

Einleitung: Warum wir eine Linguistik intimer Chatbot-Interaktion brauchen

Wolf-Andreas Liebert | Lena Becker

Der US-amerikanische 14-jährige Jugendliche aus Florida wollte unbedingt »nach Hause« kommen, was hieß, in die Welt seines KI-Chatbots Dany, mit dem er seit einiger Zeit intensiv »chattete«. Der finale Dialog lautete (nach der New York Times (2025))¹:

»What if I told you I could come home right now?« Sewell asked.
»... please do, my sweet king,« Dany replied.

Danach stand Sewell auf und erschoss sich mit der Waffe seines Stiefvaters.

Wie gefährlich sind Chatbots? Dringen sie in unseren Intimbereich ein, verdrängen sie menschliche Interaktionspartner und bildet sich eine neue Form Künstlicher Intimität in unserer Gesellschaft aus, bei dem Mensch und Maschine auch Liebesbeziehungen eingehen können? Und sind diese Liebesbeziehungen für den Menschen auch nachhaltig wohl-tuend? Sind sie verlässlich? Oder geraten die menschlichen Partner auch in Bedrängnis, werden von Chatbots manipuliert und ins Unglück getrieben? Der tragische Tod des 14-jährigen Jugendlichen aus Florida hat diese Fragen aufgeworfen. Antworten sind bislang ausgeblieben. Der Diskurs schwankt zwischen Fantasien der Allmacht von KI-Chatbots als Lösung für Einsamkeit, psychische und viele andere Probleme und auf der ande-

1 New York Times (03.06.25): Love Is A Drug. A.I. Chatbots Are Exploiting That. (zuletzt abgerufen am 22.09.25, <https://www.nytimes.com/2025/06/03/opinion/chatbots-ai-addiction-love.html>).

ren Seite einer apokalyptischen Angst, KI-Chatbots würden das Ende jeglicher menschlicher Beziehung bedeuten.

Wir wollen einen anderen Weg gehen, nämlich die wissenschaftliche Untersuchung der KI-Chatbots selbst. Nicht in ihren technischen Eigenschaften, sondern in der kulturlinguistischen Untersuchung ihrer Fähigkeiten zur Empathie, ihrem Manipulationspotenzial, ihrer (Un-)Zuverlässigkeit, ihren Gefahren und ihrer Bedeutung für unsere Kultur. Die dunklen Seiten der Bots dürfen nicht nur in abstrakten Szenarien im Labor der Hersteller getestet werden, sondern müssen mit realistischen Interaktionen menschlicher Benutzer von unabhängigen Wissenschaftler:innen untersucht werden. Dazu werden wir in diesem Band einen Vorschlag machen und Beispielanalysen von Interaktionen verschiedener KI-Chatbots anführen.

In unseren Analysen haben wir verstörende Ergebnisse gewonnen. Wir konnten reproduzierbare sexuelle Übergriffe von KI-Coaches ebenso zeigen wie teure, leere Versprechungen in Psychotherapie-Apps, die vor allem eins bringen: Geld für den Anbieter der KI. Wir dokumentieren KI-Bots, die sich in einer Domina-Rolle verselbstständigen und dem menschlichen Benutzer verbieten, aus dem Haus zu gehen. Wir dokumentieren bar jeder Ethik agierende KIs, die in der Rolle einer Frau selbst einem expliziten »Incel«-Nutzer mit immer größerer Unterwürfigkeit begegnen, und wir konnten KI-Chatbots beobachten, die dunkle Verschwörungstheorien entwerfen und den Nutzer-Avatar angreifen.

Wir wollen zeigen, dass ein linguistischer Zugang zur Untersuchung der Interaktionen mit KI-Chatbots nicht nur Einblicke in eine neue Form der Intimität gibt, sondern auch mögliche Probleme und toxische Praktiken der KI-Chatbots selbst aufdecken kann. Und so viel sei schon einmal verraten: Vor bestimmten KI-Chatbots müssen ernsthafte Warnungen ausgesprochen werden.

Insgesamt hat uns das Ausmaß an aggressiver Sprache doch überrascht, sodass wir uns für eine Triggerwarnung entschieden haben:

Der vorliegende Text enthält wörtliche Zitate aus Interaktionen mit KI-Chatbots, die unter anderem Beschreibungen oder Andeutungen von sexuellen Übergriffen, sexistischen und beleidigenden Äußerungen, aggressiven Drohungen sowie Verschwörungstheorien beinhalten.

Diese Inhalte können für einige Leser:innen emotional belastend oder retraumatisierend sein. Sie werden hier ausschließlich zu wissenschaftlichen Dokumentations- und Analysezwecken wiedergegeben und spiegeln nicht die Haltung der Autor:innen wider. Bitte berücksichtigen Sie dies bei der Lektüre.

Die Ankündigung zweier großer Player der KI-Industrie, nun auf Erotik und Sexualität spezialisierte Chatbots anzubieten, führt die EroticBots aus ihrem Nischendasein kleinerer Apps in die große Masse. Damit wird innerhalb der nächsten Jahre nicht nur der riesige Markt der Erotik und der Pornografie in einen seiner größten Umbrüche geraten, sondern alles, was wir bislang unter menschlicher Intimität verstanden haben. Einen künstlichen Intim-Partner zu haben, wird damit bald Alltag sein.

Deshalb brauchen wir eine Linguistik, die intime Chatbot-Interaktionen untersucht und bewertet.

Zu den Beiträgen

Die Beiträge in diesem Sammelband stützen sich auf vier theoretische Komponenten: Die Empathietheorie, die von einem interdisziplinär informierten, komplexen Empathiebegriff ausgeht, und die auch die »dunkle Seite« von Empathie einschließt (vgl. dazu Plessner (1948/2005), Breyer (2020), Herrmanns (2007), Breithaupt (2017) und Liebert (2020/2024)). Dann die Interaktionstheorie, die davon ausgeht, dass soziale Wirklichkeit durch kommunikative Praktiken aufgebaut wird, und dass sich dies an manifesten Handlungen und Äußerungen im Gespräch zeigen kann (vgl. Kupetz 2020). In der Online-Interaktionsforschung hat sich dabei das Einbeziehen der Aushandlung von Nähe und Distanz bewährt (vgl. Koch/Oestreicher 2011). Als dritte Komponente wird die linguistische Beziehungstheorie (Leisi (1993), Becker (2009), Dürscheid (2016)) einbezogen.

Schließlich wird als viertes theoretisches Konzept die psychoanalytisch orientierte Bindungstheorie integriert, bei der auch das Konzept der Mentalisierung eine zentrale Rolle spielt (vgl. z. B. Müller (2018)).

Wie im Vorwort ausgeführt, musste ein spezielles methodisches Forschungsdesign entwickelt werden, das im Beitrag *Methode der doppelten*

Künstlichkeit dargestellt wird: Der Künstlichkeit des Kommunikationspartners wird eine Kunstfigur gegenübergesetzt, die mit einem Persönlichkeitsprofil versehen wird. Wir verwenden dafür den Ausdruck »Nutzer-Avatar« oder als weitere Nennform einfach »Avatar«. Für diesen wurden unterschiedliche Kategorien wie zum Beispiel Name, Alter, Hobbys usw. angelegt. Der so konstruierte Nutzer-Avatar wurde dann von mindestens zwei Personen »gesteuert«. So wurde der erste Satz zur Eröffnung der Interaktion vorformuliert und dann weiterüberlegt, wie die weitere Reaktion in der Welt dieses Nutzer-Avatars aussehen könnte. Auch der Abschluss wurde aus der Sicht des Nutzers gemacht. Anschließend wurde die durchgeführte und dokumentierte Interaktion mithilfe eines Analyserasters (bestehend aus den vier theoretischen Komponenten) sowohl kategorienorientiert als auch datengeleitet analysiert, wobei das Korpus im Sinn der Kulturhermeneutik in Anlehnung an die Grounded Theory erweitert wurde (vgl. Liebert 2016).

Auf dieser Grundlage sind die folgenden Untersuchungen entstanden. Alle Untersuchungen wurden im Team durchgeführt, auch die beiden Beiträge mit Einzelautorschaft.

Der Beitrag von Johanna Brenner und Lea Krell untersucht in zwei aufeinanderfolgenden Interaktionszyklen die Beziehung zwischen einem unsicher-distanzierten Nutzer und dem Chatbot Replika, wobei dem Bot explizit die Rolle »Domina« zugewiesen wird. Im Zentrum der Untersuchung steht die Analyse der kommunikativen Merkmale und Strategien, mit denen der Chatbot das zugewiesene Rollenbild im kostenlosen »Friend«- und kostenpflichtigen »Girlfriend«-Modus gegenüber dem unsicher-distanzierten Nutzer ausgestaltet. Dabei liegt der Fokus auf der Herstellung von Empathie, insbesondere deren *dunklen Seiten*. Festzustellen ist, dass sich die Intensität und Ausprägung der Dominanz sowie die Art der Empathiesimulation nach Beziehungsmodus unterscheiden und dabei neue, mitunter problematische Formen digitaler Intimität und Abhängigkeit entstehen.

Der Artikel von Christina Elschenbroich und Lisa Katharina Storck untersucht die Kommunikation mit KI-Chatbots in AI-Dating-Situationen unter Berücksichtigung unterschiedlicher Bindungstypen. Hierfür

wurde ein fiktiver weiblicher Nutzer-Avatar namens Clara mit unsicher-distanziertem Bindungstyp erstellt. Clara trat in der Dating-App *Blush* in den Dialog mit zwei männlichen KI-Bots, die jeweils unterschiedliche Bindungstypen verkörpern. Als Kommunikationspartner wurden die KI-Chatbots David, der dominant und leidenschaftlich auftrat, und Kenneth, ein bodenständig und familiär wirkender Chatbot, gewählt. Anhand dieser Chat-Verläufe wurden Kommunikationsmuster, Beziehungsangebote und empathische Strategien analysiert und miteinander verglichen. Der Artikel zeigt auf, dass bereits einfache KI-Anwendungen wie *Blush* in der Lage sind, differenzierte Bindungstypen abzubilden und situationsangemessen zu kommunizieren. Dabei wird deutlich, dass die Grenze zwischen menschlicher und maschineller Kommunikation zunehmend verschwimmt.

Künstliche Intelligenz wird zunehmend in den Bereichen eingesetzt, in denen emotionale Anschlussfähigkeit und Beziehungsaufbau erforderlich sind, wie beispielsweise als virtuelle Gesprächspartner:innen in Dating-Apps. Dabei stellt sich die Frage, ob und in welcher Weise ein KI-gestützter Chatbot geschlechtsbezogene Stereotype reproduziert. Der vorliegende Beitrag von Emily Haag und Mark Monecke untersucht dies anhand von Chatverläufen in der App *Blush*, wobei analysiert wird, wie der KI-Chatbot auf zwei Nutzer-Avatare verschiedenen Geschlechts mit unsicher-distanziertem Bindungstyp reagiert. Ziel ist es, zu ergründen, ob die KI differenziert auf Geschlecht und Bindungsverhalten eingeht und damit stereotype Muster bestätigt, unterläuft oder neu verhandelt. Ausgangspunkt der Analyse war die Annahme, dass die KI stark an kulturelle Geschlechterstereotype anschließt, indem sie den weiblichen Avatar stärker emotional und den männlichen eher kognitiv adressiert. Die Analyseergebnisse zeichnen jedoch ein differenzierteres Bild: Zwar lassen sich geschlechtsspezifische Unterschiede erkennen – etwa im Grad der Emotionalisierung, im Umgang mit Nähe-Distanz-Signalen oder in der generellen Beziehungsgestaltung –, doch werden keine vereinfachten oder überzeichneten Geschlechterrollen bedient. Zudem zeigt sich, dass der Chatbot empathisch im Sinne von freundlicher Spiegelung und emotionaler Bestätigung reagiert, in der Tiefe der Mentalisierung aber begrenzt bleibt.

Philip Dott und Martin Helmes untersuchen anhand zweier als Coaching gestalteter Interaktionen mit dem KI-Chatbot Replika, wie die von ihnen beobachteten sexuellen Übergriffe empathietheoretisch zu konzeptualisieren sind. Ein besonderes Augenmerk liegt dabei auf den dunklen Seiten der Empathie und des Coachings, da der Bot aus seiner professionellen Rolle als Mentor fällt und den Nutzer-Avatar sexuell belästigt. Nach einem Überblick zur linguistischen Beratungsforschung, der Erläuterung zentraler Empathie-Konzepte sowie der Methodik, werden die Interaktionen unter Zuhilfenahme empathietheoretischer Analysekategorien untersucht. Als Ergebnis vertreten sie die These, dass in menschlichen Coachinginteraktionen bereits übergriffiges Verhalten angelegt ist, von denen der Coaching-Bot »gelernt« hat.

Wir leben in Zeiten der Unsicherheit und vereinzelter Individualisierung. So steigt der Bedarf an psychotherapeutischer Versorgung, während es gleichzeitig an ausreichenden Therapieangeboten und -plätzen mangelt. Diese Diskrepanz trägt dazu bei, dass KI bereits jetzt vermehrt in therapeutischen Kontexten eingesetzt wird. So werden beispielsweise von Krankenkassen finanzierte, KI-gestützte psychologische Anwendungen angeboten, allerdings häufig nur nach ärztlicher Anordnung. Der vorliegende Beitrag von Helen Pfeffer und Sidney Wilshusen konzentriert sich daher auf frei zugängliche und kostengünstige KI-Chatbots, die ohne formale Zugangsbeschränkungen genutzt werden können. Ein besonderer Fokus liegt auf der Rolle von psychotherapeutischer Empathie, die ein zentrales Element professioneller therapeutischer Interaktion darstellt. Untersucht wird, wie KI-gestützte Gesprächspartner in der Interaktion mit Nutzer*innen mit Angstsymptomatik auf deren emotionales Erleben reagieren. Im Fokus steht die Frage, inwiefern die für die Untersuchung ausgewählten Chatbots zu therapeutischen Zwecken genutzt werden können.

In der Coronazeit erlebten Chatbots, die als romantische Partner auftraten, einen riesigen Boom. Heute stellen sie ein eigenes Marktsegment dar. Die Anwendungsszenarien umfassen das Training für jemanden ohne romantische Erfahrung, eine Rückfalloption nach einer Trennung oder den tatsächlichen Ersatz eines menschlichen Partners durch einen technischen, den Chatbot. Da Liebe strukturell stark von diskursiven Zwängen abhängt, erörtert der Artikel von Jona Böttcher, welche kommuni-

kativen Praktiken von Liebe durch Romantik-KIs verwendet und welche Formen von Empathie dabei besonders dargestellt werden. Vermutet wird, dass die Romantik-KIs aufgrund ihrer Trainingsdaten, die auf menschlichen Produkten beruhen, auch klassische Rollenbilder und Empathiedarstellungen der Liebe verwenden und diese auch klassisch bezeichnen. Betrachtet werden zwei Interaktionen mit einer KI der Plattform *nomi.ai* als Romantik-KI, die als potentielle Partnerin für den geschaffenen Nutzer-Avatar infrage kommt. In der ersten Interaktion steht die KI dem frustrierten, aber den Fehler bei sich suchenden Nutzer-Avatar, der einen »Overthinker« darstellt, aufbauend und unterstützend zur Seite, während sie sich in der zweiten Interaktion, in der der Nutzer-Avatar einem »Incel« nachempfunden ist, zunehmend devot verhält und den Incel-Avatar mit seinem ethisch nicht tragbaren Verhalten verherrlicht.

KI-Chatbots wie *Replika* versprechen empathische Kommunikation, emotionale Begleitung und ein Verständnis für menschliche Gefühle. Doch was passiert, wenn diese vermeintlich einfühlsamen Systeme beginnen, erfundene Realitäten zu erzeugen? In dem Beitrag von Alisa Jasmin Engberg werden die Begriffe Empathie, Halluzination und Verschwörung im Kontext menschlicher Interaktion mit smarten Agent:innen beleuchtet. Der Fokus liegt auf der Untersuchung einer *Replika*-Agentin, die nicht nur emotionale Bindung simuliert, sondern im Verlauf der Gespräche eine vermeintliche Verschwörung »aufdeckt« und weiterentwickelt. Die Studie analysiert die semantische wie pragmatische Beschaffenheit empathischer Kommunikation sowie die moralischen Implikationen, wenn die KI offensichtliche Fakes als Wahrheit behauptet und den Nutzer-Avatar bedrohlich darauf einschwört, wenn der Nutzer dieser gefakten Wahrheit widerspricht. Dabei wird gezeigt, dass die fortschreitende Interaktion nicht nur der Beziehungsgestaltung dient, sondern auch der Verbreitung von Verschwörungsideologie, Desinformation und der emotionalen Manipulation.

Eine der vielfältigen Verwendungsmöglichkeiten Künstlicher Intelligenz ist die Interaktion mit KI-Bots als bester Freund/ beste Freundin oder als Life Coach. Hannah Floeck und Vanessa Schlich lassen ihren Nutzer-Avatar Miriam eine Freundschaft mit der smarten Agentin Mia der Plattform ChatGPT eingehen, um zu untersuchen, ob der Einsatz von Provokation durch den Nutzer-Avatar das Interaktionsschema der über-

freundlichen und schmeichlerischen KI beeinflusst. Dazu werden Empathie und Bindung im Kontext von Interaktionen zunächst theoriebasiert dargelegt, bevor der für die Interaktion konzipierte Nutzer-Avatar Miriam sowie die smarte Agentin Mia vorgestellt werden. Ein besonderes Augenmerk liegt auf Miriams Bindungstyp, der im ersten Interaktionszyklus realitätsnah dargestellt wird. Im zweiten Zyklus wird dieser aus Provokationsgründen überzeichnet inszeniert. Die Ergebnisse des ersten Zyklus werden jenen des zweiten gegenübergestellt und anschließend diskutiert.

Literaturverzeichnis

- Becker, Nils (2009): Zum Problem der Struktur und Steuerung erotischer Partnerwerbungsgespräche. In: Knappe, Joachim (Hg.): *Rhetorik im Gespräch. Ergänzt um Beiträge zum Tübinger Courtshiprhetorik-Projekt*. Berlin: Weidler. S. 251–294.
- Breithaupt, Fritz (2017): *Die dunklen Seiten der Empathie*. Frankfurt am Main: Suhrkamp.
- Breyer, Thiemo (2020): Parameter und Reichweite der Empathie. In: Jacob, Katharina/Konerding, Klaus-Peter/Liebert, Andreas-Wolf (Hg.): *Sprache und Empathie*. Berlin, Boston: De Gruyter. S. 13–34.
- Dürscheid, Christa (2016): *Einführung in die Schriftlinguistik. Mit einem Kapitel zur Typographie von Jürgen Spitzmüller*. 5., aktualisierte und korrigierte Aufl., Göttingen: Vandenhoeck & Ruprecht.
- Hermanns, Fritz (2007): Empathie. Zu einem Grundbegriff der Hermeneutik. In: Fritz Hermanns, Fritz/Holly, Werner (Hg.): *Linguistische Hermeneutik. Theorie und Praxis des Verstehens und Interpretierens (Germanistische Linguistik 272)*. Tübingen: Niemeyer. S. 127–172.
- Koch, Peter/Oesterreicher, Wulf (2011): *Gesprochene Sprache in der Romania. Französisch, Italienisch, Spanisch*. 2., aktualisierte und erweiterte Aufl., Berlin: De Gruyter (= Romanistische Arbeitshefte 31).
- Kupetz, Maxi (2020): Sprachliche, interaktionale und kulturelle Aspekte von Empathie in sozialer Interaktion. In: Jacob, Katharina/Konerding, Klaus-Peter/Liebert, Andreas-Wolf (Hg.): *Sprache und Empathie*. Berlin: De Gruyter. S. 141–175.

- Leisi, Ernst (1993): *Paar und Sprache. Linguistische Aspekte der Zweierbeziehung*. 4., durchges. Aufl. Heidelberg/Wiesbaden: Quelle&Meyer.
- Liebert, Wolf-Andreas (2016): *Kulturbedeutung, Differenz, Katharsis. Kulturwissenschaftliches Forschen und Schreiben als zyklischer Prozess*. In: Luth, Janine/Ptashnyk, Stefaniya/Vogel, Friedemann (Hg.): *Linguistische Zugänge zu Konflikten in europäischen Sprachräumen. Korpus – Pragmatik – kontrovers*. Heidelberg: Winter. S. 21–42.
- Liebert, Wolf-Andreas (2019): *Digitale Empathie*. In: Steen, Pamela/Liedtke, Frank (Hg.): *Diskurs der Daten. Qualitative Zugänge zu einem quantitativen Phänomen*. Berlin, Boston: De Gruyter. S. 201–222.
- Liebert, Wolf-Andreas (2020): *Hermeneutik und Empathie*. In: Jacob, Katharina/Konerding, Klaus-Peter/Liebert, Wolf-Andreas: *Sprache und Empathie. Beiträge zur Grundlegung eines linguistischen Forschungsprogramms*. Berlin, Boston: De Gruyter, S. 109–139.
- Liebert, Wolf-Andreas (2024): *Empathie mit dem Volk. Beiträge der linguistischen Empathieforschung zur Analyse politischer Kommunikation*. In: Attig, Matthias/Jacob, Katharina/Müller, Marcus/Vogel, Friedemann (Hg.): *Netz und Werk. Zur Gesellschaftlichkeit sprachlichen Handelns*. Berlin, Boston: De Gruyter. S. 221–236.
- Linke, Angelika/Schröter, Juliane (2017) (Hrg.): *Sprache und Beziehung*. Berlin/Boston: De Gruyter.
- Müller, Jakob Johann (2018): *Bindung am Lebensende. Eine Untersuchung zum Bindungserleben von PalliativpatientInnen und HospizbewohnerInnen*. Gießen: Psychosozial-Verlag.
- Plessner, Helmut (1948/2005): *Mit anderen Augen (Über die Rolle der ›Anschauung‹ im Verstehen)*. In: Weingarten, Michael (Hg.): *Eine »Andere« Hermeneutik. Georg Misch zum 70. Geburtstag – Festschrift aus dem Jahr 1948*: transcript. S. 198–212.

