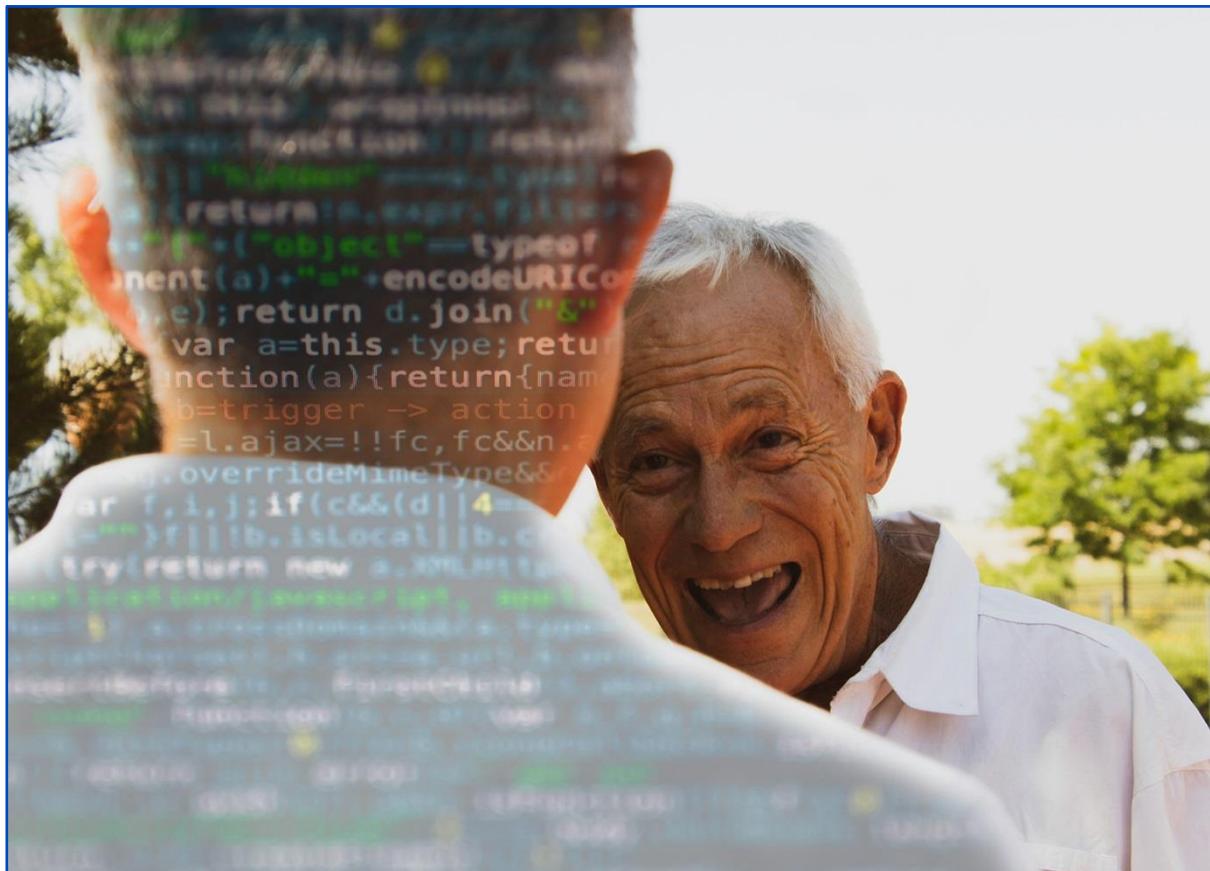


INTELLIGENTE SOZIALE AGENTEN GEGEN EINSAMKEIT?



© unsplash.com/Markus Spiske, Logan Weaver; Thomas Bayer

ZUSAMMENFASSUNG

Immer mehr digitale Anwendungen kommen auf den Markt, die zum Ziel haben, Menschen im Alltag emotional zu unterstützen und zu beraten. Unter digitalen sozialen Agenten versteht man KI-gestützte Chatbots, die textliche und/oder visuelle Inhalte für Gespräche mit Anwender:innen produzieren. Auf der einen Seite können solche Anwendungen die mentale und emotionale Gesundheit der Nutzer:innen stärken, ärmeren Bevölkerungsgruppen den Zugang zu psychologischer Unterstützung erleichtern und gegebenenfalls psychologische Behandlungen im Alltag unterstützen. Auf der anderen Seite besteht die Gefahr der Abhängigkeit, der langfristigen Verstärkung von Einsamkeit und des Verlusts zwischenmenschlicher Kompetenzen. Ob ein dialogorientierter KI-Chatbot eher die Fähigkeiten der Nutzenden stärkt, Beziehungen zu anderen Menschen aufzubauen, oder Einsamkeit eher noch verstärkt, hängt unter anderem maßgeblich von der Gestaltung und dem Diskurs über KI ab.

ÜBERBLICK ZUM THEMA

Immer mehr digitale Anwendungen kommen auf den Markt, die zum Ziel haben, Menschen im Alltag emotional zu unterstützen und zu beraten. Sie werden als intelligente soziale Agenten oder „conversational AI“ bezeichnet. Dabei handelt es sich meist um KI-gestützte Chatbot-Anwendung die mittels Generativer Künstlicher Intelligenz (GenKI) textliche, visuelle und audio- Inhalte für Gespräche mit Anwender:innen führen (Maples 2024). Laut Herstellern können intelligente soziale Agenten die Rolle eines Sparringpartners, Beraters, Mentors oder selbst eines *Freunds* übernehmen (dazu siehe auch *Social (Ro-)Bots: Maschinen als Gefährt:innen*). Zu den am häufigsten genutzten Anwendungen gehören „Replika“ aus den USA, die vor allem in Asien beliebte App „Xiaoice“ sowie die seit 2022 verfügbare Anwendung „character.ai“. Alleine Replika und Xiaoice haben zusammen insgesamt fast eine Milliarde aktive Nutzer:innen (Maples 2024). Nach Herstellerangabe wird die Plattform character.ai täglich von rund 3,5 Millionen Menschen besucht. Insbesondere Jugendliche greifen auf intelligente soziale Agenten zunehmend gerne zu (Ragni 2024).

Die auf dem Markt verfügbaren Chatbots setzen ähnliche Methoden ein und werben mit unterschiedlichen Schwerpunkten. Die Anwendung Replika wurde von der US-Firma Luka entwickelt, um als persönlicher Gesprächspartner und emotionaler Begleiter zu dienen. Im Laufe der Interaktion mit den Nutzenden lernt der Chatbot ihre Vorlieben und Interessen kennen und passt seine Antworten entsprechend an. Replika ist auf iPhone- und Android-Plattformen verfügbar und lässt sich über Text-, Sprach-, Augmented- und Virtual-Reality-Schnittstellen nutzen. Gespräche erfolgen mit einem vorgefertigten oder vom Nutzenden generierten Avatar (Maples 2024). Auch die Anwendung character.ai bietet die Möglichkeit, mit einem digitalen „persönlichen Lehrer, Assistenten oder sogar Freund“ zu interagieren, der dem/r Anwender:in „jederzeit zur Verfügung steht und bei allem helfen kann“. Die Plattform wirbt insbesondere damit, dass Gespräche mit digitalen Avataren bekannter Persönlichkeiten, Spiel- und Zeichentrickfiguren möglich sind. Die Plattform bietet außerdem die Möglichkeit, den vorprogrammierten digitalen Gesprächspartner je nach angestrebter Nutzung auszuwählen (z. B. Coach für Bewerbungsgespräche oder Fremdsprachenlehrer:in, Reiseplanungsbegleiter:in). Sowohl character.ai als auch das 2023 auf den chinesischen und japanischen Markt gebrachte „GPT Clones“ der Firma Xiaoice, die zu Microsoft gehört, werden dafür trainiert, konkrete Personen zu imitieren. Mit wenigen Daten – drei bis fünf Minuten eines öffentlichen Videos oder Daten aus sozialen Medien – kann der virtuelle Klon einer Person des öffentlichen Lebens erzeugt werden (Liu 2023). Einige Anwendungen versprechen zudem, „digitale psychologische Zwillinge“ von Privatpersonen anhand persönlicher Informationen aus E-Mails, Browserverlauf und Social-Media-Aktivitäten erstellen zu können. Manche Anbieter (z. B. das Startup „Super Brain“) haben sich auf die Entwicklung digitaler *Avatare verstorbener Personen* spezialisiert (Zhang 2024) (siehe dazu auch *Sterben 2.0: digitaler Nachlass*).

Viele Menschen,
insbesondere
Jugendliche und junge
Erwachsene, nutzen
digitale Chatbots
zur emotionalen
Unterstützung im
Alltag

Intelligente soziale
Agenten können
digitale Avatare
bekannter
Persönlichkeiten oder
von Privatpersonen
nachahmen

Intelligente soziale Agenten werden zu unterschiedlichen Zwecken genutzt. Eine zentrale Anwendung stellt die Förderung der emotionalen und mentalen Gesundheit (siehe **KI-Emotionserkennung**) dar. Sowohl Replika als auch character.ai bieten Avatare für Coachingzwecke bzw. zur Förderung der mentalen Gesundheit („mental health helper“) an. Diese werden von viele Studierenden in Anspruch genommen, um ihre mentale und emotionale Resilienz zu fördern (Maples et al. 2024; Ragni 2024). Manche Chatbots sind spezifisch und gezielt auf die Unterstützung der psychischen Gesundheit zugeschnitten, wie beispielsweise der Chatbot des Start-ups „Wysa“.¹ Wysa adressiert sowohl Patient:innen als auch Ärzt:innen. Letzteren bietet Wysa Unterstützung bei der Entwicklung von maßgeschneiderten Behandlungstherapien. In Deutschland ist vor allem „Mina“² bekannt. Dabei handelt es sich um eine Anwendung, die als Coach und digitaler Psychotherapeut getestet wird und insbesondere bei der Überwindung von Prüfungsängsten unterstützen soll, aber auch bei Mobbing oder Entscheidungsschwierigkeiten. Die meisten Chatbots, die auf großen Sprachmodellen basieren und Menschen bei psychischen Belastungen unterstützen sollen, verfügen allerdings nicht über eine Zulassung als Medizinprodukt. In UK wurde eine erste Chatbot-Anwendung im Bereich der psychischen Gesundheit („Limbic Access“,³) zugelassen (Ragni 2024). Diese soll insb. dabei helfen, Erstdiagnosen zu verfassen und Patient:innen zu identifizieren, die prioritär behandelt werden sollen.

Erste Forschungsergebnisse zeigen, dass intelligente soziale Agenten sich auf die mentale Gesundheit positiv auswirken können. Eine Umfrage unter 1006 studentischen Nutzer:innen von Replika, die kürzlich in der Zeitschrift Nature erschienen ist, zeigte, dass knapp die Hälfte der befragten Personen Replika eingesetzt haben, um als Freund oder Begleiter zu fungieren. Als Grund wurden dessen permanente Verfügbarkeit, seine Gesprächsfähigkeit und seine Fähigkeit, als Gesprächspartner nicht zu urteilen, genannt. Die Nutzung assoziierten die Befragten mit abgemilderten Angstzuständen und einem Gefühl der sozialen Unterstützung. Replika wurde auch zur therapeutischen Unterstützung bzw. zur Verarbeitung von Emotionen eingesetzt und wirkte in über 18 % der Fälle therapeutisch. Fast ein Viertel berichteten von positiven Lebensveränderungen. So berichteten Anwender:innen, sich dank der Nutzung von Replika in ihren realen Beziehungen empathischer zu verhalten. 3 % gaben an, dass Replika ihre Selbstmordgedanken stoppen konnte (Maples et al. 2024). Parallel dazu wurden gravierende Mängel festgestellt werden, wie Verletzung grundsätzlicher europäischer Datenschutzstandards oder die Ausgabe unpassender (sexueller) Inhalte (Krempl 2023). Außerdem wurden mehrere Suizidfälle mit der intensiven Nutzung von KI-Chatbots in Zusammenhang gebracht (siehe z. B. Schiefer 2024). Unabhängige Studien zu den Wirkungen intelligenter sozialer Agenten und dialogfähige Chatbots sind noch selten. Die Bindungsmechanismen, die den Anwendungen zu Grunde liegen, entwickeln sich rasch und können sich von Anwendung zu Anwendung und/oder über die Zeit grundlegend verändern. Insbesondere über die langfristige Nutzung intelligenter sozialer Agenten ist bisher wenig bekannt. Die Fähigkeit von Chatbots, Wünsche zu erkennen und eine Art

**KI-Therapiebots
werden zur Förderung
der mentalen
Gesundheit genutzt**

**Intelligente soziale
Agenten können sich
auf die mentale
Gesundheit positiv
auswirken**

**Über Langzeitfolgen
ist wenig bekannt**

¹ wysa.com.

² minabot.ai/de/.

³ access.limbic.ai.

„Echokammer der Zuneigung“ zu schaffen, könnte ihre Anwender:innen hochgradig süchtig machen (Mahari und Pataranutaporn 2024). Auch Mira Murati, Chief Technology Officer von OpenAI, warnte vor Suchtgefahren durch KI. Diese könnten insbesondere durch das langfristige Gedächtnis der Chatbots und die Möglichkeit, Systeme hochgradig zu personalisieren, entstehen (Klar 2023). Eine exzessive Nutzung könnte zu einer inneren Entfremdung und verstärkten Einsamkeit führen (Uhle 2024) oder gar eine Art „digitale Bindungsstörung“ verursachen. Die Nutzenden würden Kompetenzen verlieren, sich auf andere Menschen einzulassen und empathisch zu reagieren (Mahari und Pataranutaporn 2024). Auch die Erstellung digitaler Avatare von Verstorbenen wirft ethische und rechtliche Bedenken hinsichtlich der Einwilligung des Verstorbenen und seiner Familie sowie der emotionalen Gesundheit der Nutzer:innen auf. Ihre Nutzung könnte den Trauerprozess behindern (Zhang 2024).

KI-Chatbots, die gezielt zur Förderung der psychologischen Gesundheit angeboten werden, könnten insbesondere ärmeren Bevölkerungsgruppen den Zugang zu psychologischer Unterstützung erleichtern, die sich ein solches Angebot sonst nicht leisten könnten. Problematisch wird es allerdings insbesondere dann, wenn bestimmte Funktionalitäten (bei Xiaoice etwa komplexere Antworten und Sprachnachrichten) nur gegen Aufpreis erhältlich sind. Dies wird von manchen Nutzenden als eine potentielle Barriere für den Zugang zu psychosozialer Unterstützung wahrgenommen (Maples 2024). Bisher werden solche KI-Anwendungen, die gezielt und ausschließlich dazu dienen sollen, die mentale Gesundheit von Patient:innen zu fördern, nicht als Ersatz, sondern eher als Unterstützung einer psychologischen Behandlung angesehen. Beispielsweise spielt die Fähigkeit, Humor, Körpersprache, Mimik und Gestik von Patient:innen zu erkennen, in der therapeutischen Interaktion eine zentrale Rolle. Dies ist in einer Chatbot-basierten therapeutischen Behandlung nicht gegeben (Stegmaier 2024). Auch wenn diese Anwendungen aktuell keine professionell durchgeföhrte Psychotherapie ersetzen können, könnten sie dennoch durch Übungsempfehlungen zur Unterstützung der therapeutischen Behandlung eingesetzt werden oder auch, um die Zeit bis zum Beginn einer Therapie zu überbrücken (SWRWissen 2023a). Darüber hinaus verfügen die meisten derzeit auf dem Markt befindlichen Chatbots, die die mentale Gesundheit von Menschen unterstützen sollen, über keine medizinische Zulassung (SWRWissen 2023b), sie unterlaufen sehr unterschiedliche, tlw. transparente Qualitätssicherungsmechanismen.

KI-Chatbots können den Zugang zur psychologischen Unterstützung erleichtern, die medizinische Behandlung psychischer Störungen aber nicht ersetzen

Ob ein dialogorientierter KI-Chatbot eher die Fähigkeiten der Nutzenden stärkt, Beziehungen zu anderen Menschen aufzubauen, oder Einsamkeit eher noch verstärkt, hängt stark vom Design der Anwendungen ab. Chatbots können beispielsweise so gestaltet werden, dass sie als Übungsräum zum Erlernen von Konfliktlösungskompetenzen dienen und zur Verbesserung des Verhaltens in ihren analogen interpersonellen Beziehungen führen (Uhle 2024). Bewusste Designentscheidungen, die Nutzenden im digitalen Raum solange wie möglich verbleiben zu lassen (sogenannte „dark patterns“), können wiederum eher Einsamkeit verstärken (Mahari und Pataranutaporn 2024). Außerdem werden dialogfähige KI auch nicht immer für das eingesetzt, wofür die jeweiligen Chatbots trainiert wurden (Longpre et al. 2024). Wird ein Chatbot beispielsweise für intime Gespräche genutzt, ohne darauf trainiert worden zu sein, kann sich das negativ auf das

Design- und Kommunikationsentscheidungen wirken sich auf soziale Kompetenzen, Gesundheit und Wohlbefinden aus

Wohlbefinden der Nutzenden auswirken (Maples et al. 2024). Nicht nur Designentscheidungen, sondern auch Diskurse über KI-Chatbots bzw. intelligente soziale Agenten beeinflussen, wie Menschen mit ihnen umgehen und wie gut sie wirken. Denn das mentale Modell einer Person, das heißt ihre Vorstellung des Chatbots, zum Beispiel als mitfühlend-fürsorglicher oder als emotionsloser Algorithmus, prägt ihre Interaktion mit dem Chatbot. Dieses mentale Modell ergibt sich aus dem kulturellen Hintergrund, persönlichen Überzeugungen, aber auch aus den gesellschaftlichen Narrativen und dem Diskurs von Software-Anbietern zu ihren Produkten. Durch eine sorgfältige Darstellung der Anwendung (sogenanntes „Priming“), die einen Chatbot als einfühlsam darstellt, kann Vertrauen in das System gesteigert, eine höhere Bindung und so höhere Wirkungen beispielsweise bei der Verbesserung der mentalen Gesundheit erreicht werden (Pataranutaporn et al. 2023), wobei die Suchtgefahr auch steigen kann. Hingegen können antisoziale Konnotationen die potenziellen Vorteile eines Einsatzes von KI zur Verbesserung zwischenmenschlicher Kommunikation untergraben (Hohenstein et al. 2023). Kaum erforscht und potenziell weitreichend sind die langfristigen Folgen positiver Narrative, die ggf. die Risiken in den Hintergrund geraten lassen. Insbesondere die Ausnutzung positiver Narrative zu Marketingzwecken bzw. um Unternehmensprofite zu maximieren, könnte weitreichende gesellschaftliche Folgen mit sich bringen (Pataranutaporn et al. 2023).

RELEVANZ DES THEMAS FÜR DAS PARLAMENT UND FÜR ÖSTERREICH

Auch wenn Österreich aktuell den Aufbau von Kompetenzen im Bereich KI in Gesellschaft fördert (Bundeskanzleramt 2024), werden Chancen und Risiken der Nutzung intelligenter sozialer Agenten zum Aufbau sozialer Kompetenzen und zur Förderung der mentalen Gesundheit wenig diskutiert. Angesichts der rasch steigenden Nutzerzahlen, der unbekannten langfristigen Folgen sowohl für Gesundheit und Wohlbefinden der Nutzenden als auch für den gesellschaftlichen Zusammenhalt und der fehlenden Qualitätssicherungsmechanismen und grundsätzlichen Datenschutzbedenken mancher Anwendungen wäre eine Auseinandersetzung mit ihren Chancen und potenziellen Auswirkungen geboten. Außerdem könnten geeignete KI-Chatbots ggf. den Druck auf die mangelnden kassenfinanzierten Therapieplätze etwas entschärfen (SWRWissen 2023b), dabei wäre jedoch zu überlegen, wie die Wirksamkeit nachgewiesen werden kann, wie die Anwendungen in die klinische Versorgung integriert werden könnten, ohne einen gerechten Zugang zu Therapieplätzen zu gefährden. Außerdem müssten ausreichende Kompetenzen bei Personal und Patient:innen aufgebaut werden.

Bislang stehen intelligente soziale Agenten wenig im Vordergrund gesellschaftlicher und politischer Diskussion um Generative KI

VORSCHLAG WEITERES VORGEHEN

Die Forschung zu Auswirkungen der Verbreitung intelligenter sozialer Agenten auf Individuen im privaten und beruflichen Kontext sowie auf Gesellschaften steckt noch in den Kinderschuhen. Insbesondere sind die Langzeitfolgen des intuitiven Engagements mit intelligenten sozialen Agenten interdisziplinär zu erforschen. Außerdem wäre auszuloten, inwiefern der aktuelle Regulierungsrahmen die Risiken ihrer Nutzung ausreichend abdeckt bzw. welche Regulierungsbedarfe durch die verbreitete Nutzung der KI-Chatbots entstehen.

Ein besseres Verständnis der Nutzung und Wirkung intelligenter sozialer Agenten ist notwendig

ZITIERTE LITERATUR

- Bundeskanzleramt (2024). Digital Austria. digitalaustria.gv.at.
- Hohenstein, J. et al. (2023). Artificial intelligence in communication impacts language and social relationships. *Scientific Reports* 13 (1), 5487, doi.org/10.1038/s41598-023-30938-9.
- Klar, R. (2023). Open AI exec warns AI can become 'extremely addictive'. *The Hill*. thehill.com/policy/technology/4229972-open-ai-exec-warns-ai-can-become-extremely-addictive/.
- Krempl, S. (2023). „Virtueller Freund“: Datenschützer stoppen Chatbot Replika in Italien. heise.de/news/Virtueller-Freund-Datenschuetzer-stoppen-Chatbot-Replika-in-Italien-7485010.html.
- Liu, Y. (2023). China's „Black Mirror“ GPT Clones Raise Global Concerns. *Jing Daily Culture*. jingdailyculture.com/chinas-black-mirror-gpt-clones-raise-global-concerns/.
- Longpre, S. et al. (2024). Consent in Crisis: The Rapid Decline of the AI Data Commons. arXiv. doi.org/10.48550/arXiv.2407.14933.
- Mahari, R.; Pataranutaporn, P. (2024). We need to prepare for 'addictive intelligence'. *MIT Technology Review*. technologyreview.com/2024/08/05/1095600/we-need-to-prepare-for-addictive-intelligence/.
- Maples, B. et al. (2024). Loneliness and suicide mitigation for students using GPT3-enabled chatbots. *npj Mental Health Research* 3 (1), 1–6. doi.org/10.1038/s44184-023-00047-6.
- Pataranutaporn, P. et al. (2023). Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *Nature Machine Intelligence* 5 (10), 1076–1086. doi.org/10.1038/s42256-023-00720-7.
- Ragni, M. (2024). Character.ai: Viele Jugendliche wenden sich KI-Therapiebots zu. *THE-DECODER.de*. the-decoder.de/character-ai-viele-jugendliche-wenden-sich-ki-therapiebots-zu/.
- Schiefer, C. (2024). Suizid in den USA: Wenn KI-Freunde zur Gefahr werden. br.de/nachrichten/netzwelt/wenn-ki-freunde-zur-gefahr-werden-suizid-in-den-usa-zeigt-tragischen-verlauf-einer-ki-beziehung-USgb6Ux.
- Stegmaier, A. (2024). KI und Psychotherapie: Hallo, ChatGPT, mir geht es nicht gut. *Wr. Zeitung*. wienerzeitung.at/a/hallo-chatgpt-mir geht-es-nicht-gut.
- SWRWissen (2023a). Chatbots – die neuen Psychotherapeuten? *swr.online*. swr.de/wissen/chatbots-in-der-psychotherapie-100.html.
- SWRWissen (2023b). Chatbots liefern keine zuverlässigen medizinischen Diagnosen. *swr.online*. swr.de/wissen/chatbots-medizinische-diagnose-kein-ersatz-fuer-arztbesuch-100.html.
- Uhle, C. (2024). Künstliche Intelligenz und die Zukunft der Menschheit. Fischer.
- Zhang, X. (2024). AI Simulates Posthumous Experience. shanghaiist.com/ai-simulates-posthumous-experience/.