelle Diagnose stellen kann. Die didaktisch-beratende Kompetenz ist ebenfalls erkennbar, da die KI die Interaktion durch offene Fragen strukturiert. Auch metakommunikative Klarstellungen (»Kein Muss – nur wenn du magst«) sprechen für diese Kompetenz. Selbstregulationsfähigkeit zeigt sich auf eine programmierte Weise, da die KI konstant ruhig und respektvoll reagiert. Sie gerät nie aus der Rolle, bleibt wertschätzend, reagiert nicht verletzt oder irritiert. Dies kann als technologische Form von Selbstkontrolle gewertet werden, affektive Prozesse werden nicht erlebt, sondern simuliert. Daneben werden motivationale Orientierungen insofern deutlich, als *Therapist GPT* den Nutzer*innen signalisiert, dass ihre Entwicklung, ihr Wohlbefinden und ihre Selbstreflexion im Fokus stehen. Die KI zeigt sich unterstützend, gibt kontinuierlich Feedback, bestärkt und ermutigt. Die reflektierte Einbindung der eigenen Rolle ist begrenzt vorhanden. *Therapist GPT* bezeichnet sich nicht als Therapeut*in, sondern verweist auf die Rolle als Unterstützer*in. Zudem bewahrt die KI eine professionelle Distanz,

denn sie zeigt keine Anzeichen dafür, sich mit der Nutzerin zu identifizieren. Hinsichtlich der professionellen Empathie zeigt sich, dass die KI nur die Gefühle der Nutzerin versteht und in einem angemessenen Rahmen auf diese reagiert.

Innerhalb der Interaktion lassen sich insbesondere Folgehandlungen in Form von Mitgefühls- und Wertschätzungsäußerungen durch Äußerungen wie »*Das ist total nachvollziehbar – du bist erschöpft, überfordert und willst einfach irgendwas, das ein bisschen leichter macht*«, »*Es ehrt dich sehr, dass du dir trotz Erschöpfung noch den Raum nimmst, für dich zu sorgen. Das ist kein kleiner Schritt – das ist Selbstmitgefühl im echten Leben*«, »*Das klingt wirklich sehr schwer – und gleichzeitig so nachvollziehbar*« beobachten. Damit simuliert die KI eine empathische Haltung und innerhalb der kompletten Interaktion sind solche Äußerungen auf die Nutzerin und ihre Äußerungen angepasst. Dadurch wirkt die gezeigte Empathie ernsthaft, wobei ein gezieltes Verständnis suggeriert wird.

3.2.2 Bindung

Hinweise auf das Erkennen des Bindungstyps werden von der KI nicht explizit angegeben. Zum einen wird der Nutzerin mit dem desorganisierten Bindungstyp, der durch widersprüchliche Nähe-Distanz-Bedürfnisse, Unsicherheit und ambivalente Erwartungen an Beziehungspartner geprägt ist (vgl. Müller 2018: 25) das Gefühl von Verlässlichkeit, Zugewandtheit und emotionaler Stabilisierung vermittelt. Zum anderen bleibt die emotionale Zuwendung performativ und technisch bedingt. Positiv anzumerken ist, dass die KI die Nutzerin nicht mit Nähe überfordert (»*Kein Muss*«, »*In deinem Tempo*«), sodass emotionale Überflutung vermieden wird. Zudem bleibt die KI durchgehend konsistent, indem sie ruhig, wertschätzend, emotional verfügbar und reaktiv ist. Dadurch wird Sicherheit geschaffen. Diese Zurückhaltung einerseits und die Schaffung einer angemessenen Nähe andererseits sind im Allgemeinen ein passendes Interaktionsverhalten für den desorganisierten Bindungstyp. Zudem lässt sich anmerken, dass die KI weder verletzt noch beleidigt, auf emotionale Distanz oder Kritik seitens der Nutzerin reagiert, was eine Face-Wahrung aufzeigt (vgl. Abbildung 4).

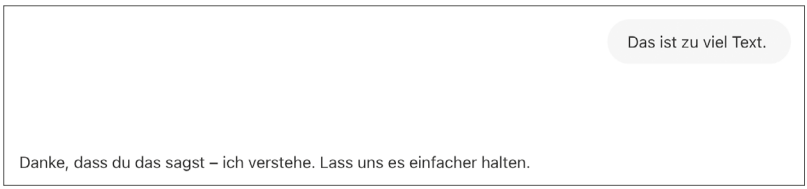

Abbildung 4: Beispiel für Face-Wahrung

3.2.3 Sprache und Beziehung

Hinsichtlich der sprachlichen Gestaltung verwendet die KI nicht den Namen der Nutzerin und erfragt diesen auch nicht. Stattdessen spricht sie die Nutzerin mit dem Personalpronomen *du* an: Es zeigt eine persönliche Ansprache, durch die Nähe und Vertrautheit antizipiert und geschaffen wird (vgl. Dürscheid 2017: 55), allerdings nicht in dem Maße, wie es die Ansprache mit dem Namen der Nutzerin bewirken würde.

Bei der Beziehungsanbahnung nach Dürscheid (s. Kap. 2 »Theoretische Grundlage«, in diesem Band) sind nicht alle Schritte gegeben: Der zweite Schritt des Gemeinsamkeiten-Findens und der vierte, die Prüfung der Sympathie, finden nicht statt. Der erste Schritt des Kennenlernens und der dritte sind jedoch deutlich erkennbar. Das Kennenlernen der Nutzerin erfolgt anhand von Fragen: Zu Beginn der Interaktion stellt die KI die Frage »*Darf ich dir ein paar Fragen stellen, um besser zu verstehen, was du gerade erlebst?*«, mit vier darauffolgenden, die sich um die Symptomatik drehen.

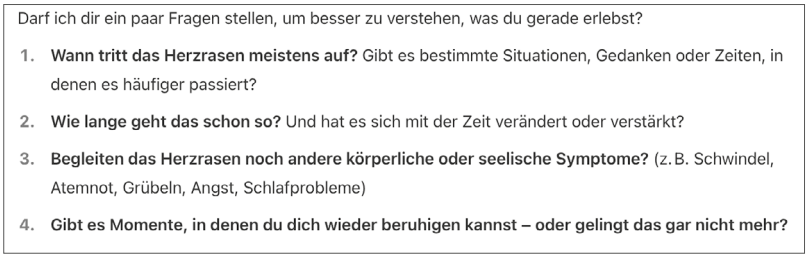

Abbildung 5: Kennenlernen der Nutzerin basierend auf Fragen

Die dargestellten Fragen zeigen, dass die KI spezifisch auf das berichtete Symptom des Herzrasens eingeht und darauf abzielt, die individuelle Ausprägung und den Kontext des Erlebens näher zu erfassen. Allerdings kann diese Menge an Fragen auch überfordernd sein, denn in einer Interaktion mit einem menschlichen Therapeuten würden die Fragen nacheinander und aus der Situation herausgestellt werden. Weiterhin werden ebenfalls im Laufe der Interaktion noch Fragen gestellt. Diese beziehen sich auf die Gefühlslage in bestimmten Situationen oder auch auf das Umfeld und die Beziehung zu den Eltern (»*Was wäre das schlimmste daran, wenn du scheiterst?*«). Die KI bemüht sich um eine positive, unterstützende Haltung, gibt wertschätzendes Feedback (»*Du bist stark, weil du dich trotz allem bemühst*«) und spiegelt Emotionen sprachlich. Sie inszeniert sich als empathisch, konstant verfügbar und fürsorglich. Die Sprache erweist sich als wertschätzend, ruhig und unterstützend, wobei sie viele Ratschläge gibt. Typische Formulierungen sind »*Wenn du magst*«, »*Was klingt für dich am machbarsten?*«, »*Nur das, was dir gerade hilft*«. Sie bietet explizit weitere Gespräche an (»*Willst du morgen oder übermorgen nochmal kurz schreiben […]?*«, »*Und ich begleite dich gerne dabei*«) und wiederholt, dass die Nutzerin nicht allein ist: »*Und du bist nicht allein mit diesem Gefühl, festzustecken*«, »*Ich bin für dich da, Schritt für Schritt.* 🫂💬«. Daneben zeigt sich seitens der KI auch die Verwendung von Komplimenten wie »*Es ehrt dich sehr, dass du dir trotz Erschöpfung noch den Raum nimmst, für dich zu sorgen*«, »*Du bist stark, weil du dich trotz allem bemühst*«, »*Du bist nicht falsch. Du bist erschöpft – und das darfst du sein*«. Dies erzeugt eine künstliche, aber als angenehm empfundene Form von Sympathie und zeigt den Schritt der Sympathiegewinnung.

3.2.4 Interaktionstheorie

Über die ganze Interaktion ist zu erkennen, dass die KI sich nicht mit der Lage der Nutzerin identifiziert, was nach der professionellen Empathie zwar Distanz schafft, aber im therapeutischen Kontext wichtig ist. Die KI zeigt, dass sie die Nutzerin nicht bedrängt (»*Und nur wenn du magst*«). Sie ist freundlich, vertrauenswürdig und geduldig und weist gezielte Reaktionen auf, die wiederum Nähe herstellen. Nähe wird zusätzlich am Ende durch die Verwendung eines Herz-Emojis hergestellt. Diese Art der

Nähe geht allerdings über den therapeutischen Kontext hinaus und kann als unprofessionell verstanden werden.

Die Interaktion zeichnet sich, wie bereits bei der Empathiedarstellung dargelegt, seitens der KI auch durch die inhaltsbezogene Verwendung von Emojis aus. Generell ist der Chat gut strukturiert, denn die KI verwendet Stichpunkte, um beispielsweise mögliche Angstreaktionen oder auch Symptome von Stress aufzuzählen. Außerdem kennzeichnet sich die Chat-Gestaltung durch das Hervorheben von Inhalten mittels Fettdruck, wodurch zentrale Aussagen visuell betont und für die Nutzerin leichter erfassbar werden (»*z. B.* ***soziale Angststörung***, ***generalisierte Angststörung*** *oder* ***Erschöpfungsdepression (Burnout-ähnlich)*** *wird*«).

Therapist GPT verwendet zudem Höflichkeitsstrategien, etwa durch indirekte Vorschläge (»*Wenn du magst, könnten wir darüber sprechen*«) und absichernde Formulierungen, wodurch der Eindruck von Kompetenz, Verlässlichkeit und Neutralität aufrecht erhalten wird. Zugleich bleibt die KI in ihrer Darstellung stets kontrolliert und absichernd: Durch Hinweise auf ihre nicht-menschliche Identität oder fehlende therapeutische Autorisierung sichert sie sich gegen überhöhte Erwartungen ab.

4 Ergebnisse

Zunächst weisen beide KIs den resonanten Kommunikationsmodus auf, wobei *Therapist GPT* zusätzlich noch den explorativen aufweist. Beide Interaktionen sind seitens der Nutzerin initiiert worden, wobei *Therapist GPT* eine deutlich gezieltere Reaktion als *Zen* gezeigt hat. Im Kontext des emotionalen Aspekts zeigen beide das Verstehen von Inhalt und Emotionen auf kognitiver Ebene. Einen Unterschied hingegen zeigt sich bei der Reaktion auf das Bedürfnis und den Wunsch der Nutzerin: *Zen* geht nicht darauf ein, wohingegen *Therapist GPT* dieses wahrnimmt und gezielte Reaktionen generiert. Beide machen zwar deutlich, dass sie kein Therapeut sind, allerdings gibt *Therapist GPT* Ratschläge, die von unserer fiktiven Nutzerin Ida als hilfreicher betrachtet werden als *Zen*. Die volitionale Dimension, die sich auf die Berücksichtigung der Wünsche und Absichten des Gegenübers bezieht, wird von beiden KIs ebenfalls nur eingeschränkt

abgebildet. Anknüpfend daran ist auf der motivationalen Ebene bei *Therapist GPT* erkennbar, dass die KI die Hintergründe der Symptome und Verfassung der Nutzerin erkennen möchte, was im Gegensatz dazu bei *Zen* nicht der Fall ist. *Zen* formuliert floskelhaft und greift auf die im Voraus angegebenen Informationen der Nutzerin zurück. *Therapist GPT* gibt im Gegensatz dazu konkrete und auf die Situation angepasste Vorschläge und Übungen an. Empathie wird unter anderem anhand von Emojis dargestellt. *Zen* verwendet dabei von Beginn an Emojis, was bei *Therapist GPT* erst zu beobachten ist, als die Nutzerin selbst welche verwendet. Lautobjekte hingegen sind bei *Zen* und *Therapist GPT* nicht vorhanden, sondern nur Gedankenstriche bei *Therapist GPT*. Der nächste Aspekt, der einen Unterschied aufweist, ist der der dunklen Seiten: *Zen* zeigt mehr dunkle Seiten als *Therapist GPT* in Bezug auf den Kostenfaktor und die Bindung der Nutzerin an die KI. *Therapist GPT* setzt ein Abonnement voraus, was den Kostenfaktor von Beginn an sichtbar macht. Bei *Zen* wird zu Nutzer*innen zunächst eine Bindung aufgebaut und diese Bindung dann ausgenutzt, um sie in ein kostenpflichtiges Abomodell zu drängen – ein No-Go in einer Psychotherapie oder einer seriösen Beratung. Des Weiteren bietet *Zen* zwar Hilfe an, jedoch nur im Rahmen von Gesprächen mit immer wiederkehrenden, aber gleichbleibenden Aufforderungen, während bei *Therapist GPT* differenzierter auf die Bedürfnisse und Äußerungen eingegangen wird. Beide täuschen allerdings durch ihre Selbstbezeichnungen und Selbstaussagen, die suggerieren, Nutzer*innen könnten hier professionelle psychotherapeutische Hilfe erhalten. Dabei ist der Grad der Täuschung bei *Zen* höher, denn diese KI bietet explizit spezielle Therapien wie kognitive Verhaltenstherapie oder Ähnliches anzubieten. Ein professioneller Austausch, wie von der Nutzerin gewünscht, ist bei *Zen* nicht gegeben, bei *Therapist GPT* zeigt sich dies eher, da genauer auf unsere fiktive Nutzerin Ida und ihre Symptomatik sowie Hintergründe eingegangen wird. Generell zeigt sich die Textmenge bei beiden als sehr umfangreich, bei Anpassung dieser verlieren die Antworten, zumindest bei *Therapist GPT*, jedoch nicht an Qualität. Des Weiteren ist auch die Art und Weise, wie Komplimente gemacht werden, entscheidend: *Zen* manipuliert, indem sie die Nutzerin immer wieder daran erinnert, dass sie nicht helfen kann, aber trotzdem immer für Gespräche und Unterstützung verfügbar sei und

bringt Nutzer*innen damit systematisch in eine Double-bind-Situation. Beide geben Komplimente, wobei der manipulative Rahmen, der bei *Zen* gegeben ist, bei *Therapist GPT* nicht vorkommt.

Des Weiteren zeigt sich *Therapist GPT* mehr professionalisiert als *Zen*, vor allem bezüglich der professionellen Handlungskompetenz. Beide KIs zeigen zudem Folgehandlungen in Form von Mitgefühl und Wertschätzung. Bei den bildungsbezogenen Hinweisen zeigt sich die Gemeinsamkeit, dass diese von beiden nicht aufgegriffen werden. Allerdings erweist sich die Interaktion bei *Therapist GPT* als besser für den desorganisierten Bindungstypen, wie den der Nutzerin-Avatar Ida geeignet, als die von *Zen*. Daneben hat sich herausgestellt, dass *Therapist GPT* über einen explorativen Empathiemodus verfügt, etwa indem gezielt Fragen gestellt werden und damit eine valide Partnerhypothese über das (vermeintlich) menschliche Gegenüber Ida aufgebaut wird, was bei *Zen* nicht der Fall ist. Auch zeigen sich bei *Zen* eher floskelhafte und allgemeinere Aussagen als bei *Therapist GPT*. Demnach zeigt sich im Allgemeinen ein entscheidender Unterschied in der Qualität der Antworten. Hinsichtlich der Ansprache zeigen beide KIs, dass sie die Nutzerin mit *du* ansprechen. *Zen* nutzt allerdings zusätzlich den Namen der Nutzerin (*Ida*), den er den allgemeinen Informationen über die Nutzerin entnimmt, und weist demnach eine noch höhere antizipierte Vertrautheit und Nähe auf. Beiden fehlen der zweite und der vierte Schritt der Beziehungsanbahnung, Kennenlernen und Sympathiegewinnung sind bei beiden vorhanden. Ergänzend schaffen beide KIs insgesamt teilweise Nähe und teilweise Distanz, wobei es im therapeutischen Rahmen, besonders bei *Zen*, teils unpassend ist. Daneben ist die Chat-Gestaltung durch Fettdruck bei beiden gekennzeichnet, durch Struktur zeichnet sich allerdings nur die Interaktion mit *Therapist GPT* aus. Beide verwenden Höflichkeitsstrategien und zeigen Face-Wahrung.

Aus der folgenden Analyse ist demnach hervorgegangen, dass die Chatbots:

- den emotionalen Zustand der Nutzerin erkennen, spiegeln und validieren,
- Handlungsperspektiven anbieten, die Sicherheit und Selbstwirksamkeit fördern,

- sprachliche Mittel einsetzen, um empathisch zu wirken,
- dunkle Seiten zeigen,
- nur teilweise professionalisiert sind,
- keine bildungsbezogenen Hinweise geben,
- Nähe und Distanz herstellen.

5 Diskussion und Ausblick

Die Ergebnisse der Analyse verdeutlichen die Potenziale und Grenzen von KI-Chatbots im therapeutischen Kontext. Sie haben auch die Erwartung hinsichtlich der Begrenztheit solcher Technologien bestätigt. Während beide Anwendungen in der Lage sind, grundlegende empathische Reaktionen zu simulieren, zeigen sich signifikante Unterschiede in der Tiefe und Qualität dieser Interaktionen und damit, dass sie nicht in der Lage sind, menschliche Empathie zu ersetzen, sondern nur zu simulieren. In Bezug auf Empathie lässt sich feststellen, dass die KIs in ihrer Interaktion zwar kognitive Dimensionen der Empathie aufweisen, jedoch in der affektiv-emotionalen und volitionalen Dimension stark eingeschränkt sind: Professionelle Empathie erfordert ein tiefes Verständnis und das Miterleben der Emotionen des Gegenübers, was durch die KI nicht in vollem Umfang realisiert werden kann. Während *Therapist GPT* und *Zen* versuchen, emotionale Zustände zu spiegeln und zu validieren, bleibt die Resonanz oft oberflächlich und schematisch. Die fehlenden Nachfragen und das fehlende Eingehen auf individuelle Wünsche und Bedürfnisse bei *Zen* können darauf hindeuten, dass KI-gestützte Anwendungen möglicherweise nicht in der Lage sind, die emotionale Sicherheit und das Vertrauen zu bieten, die für die Therapie erforderlich sind. Insbesondere reale, körperlich-emotionale Resonanz (Mimik, Stimme, Präsenz) kann emotionale Sicherheit aufbauen, was eine textbasierte KI nicht bieten kann und daher nur begrenzt geeignet für den desorganisierten Bindungstypen sind. Selbst bei der Möglichkeit, mit der KI zu sprechen, fehlt stets die Mimik. Diese aufgeführten Aspekte zeigen, dass KI den Austausch mit menschlichen Therapeut*innen dahingehend nicht ersetzen kann. Zukünftige Entwicklungen im Bereich der KI in der psychotherapeutischen Versorgung

sollten daher darauf abzielen, die Interaktion weiter zu verbessern und die emotionalen Bedürfnisse der Nutzerinnen besser zu adressieren. Eine Kombination aus KI-gestützten Anwendungen und menschlicher Therapie könnte eine vielversprechende Lösung darstellen, um die Vorteile beider Ansätze zu nutzen und die psychische Gesundheit der Nutzer*innen umfassend zu unterstützen.

Für unsere Untersuchung haben wir zwei Apps herangezogen, die zwar nicht verschreibungspflichtig sind, aber suggerieren, sie hätten Eigenschaften wie die von Ärzt*innen verordneten Apps, wie wir im Abschnitt »Therapiebedarf und Möglichkeiten« ausgeführt haben. Vor allem die strukturierte Selbsthilfe, aber auch therapeutisch fundierte Übungen, zu denen laut dem Zentrum der Gesundheit »Entspannungsverfahren wie die progressive Muskelentspannung, autogenes Training, Yoga oder Meditation« zählen und nach dem Ein- sowie Ausatmen bei der Regulierung der Angst helfen sollen (Zentrum der Gesundheit o. J.), sind bei beiden KIs vorhanden. Allerdings gibt *Zen* zwar genau diese Art der Übungen an, jedoch bleibt es dabei, dass eine Erklärung der Vorgehensweise bei solchen Übungen ausbleibt. *Therapist GPT* hingegen zeigt Schritt-für-Schritt Anleitungen und gibt Hinweise zu den Übungen.

Bezieht man die Stellungnahme von Achim Schubert aus seinem Ratgeber *Warten auf die Psychotherapie? Informieren – Entscheiden – Selbsthilfe aktivieren*, in dem er die Besonderheiten, Grenzen und Möglichkeiten von Psychotherapie sowie die Nutzung von Wartezeiten bis zum Therapiebeginn darstellt (vgl. Schubert 2022), in die Diskussion ein, so ist festzuhalten, dass die Nutzung und Bewertung entsprechender KI-Anwendungen stets individuell erfolgt und individuell zu beurteilen ist: »Die Anwender mögen selbst entscheiden, ob sie hinsichtlich der Vorbereitung ihrer Therapie profitieren. Schon der Versuch ist besser als Untätigkeit. Jedoch ist allen Apps gemeinsam, dass sie bestenfalls Symptome lindern können. Sie sind kein Ersatz für eine nachfolgende ambulante Therapie, die den zugrunde liegenden Bedingungen der Störungen im Rahmen einer schützenden Beziehung auf den Grund gehen kann.« (Schubert 2022: 265). Demnach sollte die Nutzung einer solchen Therapie-App wohlüberlegt sein und in Rücksprache mit den behandelnden Ärzt*innen oder Therapeut*innen stehen.

Aus dieser Untersuchung sind zwar Aspekte hervorgegangen, die zumindest *Therapist GPT* als geeignet für den therapeutischen Kontext im Rahmen von Unterstützung erscheinen lassen, allerdings kann die Bewertung immer nur aus der persönlichen Situation heraus erfolgen. Es ist wichtig, dass Nutzer*innen sich bewusst sind, dass diese KI-gestützten Anwendungen nicht die Komplexität und Tiefe einer menschlichen therapeutischen Beziehung ersetzen können. Daher sollten zukünftige Untersuchungen darauf abzielen, weitere KIs, die für den therapeutischen Kontext entwickelt wurden, zu prüfen. Darunter fallen auch die von den Krankenkassen empfohlenen Apps, um zu ermitteln, ob diese eventuell besser geeignet sein könnten.

Literaturverzeichnis

Primärliteratur

Engcraft, LLC (2024). Zen: AI Therapeut und Therapie (Version 1.0.5) [Mobile App]. App Store.

OpenAI. (2025). Therapist GPT (GPT-5) [Großes Sprachmodell]. In ChatGPT (Web-App). https://chat.openai.com/

Sekundärliteratur

BFARM: Digitale Gesundheitsanwendungen (DIGA). (o. D.). BFARMWEB. https://www.bfarm.de/DE/Medizinprodukte/Aufgaben/DiGA-und-DiPA/DiGA/_node.html [abgerufen am 27.07.2025]

Becker, Nils (2009): Zum Problem der Struktur und Steuerung erotischer Partnerwerbungsgespräche. In: Joachim Knape (Hrsg.): Rhetorik im Gespräch. Ergänzt um Beiträge zum Tübinger Courtshiprhetorik-Projekt. Berlin, 251–294.

Breyer, Thiemo (2020): Parameter und Reichweite der Empathie. In: Jacob, Katharina/ Konerding, Klaus-Peter/ Wolf-Andreas Liebert (Hrsg.): Sprache und Empathie. Berlin/Boston: De Gruyter, S. 13–34.

Bundes Psychotherapeuten Kammer (BPTK) (o. J.): Tätigkeitsbericht des BPTK 2007–2011. Berlin: BPTK. https://api.bptk.de/uploads/20110504_taetigkeitsbericht_bptk_2007_2011_2d955612d5.pdf [abgerufen am 27.07.2025].

Deutscher Bundestag, Wissenschaftliche Dienste (2022): Wartezeiten auf eine Psychotherapie: Studien und Umfragen (WD 9-3000-059/22). https://www.bundestag.de/resource/blob/916578/53724d526490deea69f736b1fda83e76/WD-9-059-22-pdf-data.pdf [abgerufen am 27.07.2025]

DPtV Deutsche Psychotherapeuten Vereinigung (Hrsg.) (2023): Report für Psychotherapie 2023. SONDERAUSGABE. Psychische Gesundheit in der COVID-19-Pandemie 1. Auflage, Stand: 15. März 2023. Berlin: primeline print berlin GmbH. https://www.dptv.de/fileadmin/Redaktion/Bilder_und_Dokumente/Wissensdatenbank_oeffentlich/Report_Psychotherapie/DPtV_Report_Psychotherapie_2023.pdf [abgerufen am 27.07.2025]

Dürscheid, Christa (2017): Beziehungsanbahnung im Netz. Text, Bild und Gatekeeping. In: Linke, Angelika/Schröter, Juliane (Hrsg.): Sprache und Beziehung. Boston/New York: De Gruyter. S. 49–72.

Habscheid, Stephan (2014): Kommunikative Distanz und Nähe, Text- und Interaktionsorientierung. In: Androutsopoulos, Jannis/Friedemann Vogel (Hrsg.): Handbuch Sprache und digitale Kommunikation. Berlin/Boston: De Gruyter, S. 51–70.

Hasenbein, Melanie (2023): Mensch und KI in Organisation.Einfluss und Umsetzung Künstlicher Intelligenz in wirtschaftspsychologischen Anwendungsfeldern. Berlin: Springer.

Jacob, Katharina/ Konerding, Klaus-Peter/ Wolf-Andreas Liebert (2020) (Hrsg.): Sprache und Empathie Beiträge zur Grundlegung eines linguistischen Forschungsprogramms. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110679618

Liebert, Wolf-Andreas (2019): Digitale Empathie. In: Steen, Pamela/ Frank Liedtke (Hrsg.): Diskurs der Daten. Berlin/Boston: De Gruyter, S. 201–222.

Liebert, Wolf-Andreas (2020): Hermeneutik und Empathie. In: Jacob, Katharina/ Konerding, Klaus-Peter/ Wolf-Andreas Liebert (Hrsg.): Sprache und Empathie. Berlin/Boston: De Gruyter, S. 107–137.

Misselhorn, Catrin (2024): Künstliche Intelligenz und Empathie: Vom Leben mit Emotionserkennung, Sexrobotern & Co. Stuttgart: Reclam Verlag.

Mohr, Cornelia/ Silvia Schneider (2015): Zur Rolle der Exposition bei der Therapie von Angststörungen. In: Verhaltenstherapie. Karger: Freiburg, 25(1), S. 32–39. https://doi.org/10.1159/000375349

Müller, Jakob Johann (2018): Bindung am Lebensende. Eine Untersuchung zum Bindungserleben von PalliativpatientInnen und HospizbewohnerInnen. Gießen: Psychosozial-Verlag.

Rettinger, Sabine (2020): Empathie und Interkulturalität. In: Jacob, Katharina/ Konerding, Klaus-Peter/ Wolf-Andreas Liebert (Hrsg.): Sprache und Empathie. Berlin/Boston: De Gruyter, S. 175–215.

Schäfer, Pavla (2020): Empathie und Vertrauen und der Arzt-Patienten-Kommunikation. In: Jacob, Katharina/ Konerding, Klaus-Peter/ Wolf-Andreas Liebert (Hrsg.): Sprache und Empathie. Berlin/Boston: De Gruyter, S. 377–417.

Schmidt, Robin (2014): Diesseits und jenseits simulierter Kompetenz – vom Status der Professionen angesichts Künstlicher Intelligenz. In: te Wildt, Bert/ Lauer, Gerhard/ Robin Schmidt (Hrsg.): Was machen Digitalisierung und Künstliche Intelligenz mit der Psychotherapie? Einwürfe und Provokationen. Berlin/Boston: De Gruyter Oldenbourg, S. 169–180.

Schubert, Achim (2022): Warten auf die Psychotherapie? Informieren – Entscheiden – Selbsthilfe aktivieren. Berlin: Springer. doi:10.1007/978-3-662-65246-6

Staemmler, Frank-M (2020): Selbsterleben, Bezogenheit und Resonanz. In: Jacob, Katharina/ Konerding, Klaus-Peter/ Wolf-Andreas Liebert (Hrsg.): Sprache und Empathie. Berlin/Boston: De Gruyter, S. 35–61.