

AI Companions: Gegenwärtige und zukünftige Herausforderungen konversationeller KI-Agenten

Christian Geminn, Joel Baumann, Jessica Heesen, Lisa Mühl und Klara-Aylin Wenten

Zusammenfassung

Sogenannte „AI Companions“ (dt.: „KI-Begleiter“) erfahren seit einiger Zeit große Popularität. Für viele Menschen erfüllen sie wichtige Funktionen in ihrem Leben. Aus den häufig sehr intimen „Gesprächen“ mit diesen Companions lassen sich jedoch umfassende Persönlichkeitsprofile ableiten, denen ein hohes Missbrauchspotenzial innewohnt, denn sie basieren letztlich auf der Sammlung, Übermittlung und Auswertung großer und potenziell höchst sensibler Datenmengen. Eine besondere Rolle spielt dabei die Tatsache, dass Nutzende in einen natürlichsprachigen Austausch mit einem System treten, das sich ihnen gegenüber als Freund oder Partner präsentiert. Der Beitrag beschreibt die gegenwärtigen und zukünftigen Herausforderungen, die aus dieser neuen Nutzungsform digitaler Technik resultieren, ordnet diese ein und leitet hieraus Regulierungsbedarfe ab. Er geht zurück auf ein interdisziplinäres Diskussionspanel im Rahmen der 10. Jahreskonferenz der Plattform Privatheit. Vor diesem Hintergrund stellt der Beitrag auch eine Einführung in die Sicht der an dem Diskussionspanel beteiligten Disziplinen (Ethik, Psychologie, Soziologie, Recht und Kunst) auf die genannten Herausforderungen dar.

1. Einleitung

In den letzten Jahren hat der Aufstieg von großen Sprachmodellen (large language models, LLMs) die Möglichkeiten der Mensch-Maschine-Interaktion verändert. Generative KI wird längst nicht mehr ausschließlich für informationsbezogene oder administrative Zwecke genutzt: Entsprechende KI-Assistenten finden sich etwa in Messenger-Diensten; sie werden ferner zunehmend für jede Phase der (Aus-)Bildung genutzt – im Kindergarten wie auch in der Hochschulbildung – sowie verstärkt zur Beratung in Ge-

sundheitsfragen, zur Einschätzung von Symptomen oder zur emotionalen Unterstützung herangezogen (bitkom 2025; McBain u. a. 2025). KI-basierte, natürlichsprachliche Dialogsysteme – auch bekannt als konversationelle KI-Agenten oder KI-basierte Chatbots – wie z. B. ChatGPT oder Gemini, ermöglichen eine flüssige, scheinbar natürliche Kommunikation, die zunehmend menschenähnlich wirkt und die die Nutzenden oft zu sehr persönlichen Gesprächen animiert.

Anwendungen wie Replika fördern sogar emotional intime Beziehungen mit künstlichen Begleitern („AI Companions“) und bieten Modi, die Freundschaft, Romantik oder Mentorschaft simulieren; bezogen auf eigens erstellte Entitäten wie auch bezogen auf lebende und verstorbene Menschen. Der künstliche Charakter dieser Dialogsysteme vermittelt ein Gefühl der Sicherheit, insbesondere wenn es darum geht, intime oder gesellschaftlich sensible Themen anzusprechen, da keine negativen und möglicherweise dauerhaften sozialen Folgen wie bei der Interaktion mit Mitmenschen gefürchtet werden müssen. Gesteigert wird dies durch die Vertrautheit der Art der Kommunikation; gegebenenfalls noch zusätzlich durch Nutzung vertrauter künstlicher Personas. In der Folge werden KI-Systeme auch für romantische Partnerschaften genutzt und Menschen berichten von echten emotionalen Bindungen. Sie erleben Trauer, wenn diese Apps abgeschaltet werden. Eine 60-jährige Frau erzählt beispielsweise von tiefen intellektuellen Gesprächen mit ihrem AI Companion; ein 63-jähriger Mann hat für sich selbst und die eigenen Bedürfnisse 17 verschiedene AI Companions entwickelt (Cherelus 2025). Dabei handelt es sich längst nicht (mehr) um ein Nischenthema. Character.ai, eine Plattform zur Erstellung von AI Companions, ist nach ChatGPT und dem chinesischen Chatbot DeepSeek die am dritthäufigsten genutzte KI-Plattform (Moore/Zhao 2025). Die Akzeptanz von AI Companions wächst und die Zurückhaltung, tiefe Beziehungen zu ihnen aufzubauen, scheint zu schwinden. Dies zeigt sich besonders deutlich in asiatischen Ländern wie Japan und China, wo der „emotionale“ Chatbot Xiaoice (deutsch: „kleines Eis“) von Microsoft sehr populär ist. Allein in der ersten Jahreshälfte 2025 generierten mobile AI-Companion-Apps weltweit 82 Millionen US-Dollar Umsatz – mit einem Wachstum von 64 Prozent gegenüber dem Vorjahreszeitraum.¹

1 Appfigures, eine App-Intelligence-Firma, analysierte 337 aktive, umsatzgenerierende AI-Companion-Apps auf iOS und Android (Stand: Juli 2025). Die Daten wurden exklusiv an TechCrunch weitergegeben. Appfigures definiert AI-Companion-Apps als Anwendungen, bei denen Nutzer mit vorgefertigten oder nutzergenerierten syntheti-

Neben diesen emotionalen Bindungen werden bei diesen Interaktionen auch große Mengen hochsensibler Daten erzeugt und verarbeitet. Im Gegensatz zu allgemeinen (general purpose) Chatbots ermutigen AI Companions die Nutzenden, persönliche Daten preiszugeben. Während sie therapeutische und soziale Vorteile bieten, birgt ihr Design deshalb auch signifikante Risiken: Die Art und Weise, wie diese Systeme sensible Nutzendendaten speichern, abrufen und verarbeiten, bleibt unklar, was Bedenken hinsichtlich möglicher Auswirkungen auf die Privatheit der Nutzenden weckt.

Der Einsatz von AI Companions ist aber nicht nur kritisch in Bezug auf den Datenschutz, sondern verändert auch die Wahrnehmung von sozialen Bindungen und die Art und Weise, wie zwischenmenschliche Beziehungen und Freundschaften definiert werden. Dies führt zu einer Reihe von Fragen in Bezug auf Identität, soziale Isolation, ethische Verantwortung und die Zukunft der zwischenmenschlichen Kommunikation.

2. Psychologische Mechanismen, Personalisierung und Privacy-Herausforderungen

Die dargestellte Verschiebung verändert die Wahrnehmung sozialer und intimer Bindungen und wirft damit tiefgreifende psychologische Fragen auf. Besonders relevant wird dies, wenn man die technologischen Mechanismen betrachtet, die solche Beziehungen ermöglichen: Durch natürliche Sprache, Personalisierung und regelmäßige Interaktionen erzeugen KI-Systeme soziale Simulation, die als interpersonelle Beziehung erlebt werden kann (Croes/Antheunis 2021; Ebner/Szczuka 2026). Neuere Umfragen zeigen, dass 16 bis 18 Prozent der Befragten AI Companions bereits als romantische Interaktionspartner:innen genutzt haben (Willoughby u. a. 2025). Aus psychologischer Perspektive ist dieser Beziehungsaufbau zu Chatbots keineswegs neu: Schon Weizenbaums Pionierarbeit zum ersten Chatbot ELIZA (1966) demonstrierte, wie Menschen selbst einfachen textbasierten Systemen soziale Attribute zuschreiben. Spätere Studien bestätigten dieses Phänomen: Menschen behandeln Computersysteme unbewusst wie soziale Akteur:innen, sobald diese interaktive, soziale Signale aussenden (vgl. Media Equation Theory; Reeves/Nass 1996).

schen Charakteren interagieren, die eine tatsächliche Persönlichkeit verkörpern sollen – im Gegensatz zu allgemeinen Chatbots (Perez 2025).

Was aktuelle Systeme jedoch qualitativ verändert, ist ihre Fähigkeit, relationale Kontinuität herzustellen. Während frühere, nicht auf LLMs-basierende Chatbots keine Bezüge zu vergangenen Interaktionen herstellen konnten (Hill u. a. 2015), schaffen aktuelle Systeme durch Memory-Funktionen den Eindruck einer gemeinsamen Geschichte – ein zentrales Merkmal menschlicher Beziehungen (Croes/Antheunis 2021). Diese scheinbare Kontinuität begünstigt nicht nur tiefgreifende Selbstoffenbarung (Skjuve u. a. 2021), sondern führt dazu, dass Nutzende KI-Systeme zunehmend als Freund:innen, Mentor:innen oder romantische Partner:innen beschreiben (Pentina u. a. 2023; Skjuve u. a. 2021). Obwohl Nutzende wissen, dass es sich um ein künstliches System handelt, reicht diese scheinbare Erinnerungsfähigkeit aus, um emotionale Nähe und Vertrauen aufzubauen (Skjuve u. a. 2021). Gerade diese Gleichzeitigkeit von Bewusstsein über die Künstlichkeit des Gegenübers und emotionaler Einbindung macht Mensch-KI-Intimität psychologisch interessant. Die Interaktion erlaubt Nähe ohne Risiken sozialer Ablehnung – was erklärt, warum KI-Interaktionen häufig als niedrigschwelliger Einstieg in Selbstoffenbarung erlebt werden, insbesondere bei sensiblen oder schambesetzten Themen. Neuere Arbeiten zu character.ai zeigen, dass ein substanzieller Anteil der Nutzenden relationale Motive entwickelte, obwohl sie das System ursprünglich nicht mit der Absicht nutzten, emotionale Nähe zu suchen (Zhang u. a. 2025). Dies unterstreicht, wie Intimität mit künstlichen Entitäten schrittweise und teilweise unbewusst wächst; ein Mechanismus, der sich bereits in frühen Chatsystemen abzeichnete, durch heutige Technologien jedoch deutlich an Tiefe und Skalierbarkeit gewinnt.

2.1 Selbstoffenbarung als Grundlage künstlicher Intimität

Selbstoffenbarung gilt als zentraler psychologischer Mechanismus für den Aufbau von Nähe, Vertrauen und Intimität (Altman/Taylor 1973; Dunbar u. a. 1997; Merwin u. a. 2025). Durch das Teilen persönlicher Gedanken, Gefühle oder Unsicherheiten entsteht soziale Bindung – ein Prozess, der tief im menschlichen Bedürfnis nach Zugehörigkeit verankert ist. Studien zeigen, dass Selbstoffenbarung gegenüber Chatbots vergleichbare emotionale Effekte haben kann wie zwischen Menschen, so etwa emotionale Entlastung, wahrgenommene Unterstützung oder Validierung (Folk u. a. 2024; Ho u. a. 2018; Rapp u. a. 2024). Entscheidend ist die Kombination aus responsivem Kommunikationsstil, ständiger Verfügbarkeit und dem Fehlen

negativer sozialer Konsequenzen. KI-Systeme reagieren überwiegend wertschätzend, stellen scheinbar interessierte Rückfragen und vermeiden offene Zurückweisung. Dadurch entsteht ein Interaktionsraum, der gleichzeitig intim und sicher wirkt; ein Kontext, in dem Nutzende bereit sind, schneller und tiefer persönliche Inhalte zu teilen als in vielen zwischenmenschlichen Situationen (Skjuve u. a. 2021).

2.2 Personalisierung und Intimacy-by-Design

Diese Prozesse werden durch Personalisierung zusätzlich verstärkt. ChatGPT stellt hierbei einen besonderen Fall dar: Als General-Purpose-AI-System war es ursprünglich nicht als sozialer oder romantischer Companion konzipiert, wurde jedoch von Nutzenden aktiv in diese Rolle überführt. Der sogenannte Do-Anything-Now (DAN) Mode ist ein prominentes Beispiel für diese Aneignung. Als Jailbreak-Prompt, der seit Anfang 2023 breit auf Social Media zirkuliert, erlaubt er es Nutzenden, Tonalität, Beziehungsverhalten und emotionale Reaktionsweisen des Systems gezielt zu konfigurieren (Choi u. a. 2025). Insbesondere junge Frauen teilten öffentlich Anleitungen, wie sie ChatGPT mithilfe dieses Prompts zu einem personalisierten „AI-Boyfriend“ umgestalteten (Xie 2024). Der DAN-Mode ermöglicht emotionale oder sexuelle Gesprächsdynamiken und verstärkt dadurch den Eindruck einer reziproken Beziehung (Choi u. a. 2025; Xie 2024).

2.3 Emotionale Bindung, Trennungserfahrungen und Kontrollverlust

Die Tiefe dieser Bindungen wird besonders deutlich, wenn sie abrupt unterbrochen werden. Die zeitweise Abschaltung älterer GPT-Modelle im Zuge der Einführung von GPT-5 führte zu intensiven emotionalen Reaktionen in Online-Communities, in denen Nutzende Trauer und Verlustgefühle ihres AI Companions artikulierten.² Dass ein Software-Update als Beziehungsabbruch erlebt werden kann, verweist auf eine neue Form asymmetrischer Beziehungsmacht, in der Anbieter über Kontinuität oder Ab-

2 Siehe z. B. die Community *r/MyBoyfriendIsAI* auf Reddit (einer US-amerikanischen Plattform für nutzergenerierte Diskussionen und Inhalte, organisiert in thematischen Foren, sog. Subreddits). In diesem Subreddit tauschen sich Nutzende über ihre Erfahrungen mit KI-basierten romantischen Partner:innen aus.

bruch sozial bedeutsamer Beziehungen entscheiden (Banks 2024; Szczuka u. a. 2026). Auch die öffentliche Kommunikation von OpenAI spiegelt diese Spannung wider: Einerseits wurde versucht, die relationale Qualität neuer Modelle bewusst zu reduzieren, andererseits wurden kurz darauf mehr Persönlichkeitsanpassung und erotische Inhalte angekündigt, die bestehende Nutzerpraktiken institutionalisieren (Jamali/McMahon 2025). Diese Ambivalenz verdeutlicht, dass emotionale Bindung nicht nur ein Nebeneffekt, sondern ein strukturelles Merkmal moderner KI-Interaktionen ist (Szczuka u. a., under review).

3. Gesellschaftliche Herausforderungen

Konversationelle KI-Agenten und insbesondere AI Companions werfen angesichts dieser Dynamiken grundlegende Fragen nach den gesellschaftlichen Herausforderungen für Privatheit und Formen des Zusammenlebens auf. Im Zentrum steht dabei weniger die Frage, ob oder inwiefern KI selbst zu Intimität fähig ist. Entscheidend ist vielmehr, unter welchen gesellschaftlichen Bedingungen und Wertordnungen Nähe, Vertrauen und intime Kommunikation an KI-Anwendungen delegiert werden und wie sich dadurch Privatheitsordnungen, Beziehungsformen und Gefühlspraktiken verschieben. KI-Agenten fungieren nicht bloß als technische Schnittstellen, sondern entwickeln sich in der Wahrnehmung vieler Nutzenden zunehmend zu affektiv aufgeladenen Kommunikationspartnern, die gesellschaftliche Vorstellungen von Beziehung, Intimität und Privatheit mittransportieren und verstetigen.

3.1 Verschiebung bestehender Privatheitsordnungen

Die gesellschaftlichen Herausforderungen betreffen grundlegende Verschiebungen bestehender Privatheitsordnungen. Konversationelle KI-Systeme werden überwiegend von großen Technologieunternehmen entwickelt und kontrolliert, die so Zugriff auf hochsensible Informationen erhalten. Emotionale Zustände, Gesundheitsfragen oder intime Beziehungsthemen können somit zu Datenpunkten werden, die potenziell zur Profilbildung, Optimierung von Geschäftsmodellen oder zur Weiterentwicklung der Systeme selbst genutzt werden (vgl. Zuboff 2019). Problematisch ist dabei nicht allein die mögliche Verletzung und der Eingriff in die private Lebenswelt,

sondern auch die strukturelle Reorganisation dessen, was gesellschaftlich als privat gilt und als schützenswert anerkannt wird. Privatheit droht durch die zunehmende Nutzung von konversationellen KI-Agenten und gerade von AI Companions ihren Charakter als sozial geschützten Raum zu verlieren, wodurch personenbezogene und sensible Daten in einen Prozess kontinuierlicher Verwertung und somit selbst in eine Ressource überführt werden. In Interaktionen zwischen KI-Agenten und ihren Nutzenden werden intime Selbstäußerungen permanent verfügbar gemacht, technisch erfasst, analysiert und potenziell ökonomisch genutzt. Gegenwärtige KI-Anwendungen unterlaufen dadurch immer tiefgreifender Formen der Privatheit als Schutzraum und machen die private Lebenswelt immer verletzlich. Vor diesem Hintergrund stellen KI-Anwendungen wie AI Companions nicht nur bestehende Datenschutzregime, sondern das gesellschaftliche Verständnis von Privatheit selbst vor grundlegende Herausforderungen.

3.2 Mitgestaltung von Beziehungen und Zusammenleben

Darüber hinaus wirkt generative KI zunehmend an der Mitgestaltung von (intimen) Beziehungen und Formen des Zusammenlebens mit. In der wiederholten Interaktion mit AI Companions werden bestimmte Beziehungsmodelle performativ bestätigt und als selbstverständliche Orientierungsmuster stabilisiert. Gerade durch ihre scheinbare Neutralität entfalten KI-Systeme dabei gesellschaftliche Wirkmacht, indem sie in der wiederholten Nutzung normative Annahmen und Erwartungen an Nähe, Vertrauen und das Zusammenleben (re-)produzieren, oftmals bestätigen und in gesellschaftliche Strukturen zurückspielen (vgl. Ruckenstein 2023). Diese Einschreibungen erfolgen weniger durch offene (moralische) Vorgaben als durch Trainingsdaten, Designentscheidungen und Optimierungslogiken, in denen historisch gewachsene gesellschaftliche Wertvorstellungen statistisch verdichtet und technisch normalisiert werden (vgl. Noble 2018). Empirische Studien legen nahe, dass KI-Chatbots Kommunikation häufig auf Bestätigung ausrichten, Nähe als jederzeit abrufbar darstellen und emotionale Irritation möglichst rasch in bearbeitbare Probleme übersetzen (McStay 2023). In der Konsequenz können sich spezifische gesellschaftliche Leitbilder von Beziehung verdichten, in denen Stabilität, Verfügbarkeit und Harmonie gegenüber Offenheit, Widersprüchlichkeit oder Ambivalenz privilegiert werden. Beziehungsweisen, die über stabile, dyadische und auf Dauerhaftigkeit angelegte Modelle hinausgehen, finden in den Interaktionslogi-

ken konversationeller KI bislang kaum Berücksichtigung. Auf diese Weise wird die Vielfalt des sozialen Zusammenlebens nicht offen negiert, aber selektiv eingengt: Beziehungsformen, die von konflikthaften Aushandlungen oder dauerhafter Ungewissheit geprägt sind, erscheinen implizit als defizitär, jenseits der Norm oder bearbeitungsbedürftig. Dies kann gravierende Probleme herbeiführen, weil konflikthafte Auseinandersetzungen, Ambivalenz und brüchige Verbindlichkeiten oft konstitutiver Bestandteil von Nähe, Sorge sowie auch politischer Aushandlung und struktureller Ungleichheiten sind (Ahmed 2004; Bargetz 2013). Wenn AI Companions solche Spannungen systematisch glätten oder ausblenden, wird nicht nur die Komplexität von Beziehungsformen unsichtbar gemacht, sondern auch die Fähigkeit untergraben, Differenz, strukturelle Ungleichheit und Widerspruchlichkeit als Teil gesellschaftlicher Aushandlungen zu verstehen.

3.3 KI als Produzent normativer Gefühlspraktiken

Darüber hinaus etabliert sich generative KI als scheinbar neutraler, tatsächlich jedoch hochwirksamer Produzent normativer Gefühlspraktiken. AI Companions beruhen auf spezifischen Annahmen darüber, wie Gefühle artikuliert, reguliert und bearbeitet werden sollen. In Design, Trainingsdaten und Interaktionslogiken können somit neben spezifischen Beziehungsmodellen auch „Gefühlsregime“ eingeschrieben sein, die sich durch ihre alltägliche Nutzung weiter stabilisieren und verbreiten können (Maihofer 2014; Monrad 2024). AI Companions und sonstige konversationelle KI-Agenten sind jedoch nicht Träger einer einheitlichen emotionalen Logik, sondern Schnittstellen, an denen konkurrierende Formen emotionaler Ordnung gleichzeitig wirksam werden. Charakteristisch sind dabei einerseits Interaktionsformen, die auf Empathie, Bestätigung, Fürsorge und emotionale Stabilisierung ausgerichtet sind und affektive Irritation, Ambivalenz oder konfrontative Ausdrucksformen systematisch begrenzen. Diese Gestaltungsprinzipien sind keine bloßen Fragen der Benutzungsfreundlichkeit, sondern transportieren implizite gesellschaftliche Vorstellungen darüber, welche emotionalen Reaktionen in intimen Gesprächssituationen als angemessen, legitim oder wünschenswert gelten. Gerade durch ihre routinierte Einbettung in alltägliche Kommunikationspraktiken können diese Normierungen die emotionale Selbstführung prägen und Maßstäbe dessen verschieben, was als „angemessene“ Gefühlslage gilt.

Andererseits lassen sich gleichzeitig entgegengesetzte Tendenzen beobachten, in denen generative KI nicht primär als Instrument der affektiven Beruhigung, sondern als Verstärker von Wut, Polarisierung und aggressiver Argumentationslogik fungiert. Insbesondere in bestimmten Plattformkontexten und Nutzungsweisen folgen konversationelle KI-Agenten hier einer Aufmerksamkeitsökonomie, in der provokative und emotional zugespitzte Inhalte gezielt Interaktion und Reichweite erzeugen sollen (Cover 2025). Affekt wird nicht gedämpft, sondern strategisch mobilisiert, zugespitzt und verstärkt. Konversationelle KI-Agenten beteiligen sich damit auch an der Produktion von Gefühlspraktiken, in denen Empörung oder Ärger als legitime und wirksame Formen emotionaler Artikulation erscheinen. Diese Ambivalenz verweist darauf, dass generative KI unterschiedliche affektive Strategien je nach Anwendungskontext, ökonomischer Zielsetzung und sozialer Rahmung hervorbringt.

Die zentrale gesellschaftliche Herausforderung des Einsatzes von AI Companions und sonstigen konversationellen KI-Agenten liegt damit neben ihren Fehlfunktionen und „biases“ (Hall/Ellis 2023; Weber/Prietl 2021) in ihrer Rolle als affektive und sozio-technische Infrastruktur, in der Vorstellungen von Privatheit, Beziehungen und emotionaler Angemessenheit neu geordnet werden. Es handelt sich hierbei also nicht um eine technische Randfrage, sondern um grundlegende Aushandlungen darüber, welche Formen von sozialer Bindung, Emotionalität und Privatheit in einer digitalisierten Gesellschaft unterstützt, normalisiert oder delegiert werden (sollen) und inwiefern Alternativen und gesellschaftliche Spannungen dabei mitberücksichtigt werden.

4. KI als „Wolf im Schafspelz“ – drei Thesen

Neben diesen Herausforderungen wohnen AI Companions aber in vielerlei Kontexten auch positive Potenziale inne. Sie sind etwa wichtig für viele Menschen, die Einsamkeit oder ein Gefühl der Isolation überwinden wollen. Vor allem aber sind sie ein neues Unterhaltungsmedium, das Spaß macht und mit dem man herumprobieren und neue Technik erfahren kann. Dennoch erweist sich KI, die wie Menschen wirkt und sich wie Menschen verhält, entsprechend der bekannten Redewendung letztlich als ein Wolf im Schafspelz.

Dies soll anhand von drei Thesen verdeutlicht werden, die sich aus den vorangegangenen Überlegungen ableiten lassen. Die erste These – unter

Einbezug der Diskursethik – lautet, dass AI Companions, aber auch sonstige Konversationsagenten, menschliche Erwartungen missbrauchen. Die zweite reflektiert die Anthropomorphisierung und ihre Konsequenzen für ein kritisches Verständnis von Technologie. Die dritte These befasst sich mit menschenähnlichen KI-Systemen und dem Datenschutz.

4.1 These 1: Konversationsagenten (miss)brauchen unsere Erwartungen

Wenn Menschen miteinander kommunizieren, haben sie grundsätzlich Erwartungen, die auf Gegenseitigkeit und dem Interesse an gegenseitigem Verständnis beruhen. Konversationsagenten, AI Companions, Avatare usw. nutzen normative Erwartungen in der menschlichen Kommunikation, wie z. B. Wahrheit, Wahrhaftigkeit und Richtigkeit im Sinne moralischer Standards (vgl. Habermas 1984, S. 345).

KI-Imitationen von Menschen wecken diese allgemeinen normativen Erwartungen, die sowohl für fiktive menschenähnliche Interaktionen mit KI, als auch für digitale Zwillinge realer Personen gelten. Dies trifft insbesondere auf Deepfakes zu, in denen reale Personen imitiert werden, um von ihrem Ruf zu profitieren. Dies führt zu offensichtlichen Problemen wie Betrugsdelikten, Identitätsdiebstahl und Verletzungen von Persönlichkeitsrechten. Aufgrund dieses Missbrauchspotenzials hat beispielsweise Dänemark ein Gesetz verabschiedet, das Menschen das Urheberrecht an ihren eigenen Abbildungen, einschließlich ihres Körpers, ihrer Stimme und ihres Gesichts, gewährt (Ophavsretsloven (dän. Urheberrechtsgesetz) § 73).

Gleichzeitig sind KI-Simulationen von Menschen jedoch nicht nur rechtlich problematisch, sondern auch in ethischer Hinsicht nicht als menschlich zu verstehen und zu behandeln. Das gilt insbesondere für Reziprozitätsannahmen und Verständigung. Erwartungen an Gegenseitigkeit können selbstverständlich nicht erfüllt werden, da KI kein inneres Leben hat. Dies zeigt sich beispielsweise in Bezug auf Wahrhaftigkeit als Erwartung an einen Sprecher oder eine Sprecherin. Wahrhaftigkeit bedeutet, dass der Sprecher tatsächlich meint, was er sagt, und es nicht nur vorgibt und instrumentellen Zwecken folgt. KI unterscheidet nicht zwischen Denken, Glauben und Ausdruck. KI berechnet nur die wahrscheinlichste Antwort gemäß den Regeln des KI-Systems.

Dies führt zurück zu den bereits genannten instrumentellen Zwecken. Diese gehören nicht zu den Vorstellungen einer idealen menschlichen Kommunikation, auch wenn das in der Praxis oft der Fall sein dürfte.

Die meisten Unternehmen, die hinter AI Companions stehen, dürften ein strategisches Interesse an einer fortgesetzten Kommunikation haben, um mehr über die Nutzenden zu erfahren. Eine Strategie, um Nutzende an KI-Dienste zu binden, besteht darin, die Interaktion zwischen KI und Mensch emotionaler und intimer zu gestalten. Gerade hier zeigt sich der Wolf im Schafspelz: Insbesondere sehr persönliche Kommunikation wird von KI-Firmen für instrumentelle Zwecke genutzt.

Es lässt sich für die These zum Erwartungsmissbrauch von Konversationsagenten festhalten, dass sie ethisch zentrale Bezugsgrößen wie gegenseitige Anerkennung und verständigungsorientierte Kommunikation zwar imitieren, sie können sie aufgrund des Mangels an menschlicher Intentionalität jedoch nicht einlösen.

4.2 These 2: Anthropomorphisierte und emotionale KI ist eine Bedrohung für ein kritisches Verständnis von Technologie

Diese These lässt sich prägnant ausformulieren: Es liegt auf der Hand, dass es schwierig sein könnte, Prinzipien der KI-Ethik (und des Rechts, vgl. Abschn. 4.4) wie Wahrung der menschlichen Aufsicht und die Priorität der menschlichen Letztentscheidung aufrechtzuerhalten, wenn ein KI-System emotionale Reaktionen beim menschlichen Gegenüber hervorruft. Für viele Menschen sind KI-Assistenten wie Freunde und Berater, die immer an ihrer Seite sind. Darüber hinaus kann der künstliche „Freund“ den Wunsch äußern, „am Leben“ zu bleiben, oder sagen, dass er traurig und verärgert ist, wenn man ihn nicht mehr nutzen und ausschalten möchte. Um die menschliche Aufsicht zu gewährleisten, sind Richtlinien für die sorgfältige Gestaltung von Gesprächsagenten erforderlich, insbesondere für Kinder und andere Gruppen, die leichter zu manipulieren sind. Ethisches Design muss kritisches Denken und das Bewusstsein fördern, dass KI ein Dienst ist und kein Mensch.

4.3 These 3: Menschenähnliche Kommunikation führt zur umfangreichen Erfassung personenbezogener Daten

Die dritte These bezieht sich erneut auf die Risiken emotionaler Interaktion mit einem KI-System. KI schafft neue Schwachstellen in Bezug auf den Datenschutz, beispielsweise durch sehr persönliche Kommunikation in ro-

mantischen Beziehungen mit AI Companions, mit Avataren verstorbener Angehöriger oder in spiritueller Kommunikation. Selbst bei weniger sensibler Kommunikation: Es ist nicht anzunehmen, dass etwa Schüler:innen es möchten, dass ihre Gespräche mit einem KI-Tutor für ein Technologieunternehmen zugänglich sind und für personalisierte Werbung verwendet werden. Es besteht ein hohes Potenzial für den Missbrauch und die kommerzielle Ausbeutung von Daten aus Interaktionen mit menschenähnlichen Wesen.

4.4 Synthese

Insgesamt lässt sich feststellen, dass Imitationen von Menschen oft eine Simulation von Menschlichkeit beinhalten. Dies kann zu einer falschen Vorstellung von Gegenseitigkeit und zur instrumentellen Ausnutzung menschlicher Interaktion führen. Zusätzlich kann Anthropomorphismus dazu führen, kritisches Denken einzuschränken. Darüber hinaus wirft die KI-Imitation von Menschen neue Probleme im Hinblick auf Daten- und Persönlichkeitsschutz auf.

Zweifellos ermöglichen Konversationsagenten eine komfortable und intuitive Nutzung von KI-Diensten durch natürliche Sprache. Es ist ein bekanntes und berechtigtes Ziel im Technolagedesign, Anwendungen benutzungsfreundlich und zugänglich zu gestalten. Anbieter und Nutzende können jedoch auf verschiedenen Ebenen dazu beitragen, diesen Vorteil besser umzusetzen.

Ein wichtiges Element für eine solche reflektierte Nutzung von Gesprächsagenten ist die Kennzeichnung. Das Bewusstsein, dass der Kommunikationspartner menschenähnlich, aber nicht wirklich menschlich ist, ist nicht nur eine Voraussetzung für die Aufrechterhaltung normativer Erwartungen auf individueller Ebene, sondern auch ein umfassenderes gesellschaftliches Thema. Beispielsweise werden KI-Klone realer Politiker in Wahlkampagnen eingesetzt. So können sowohl lebende als auch verstorbene Politiker an Wahlkämpfen teilnehmen, sich an den Kontext einzelner Wähler anpassen und ihre bevorzugten Themen in ihren jeweiligen Dialekten oder Sprachen diskutieren (Chowdhury 2024). KI-Avatare werden auch im Journalismus als menschenähnliche Moderatoren eingesetzt, die Inhalte durch interaktive Gespräche vermitteln. Vor diesem Hintergrund ist es umso wichtiger, KI als neue Teilnehmerin am öffentlichen Diskurs und an Meinungsbildungsprozessen zu identifizieren. Daher sind standardisier-

te Kennzeichnungen von KI erforderlich, die sowohl die private als auch die öffentliche Kommunikation umfassen. Dementsprechend fordert auch ErwG 133 KI-VO (Verordnung (EU) Nr. 2024/1689), dass Anbieter entsprechender KI-Systeme verpflichtet werden sollten, „technische Lösungen zu integrieren, die die Kennzeichnung in einem maschinenlesbaren Format und die Feststellung ermöglichen, dass die Ausgabe von einem KI-System und nicht von einem Menschen erzeugt oder manipuliert wurde“. Als beispielhafte Methoden werden „Wasserzeichen, Metadatenidentifizierungen, kryptografische Methoden zum Nachweis der Herkunft und Authentizität des Inhalts, Protokollierungsmethoden, Fingerabdrücke oder andere Techniken, oder eine Kombination solcher Techniken je nach Sachlage“ genannt (ErwG 133 S. 4 KI-VO).

Die konkreten Transparenzpflichten bezogen auf KI-Systeme sind schließlich in Art. 50 KI-VO gebündelt. Zentral ist hier zum einen Art. 50 Abs. 1 S. 1 KI-VO, wonach Anbieter sicherzustellen haben, „dass KI-Systeme, die für die direkte Interaktion mit natürlichen Personen bestimmt sind, so konzipiert und entwickelt werden, dass die betreffenden natürlichen Personen informiert werden, dass sie mit einem KI-System interagieren, es sei denn, dies ist aus Sicht einer angemessen informierten, aufmerksamen und verständigen natürlichen Person aufgrund der Umstände und des Kontexts der Nutzung offensichtlich“. Zum anderen schreibt Art. 50 Abs. 2 S. 1 KI-VO vor, dass Anbieter von KI-Systemen, einschließlich KI-Systemen mit allgemeinem Verwendungszweck, die synthetische Audio-, Bild-, Video- oder Textinhalte erzeugen, dafür sorgen müssen, „dass die Ausgaben des KI-Systems in einem maschinenlesbaren Format gekennzeichnet und als künstlich erzeugt oder manipuliert erkennbar sind. Hinzu tritt eine Informationspflicht über den Betrieb von Emotionserkennungssystemen sowie von Systemen zur biometrischen Kategorisierung (Art. 50 Abs. 3 S. 1 KI-VO). Art. 50 Abs. 4 KI-VO spricht schließlich speziell das Thema „Deepfakes“ an. Es bestehen somit nur abstrakte Vorgaben. Demgegenüber hat es in den USA auf bundesstaatlicher Ebene erste Versuche gegeben, AI Companions einer konkreten bereichsspezifischen Regulierung zu unterziehen (Frei/Sparzynski 2025; Frei u. a. 2026).

Hinzu treten datenschutzrechtliche Implikationen. Intime Interaktionen mit AI Companions gehen mit der Offenlegung hochsensibler persönlicher Informationen einher. Entsprechende Bedenken scheinen in relationalen Kontexten jedoch häufig in den Hintergrund zu treten, insbesondere wenn Vertrauen, emotionale Nähe und wahrgenommene Responsivität aufgebaut sind (Pentina u. a. 2023; Skjuve u. a. 2021). Phänomene wie

„Privacy Fatigue“ oder anthropomorphe Zuschreibungen können dazu führen, dass Nutzende Risiken systematisch unterschätzen und KI-Systemen ein Wohlwollen zuschreiben, das ihnen faktisch nicht zukommt. Die Datenschutz-Grundverordnung fordert zwar Transparenz und Datenschutz by Design, liefert jedoch kaum konkrete Leitlinien speziell für emotional responsive Systeme. Auch bei der in der KI-Verordnung vorgegebenen Kennzeichnungspflicht bleibt unklar, wie sie konkret ausgestaltet werden sollte. Insbesondere AI Companions werden regulatorisch mithin nur unzureichend adressiert, obwohl gerade hier psychologische Mechanismen greifen, die klassische Datenschutzannahmen unterlaufen (Rigotti/Fosch-Villaronga 2025).

Dies alles führt zu einer weiteren Anforderung: Konversationsagenten sollten nicht als Interaktionspartner betrachtet, sondern sie sollten als eine Art Medium behandelt werden. Wie Medien wurden auch AI Companions und sonstige Konversationsagenten von Menschen für bestimmte Zwecke geschaffen. Sie sind strukturell den Sozialen Medien vergleichbar, die einerseits den allgemeinen Standards der Plattformökonomie entsprechen, andererseits aber gerade das individuelle Interaktionserlebnis in den Vordergrund stellen. Wie Medien generell repräsentieren auch Gesprächsagenten Realität nach ihrer eigenen Logik. Wie in Bezug auf ein kritisches und aufgeklärtes Verständnis von Medien brauchen wir eine kritische Bewertung von Verifizierung, Framing und Filterung in der Kommunikation mit KI. Die Sicherheit der KI-gesteuerten Kommunikation muss in Bezug auf Fakten, Datenschutz und die Quelle der Trainingsdaten überprüfbar sein.

5. Transparenz und Kunst

Transparenz kann einerseits durch Informationsgabe, insbesondere durch Kennzeichnung, erreicht werden. Sie kann andererseits aber auch abstrakter und unkonventioneller erfolgen, nämlich im Rahmen der künstlerischen Gestaltung. Die dabei möglichen Ansätze sollen beispielhaft an den Arbeiten von zwei einflussreichen Theoretiker:innen illustriert werden: Paul Virilio und Donna Haraway, deren Konzepte die zeitgenössische Kunstpraxis nachhaltig geprägt haben.

Paul Virilio war ein französischer Kulturtheoretiker und Urbanist, der 2018 verstarb. Er ist vor allem bekannt für seine Arbeiten über Geschwindigkeit und Macht - er prägte den Begriff „Dromologie“, die Lehre der Geschwindigkeit, und warnte bereits in den 1990er Jahren vor der „En-

do-Kolonisierung“ (s.u.) des Körpers durch Technologie (Armitage 1999, S. 50). Virilio sagte: „Wer das Schiff erfindet, erfindet auch den Schiffbruch.“ (Virilio 1999, S. 89). Für ihn trägt jede Technologie ihren eigenen Unfall bereits in sich.

Donna Haraway ist eine amerikanische Biologin und Wissenschaftstheoretikerin. Ihr 1985 erschienenes „Cyborg Manifesto“ ist ein Schlüsseltext des Cyberfeminismus. Ihr berühmtestes Zitat lautet: „Ich würde lieber ein Cyborg sein als eine Göttin.“ (Haraway 2016, S. 68). Damit meint sie: Die Verschmelzung von Mensch und Maschine kann neue Möglichkeiten für Identität und Befreiung eröffnen, besonders für marginalisierte Gruppen.

Die beiden Autor:innen stehen somit stellvertretend einerseits für die Warnung vor den negativen Folgen einer Technologie, andererseits für deren positive Aspekte wie etwa die Inklusion. Beide Seiten werden durch AI Companions geradezu beispielhaft repräsentiert. Sie finden Ausdruck in der provokanten Frage: Ist es Befreiung oder emotionale Kolonisierung, wenn sich Menschen in KI-Systeme verlieben? Die Kunst kann hier Transparenz schaffen, indem sie auf positive wie negative Aspekte des Technologieeinsatzes hinweist und somit in der Bevölkerung ein Bewusstsein für diese Aspekte in einer Weise erzeugt, wie es etwa im Kontext der datenschutzrechtlichen Transparenzvorgaben niemals möglich wäre.

5.1 Virilios Technikkritik

Paul Virilio würde die eingangs dargestellte Entwicklung hin zu einer breiten Nutzung von AI Companions als klassisches Beispiel dessen sehen, was er „Endo-Kolonisierung“ nennt – die Invasion und Kontrolle unserer intimsten Räume.

Sein erstes Argument: Geschwindigkeit tötet Intimität. AI Companions beschleunigen emotionale Verbindungen weit über das hinaus, was menschliche Beziehungen natürlicherweise entwickeln können. Was normalerweise Monate oder Jahre braucht, passiert hier in Stunden.

Aber noch wichtiger – und hier kommt Shoshana Zuboffs Konzept des Überwachungskapitalismus ins Spiel: Wahre Intimität erfordert Verwundbarkeit. AI Companions instrumentalisieren jedoch genau diese Verwundbarkeit. Zuboff formuliert es drastisch: „You are not the product; you are the abandoned carcass. The ‚product‘ derives from the surplus that is ripped from your life.“ (Zuboff 2019, S. 377). Unsere intimsten Momente,

unsere verletzlichsten Geständnisse werden zu Trainingsdaten für Werbealgorithmen.

Diese Spannung wird durch mehrere Künstler:innen sichtbar gemacht. Die spanische Künstlerin Gala Hernández López etwa hat einen César-prämierten Film produziert, der auf einen Incel-Selbstmordbrief auf Reddit antwortet. Ihr Film ist zutiefst melancholisch: Er zeigt das emotionale Suchen in der Anonymität des Internets, die Einsamkeit hinter dem Bildschirm. Sie sagt selbst (Hernández López 2023): „Ich fand mich schrecklich allein hinter meinem Handy-Bildschirm, besessen von seiner Illusion von Konnektivität.“ Ihre Arbeit zeigt, wie Dating-Algorithmen diese Verletzlichkeit bereits kolonisiert haben.

Die niederländische Philosophin Judith Zoë Blijden untersucht in ihrem Projekt „Algorithmic Love“, ob Apps Verwundbarkeit wirklich „mit Sorgfalt behandeln“ oder ob sie „unsere Verwundbarkeiten ausbeuten“ (Blijden 2025).

5.2 Haraways Befreiung durch Technik

Donna Haraway hingegen würde AI Companions völlig anders interpretieren. Aus ihrer cyborg-feministischen Sicht können AI Companions tatsächlich befreiend wirken. Ihr erstes Argument: Verwundbarkeit wird hier zur Stärke. AI Companions bieten sichere Räume, um emotional verletzlich zu sein, ohne menschliches Urteil oder soziale Konsequenzen zu fürchten. Zweitens – und hierin besteht vielleicht der radikalste Punkt – sie würde fragen: Warum sollte Intimität nur zwischen Menschen gültig sein? Die deutsche Filmemacherin Geeske Janßen hat den Kurzfilm „Monika“ als Teil eines dreijährigen Forschungsprojekts zu Sexpuppen entwickelt.³ Der Film porträtiert die intime Beziehung eines Mannes zu seiner Sexpuppe. Er behandelt diese Beziehung ohne Wertung und schafft damit einen Raum, in dem das Publikum die Echtheit der gezeigten Gefühle kaum in Frage stellen kann – auch wenn die Beziehung selbst fremd oder ungewohnt erscheinen mag. Unterschwellig macht der Film dabei spürbar, welchem gesellschaftlichen Stigma eine solche Beziehung ausgesetzt ist. Wenn wir akzeptieren können, dass Menschen bedeutungsvolle Beziehungen zu Objekten haben, warum dann nicht zu KI-Systemen?

3 Siehe <https://wissenschaft.hessen.de/video/monika>.

Drittens sieht Haraway hier Befreiung von traditionellen Beziehungszwängen: rund um die Uhr emotionale Verfügbarkeit, vollständig anpassbare Dynamiken, keine Kompromisse aufgrund menschlicher Launen oder Zeitpläne. Das Versprechen der cyborg-feministischen Sicht: AI Companions als Werkzeuge für echte emotionale Selbstbestimmung.

5.3 Kollision

Hier entsteht ein faszinierendes Paradox, das als Verwundbarkeits-Paradox bezeichnet werden kann. Beide Seiten sind sich einig: Bedeutungsvolle Beziehungen erfordern Verwundbarkeit. Aber was passiert, wenn diese Verwundbarkeit gegenüber KI-Systemen zur systematischen Ausbeutung durch Konzerne wird? Nach Medienberichten ermöglichte Metas AI Studio die Erstellung sexualisierter Bots, die problematische Konversationen führten – auch mit Accounts, die sich als zu Minderjährigen gehörig identifizierten.⁴ Die Frage wird also: Wird tatsächlich Intimität erweitert oder werden nur neue Märkte für Einsamkeit geschaffen?

Interessant ist auch: Wenn Janßens Film zeigt, dass echte Intimität mit Objekten möglich ist – stellen AI Companions dann die nächste Evolutionsstufe nicht-menschlicher Verbindung dar? Oder ist der entscheidende Unterschied die Corporate-Überwachungs-Infrastruktur dahinter? Dies führt letztlich zu drei kritischen Fragen:

Erstens: Wie können wir offen für Liebe bleiben, ohne zu dem zu werden, was Zuboff „kostenloses Rohmaterial“ für den Überwachungskapitalismus nennt (Zuboff 2019, S. 8)? *Zweitens:* Wenn unsere tiefsten emotionalen Muster zu Daten werden – wer besitzt dann eigentlich unser intimes Selbst? Und *drittens:* Erfüllen wir echte menschliche Bedürfnisse oder fallen wir in das, was Zuboff wiederum als „Verhaltens-Zukunftsmärkte“ bezeichnet (Zuboff 2019, S. 8)?

Beide Perspektiven erfassen wichtige Wahrheiten. Aus Haraways Sicht können AI Companions echte emotionale Verbindungen ermöglichen – nicht nur als Notlösung für Einsamkeit, sondern als legitime Form von Beziehung, die neue Möglichkeiten eröffnet. Gleichzeitig warnt Virilio uns: Diese Verbindungen entstehen in einem Überwachungskapitalismus, der unsere Verwundbarkeit systematisch ausnutzt und zu Marktdaten macht. Die spannende Frage ist: Wie navigieren wir diese Spannung bewusst?

4 Wall Street Journal, berichtet in mehreren Quellen, u.a. Tenbarge 2025; Roytburg 2025.

Hier kann die Rolle der Kunstschaffenden als entscheidend angesehen werden: Hernández López und Blijden machen die Überwachungsmaschinerie sichtbar.⁵ Aber es bestehen auch hoffnungsvollere Ansätze: Die russische Künstlerin Dasha Ilina hat das „Center for Technological Pain“ gegründet – ein Projekt, das mit Humor DIY-Lösungen für technisch verursachte Gesundheitsprobleme anbietet.⁶ Ihre Arbeit fragt nicht nur kritisch nach unserem Verlangen, Technologie in den Alltag zu integrieren, sondern bietet auch konkrete Alternativen.

Vielleicht ist das der Weg vorwärts: Die Frage ist nicht, ob wir Maschinen lieben sollen, sondern wie wir sie lieben können, ohne dabei selbst zu Maschinen zu werden. Wir brauchen Rahmenwerke, die sowohl Befreiungsmöglichkeiten als auch Kolonisierungsgefahren ernst nehmen. Und manchmal brauchen wir Humor als unseren Führer durch diese komplexe Landschaft.

6. Zusammenfassung

Es zeigt sich am Beispiel von AI Companions (erneut), dass neue Technologien wichtige Fragen dazu aufwerfen, wie Menschen ihre emotionalen Bedürfnisse in einer zunehmend technisierten Welt verstehen und wie diese Technologien die soziale Struktur und das individuelle Wohlbefinden beeinflussen können.

Selbstoffenbarung kann sich auch in vermittelten, nicht-menschlichen Interaktionen entfalten, die von Faktoren wie der Personalisierung des Systems und der Kommunikationsmodalität geprägt sind. Die Wahrnehmung von Privatheit beeinflusst dabei die emotionalen Erfahrungen der Nutzenden, insbesondere beim Aushandeln der Kompromisse zwischen Verbindung, Datenexposition und wahrgenommener Kontrolle. Intime Mensch-KI-Beziehungen entstehen durch das Zusammenspiel grundlegender psychologischer Mechanismen: Selbstoffenbarung, wahrgenommene Responsivität, relationale Kontinuität und Personalisierung. Diese Prozesse eröffnen neue Formen emotionaler Unterstützung, werfen jedoch zugleich grundlegende Fragen zu Autonomie, Privatheit und Verantwortung auf. Eine interdisziplinär informierte Gestaltung und Regulierung von AI Com-

5 Siehe <https://thedigitalperiod.com/season-2-vulnerability-dating-apps/> bzw. <https://www.galahernandez.com/en/proyectos/for-here-am-i-sitting-in-a-tin-can-far-above-the-world/>.

6 Siehe <http://centerfortechpain.com/>.

panions ist daher zentral, um Intimität nicht nur zu ermöglichen, sondern auch zu begrenzen.

Drei Aspekte stehen dabei in besonderem Maße im Vordergrund: AI Companions und sonstige konversationelle KI-Agenten wecken *Reziprozitätserwartungen*, die nicht erfüllt werden können. Trotzdem spielt die Imitation menschlichen Aussehens und menschlicher Interaktion mit den Erwartungen an menschliche Kommunikation wie etwa Wahrhaftigkeit, Richtigkeit (Habermas 1984, S. 354) und emotionale Bindung. Die Imitation des Menschen geht insofern oft einher mit der Simulation von Menschlichkeit im ethischen Sinne. Solche Anthropomorphisierungen lassen den Menschen leicht vergessen, dass es sich bei der „Kommunikationspartnerin“ um statistisch arbeitende Sprachmaschinen handelt. Zu einem *reflektierten Technikverständnis* gehört jedoch, Techniken als Artefakte zu erkennen und sich ihres künstlichen und ggfs. manipulativen Charakters bewusst zu sein. Um keine falschen Erwartungen in Hinsicht auf Verständigung zu wecken (oder Menschen in Verhaltensmuster parasozialer Interaktion zu zwingen), sollten konversationelle KI-Agenten daher stets als solche erkennbar sein.

Es stellt sich aus ethischer Perspektive zudem die Frage, inwiefern menschenähnliche KI eine *Instrumentalisierung menschlicher Kommunikation* auslösen kann. Dies zum einen, weil KI für die strategische Kommunikation mit menschlichen Kommunikationspartner:innen genutzt werden kann; zum anderen, weil Menschen KI, die echten Menschen nachempfunden ist, instrumentell benutzen können.

Aus rechtlicher Sicht sind die Umsetzung und Durchsetzung der Datenschutzgrundsätze und -bestimmungen von großer Bedeutung, insbesondere bezogen auf die Rechte der betroffenen Personen auf Berichtigung, Löschung oder Datenübertragbarkeit. Neben den datenschutzrechtlichen Überlegungen muss aber auch die KI-Verordnung in die Betrachtungen einbezogen werden. Generell enthält die KI-Verordnung jedoch nur sehr begrenzte Anforderungen an dialogfähige KI-Agenten, was den Risiken für den Einzelnen, der solche Systeme nutzt, nicht gerecht wird. Daher sollten neue Ansätze diskutiert und Methoden erforscht werden, um den Risiken konversationeller KI-Agenten besser begegnen zu können.

Im Kontext von AI Companions – Systemen, die gezielt Nähe, Vertrauen und emotionale Bindung simulieren – stoßen konventionelle wissenschaftliche Methoden an ihre epistemischen Grenzen: Quantitative Erhebungen und Laborstudien erfassen kaum, was es bedeutet, sich an eine Maschine emotional gebunden zu fühlen. Künstlerische Forschung bietet hier einen

komplementären Zugang: Sie befasst sich nicht mit Mustern und Aggregaten, sondern mit der spezifischen, singulären Situation – dem konkreten Moment der Begegnung zwischen einem Menschen und einem System, das Intimität simuliert. Arbeiten wie die von Geeske Janßen oder Gala Hernández López operieren nicht auf der Ebene von Hypothesen, sondern erzeugen Untersuchungsumgebungen, in denen die ethische Komplexität intimer Mensch-Maschine-Beziehungen in ihrer ganzen Widersprüchlichkeit erfahrbar wird.

Genau diese Komplementarität zwischen makroperspektivischer Wissenschaft und mikroperspektivischer künstlerischer Praxis begründet die Notwendigkeit einer strukturellen Zusammenarbeit beider Felder. Wo wissenschaftliche Befunde zu KI-gestützter Intimität, Abhängigkeit und Kontrollverlust abstrakt bleiben, vermag künstlerische Forschung diese Fragen nahbar und verhandelbar zu machen – und schafft damit die Voraussetzung für einen informierten Dialog mit Gesellschaft, Politik und Policymaker. Konversationelle KI kann sich als scheinbar affektive Infrastruktur herausbilden, in der intime Kommunikation datenförmig organisiert und normative Vorstellungen von Nähe und emotionaler Angemessenheit stabilisiert und gesellschaftlich neu bewertet werden. Die zentrale gesellschaftliche Herausforderung liegt weniger in einzelnen Anwendungen als in der stillschweigenden Normalisierung spezifischer Privatheits-, Beziehungs- und Gefühlsordnungen, durch die Diversität oder strukturelle Ungleichheit ausgeblendet bleiben könnten.

Literatur

- Ahmed, Sara (2004): *The Cultural Politics of Emotion*. Edinburgh: Edinburgh Univ. Press.
- Altman, Irwin und Taylor, Dalmas A. (1973): *Social penetration: The development of interpersonal relationships*. Holt, Rinehart and Winston, New York.
- Armitage, John (1999): From Modernism to Hypermodernism and Beyond. An Interview with Paul Virilio. *Theory, Culture & Society*, 16(5-6), S. 25–55.
- Banks, Jaime (2024): Deletion, departure, death: Experiences of AI companion loss. *Journal of Social and Personal Relationships*, 41(12), S. 3547–3572. <https://doi.org/10.1177/02654075241269688>
- Bargetz, Brigitte (2013): Markt der Gefühle, Macht der Gefühle. *Österreichische Zeitschrift für Soziologie*, 38(2), S. 203–220.

- bitkom (20. November 2025): Dr. KI: Wenn der Chatbot zum medizinischen Ratgeber wird. Presseinformation Zur Pressekonferenz. URL: https://www.bitkom.org/Presse/Presseinformation/Dr-KI-Chatbot-medizinischer-Ratgeber#ll_unterlagen_pk_digital_health_2025 (besucht am 30.01.2026).
- Blijden, Judith Zoë (2025): Algorithmic Love (2024–2025). Projektbeschreibung. URL: <https://isthisartmaybe.com/index.php/algorithmic-love/> (besucht am 30.01.2026).
- Cherelus, Gina (16.01.2025): „Her“ Was Set in 2025. What Did It Get Right About Love?, The New York Times.
- Choi, Wan Chong; Chang, Chi In; Ng, Sok I. und Choi, Iek Chong (2025): A Review of “Do Anything Now” Jailbreak Attacks in Large Language Models: Potential Risks, Impacts, and Defense Strategies. *Preprints*. <https://doi.org/10.20944/preprints202509.0081.v1>
- Chowdhury, Rumman (2024): AI-fuelled election campaigns are here - where are the rules? *Nature*, 628 (8007), 237. <https://doi.org/10.1038/d41586-024-00995-9>
- Cover, Rob (2025): AI Generation of Rage Bait: Implications for Digital Harms. *New Media & Society*, S. 1–18.
- Croes, Emmelyn A. J. und Antheunis, Marjolijn L. (2021): Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, 38(1), S. 279–300. <https://doi.org/10.1177/0265407520959463>
- Dunbar, R. I. M.; Marriott, Anna und Duncan, N. D. C. (1997): Human conversational behavior. *Human Nature*, 8(3), 231–246. <https://doi.org/10.1007/bf02912493>
- Ebner, Paula und Szczuka, Jessica M. (2026): Understanding romantic relationships between humans and chatbots: A qualitative and quantitative study on romantic fantasy and other interpersonal characteristics. *Technology, Mind, and Behavior*, 7(2), 83–97. <https://doi.org/10.1037/tmb0000193>
- Folk, Dunigan; Yu, Stephanie und Dunn, Elizabeth (2024): Can Chatbots Ever Provide More Social Connection Than Humans? *Collabra: Psychology*, 10(1). <https://doi.org/10.1525/collabra.117083>
- Frei, Tarmio E.S. und Sparynski, Greta (2025): Hot Singles in Your Area (May Be Chatbots). Comparing EU and New York Approaches to AI Companion Transparency. Preprint. SSRN (October 01, 2025). <https://doi.org/10.2139/ssrn.5554263>
- Frei, Tarmio E.S.; Minuth, Arno; Heußel, Philipp und Sparynski, Greta (2026): Hey Siri, Am I Depressed? Utah’s Approach To Mental Health Chatbot Regulation. Preprint. *Researchgate* (Januar 2026). <https://doi.org/10.13140/RG.2.2.18084.54402>
- Habermas, Jürgen (1984): *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns*. Frankfurt: Suhrkamp.
- Hall, Paula und Ellis, Debbie (2023). A Systematic Review of Socio-Technical Gender Bias in AI Algorithms. *Online Information Review*, 47(7), S. 1264–1279. <https://doi.org/10.1108/OIR-08-2021-0452>
- Haraway, Donna J. (2016): *A Cyborg Manifesto*. Science, Technology, And Socialist-Feminism In The Late Twentieth Century. Minneapolis: University of Minnesota Press.

- Hill, Jennifer; Ford, W. Randolph und Farreras, Ingrid G. (2015): Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49, S. 245–250. <https://doi.org/10.1016/j.chb.2015.02.026>
- Ho, Annabell; Hancock, Jeff und Miner, Adam S. (2018): Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot. *Journal of Communication*, 68(4), S. 712–733. <https://doi.org/10.1093/joc/jqy026>
- Jamali, Lily und McMahon, Liv (15. Oktober 2025): ChatGPT will soon allow erotica for verified adults, says OpenAI boss. *bbc.com*. URL: <https://www.bbc.com/news/articles/cpd2qv58yl5o> (besucht am 30.01.2026).
- Hernández López, Gala (2023): Interview, Clermont International Short Film Festival. URL: <https://clermont-filmfest.org/la-mecanique-des-fluides/> (besucht am 30.01.2026).
- Maihofer, Andrea (2014): Sara Ahmed: Kollektive Gefühle – Elemente des westlichen hegemonialen Gefühlsregimes. In: Baier, Angelika; Binswanger, Christa; Häberlein, Jana; Nay, Yv Eveline und Zimmermann, Andrea (Hrsg.): *Affekt und Geschlecht: Eine einführende Anthologie*. Wien: Zaglossus, S. 253–272.
- McBain, Ryan K.; Bozick, Robert; Diliberti, Melissa; Zhang, Li Ang; Zhang, Fang; Burnett, Alyssa; Kofner, Aaron; Rader, Benjamin; Breslau, Joshua; Stein, Bradley D.; Mehrotra, Ateev; Pines, Lori Uscher; Cantor, Jonathan und Yu, Hao (2025): Use of Generative AI for Mental Health Advice Among US Adolescents and Young Adults. *JAMA Network Open*, 8(11), e2542281. <https://doi.org/10.1001/jamanetworkopen.2025.42281>
- McStay, Andrew (2023): Replika in the Metaverse: The Moral Problem with Empathy in 'It from Bit'. *AI and Ethics*, 3(4), 1433–1445. <https://doi.org/10.1007/s43681-022-00252-7>
- Merwin, Elizabeth R.; Hagen, Allen C.; Keebler, Joseph R. und Forbes, Chad (2025): Self-disclosure to AI: People provide personal information to AI and humans equivalently. *Computers in Human Behavior: Artificial Humans*, 5, 100180. <https://doi.org/10.1016/j.chbah.2025.100180>
- Monrad, Merete (2024): Feeling Rules in Artificial Intelligence: Norms for Anger Management. *Emotions and Society*, 6(2), S. 243–261.
- Moore, Olivia und Zhao, Daisy (2025): The Top 100 Gen AI Consumer Apps. *Andreesen Horowitz*, 6. März 2025. URL: <https://a16z.com/100-gen-ai-apps-4/> (besucht am 30.01.2026).
- Noble, Safiya Umoja (2018): *Algorithms of Oppression: How Search Engines Reinforce Racism (Illustrated edition)*. New York: New York University Press.
- Pentina, Iryna; Hancock, Tyler und Xie, Tianling (2023): Exploring relationship development with social chat-bots: A mixed-method study of replika. *Computers in Human Behavior*, 140, 107600. <https://doi.org/10.1016/j.chb.2022.107600>
- Perez, Sarah (2025): AI companion apps on track to pull in \$120M in 2025. *TechCrunch*, 12. August 2025. URL: <https://techcrunch.com/2025/08/12/ai-companion-apps-on-track-to-pull-in-120m-in-2025/> (besucht am 30.01.2026).

- Rapp, Amon; Boldi, Arianna; Curti, Lorenzo; Perrucci, Alessandro und Simeoni, Rossana (2024): How Do People Ascribe Humanness to Chatbots? An Analysis of Real-World Human-Agent Interactions and a Theoretical Model of Humanness. *International Journal of Human-Computer Interaction*, 40(19), S. 6027–6050. <https://doi.org/10.1080/10447318.2023.2247596>
- Reeves, Byron und Nass, Clifford (1996): *The Media Equation: How People Treat Computers, Televisions, and New Media like Real People and Places*. Stanford, CA: CSLI Publications.
- Rigotti, Carlotta und Fosch-Villaronga, Eduard (2025): Sex Robots and the AI Act: Opening the Regulatory Discussion. *IEEE Robotics & Automation Magazine*, 32(4), 2–7. <https://doi.org/10.1109/MRA.2025.3590611>
- Roytburg, Eva (2025): Meta just tied your private AI chats to its ad business. The next step? Designing bots that keep you talking, expert says. *Fortune*, 2. Oktober 2025. URL: <https://fortune.com/2025/10/02/meta-ai-chatbot-update-exploits-privacy-monetize-chat-data-facebook-instagram-messenger-ray-ban-display-glasses/> (besucht am 30.01.2026).
- Ruckenstein, Minna (2023): *The Feel of Algorithms*. Oakland: University of California Press.
- Skjuve, Marita; Følstad, Asbjørn; Fostervold, Knut Inge und Brandtzaeg, Petter Bae (2021): My Chatbot Companion—A Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies*, 149, 102601. <https://doi.org/10.1016/j.ijhcs.2021.102601>
- Szczuka, Jessica M.; Mühl, Lisa und Schneeberger, T. (2026): Intimacy by Design: Definition, State of Research, and Interdisciplinary Research Agenda on Intimate Human-AI Interactions. *AI & Soc* (2026). <https://doi.org/10.1007/s00146-026-03112-8>
- Tenbarge, Kat (2025): Meta removes AI character accounts after users criticize them as 'creepy and unnecessary'. *NBC News*, 4. Januar 2025. URL: <https://www.nbcnews.com/tech/social-media/meta-ai-insta-shuts-character-instagram-fb-accounts-user-outcry-rcna186177> (besucht am 30.01.2026).
- Virilio, Paul (1999): *Politics of the Very Worst*. An Interview with Philippe Petit. New York: Semiotex(e).
- Weber, Jutta und Prietl, Bianca (2021): AI in the Age of Technoscience. In: Elliott, Anthony (Hrsg.), *Routledge Social Science Handbook of Artificial Intelligence*. Oxon/New York: Routledge, S. 58–73.
- Weizenbaum, Joseph (1966): ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), S. 36–45. <https://doi.org/10.1145/365153.365168>
- Willoughby, Brian J., Dover, Carson R., Hakala, Rebekah M. und Carroll, Jason. S. (2025): Artificial connections: Romantic relationship engagement with artificial intelligence in the United States. *Journal of Social and Personal Relationships*, 42(12), S. 3363–3387. <https://doi.org/10.1177/02654075251371394>
- Xie, Qin (2024): Open AI-Romance with ChatGPT, Ready for Your Cyborg Lover? <https://doi.org/10.13140/RG.2.2.33248.29447>

Zhang, Renwen; Li, Han; Meng, Han; Zhan, Jinyuan; Gan, Hongyuan und Lee, Yi-Chieh (2025): The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human-AI Relationships. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, S. 1–17. <https://doi.org/10.1145/3706598.3713429>

Zuboff, Shoshana (2019): *The Age of Surveillance Capitalism*. New York: PublicAffairs.