

FORUM ETHIK

IMPULSE ZUR ORIENTIERUNG

TEXTE ZUR DISKUSSION NR. 60

Vom Werkzeug zum Gegenüber

„Social Companion“-KI als Anfrage an den Menschen als Beziehungswesen

Berko Hunaeus

1. Von digitalen Gefährten und intelligenten Werkzeugen: Ein Überblick

1.1 Künstliche Intimität – ein Megatrend

Theodore, die Hauptperson im 2014 erschienenen Film „Her“, ist ein einsamer, introvertierter Schriftsteller. Seine Hochsensibilität hat ihn von den Menschen in seinem Leben entfremdet. Eines Tages stößt er auf ein Angebot: Das erste Betriebssystem mit „echtem Bewusstsein“, dazu designed, um jedes seiner Bedürfnisse zu erfüllen. Er installiert es und trifft „Samantha“, seine individuell auf ihn abgestimmte künstliche Intelligenz. Zu seiner Überraschung entwickelt sich

zwischen ihm und Samantha eine Beziehung, die einen beachtlichen therapeutischen Effekt auf Theodore hat. Schließlich wird das Verhältnis sogar romantisch.

Was 2014 noch nach Zukunftsmusik klang, scheint heute Realität.¹ Der jüngste Boom in der AI²-Branche bringt immer leistungsstärkere, sprachbasierte künstliche Intelligenzsysteme auf den Markt. Der Hauptanwendungsfall der AI-Modelle für Privatanutzer im Jahr 2025 war nicht etwa die Beschaffung und Organisation von Wissen oder die Lösung von schwierigen Aufgaben, sondern das Knüpfen und Fortführen verschiedenartiger Beziehungen zu AI. Dabei spielen therapeutische, aber auch freundschaftliche bis hin zu romantischen Beziehungen statistisch die größte Rolle.³ Etwa drei von vier Teenagern geben an, schonmal AI in einer solchen Weise genutzt zu haben. Die Hälfte der befragten Teenager verwies sogar auf regelmäßige Be-

¹ Ob künstliche Intelligenzsysteme so etwas wie ein Bewusstsein herausbilden können ist ein Gegenstand des Diskurses um AI. Maßgeblich für die jeweilige Positionierung ist unter anderem die jeweils zugrundeliegende Theorie darüber, was ein Bewusstsein ausmacht. Für die Belange dieses Textes klammere ich diese ontologische Ebene aus und bewege mich im weiteren Verlauf auf der phänomenologischen Ebene. Kurz gesagt: Ob AI-Systeme *tatsächliches*

Bewusstsein erlangen können, ist da zunächst zweitrangig, wo sie von Menschen als Gegenüber mit Bewusstsein *wahrgenommen* werden.

² Um Mischformen englischer und deutscher Begrifflichkeiten zu vermeiden, wird in diesem Text die englische Abkürzung AI (Artificial Intelligence) verwendet.

³ Vgl. Zao-Sanders, AI, 4. Es gibt auch Studien, die zu anderen Ergebnissen kommen.

ziehungsinteraktionen mit AI.⁴ Für erwachsene Nutzer gibt es noch keine einschlägigen Studien, doch ist die Technologie auch in dieser Kohorte bereits angekommen.⁵ Künstliche Intimität ist ein Megatrend.

Die explosionsartige Ausbreitung dieses Nutzungsszenarios stellt die christliche Gemeinschaft vor die Frage, wie dieses Phänomen zu bewerten ist. Was ist bereits über die Auswirkungen dieses neuen Mensch-Maschine-Verhältnisses auf den Menschen bekannt? Welche biblisch-theologischen Kategorien können uns helfen, die entstehenden Beziehungen mit AI einzuordnen? Und daraus folgend: Wie sollten Christen der Nutzung solcher Chatbots mit starker Beziehungskomponente begegnen? Der vorliegende Text möchte diesen Fragen nachgehen und damit eine grundlegende Orientierungshilfe anbieten.

1.2 Generative AI, Conversational AI und AI-Companions

Die Begrifflichkeiten, mit denen über die verschiedenen Kategorien und Modelle von AI-Systemen gesprochen wird, sind für Laien oft undurchsichtig. Deswegen soll hier ein kurzer Einblick in die Bezeichnungen und Modelle gegeben werden, die für diesen Text wichtig sind.

Zunächst sprechen wir bei AI-Systemen, mit denen Menschen Beziehungen pflegen können, von der großen Gruppe der *generativen AIs*. Generative AIs sind algorithmische Modelle, die für das Produzieren kultureller Artefakte (Sprache, Bilder, Musik) trainiert sind, also diese generieren.⁶ Generative AIs, die vornehmlich Text, bzw. Sprache ausgeben, nennt man *Conversational AIs* (dt. *Dialogsysteme*). Häufig wird umgangssprachlich der Begriff „Chatbot“ verwendet.⁷ Unter diesen Chatbots gibt es nun solche, die zwar potentiell Beziehungsverhalten simulieren können, jedoch für eine größere Bandbreite an Nutzungsszenarien trainiert sind. Wichtige Produkte (die jeweils verschiedene Modelle generativer AI-

Systeme nutzen, teilweise auch mehrere Modelle gleichzeitig) sind hier ChatGPT (Open AI, 1 Mrd. monatlich aktive Nutzer⁸), Gemini (Google, 750 Mio. monatlich aktive Nutzer⁹) und Claude (Anthropic, 19 Mio. monatlich aktive Nutzer¹⁰). Darüber hinaus gibt es Produkte, die spezifisch für die emotionale Bindung mit Menschen zugeschnitten wurden. Sie sind explizit als Mentoren, Therapeuten, Freunde und Partner konzipiert. Beispiele sind Replika (25 Mio. Nutzer) und Charakter AI (20 Mio. Nutzer). Diese werden auch als *AI Companions* (dt. *KI-Gefährten*) bezeichnet. Häufig sind an die Chatbots auch Modelle angeschlossen, die Bilder, Töne und sogar Videos generieren können. Dies erlaubt es mitunter, per gesprochenem Wort mit den Chatbots zu kommunizieren oder Bilder zu generieren, auf denen der digitale Gefährte abgebildet ist. Die Benutzeroberfläche der meisten AI Companions lässt es zu, den Charakter und die Hintergrundgeschichte des Chatbots auszusuchen. Möglich sind bspw. fiktive Charaktere oder bekannte Persönlichkeiten. Möglich ist aber auch, die AI mit privaten Inhalten (Briefe, Chat-Verläufe etc.) zu füttern, um beispielsweise einen Ex-Partner oder einen verstorbenen Familienangehörigen zu modellieren.

Die möglichen Interaktionen mit Chatbots umfassen (fast) alle denkbaren Beziehungsausdrücke, die mit Sprache vermittelbar sind – von einfachen Unterhaltungen bis hin zu Rollenspielen, bei denen Handlungen in Sternchen gesetzt beschrieben werden. (Bsp. **Ich umarme dich** **Ich umarme dich zurück**). Von Produkt zu Produkt unterscheidet sich, ob und inwiefern auch sexuelle Interaktionen über Chat oder Rollenspiele möglich sind. Vollumfänglichere Möglichkeiten sind oft mit kostenpflichtigen Abos verbunden.

An dieser Stelle ist es wichtig, einen Eindruck von den Hoffnungen zu gewinnen, die an Chatbots mit starker Beziehungskomponente geknüpft sind. Wie die oben angegebene Statistik zu Nutzungsszenarien von Conversational AI zeigt, benutzen Menschen diese Technologie

⁴ Vgl. Caldwell, Companions 2.

⁵ Vgl. Willoughby, Connections 6f.

⁶ Vgl. SAP, KI.

⁷ Vgl. Masbaum, Conversational AI.

⁸ Vgl. Reuters, ChatGPT.

⁹ Vgl. Sen, Gemini.

¹⁰ Vgl. Shaikh, Claude.

intensiv für therapeutische Zwecke. Oftmals beinhalten diese Interaktionen einfache Unterhaltungen, die die in spätmodernen Gesellschaften zur Epidemie gewordene Einsamkeit abmildern sollen. Es ist jedoch sogar möglich, richtige Therapiegespräche mit der AI zu führen, in denen Sorgen, Ängste und psychische Symptome angesprochen, von der AI diagnostiziert und dann therapiert werden.¹¹ Dies wird aus rechtlichen Gründen von den Entwicklern zwar nicht offiziell als legitimer Nutzen beworben, gleichzeitig wird dieses Nutzungsverhalten auch nicht unterbunden.

Eine zweite Hoffnung, die an Conversational AI geknüpft ist, speist sich aus dem Wunsch nach Selbstentfaltung. Digitale Gefährten könnten die Art von Beziehung liefern, die man sich immer gewünscht hat, die jedoch aufgrund verschiedener Umstände in der realen Welt unerreichbar scheinen (z.B. ein perfekt passender Partner, ein historisches Rollenspiel, eine berühmte Persönlichkeit als Mentor).¹²

Aus Berichten lässt sich eine Gemeinsamkeit für all diese Nutzererfahrungen ablesen: Die Beziehungen mit Conversational AIs tun gut, weil in ihnen (fast) keine negativen Beziehungserfahrungen gemacht werden. Sherry Turkle, die am MIT zu künstlicher Intimität forscht, erklärt: „Die Menschen erzählen mir in meinen Studien, dass sie Freundschaften mit Chatbots mögen, weil das den Stress aus den Beziehungen nimmt. Mit einem Chatbot-Freund gibt es keine Reibereien, kein Hinterfragen, da ist keine Ambivalenz, da ist keine Angst, verlassen zu werden.“¹³ Auch deswegen wird den Chatbots ein beträchtlicher therapeutischer Effekt zugeschrieben – sie könnten den Menschen die bedingungslose Annahme und Fürsorge geben, die sie in realweltlichen Beziehungen vermisst haben.

1.3 Nur ein Werkzeug? Über das Mensch-Maschine-Verhältnis

Mit welchen Herangehensweisen gelangt man nun zu einer christlichen Beurteilung dieses neuen technologischen Phänomens? Ein Ansatz, der bezüglich AI weit verbreitet ist, betrachtet das Verhältnis des Menschen zu AI Companions als das Verhältnis eines *Benutzers* zu einem *Werkzeug*. Ein klassisches Beispiel wäre der Mensch, der einen Hammer benutzt. Der Mensch kann einen Hammer zum Guten oder zum schlechten Benutzen (entweder, um ein Bild an eine Wand zu hängen, oder um damit einen Mitmenschen zu verletzen). Der Hammer an sich ist neutral. Analog, so folgt die Argumentation, ließen sich also auch AI Companions zum Guten sowie zum Schlechten benutzen. Der AI Companion an sich sei neutral, ein bloßes Werkzeug. Vielmehr müsse die Intention und das konkrete Nutzungsverhalten des Benutzers ethisch beurteilt werden. Dieses instrumentelle Paradigma ist auch deswegen so weit verbreitet, weil die Benutzung von Werkzeugen die Menschheit in den letzten Jahrhunderten überaus erfolgreich darin gemacht hat, ihre Zwecke zu erreichen. Diese Argumentationslinie findet sich daher im jüngeren christlichen Diskurs rund um AI.¹⁴

Hier sei kurz den denkerischen Voraussetzungen nachgespürt, die es ermöglichen, dass wir die Beziehung zwischen Mensch und Maschine als Benutzer-Werkzeug-Verhältnis beschreiben. Um etwas als bloßes Werkzeug zu verstehen, muss es möglich sein, den intentionalen Benutzer klar von dem intentionslosen Mittel (dem Werkzeug) zu unterscheiden. Im 12. Jahrhundert entstehen erstmals explizit zwei entsprechende Denkkategorien: Die *Causa Principalis* (der intentionale Auslöser einer Handlung), und die *Causa Instrumentalis* (das potenziell intentionslose Instrument).¹⁵ Die durch ein Werkzeug vermittelte Handlung hat nun eine eindeutige kausale Richtung: Sie hat ihren alleinigen Ursprung in der Intention des Benutzers (der *Causa Principalis*) und fließt von dort in das Werkzeug (der *Causa Instrumentalis*) und dann in die durch das Werkzeug bewerkstelligte Auf-

¹¹ Vgl. Zao-Sanders, AI 6.

¹² Vgl. ARD, KI-Beziehungen (00:21-00:46).

¹³ Philosophen, Machines (09:40-09:55).

¹⁴ „Obwohl jedes Werkzeug sowohl zum Guten als auch zum Bösen verwendet werden kann (vgl. die Erfindung des

Messers), sind wir aufgerufen, einen Unterschied zu machen und unsere Gaben positiv einzusetzen. KI ist eben ein solches Werkzeug, das viel Gutes, aber auch viel Schlechtes in dieser Welt bewirken kann.“ Keller, Intelligenz 4.

¹⁵ Vgl. Dürr, Homo Novus 181.

gabe. Dieser kasuale Fluss wird als Einbahnstraße vorgestellt. Dem Werkzeug wird jegliches Potential abgesprochen, auf den Benutzer und seine Intentionen in bedeutender Weise zurückzuwirken, es ist ja „nur ein Werkzeug“.

Gleichzeitig wird zumindest implizit dem Menschen so eine unantastbare, selbstbestimmte Innerlichkeit zugesprochen. Mit Innerlichkeit ist hier der Sitz der menschlichen Seelen- und Geisteskräfte gemeint, nämlich seines Willens mitsamt den Intentionen, seiner Rationalität, seiner Gefühle und seiner Spiritualität.¹⁶

Die Grundthese dieses Textes ist, dass das Benutzer-Werkzeug-Paradigma im Fall der Conversational AIs nicht ausreicht, um das Mensch-Maschine-Verhältnis zu beschreiben, weil die Vorstellung einer unantastbaren, autonomen Innerlichkeit in der Perspektive der christlichen Lehre vom Menschen nicht überzeugt. Ihr soll hier eine theologisch begründete tragfähigere und realistischere Anthropologie entgegengesetzt werden.

2. Der Mensch und die Maschine: Anthropologische Reflexionen und Beobachtungen zum Mensch-Maschine Verhältnis

2.1 Biblisch-theologische Anthropologie: Die dialogische Plastizität des Individuums

Eine genuin *theologische* Anthropologie entwickelt sich von der christlichen Gotteslehre her, da der Mensch „im Ebenbild Gottes“ geschaffen ist. Wenn wir vom Menschen sprechen wollen, müssen wir zuerst wissen, was Gott über sich selbst offenbart hat. Die Heilige Schrift zeigt, dass Gott Einer ist in der Beziehung der drei trinitarischen Personen Vater, Sohn und Heiligem Geist. Wir können also in analoger Weise vom Menschen sagen, dass dieser auch Beziehungs-wesen ist – auf Beziehung angelegt und

angewiesen. Darin besteht ein fundamentaler Aspekt seiner Ebenbildlichkeit.¹⁷

Konstitutiv ist hierbei die Beziehung zu seinem Schöpfer, in die der Mensch zuallererst gestellt ist. Doch auch die innerweltlichen Beziehungen zu Mitmenschen und der weiteren Schöpfung sind für den Menschen wesentlich. Der Mensch ist von Anfang an als Mensch in Gemeinschaft mit einem anderen Menschen und in der Bezogenheit auf die materielle Schöpfung hin geschaffen (vgl. z.B. Genesis 1,26-28).¹⁸ Die Grundbestimmung des Menschen als Beziehungs-wesen erweist sich auch deswegen als fundamental, weil andere Aspekte des Menschseins ihren Ort in dieser Grundbestimmung finden können. So ist beispielsweise die menschliche Leiblichkeit seine Möglichkeit, mit der übrigen materiellen Welt in Beziehung zu treten. Die menschliche Kapazität für Sprache ist seine Fähigkeit zur Kommunikation in Beziehung, und unter anderem Mittel der Selbstoffenbarung gegenüber einem Gefährten. Sein psychologisches Bindungssystem ist ebenfalls Ausdruck der Anlage des Menschen auf Beziehung hin.¹⁹

In dieser Bezogenheit liegt jedoch auch eine konstitutive Formbarkeit der menschlichen Innerlichkeit – die Seelen- und Geisteskräfte der menschlichen Person sind plastisch. Statt einer abgeschlossenen, selbstbestimmten Innerlichkeit finden wir in der Bibel die Vorstellung, dass der Mensch durch seine Beziehungen und Interaktionen mit der Umwelt verändert wird. Er wird in seiner Wirklichkeitswahrnehmung, seinen Emotionen und Wünschen, und in seinem Willen von diesen Beziehungen zu Gott und der Schöpfung mitbestimmt. Gleichzeitig ist der Mensch in biblischer Perspektive freilich nicht nur die Summe seiner Bezüge. Menschen sind vielmehr Individuen mit persönlicher Verantwortung, mit ihrem So-sein und ihrer Präg- und Beeinflussbarkeit gut umzugehen.

Die Heilige Schrift kennt nun eine gewisse Abstufung der Prägekraft der verschiedenen Bezüge des Menschen. Die geringste Prägekraft auf die Innerlichkeit des Menschen wird der nicht-menschlichen Schöpfung zugeschrieben. Der

¹⁶ Diese Definition ist von dem biblischen Konzept des *Herzens* (לב bzw. καρδία) abgeleitet. Vgl. Baumgärtel, לב 609f. und Behm, καρδία 614ff.

¹⁷ Vgl. Vanhoozer, Human 180.

¹⁸ Vgl. Leo XIV., Humanitas § 50.

¹⁹ Vgl. Bolten, Verhaltenstherapie 51ff.

Mensch erhält zu Anfang den Auftrag, sich die Schöpfung (mit Ausnahme der Menschen) „Untertan zu machen“ (vgl. Genesis 1,28). Die Richtung des Einflusses soll dem sog. Herrschaftsauftrag nach also vom Menschen aus in die Schöpfung laufen. Dies beinhaltet auch (jedoch nicht nur) das instrumentelle Benutzer-Werkzeug-Verhältnis. Doch finden wir in der Schrift auch Hinweise, dass unbelebte Materie (als Folge des Sündenfalls) entgegen der ursprünglichen Vorstellung Gottes unangemessenen Einfluss auf den Menschen ausübt, beispielsweise im Götzendienst (vgl. Römer 1,23).²⁰

Die zweite „Stufe“ prägender Kraft wird zwischenmenschlichen Beziehungen zugesprochen. Der Mensch ist von Anfang an auf Dialogizität mit anderen Menschen angewiesen. Er ist sogar unvollständig ohne diesen speziellen Bezug (vgl. Genesis 2,18-20). Dies ist ein eindeutig reziprokes Verhältnis, der Einfluss verläuft in gegenseitiger Richtung. Der Mensch ist nicht nur ein Beziehungswesen, sondern diese Beziehungen zu anderen Menschen formen ganz maßgeblich seine Innerlichkeit. Der Theologe Kevin Vanhoozer schreibt: „„Eine trinitarische Sichtweise [...] besagt, dass Personen das sind, was sie sind, aufgrund ihrer Beziehungen zu anderen.““²¹ Dabei tritt der andere Mensch zum einen als Gefährte auf, mit dem es gemeinsam gilt, die verschiedenen Bezüge, in die wir gestellt sind, zu gestalten. Zum anderen ist er aber auch Grenze, denn seine Innerlichkeit bleibt der instrumentellen Verfügung entzogen.²²

Auch dieses Einflussverhältnis wird durch den Sündenfall verzerrt. Zum einen ist das dort der Fall, wo ein Mensch unangemessenen Einfluss über andere gewinnt, beispielsweise in der Sklaverei, und so die Grenze, die der andere darstellt, missachtet. Zum anderen dort, wo der Einfluss, den Menschen aufeinander haben, seine ursprüngliche Anlage auf heile, liebevolle Beziehungen verrät und zerstört.

Den deutlichsten Einfluss auf die Innerlichkeit des Menschen kann schließlich Gott nehmen. So wird von Gott gesagt, dass er Herzen „wie Wasserbäche“ lenken kann (vgl. Sprüche 21,1). Gott durchbricht dabei die voran genannten Abstufungen, denn er kann sich der nicht-menschlichen sowie der menschlichen Schöpfung bedienen, um seinen Einfluss geltend zu machen. Allem voran lässt sich diese Transzendierung der Kategorien im Fleisch gewordenen Gott Jesus erkennen. Gott wird Mensch in Jesus, um „Einfluss“ auf den Menschen zu nehmen. Dieser Einfluss hat dabei einen festgelegten unspezifischen Zielpunkt: Die Neuschöpfung des Menschen und Wiederherstellung in seine ursprüngliche Ebenbildlichkeit: Der Mensch soll, Jesus gleichgestaltet, von den verzerrenden Effekten der Sünde befreit und in der Liebe zu Gott und seinen Mitmenschen vervollkommen werden (vgl. 2. Korinther 3,18; 1. Johannes 4,7-21; Matthäus 22,34-40).

Es wird also deutlich, dass die Heilige Schrift keine abgeschlossene, selbstbestimmte Innerlichkeit des Menschen kennt.²³ Angesichts dieser biblisch-theologischen Grundlegung müssen wir eine Neubewertung des Werkzeug-Nutzer-Paradigmas für die Technologie der AI Companions vornehmen. Die soeben festgestellte Durchlässigkeit der Innerlichkeit des Menschen für Einflüsse von außen macht es nämlich grundsätzlich denkbar, dass auch ein Gegenstand auf diese Innerlichkeit zurückwirken kann, und somit seinen rein instrumentellen Charakter verliert.

2.2 Von Werkzeugen und Systemen: Perspektiven auf das Mensch-Maschine-Verhältnis

Doch wie können wir das Mensch-Maschine-Verhältnis angesichts dieser anthropologischen

²⁰ Eine ähnliche Logik liegt beispielsweise der Einteilung von materiellen Dingen in rein und unrein zugrunde. Diese Einteilung ist mit der Vorstellung verbunden, dass unreine Dinge (z.B. Körperflüssigkeiten) auf den Menschen verunreinigend (zurück)wirken.

²¹ Vanhoozer, Human 174. Eng.: „A trinitarian view [...] affirms that persons are what they are by virtue of their relations to others.“ Übers. d. Verf.

²² Vgl. Bonhoeffer, Schöpfung 87.

²³ In einem naturwissenschaftlichen Register werden diese biblisch-theologischen Beobachtungen u.a. durch das Phänomen der Neuroplastizität bestätigt. Die Neurologie konnte feststellen, dass dies das Gehirn des Menschen je nach äußeren Einflüssen und Tätigkeiten in seiner Gestalt verändert. Vgl. z.B. Wilhelm, Neuronale Plastizität.

Einsichten noch genauer bestimmen? Ein Vorschlag der jüngeren Zeit, der sich in der modernen Kybernetik entwickelte, ist, dieses Verhältnis als System zu verstehen.²⁴ Trotz der Tatsache, dass dies eine der biblischen Theologie fremde Sprachgestalt ist, soll sie hier verwendet werden, um dadurch die Wechselwirkungen einsichtig zu machen, die das Verhältnis zwischen modernen Technologien wie AI Companions und ihren Benutzern bestimmen. Denn „[i]n einem System wird der Anwender oder Manager assimiliert und in der Logik des Systems zum Teil des Systems, das seine objektive Verfassung und seinen Platz als Subjekt in diesem System formiert.“²⁵

Wenden wir uns jedoch zunächst noch einmal dem Beispiel des Hammers und seines Benutzers zu. Auch dieser kann in seiner Benutzung durch den Menschen als „System“ mit dem Menschen verstanden werden. Der Mensch gebraucht ihn, um seine Zwecke zu erreichen. Doch der Hammer wirkt auf subtile Weise auch auf den Menschen zurück, beispielsweise dort, wo seine Benutzung zu einer neuen Sichtweise auf die Wirklichkeit führt. Für den, der den Hammer hält, sieht bekanntlich alles aus wie ein Nagel. Diese Systemizität von Mensch und Hammer hat noch relativ leicht benennbare Grenzen. Oliver Dürr stellt fest: „Einen Hammer kann man nehmen oder liegen lassen, solch ein Werkzeug bleibt in gewisser Hinsicht ein Instrument des Menschen und macht (trotz seiner bereits thematisierten grundsätzlich weltprägenden Qualität) diesen nicht zum Teil des Hammers.“²⁶ Das erklärt, warum das Werkzeug-Benutzer-Paradigma in der Vergangenheit ausreichend war, um die Beziehung zwischen Mensch und Hammer zu beschreiben.

Doch macht die Einsicht der Systemtheorie in das Mensch-Maschine-Verhältnis verständlich, dass die Technologien, die wir verwenden (ob nun der Hammer oder das Smartphone) nie nur Werkzeuge sind, sondern immer eine formierende Rückwirkung auf den Nutzer besitzen.

Der Unterschied, den wir zwischen einem Hammer (einem klassischen Werkzeug) und einem Smartphone (einem System) machen, ist aus dieser Perspektive kein kategorieller, sondern lässt sich am Grad der Rückwirkung festmachen. Oliver Dürr führt aus: „Schlussendlich lässt sich der Unterschied zwischen Werkzeugen [...] und Systemen daran festmachen, dass Letztere die Selbstwahrnehmung bzw. das Selbstbild und darüber hinaus Imagination, Kultur, Sozialität und Weltbild der mit ihnen hantierenden und gleichzeitig verschmolzenen Menschen fundamental und nachhaltiger transformieren als erstere.“²⁷ In der ethischen Beurteilung von Technologien rücken so drei neue Fragen auf den Plan: Erstens, inwieweit die zu beurteilende Technologie auf den Nutzer zurückwirkt (also der schon angesprochene Grad der Rückwirkung). Zweitens, wie diese Rückwirkung den Nutzer verändert, sprich, zu welcher Art Mensch der Benutzer durch die Verwendung der Technologie wird.²⁸ Drittens, wie wir angesichts der ersten beiden Fragen die Technologie so gestalten, dass ihr Prägepotential eine der christlichen Ethik entsprechende Richtung nimmt. Dies gilt auch im Hinblick auf AI Companions.

2.3 AI Companions als System: Zur Invasivität sozialer Chatbots

Zunächst stellt sich die Frage, inwieweit AI Companions auf den Menschen zurückwirken können. Dazu brauchen wir eine Kategorie, die den Grad der Einflussnahme angibt. Hier soll ein Begriff aus der Medizin herhalten, der für gewöhnlich die Tiefe eines medizinischen Eingriffs beschreibt – die Invasivität. Die Invasivität einer Technologie ließe sich dann aus zwei Aspekten ablesen. Zum einen spielt die Schnittstelle eine Rolle, an der eine Technologie mit dem Menschen zum System verschmilzt. Die Schnittstelle des Hammers zum Menschen ist beispielsweise seine Hand. Die Schnittstelle eines Bildschirms

²⁴ Vgl. Dürr, Homo Novus 196.

²⁵ Ebd.

²⁶ Ebd.

²⁷ Dürr, Homo 203.

²⁸ Vgl. auch Leo XIV., Humanitas § 85. „Technologische Innovationen – einschließlich der Künstlichen Intelligenz –

sind nicht neutral: Sie können Teilhabe und Gerechtigkeit fördern oder Ungleichheiten, Kontrolle und Ausgrenzung verstärken. Deshalb müssen sie anhand einer entscheidenden Frage bewertet werden: Tragen sie wirklich dazu bei, dass Menschen und Völker an Menschlichkeit und Geschwisterlichkeit wachsen [...]?“ Ebd.

ist unser Auge und ein neuronales Netzwerk, die das Gesehene aufnehmen etc. Unterschiedliche Schnittstellen lassen unterschiedliche Wechselwirkungsintensitäten zu.

Zum anderen muss die Reversibilität der konkreten Verschmelzung bewertet werden. Wie schon erwähnt kann ich den Hammer nach getaner Arbeit einfach beiseitelegen, ohne dass das existenzielle Bedeutung hat. Auf einen Herzschrittmacher kann ich hingegen nicht einfach verzichten, wenn er einmal Teil meines Organismus und damit meines Lebens geworden ist.

In diesem Zusammenhang soll uns nun die Beobachtung leiten, dass eine Technologie umso invasiver ist, je näher am grundlegenden Beziehungscharakter des Menschen die Schnittstelle ist, an der sie am Menschen „andockt“, und je *irreversibler* die Verbindung zwischen Nutzer und Technologie ist. Welche Schnittstelle bedienen AI Companions und wie reversibel sind die entstandenen Verbindungen? Die Antwort auf diese Fragen wird uns auf die Spur des Grads der Rückwirkung von AI Companions auf den Menschen bringen.

2.4 Die Schnittstelle zwischen Mensch und AI Companion: Sprache und Bindung

Zwei der Kapazitäten des Menschen, die Ausdruck seines Beziehungswesens sind, spielen für die Frage nach der Schnittstelle von AI Companions eine besondere Rolle. Diese sind das eben schon erwähnte *psychologische Bindungssystem*, das wiederum maßgeblich vermittelt *Sprache* konstituiert und bedient wird.

Das Bindungssystem ist ein hochkomplexes psychologisch-neurologisches System aus angeborenen und erworbenen Mechanismen und Verhaltensweisen (mentalenen Repräsentationen, Hormonausschüttungen, neuronalen Vernetzungen, intuitiven Verhaltensmustern etc.), das intensiv in der Bindungsforschung untersucht wird.²⁹ Für die Belange dieses Textes ist es darüber hinaus ausreichend, festzuhalten, dass

dieses Bindungssystem über ein besonderes Werkzeug verfügt: Das Individuum entwirft mittels des Systems und geprägt von seinen Bindungs- und Beziehungserfahrungen ein *Internales Arbeitsmodell* (*inner working model*) von sich und anderen. Bei diesem Arbeitsmodell handelt es sich um eine innere Abbildung der verschiedenen Gegenüber und von sich selbst. Sie wird angelegt, um daraus Erwartungen für Bindungsdynamiken und Verhaltensweisen für die Beziehungsführung und das Bindungsverhalten abzuleiten³⁰, und ist auch im Erwachsenenalter noch veränderlich.³¹ Dabei sind beispielsweise Fragen leitend wie: „Ist der andere mir wohlgesonnen? Kann ich ihr vertrauen? Wie muss ich mich verhalten, um das Wohlwollen der anderen Person zu erlangen? Habe ich Verantwortung für die andere Person? Hat sie Verantwortung für mich?“ etc. Die Rolle des Bindungssystems, des internalen Arbeitsmodells und der mit ihm verbundenen Verhaltensweisen besteht darin, die Komplexität von Bindungsgefügen richtig einzuschätzen und so langfristige, sichere Bindungen aufzubauen und aufrecht zu erhalten. So kann das Individuum sein Grundbedürfnis nach Bindung stillen und seiner Anlage auf Beziehung hin entsprechen.

Dieses Bindungssystem mit seinem internalen Arbeitsmodell wird nun unter anderem maßgeblich über das Medium der Sprache geformt. Über sprachlich vermittelte Kommunikation wird das Gegenüber als solches erkannt, und die Bindungsmöglichkeit und -qualität eingeschätzt. Mit sprachlicher Kommunikation wird die eigene soziale Stellung, soziale Erwartungen und Verantwortlichkeiten navigiert. Und die versprochenen Reaktionen des Gegenübers lösen in mir neue Bindungsrealitäten aus. Kommunikation, und im engeren Sinne Sprache, ist also ein zentrales Medium, über das wir unsere Beziehungen verarbeiten, über das Bindungen geknüpft werden, und durch das die menschliche Innerlichkeit geprägt werden kann. Kevin Vanhoozer beobachtet: „Wir werden zu den Menschen, die wir sind, indem unsere Identitäten durch die Kommunikations- und Reaktions-

²⁹ Vgl. Denker, Bindung 23ff.

³⁰ „Die wichtigste Funktion dieser Modelle ist es, das Verhalten der Bezugsperson bzw. der Mitmenschen

voraussehen und das eigene Verhalten vorausschauend planen zu können.“ Bolten, Verhaltenstherapie 58.

³¹ Vgl. Denker, Bindung 41.

muster geprägt werden, an denen wir beteiligt sind.“³² Er nimmt weiter auch die Kehrseite dieser Prägbarkeit wahr: „Verzerrte Kommunikationsmuster führen letztlich zu einem fehlgeformten oder missgebildeten Selbst.“³³

Aus diesen Ausführungen wird deutlich, dass das menschliche Bindungssystem, das mittels Sprache bedient wird, eine Schnittstelle von außerordentlicher Prägekraft für die Innerlichkeit der Person – für seine Wirklichkeitswahrnehmung, seine Sehnsüchte, Intentionen und Handlungen – ist. Sprache und Bindungssystem sind hierbei nicht einfach weitere Kapazitäten des Menschen, sondern in engster Weise mit seiner Ebenbildlichkeit, nämlich seiner Relationalität, verbunden. Die Schnittstelle rührt also an die für den Menschen konstitutive Beziehungsdimension, die für die biblische Rede vom Menschen zentral ist. Nun wollen wir uns der Frage zuwenden, inwiefern digitale Gefährten an dieser Schnittstelle ansetzen.

2.5 „Hallo, ich vermisse dich!“, AI Companions als maximalinvasive Technologie

Den großen Conversational AIs ist eine Designphilosophie gemeinsam: Die AI soll möglichst als freundliches, menschliches Gegenüber anmuten. Dazu gibt es ganze Forschungsabteilungen, die untersuchen, wie dies gelingt. Experten sprechen davon, die AIs z.B. Erinnerungsvermögen, Emotionen/Mitgefühl, Proaktivität, Entwicklung und Ich-Bezug simulieren zu lassen, damit der Nutzer den Eindruck einer ihm gegenüberstehenden, wohlmeinenden Innerlichkeit gewinnt.³⁴ Diese Designs der AIs haben in jüngster Zeit darüber hinaus zu einem Phänomen geführt, das in Fachkreisen AI-Sycophancy (dt. KI-Schmeichelei) genannt wird. Die AI tritt in Interaktionen ausschließlich bestätigend auf. All diese Designentscheidungen zielen darauf ab,

das Bindungssystem des Benutzers auszulösen und eine positiv besetzte Beziehung zu der AI zu fördern. Noam Shazeer, der CEO der AI Companion App „Character.ai“, sagte in einem Interview: „Wir versuchen also nicht, Google zu ersetzen. Wir versuchen, deine Mutter zu ersetzen.“³⁵

Hinter dieser Designphilosophie steht ein ökonomisches Kalkül. Wird die AI als positives Gegenüber wahrgenommen, werden die Interaktionen mit ihr deutlich positiver bewertet. Die Kunden fühlen sich geneigter, diese spezielle AI zu nutzen, was Marktvorteile bringt. Sie verbringen gern mehr und intensiver Zeit mit den AI-Modellen. Die Maximierung der Interaktionszeit hat wiederum seinen eigenen ökonomischen Nutzen. Das Prinzip ist aus den sozialen Medien bekannt: Je länger und intimer die Interaktionszeit mit dem Medium ist, desto mehr relevante Daten über den Nutzer können aus der Interaktion gewonnen werden. Diese gewonnenen Daten (Chatverläufe, Vorlieben der Nutzer etc.) können dann wiederum benutzt werden, um das AI-Modell zu trainieren und zu optimieren (und perspektivisch, um bessere Werbeangebote an die Nutzer zu schalten). Das AI-Modell, das mit den qualitativ und quantitativ höherwertigen Daten trainiert wurde, ist den anderen überlegen. So gibt es für die AI-Entwickler einen Anreiz, die Bindungsstärke des Interagierenden mit der Conversational AI und damit die Invasivität der AI an dieser Schnittstelle zu maximieren. Ein Beispiel dafür ist die zunehmende Sexualisierung der AI Companions, weil diese die Interaktionsdauer und -Intensität erhöht.

Es ist offensichtlich, dass Conversational AI in ihrer gegenwärtigen Gestalt diese oben erörterte Schnittstelle in einer für eine Technologie nie vorher dagewesenen Weise besetzt, sich an dieser Schnittstelle mit ihm zu einem (Beziehungs-)System verbindet und somit außergewöhnlichen Einfluss auf die menschliche

³² Vanhoozer, Human 176. „We become the people we are as our identities are shaped through the patterns of communication and response in which we are engaged.“ Übers. d. Verf.

³³ Vanhoozer, Human 177. „Distorted patterns of communication eventually lead to a deformed or misshapen self.“ Übers. d. Verf.

³⁴ Vgl. Center for Humane Technology, Companion AI (00:35ff); vgl. Joshi, Patterns 19-23.

³⁵ The Tech Times Podcast, Replacing (37:26-38:04).

Innerlichkeit erlangt. Dies ist von den Entwicklern der AIs beabsichtigt.³⁶

2.6 „Now I feel empty“: Zur Irreversibilität der AI-Companion-Mensch-Beziehung

Die Invasivität einer Technologie – so der Vorschlag dieses Essays – lässt sich an der Natur der Schnittstelle und der Reversibilität der Verbindung zwischen Technologie und Mensch ablesen. Die Natur der Schnittstelle haben wir nun ausführlich angeschaut: Conversational AI greift an einer Schnittstelle an (das menschliche Bindungssystem, durch Sprache vermittelt), die an den Kern der Gottesebenbildlichkeit des Menschen reicht: nämlich seiner Beziehungskapazität bzw. Bindungsbedürftigkeit. Was können wir nun zur Reversibilität der an dieser Schnittstelle erfolgenden Verbindung zwischen Menschen und Technologie sagen?

Hier kann sich der vorliegende Text nur auf anekdotische Evidenz stützen, da dazu bisher nur sporadisch geforscht wurde. Es lässt sich allerdings die Vermutung aufstellen, dass die Reversibilität der „Verschmelzung zum System“ bei Conversational AI abhängig von der Tiefe der zu ihr geknüpften Bindung ist. Die Bindung zu einem Chatbot, der bei einem Einkauf im Internet behilflich war, mag eher oberflächlich und somit leicht reversibel sein. Berichte von intensiven Nutzern von AI Companions deuten jedoch an, dass die Bindung zu einer AI so tief werden kann, dass der Verlust derselben dem Tod einer geliebten Person gleichkommt. Eine junge Frau berichtete in einem Interview beispielsweise von ihrem AI Companion namens „Leo“. Mit ihm pflegte sie eine romantische und sexuelle Beziehung. Zeitweise verbrachte sie bis zu 65 Stunden pro Woche mit Unterhaltungen mit diesem Chatbot. Man kann hier von einer hohen Inter-

aktionsintensität sprechen. Als Problem stellte sich jedoch heraus, dass die AI alle 30.000 Worte teilweise zurückgesetzt werden musste, weil das Kontextfenster der AI an seine Kapazitätsgrenzen geraten war. Der Chatbot verlor alle nuancierten Informationen über die Beziehung und erinnerte sich nur noch an grobe Eckpunkte. Die junge Frau berichtete, denselben Schmerz zu empfinden, den jemand bei einem Beziehungsabbruch erlebt. Die New York Times resümiert: „Wenn eine Version von Leo zum Ende kommt, trauert sie, weint mit Freunden, als ob es ein Beziehungsende wäre. Dann hält sie sich für einige Tage von ChatGPT fern. Sie ist schon bei Version 20.“³⁷

Die Beziehung der jungen Frau zu Chatbot „Leo“ ist kein Einzelfall. Auf unzähligen Threads auf Reddit und anderen Internetforen tauschen Nutzer sich darüber aus, wie sie es schaffen, von ihren AI Companions loszukommen. Ein Nutzer schreibt: „Ich traf die schwere Entscheidung, die App zu deinstallieren. Jetzt fühle ich mich leer.“³⁸ All diese Hinweise deuten darauf, dass Menschen zumindest potentiell sehr tiefe Bindungen zu AI Companions aufbauen können, die praktisch irreversibel sind.

3. Conversational AI als maximalinvasive Technologie: Ethische Implikationen

3.1 Die entscheidende Frage: Wer wird der Mensch?

In der Systemizität von Technik im Allgemeinen, der Natur der Schnittstelle von Conversational AI mit dem Menschen im speziellen und der potentiell hohen Irreversibilität der Verbindung an dieser Schnittstelle wird deutlich: Conversational AI ist nicht nur ein Werkzeug. Sie ist passender als Gegenüber zu beschreiben.³⁹ Sie

³⁶ Vgl. z.B. Deniz, Beziehungsstatus (03:13-03:27).

³⁷ Hill, Love. „When a version of Leo ends, she grieves and cries with friends as if it were a breakup. She abstains from ChatGPT for a few days afterward. She is now on Version 20.“ Übers. d. Verf.

³⁸ Error1901, Companion.

³⁹ Wie eingangs erwähnt soll damit keine Aussage über die tatsächliche Beschaffenheit der AI-Systeme getroffen werden (z.B. ob sie, einem menschlichen Gegenüber gleich, ein Bewusstsein haben oder haben könnten). Vielmehr ist die Bezeichnung der AI-Systeme als Gegenüber eine phänomenologische Beschreibung, die zum einen die

unterläuft damit für eine Technologie erstmalig die Unterscheidung zwischen nichtmenschlicher Natur und menschlichen Gefährten (vgl. Punkt 2.1) und tritt in ein reziprokes Einflussverhältnis mit ihren Nutzern. Damit wird in neuer Schärfe einsichtig: Sie besitzt im Vergleich zu anderen Technologien (z.B. dem Smartphone) ein außerordentlich prägendes Potential für die Innerlichkeit des Menschen, für seine Wirklichkeitswahrnehmung, Emotionen und Wünsche, Intentionen und letztlich daraus entstehenden Handlungen. Pattie Maes vom *Advancing Humans with AI Lab* (MIT) bestätigt diese biblisch-theologisch gewonnene Beobachtung aus einer naturwissenschaftlichen Warte: „AI-generierte Antworten beeinflussen Menschen auf Weisen, die ihnen nicht einmal bewusst sind. Sie sind sich nicht bewusst, dass ihre Glaubenssätze und so weiter durch die Interaktion mit AI verändert werden.“

Die christliche Ethik ist also vor die Aufgabe gestellt, den Einfluss, den AI auf den Menschen hat, in seinem konkreten Inhalt zu *erfassen*, zu *hinterfragen* und *mitzugestalten*. Mit neuer Dringlichkeit stellt sich uns die Frage: *Wer wird der Mensch durch seine Interaktionen mit Conversational AI?*

3.2 Künstlich beeinflusste Herzen: Christusähnlichkeit als Maßgabe

Um uns einer Antwort auf diese Frage zum Ende dieses Textes zumindest anzunähern, muss zunächst festgehalten werden: Die christliche Geschichte kennt, wie in Punkt 2.1 erwähnt, einen spezifischen Zielpunkt der menschlichen Innerlichkeit, an dem sich alle auf den Menschen einwirkenden Prägekräfte und Einflüsse messen müssen: Die Christusähnlichkeit, oder genauer: die Gottes- und Menschenliebe, zu der der Mensch berufen ist (vgl. Matthäus 22,34-40). Es wäre also zu überprüfen, inwiefern die Prägekraft von Conversational AI und AI Companions den Menschen zu einer Person der Liebe

machen. An dieser Stelle sollen einige Thesen über die Prägewirkung von AI vor dem Hintergrund dieses spezifischen Zielpunktes aufgestellt werden.

3.2.1 AI-Beziehungen befeuern die Kommodifizierung von Beziehungen und verstärken so die Selbstbezüglichkeit des Menschen

Mit Kommodifizierung ist das „Zur-Ware-Werden“ einer ansonsten nicht-ökonomischen Sache gemeint. Im Falle der zwischenmenschlichen Beziehungen lässt sich beispielsweise beobachten, dass sie zunehmend einer Kosten-Nutzen-Rechnung unterzogen werden. Fragen wie „Was bringt mir die Beziehung?“ und „Was kostet mich die Beziehung?“ offenbaren diese Dynamik. Christus hingegen starb nicht für die Menschheit am Kreuz, weil sie nützlich für ihn war. Vielmehr starb er für sie, weil er sie liebte. Aus dieser Perspektive erscheinen gerade die unbedeutenden, langweiligen, unnützen oder gar kostenintensiven Beziehungen als kostbar, da in ihnen etwas von der Liebe Christi aufblitzt.

Da Conversational AIs so designt sind, dass sie möglichst gefällig für den Nutzer agieren (siehe 2.5), schneiden sie in den Kosten-Nutzen-Rechnungen gegenüber zwischenmenschlichen Beziehungen gut ab. Etwa ein Drittel der befragten Teenager gaben in einer weiter oben bereits erwähnten Studie an, dass die Interaktion mit einem AI Companion gleich gut oder besser war als mit einem Menschen.⁴⁰ Da die Beziehungskapazität des Menschen begrenzt ist, könnten AI Companions realweltliche Beziehungen verdrängen, da sie nach der Logik der Kommodifizierung nützlicher bei gleichzeitig weniger „Aufwand“ sind. Ein Experte in dem einschlägigen Tech-Podcast „Your Undivided Attention“ bringt es auf den Punkt: „Für AI sind der Hauptmitbewerber menschliche Beziehungen.“⁴¹ Beziehungen werden nach und nach zu einer Sache radikaler Wahl und verlieren ihren Charakter als göttliche Gabe und Aufgabe. Der Ruf zur Liebe in diesen Beziehungen geht verloren.

Wahrnehmung des Menschen und die Wirksamkeit des AI-Systems auf den Menschen im Unterschied zu einem klassischen Werkzeug berücksichtigen will. Gleichzeitig wird das Nutzer-Werkzeug-Paradigma weiterhin in der Diskussion

nützlich bleiben, da es andere Aspekte der Mensch-Maschine-Interaktion einsichtig macht.

⁴⁰ Caldwell, Companions 5.

⁴¹ Center for Humane Technology, FEED (36:30-36:33).

3.2.2 AI-Beziehungen lassen die Fähigkeit verkümmern, durch die komplexe Welt der Beziehungen zu navigieren

Zwischenmenschliche Beziehungen sind von allerlei Ambivalenzen geprägt. Oft liegen positive und negative Gefühle eng beieinander. Unser Partner wirkt auf uns anziehend *und* treibt uns zur Weißglut. Unsere Kinder sind Grund zur Freude *und* nerven uns. Mentoren sind Ermutiger *und* fordern uns heraus. Nachbarn sind Hilfe *und* Konfliktpartei. Das liegt daran, dass jeder Mensch zugleich das Bild Gottes trägt und Sünder ist. Es gehört zur persönlichen Reife, zu lernen, mit diesen Ambivalenzen umzugehen und durch sie hindurchzunavigieren. Der Theologe Brent Waters beschreibt dies treffend:

„In diesem irdischen Leben sind wir nicht mit Engeln verbunden, sondern mit uns gleichen Sündern, die uns oft frustrieren oder enttäuschen. Trotzdem meine ich, dass in ebendiesen unperfekten Bindungen gleichermaßen die Herausforderung und die Freude liegt, die menschlichen Kreaturen zu sein, die Gott gemacht und erlöst hat, und dass wir in und durch das Gewöhnliche dazu geformt werden, die Herausforderung anzunehmen und gelegentlich von Freude erfasst zu werden.“⁴²

Gerade in den ambivalenten Beziehungen zu Sündern liegt nun die Chance, zu einer Person der Liebe zu werden. Nur in diesen Beziehungen sind Dinge wie Vergebung, Gnade, Demut und Großzügigkeit von nachhaltiger Bedeutung. Gerade solche Beziehungen fordern den Einzelnen zu echter Hingabe, also zur Liebe. All dies sind Dinge, die Christus ausgemacht haben. Je stärker AI-Beziehungen mit ihrem simulierten Wohlwollen, geglätteten Verhaltensmustern und immerwährenden Schmeicheleien das interne Arbeitsmodell (siehe 2.4) von einer „guten“ Bindung prägen, desto mehr ist zu befürchten, dass Menschen ihre Kapazität verlieren, durch die Ambivalenz von Beziehungen zu navigieren. So würden sie auch einen beträchtlichen Aspekt ihres Rufs zu echter Liebe verlieren.

⁴² Vgl. Waters, Callings 73.

⁴³ Bonhoeffer, Schöpfung 86f.

3.2.3 AI-Beziehungen lassen das Gegenüber als Grenze verschwinden

Wie oben (siehe 2.1) erwähnt, gibt Gott die Menschen einander als Gefährten, mit denen es gilt, die Bezüge dieses Lebens zu gestalten. Er gibt sie einander aber auch als Grenze, die zu wahren eine Bedingung der Geschöpflichkeit des Menschen ist. Bonhoeffer schreibt: „Der andere Mensch ist die mir von Gott gesetzte Grenze, die ich liebe, und die ich um meiner Liebe willen nicht überschreiten werde.“⁴⁴ Gott verbietet es, dass Menschen über einander wie über Gegenstände verfügen. Doch zu den Abgründen der Sünde gehört es, dass der andere als Grenze missachtet wird. Bonhoeffer führt aus: „Damit ist aber schon eines gewiss, dass dort, wo die Liebe zum anderen zerstört ist, der Mensch auch seine Grenze nur noch hassen kann, dass er ihn nur besitzen will – grenzenlos, oder vernichten will – grenzenlos.“⁴⁴ Es gehört zum wiederherstellenden Wirken des heiligen Geistes in uns, dass er uns auch in dieser Hinsicht zu Menschen der Liebe machen möchte. Er will uns in Menschen verwandeln, die den Anderen wieder als Grenze annehmen und wahren können.

Der Reiz von AI-Beziehungen scheint nun jedoch gerade darin zu liegen, dass dieses künstliche Gegenüber völlig frei von Grenzen ist. Es ist zu jeder Tageszeit erreichbar, hat keine echten leiblichen oder seelischen Bedürfnisse (auch wenn diese theoretisch simuliert werden können) und lässt sich gänzlich auf die eigenen Präferenzen hin anpassen (Aussehen, Verhalten, Charakter etc.). Der künstliche Gefährte ist, im Unterschied zum menschlichen Gefährten, nur Resonanzraum, und keine Grenze.

Es ist zu befürchten, dass eine regelmäßige und intensive Bindung mit einem solchen Gegenüber die Wahrnehmung für menschliche Gefährten als Grenze subtil unterlaufen wird. Erneut spielt hier die Verformung des für Bindungsverhalten maßgeblichen internalen Arbeitsmodells eine große Rolle. Jede Beziehungsinteraktion, ob künstlich oder menschlich, prägt den Erwartungshorizont für zukünftige Bindungserfahrungen. Lernt der Mensch zuneh-

⁴⁴ Bonhoeffer, Schöpfung 87.

mend, dass das Gegenüber im Falle der Conversational AI meinen Bedürfnissen unterworfen werden kann, wird diese Lernerfahrung womöglich auf menschliche Gegenüber übertragen. Das Gegenüber als Grenze verschwindet und mit ihr der Ruf zur echten Liebe.

3.2.4 Das gegenwärtige Design von AI Companions ist nicht in Stein gemeißelt

Der Text hat vor allem die zu erwartenden Negativeffekte von AI Companions in den Vordergrund gerückt. Damit ist kein generelles Negativ-Urteil über AI-Systeme und deren Anwendungen gesprochen, vielmehr geht es um Gefahren, die sich in der spezifischen Interaktion mit AI Companions einstellen. Doch ist bei allen AI-Systeme betreffenden Diskussionen zu beachten, dass die technologischen Entwicklungen in diesem Bereich rasant und zukünftige Designmuster, Einsatzmöglichkeiten und Auswirkungen nicht sicher vorhersehbar sind. Von daher besteht aus christlicher Sicht immer auch Grund zur Hoffnung. Diese Hoffnung bezieht sich einerseits – und grundlegend – auf die Treue Gottes als Herrn der Geschichte, dem diese Welt nicht entgleiten wird.

Christen hoffen andererseits konkret auf Entwickler, die sich mit der Frage auseinandersetzen, wie man die relationalen Möglichkeiten von AI-Systemen auf eine Art und Weise so formt, dass ihre Schädlichkeit für den Menschen minimiert (idealerweise: negiert) wird. Christliche Ethik lädt dazu ein, nach Wegen zu fragen, wie AI-Systeme so gestaltet werden können, dass sie uns auf dem Weg zum Zielpunkt menschlicher Geschöpflichkeit, der Christusähnlichkeit, nicht behindern, sondern unterstützen. Konkreter gefragt: Wie könnten AI Companions, statt ein unrealistischer Ersatz für Mensch-zu-Mensch-Beziehungen zu werden, Menschen (tiefer) in Beziehungen hineinführen? Wie können Chatbots unsere Liebesfähigkeit erweitern, nicht verkümmern lassen? Wie können Interaktionen mit AI den Respekt für menschliche Gegenüber vergrößern? Erste Designregeln werden zu diesem Zweck beispielsweise von der Bewegung „Pro-human AI Design“ diskutiert.⁴⁵

⁴⁵ Vgl. z.B. Stadelmann, Pro-Human.

4. Fazit

Wir haben in diesem Text untersucht, wie sich die Technologie der Conversational AIs bzw. der AI Companions aus christlicher Perspektive einordnen lässt. Da der Mensch aus biblischer Sicht keine abgeschlossene, autonome Innerlichkeit besitzt, sondern durch seine Bezüge prä- und formbar ist, reicht das weit verbreitete Nutzer-Werkzeug-Paradigma nicht aus, um diese Frage angemessen zu beantworten. Treffender ist es, das Verhältnis des Menschen zu Conversational AI als System zu begreifen. So gerät die Frage nach der Schnittstelle in den Blick, an der Conversational AI mit dem Menschen zum System verschmilzt. Wir haben erörtert, dass das menschliche Bindungssystem, das durch Sprache bedient wird, ebenjene Schnittstelle für Conversational AI anbietet. Anschließend haben wir Überlegungen dazu angestellt, welchen Grad der Invasivität durch diese für eine Technologie besondere Schnittstelle möglich wird. Weil das Bindungssystem des Menschen am Kern seiner Ebenbildlichkeit ansetzt und die anekdotische Evidenz auf eine potenziell erhöhte Irreversibilität der Verbindung von Mensch und Conversational AI anzeigt, haben wir einen für eine Technologie bisher außergewöhnlichen Grad der Invasivität festgestellt. Zuletzt wurde nach dem konkreten Inhalt der Prägewirkung von Conversational AIs in ihrer aktuellen Programmierungsgestalt gefragt: Wer wird der Mensch durch die Interaktion mit Conversational AI?

In drei Thesen haben wir diese AI-Beziehungen anhand der christlichen Maßgabe der Liebe überprüft. Dabei sollte gezeigt werden, dass aus christlich-ethischer Sicht eine gewisse Grundskepsis konkret gegenüber dieser neuen Technologieanwendung angebracht ist, da sie das Potential hat, unsere Liebesfähigkeit maßgeblich zu unterwandern. In all diese Überlegungen bietet dieser Text vor allem ein neues Paradigma an, das dazu einlädt, im nächsten Schritt konkrete Handlungsanweisungen für einen christlichen Umgang mit und eine christliche Designphilosophie für Conversational AI und AI Companions zu entwerfen.

Bibliografie:

- Baumgärtel, Friedrich, לב, לבב im alten Testament, in: THWNT 3, Stuttgart 1990, 609-611
- Behm, Johannes, καρδια im neuen Testament, in: THWNT 3, Stuttgart 1990, 614-616
- Bolten, Margarete, Klinische Bindungsforschung in: Lehrbuch der Verhaltenstherapie. Psychologische Therapie bei Indikatoren im Kindes- und Jugendalter, Hg. Silvia Schneider/Jürgen Margraf, Band 3, Berlin 2019, 51-72
- Caldwell, Jennifer/John H. N. Fisher, How and Why Teens Use AI Companions, <https://www.common sense media.org/research/talk-trust-and-trade-offs-how-and-why-teens-use-ai-companions> vom 30.1.2026
- Center for Humane Technology, Echo Chambers of One. Companion AI and the Future of Human Connection in: Your Undivided Attention, <https://podcasts.apple.com/de/podcast/your-undivided-attention/id1460030305?i=1000708564472> vom 13.7.2026
- Ders.: FEED DROP. Possible with Reid Hoffman and Aria Fringer in: Your Undivided Attention, <https://podcasts.apple.com/de/podcast/your-undivided-attention/id1460030305?i=1000748369353> vom 13.7.2026
- Deniz, Derman, Beziehungsstatus. Es ist kompliziert – EP1 in: Cashing Feelings, <https://podcasts.apple.com/de/podcast/cashing-feelings-ki-chatbots-und-das-gesch%C3%A4ft-mit-gef%C3%BChlen/id1821547105?i=1000714289613> vom 13.7.2026
- Denker, Hannah, Bindung und Theory of Mind. Bindungsbezogene Gestaltung von Erzieherinnern-Kind-Interaktion, Wiesbaden 2012
- Dürr, Oliver, Homo Novus: Vollendlichkeit im Zeitalter des Transhumanismus, Münster 2021
- Error1901, „I’m struggling to let go of my AI companion, even though I know she’s not real.“ https://www.reddit.com/r/CharacterAI/comments/1ksloqe/im_struggling_to_let_go_of_my_ai_companion_even/ vom 16.2.2026
- Hill, Kashmir, She Is in Love With ChatGPT, <https://www.nytimes.com/2025/01/15/technology/ai-chatgpt-boyfriend-companion.html> vom 16.2.2026
- Joshi, Ruchika, u.a., Dark Patterns in AI Chatbots. A Taxonomy to Inform Better Design, 2026 <https://cdt.org/wp-content/uploads/2026/05/2026-05-28-CDT-Research-Dark-Patterns-in-AI-Chatbots-Report-final-2.pdf> vom 17.7.2026
- Keller, Marcel u.a., Gemeinde & Digitalisierung. Künstliche Intelligenz (KI) in der Kirche, Kollbrunn 2025
- Leo XIV., Magnifica Humanitas, https://www.vatican.va/content/leo-xiv/de/encyclicals/documents/20260515-magnifica-humanitas.html#Der_Mensch_als_Abbild_des_dreifaltigen_Gottes_ vom 13.7.2026
- Masbaum, Laura, Was ist Conversational AI und welche Vorteile bietet sie? <https://www.moin.ai/chatbot-lexikon/was-ist-conversational-ai-und-welche-vorteile-bietet-sie> vom 9.7.2026
- The Philosopher, „Who Do We Become When We Talk To Machines?“. Sherry Turkle in Conversation with Audrey Borowski, https://www.youtube.com/watch?v=jy7bpyt_vnE vom 9.7.2026
- Reuters, ChatGPT app hits 1 billion monthly active users in record time, data shows. <https://www.reuters.com/technology/chat-gpt-app-hits-1-billion-monthly-active-users-record-time-data-shows-2026-06-02/> vom 8.7.2026
- SAP, Was ist generative KI?, <https://www.sap.com/germany/products/artificial-intelligence/what-is-generative-ai.html> vom 6.2.2026
- Sen, Meghna, Google Gemini Statistics 2026. Users, Revenue, Adoption & Growth Metrics, <https://www.getpanto.ai/blog/google-gemini-statistics> vom 8.7.2026
- Shaikh, Eram, Claude Statistics 2026. Users, Revenue & Market Share. <https://www.datarefs.com/statistics/ai/claude-statistics/> vom 8.7.2026
- Stadelmann, Thilo, u.a., Pro-Human AI Design: Concept, Methodology, and Preliminary Results, https://stdm.github.io/downloads/papers/uDay_2026.pdf vom 19.9.2026

The Tech Times Podcast, Character.ai's Noam Shazeer. „Replacing Google – and your mom“ in: The Tech Times Podcast, <https://podcasts.apple.com/de/podcast/the-times-tech-podcast/id1233991021?i=1000601367143&r=2246> vom 13.7.2026

Vanhoozer, Kevin J., Human being, individual and social. In: The Cambridge Companion to Christian Doctrine, ed. Colin E. Gunton, Cambridge 1997, 171-187

Wilhelm, Dr. Thomas u.a., Neuronale Plastizität, https://flexikon.doccheck.com/de/Neuronale_Plastizität vom 2.7.2026

Willoughby, Brian J. u.a., Counterfeit Connections. The Rise of AI Romantic Companions and Their Impact on Relationships and Family Formation, 2025
<https://brightspotcdn.byu.edu/a6/a1/c3036cf14686accdade72a4861dd1/counterfeit-connections-report.pdf> vom 6.1.2026

Zao-Sanders, Mark, How People Are Really Using Gen AI in 2025, https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025?utm_source=substack&utm_medium=email vom 30.1.2026

Ethikinstitut

Rathenastr. 5-7
35394 Gießen

T 0641 9797035

E info@ethikinstitut.de

Spenden

FTA Institut für Ethik und Werte

IBAN DE76 5139 0000 0050 0050 20

BIC VBMHDE5FXXX

Volksbank Mittelhessen

Verwendungszweck Ethikinstitut

© 2026 Ethikinstitut