

«Verkörperte KI» – Soziale Roboter in der ambulanten Suchthilfe?

2024-2
Jg. 50
S. 30 - 34

Stellen Sie sich vor, ein sozialer Roboter könnte künftig nicht nur das Zuhause sauber halten, sondern auch eine tiefe, emotionale Hilfe für Klient:innen einer ambulanten Suchttherapie bieten: Ohne Freunde, isoliert und mit nur einer wöchentlichen Beratungssitzung und einem monatlichen Gruppenmeeting als Unterstützung wird der soziale Roboter zu einem Gefährten, der die stille Leere zu Hause mit Verständnis füllt und mit seiner kontinuierlichen Präsenz die ambulante Beratung ergänzt. Eine Utopie? Einblicke in ein geplantes Forschungsprojekt.

STEFAN M. ADAM

Prof., Leiter Institut Beratung, Coaching und Sozialmanagement, Hochschule für Soziale Arbeit, Fachhochschule Nordwestschweiz, Riggensbachstrasse 16, CH-4600 Olten, Tel. +41 (0)629 572 01 2, stefan.adam@fhnw.ch

PHILLIP GACHNANG

MSc., Wissenschaftlicher Mitarbeiter, Institut für Wirtschaftsinformatik im Kompetenzschwerpunkt Systems Engineering, Hochschule für Wirtschaft, Fachhochschule Nordwestschweiz, Riggensbachstrasse 16, CH-4600 Olten, Tel. +41 (0)629 572 88 9, phillip.gachnang@fhnw.ch

SUSANNE LEIBERG

Dr. rer. nat., Psychologin, Co-Leiterin Ambulatorium Lenzburg, Klinik im Hasel AG, Niederlenzer Kirchweg 1, CH-5600 Lenzburg, Tel. +41 (0)627 386 39, sleiberg@kih.ch

HARTMUT SCHULZE

Prof. Dr., Psychologe, Institut für Kooperationsforschung und -entwicklung, Hochschule für Angewandte Psychologie, Fachhochschule Nordwestschweiz, Riggensbachstrasse 16, CH-4600 Olten, Tel. +41 (0)629 572 41 9, hartmut.schulze@fhnw.ch

JAN WILLEM NIEUWENBOOM

Dr., Gesundheitspsychologe FSP, Institut Soziale Arbeit und Gesundheit, Hochschule für Soziale Arbeit, Fachhochschule Nordwestschweiz, Riggensbachstrasse 16, CH-4600 Olten, Tel. +41 (0)629 572 10 9, wim.nieuwenboom@fhnw.ch

Alle Menschen streben nach engen, sicheren und tiefen emotionalen zwischenmenschlichen Beziehungen. Dieser Wunsch nach Bindung wird in verschiedenen Konzeptionen menschlicher Motivation als Grundbedürfnis angesehen (Grawe 1998; Young 2005). Wenn Menschen während ihres Aufwachsens dauerhaft belastende, traumatisierende Bindungserfahrungen machen, bedeuten Beziehungen für sie Unsicherheit und Gefahr. Um ihre Gefühle «gefährlos» regulieren zu können und so etwas wie Nähe und Wärme spüren zu können, greifen manche von ihnen zu Suchtmitteln oder süchtigen Verhaltensweisen und eine Abhängigkeit kann sich entwickeln (Brisch 2015). Das Suchtmittel wird hier zur «Ersatz-Bindungsperson».

Die Förderung positiver Bindungserfahrungen in einem geschützten Rahmen ist besonders gut während einer stationären Behandlung möglich, doch sind diese Aufenthalte, richtigerweise, zeitlich begrenzt, damit die Klient:innen das Gelernte im Alltag umsetzen lernen.

Restriktionen ambulanter Therapie und Beratung

Nach einem stationären Aufenthalt findet die Nachsorge meist in Form einer ambulanten Psychotherapie oder Beratung statt. Auch hier ist die konstante, belastbare zwischenmenschliche Beziehung, sei es in der Einzel- und/oder Gruppentherapie, ein wichtiges Ziel. Jedoch sind Wartezeiten für ambulante Therapieplätze lang und Therapiesitzun-

gen finden meist nur einmal pro Woche und werktags statt. Weil soziale und berufliche Strukturen teilweise nur spärlich etabliert werden konnten, beziehungsweise durch die Erkrankung weggebrochen sind, erleben viele Klient:innen Einsamkeit und Isolation in den Zeiten zwischen den Sitzungen. Das erhöht das Risiko eines Rückfalls und verlängert somit die Therapie. Um Klient:innen zwischen den Therapiesitzungen niederschwellig therapeutische Unterstützung anbieten zu können, kommen mehr und mehr Online-Therapieprogramme als «blended psychotherapy» zum Einsatz (Bielinski et al. 2021). In Online-Therapieprogrammen werden meist die Inhalte der vorherigen Therapiesitzung mittels Übungen vertieft. Jedoch können

positive Bindungserfahrungen hier allenfalls angeregt, aber nicht «gemacht» werden, zum Beispiel durch die Aufgabe, eine:n Freund:in anzurufen.

Hier sieht ein interdisziplinäres Team der Fachhochschule Nordwestschweiz Chancen für den Einsatz sozialer Roboter: Sie könnten die Übergänge zwischen Therapiesettings als «soziale Begleiter» emotional vereinfachen sowie eine «verkörperte» Bindung zwischen den Therapiesitzungen ermöglichen. Eine explorative Studie des Forschungsteams zeigte bereits eine sehr hohe Aufgeschlossenheit der Klient:innen sowie der Beratenden gegenüber einem sozialen Roboter (Bestgen et al. 2022).

Im vorliegenden Beitrag geben wir einen Einblick in ein geplantes Forschungs- und Entwicklungsprojekt (voraussichtlich September 2024 bis März 2027). Zunächst klären wir den Begriff sozialer Roboter, zeigen den aktuellen Forschungsstand auf und leiten daraus psychologische und technische Anforderungen sowie das Forschungsdesign ab. Ziel des Projekts ist herauszufinden, ob soziale Roboter die Aufrechterhaltung positiver Bindungserfahrungen unterstützen sowie zu niedrigeren Rückfallquoten und zu kürzerer Therapiedauer beitragen können.

Was ist ein sozialer Roboter?

Soziale Roboter (engl. «social robots», auch «socially interactive robots» oder «socially assistive robots») stellen einen eigenen Robotertyp dar, der spezifisch für die Kommunikation und die Zusammenarbeit mit Menschen entwickelt wurde (Bendel 2021; Henschel et al. 2021). Bekannt sind beispielsweise Pepper, QTrobot und Paro, welche auch als «Therapieroboter» genutzt und untersucht werden (Akyazi & Baştumur 2024; Silvera-Tawil 2024). Aufgrund von künstlicher Intelligenz (KI) sind sie zu Handlungen, Bewegungen, aber auch zur Erfassung verbaler und nonverbaler Kommunikation und zu angemessenen und zielführenden Handlungen und Interaktionen fähig (Schulze et al. 2021). Chatbots und soziale Roboter haben gemeinsam, dass sie auf die verbale und/oder textliche Kommunikation mit Menschen ausgelegt sind. Im Gegensatz zu Chatbots haben soziale Roboter eine

dinghafte, tier- oder menschenähnliche physische Gestalt, sind berührbar und haben in der Regel verschiedene Sensoren, mit denen sie ihre Umwelt, aber auch Emotionen ihrer Gegenüber erkennen und selbst ebenfalls emotional kommunizieren können. Insbesondere aus der Perspektive der Suchttherapie scheint die verkörperte KI in Form eines sozialen Roboters sehr bedeutsam. Denn Studien zeigen, dass ein sozialer Roboter gegenüber einem Chatbot infolge der Repräsentanz bevorzugt und als lustiger wahrgenommen wird. Zudem kann zu ihm eine höhere Compliance und damit eine stärkere Bindung entwickelt werden (Scoglio et al. 2019).

Stand der Forschung

Auf Basis eines systematischen Literaturreviews kamen Scoglio et al. (2019) zum Schluss, dass die Forschung im Bereich der psychischen Gesundheit mit sozialen Robotern erst im Entstehen ist (Scoglio et al. 2019: 9). Die in den Review eingeschlossenen zwölf Studien umfassten den Einsatz von fünf verschiedenen sozialen Robotern zur Beeinflussung verschiedener Kenngrößen der psychischen Gesundheit. Die meisten Studien kon-

zentrierten sich nach den Autor:innen darauf, den Studienteilnehmenden Trost und Gesellschaft zu bieten (zum Beispiel die Roboter-Robbe Paro). Eine Minderheit der Studien nutzte soziale Roboter zur Durchführung einer spezifischen Intervention (zum Beispiel Motivationsgespräche mit dem Roboter NAO). Die Auswirkungen von Interventionen und Interaktionen mit sozialen Robotern auf die psychische Gesundheit reichten von allgemein positiv bis gemischt, wobei einige Studien positive Veränderungen der Stimmung und der Lebensqualität nach Roboterinteraktionen feststellten. Diese Übersicht zeigt, dass die vorhandenen Studien zu den potenziellen Auswirkungen von sozialen Robotern auf das psychische Wohlbefinden im Gesundheitsbereich von ihrer empirisch gesicherten Verallgemeinerbarkeit her noch sehr eingeschränkt sind. Denn fast alle in diesem Bereich erfolgten Studien wurden in stationären Einrichtungen der Altenpflege oder in Laboren durchgeführt. Auch eine kürzlich veröffentlichte bibliometrische Analyse technologiebasierter Interventionen in der klinischen Gesundheitspsychologie (Flujas-Contreras et al. 2023) zeigte ein ähnliches Bild. Die Veröffent-

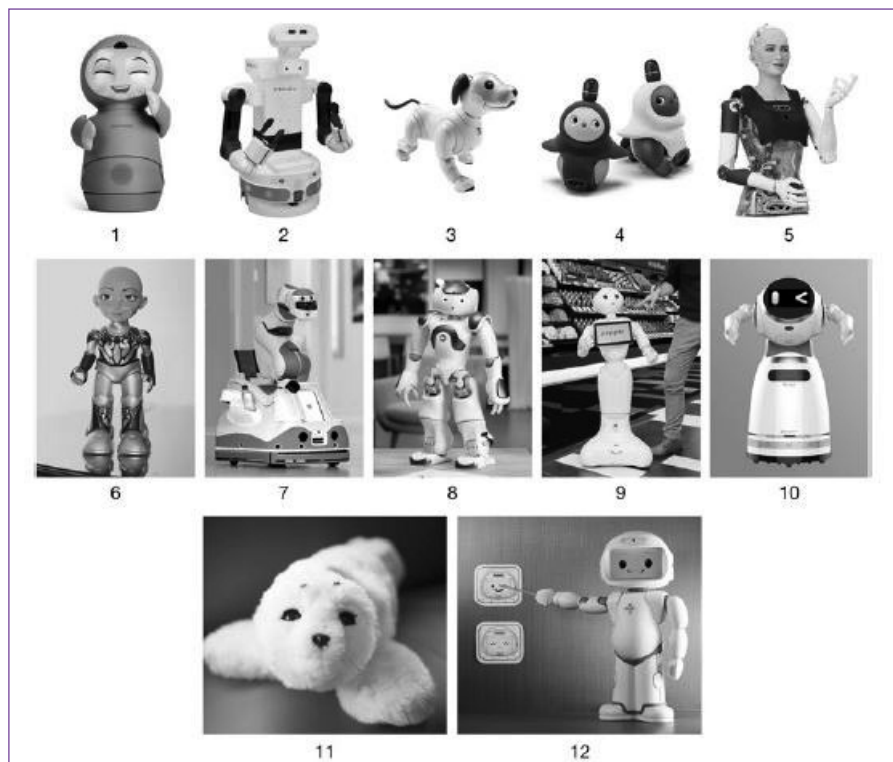


Abbildung 1: Übersicht über Robotermodelle mit sozialen Funktionen (Bendel 2021).
 [1] Moxie,¹ [2] Tiago,² [3] Aibo,³ [4] Lovot,⁴ [5] [6] Sophia u. Little Sophia,⁵ [7] LiO,⁶ [8] [9] NAO u. Pepper,⁷ [10] Cruzr,⁸ [11] Paro,⁹ [12] QTrobot¹⁰.

lichungen zu sozialen Robotern machten ca. 5 % aller gefundenen Studien aus und blieben auf relativ kleinem Niveau konstant (ca. 4–6 Studien mit guter psychometrischer Güte pro Jahr).

Anforderungen an soziale Roboter als Therapieroboter

Auf der Basis der wissenschaftlichen Literatur können die folgenden Anforderungen an einen erfolgreichen und akzeptanzfähigen Einsatz von sozialen Robotern im Therapiekontext abgeleitet werden:

Anforderung 1: Von der Dyade zur Triade: Häufig wird die «Mensch-Roboter-Interaktion» gedacht als die Interaktion eines einzelnen Menschen (Libin & Libin 2004). Auch die in Evaluationsstudien zunehmend abgefragte Variable «Working Alliance» wird in der Regel bezogen auf die Zusammenarbeit mit dem Roboter ohne Einbezug der Therapeut:innen verstanden (Prochaska et al. 2023). Zwar werden in der Diskussion um Therapieroboter mit den Rollen «Robo-Mediator» und «Robo-Assistent» zwei Rollen unterschieden, die nur in Kooperation mit Psychotherapeut:innen denkbar sind. Die Rolle «Robo-Therapeut» wird jedoch vor allem als Ersatz für menschliche Therapeut:innen konzipiert, auch wenn dies nur kurzzeitig der Fall ist, weil beispielsweise menschliche Therapeut:innen nicht verfügbar sind (Baisch & Kolling 2021). Die Vorstellung eines Ersatzes von Therapeut:innen durch Roboter könnte nach Baisch & Kolling (2021) «z. B. bei Menschen mit Demenz oder psychischen Erkrankungen, zu einer Verschlechterung des Befindens führen, da «echte» Kontakte verringert werden» (ebd.: 433). Für das geplante Forschungsprojekt bedeutet dies, den Einsatz von Robotern in der Therapiesituation konsequent als Netzwerk bzw. Triade, bestehend aus Klient:in, Roboter und Therapeut:in zu konzipieren.

Anforderung 2: Bindung zwischen Klient:in und Roboter: Als Erfolgsfaktor für eine gelingende Therapie von Abhängigkeitserkrankungen hat sich der Aufbau einer vertrauensvollen, stabilen Beziehung zwischen Klient:in und Therapeut:in, wie eingangs aufgezeigt, herausgestellt (Khawaja & Bélisle-Pipon 2023). Dafür ist Empathie ein wesentlicher Bau-

stein und es wird häufig gefragt, ob und in welcher Weise Roboter zu Empathiefähig sind oder dies zumindest simulieren können. Das Team geht in diesem Zusammenhang davon aus, dass die Beziehung zwischen Klient:in und Roboter von der Bindung zwischen Klient:in und Therapeut geprägt sein wird. Die Beziehung zwischen Klient:in und Roboter wird quasi ein Stück weit emotional und symbolisch bedeutsamer, wenn es auch Sitzungen zu dritt gibt. Auch die physische Gestalt von Robotern spielt hier eine Rolle, da die Präsenz eines Roboters im Raum in der Regel als grösser erlebt wird als bei einem virtuellen Chatbot oder einem Avatar (Baisch & Kolling 2021).

Ebenfalls verhalten sich Menschen gegenüber sozialen Robotern zum Teil offener und vertrauen ihnen Inhalte an, die sie zum Beispiel aufgrund von sozialen Normen und Konventionen selbst einem Therapeuten / einer Therapeutin nicht ohne Weiteres anvertrauen würden. Baisch & Kolling (2021) vermuten, dass «eine subjektiv wahrgenommene Unvoreingenommenheit von Robotern im Vergleich zu Therapeut:innen das Schamempfinden mancher Betroffenen über ihre Defizite mindern könne» (ebd.: 420). Daraus folgt, dass der physischen Gestalt des Roboters ein wichtiger Stellenwert zukommt, der im geplanten Projekt zu berücksichtigen, zu variieren und dessen Wirkung zu messen ist.

Anforderung 3: Fokussierung des Roboters auf Verhaltensänderungen: Die Ergebnisse einer Meta-Analyse von Costescu et al. (2014) zur Robot-enhanced Psychotherapy ergaben einen mittelgrossen signifikanten Effekt auf Verbesserungen hinsichtlich der vereinbarten Therapieziele: 69 % der Teilnehmenden ohne Roboter erreichten eine weniger gute Verbesserung als die durchschnittlichen positiven Veränderungen der Teilnehmenden in der robotergestützten Therapiebedingung. Die Autor:innen konnten den übergeordneten positiven Effekt in der Analyse auf den signifikanten Effekt der robotergestützten Therapie auf die Verbesserung der Leistungen auf der Verhaltensebene zurückführen (Costescu et al. 2014). Keinen Effekt fanden sie bezogen auf die kognitive Dimension. Entsprechend kann daraus

gefolgert werden, dass sich die Aufgaben des Roboters in der Therapie von Abhängigkeitserkrankungen vor allem auf die verhaltensbezogene Ebene beziehen sollten. Dies schliesst ebenfalls die Vermeidung von Verhaltensweisen wie zum Beispiel den Griff zum Suchtmittel ein.

Technologische Anforderungen

Künstliche Intelligenz bezeichnet die Fähigkeit von Computern, komplexe Aufgaben auszuführen, die normalerweise menschliche Intelligenz erfordern. Heutzutage wird der Begriff oft mit maschinellem Lernen assoziiert, einem Teilgebiet der künstlichen Intelligenz, das sich auf das Training von Computern konzentriert, um aus vergangenen Daten zu lernen und sich zu verbessern (siehe den Beitrag von Robert Lehmann in dieser Ausgabe). KI wird von der Expert:innengruppe der EU-Kommission (European Commission 2018) als System definiert, das seine Umgebung analysiert und autonom handeln kann, ohne dass jede Aktion vorher festgelegt sein muss. Demnach spricht man von künstlicher Intelligenz, wenn für einen Computer nicht jeder einzelne Schritt im Voraus programmiert sein muss, um eine spezifische Aufgabe zu lösen.

Aus technologischer Sicht stellt die Entwicklung eines sozialen Roboters für die Therapie mit Klient:innen ganz besondere Anforderungen an die Hardware, die Software und die Ethik. Die Komplexität besteht darin, dass die Sensorik in der Lage sein muss, Gesichter, Stimmen, Emotionen und Situationen auch in unvorhersehbaren Umgebungen zuverlässig zu erkennen. Eine der grössten Herausforderungen dabei ist, einem Roboter ein situatives Wissen zu vermitteln, welches ermöglicht, akute Situationen automatisch zu erfassen. Dies erfordert die nahtlose Integration verschiedener künstlicher Intelligenzen wie Gesichtserkennung, Stimmerkennung, Emotionserkennung und ein gutes Verständnis gesprochener Sprache in unbekannter Umgebung, was mit einer enormen Rechenkapazität und Datensammlung verbunden ist. Es gibt heute bereits Systeme, die einzelne dieser Herausforderungen lösen können. Allerdings gibt es noch kein umfassendes System, das all diese Probleme zusammen löst

und speziell auf die sensible Domäne der Suchttherapie ausgerichtet ist. Um ein solches System zu entwickeln, werden Sensordaten benötigt, welche akute Situationen in der Suchttherapie abbilden. Diese Daten können nur unter Laborbedingungen gesammelt werden, da die Beschaffung in der sensiblen Domäne der Suchttherapie eine Herausforderung in sich selbst darstellt. Zusätzlich sind Labordaten in der Anwendung oft unzureichend, um alle Facetten der Realität erfolgreich zu erfassen. Das Projekt zielt darauf ab, verschiedene künstliche Intelligenzen miteinander zu verknüpfen und unter Laborbedingungen für die spezielle Anwendung in der Suchttherapie zu trainieren. Das langfristige Ziel ist die Schaffung einer umfassenden künstlichen Intelligenz für einen Roboter in der Suchttherapie. Einmal trainiert, kann der soziale Roboter zwar mit weniger Rechenleistung verschiedene Aufgaben lösen, benötigt jedoch immer noch einen leistungsstarken Prozessor, um in Echtzeit reagieren zu können, welcher über Akkus betrieben werden müsste. Weiter kommt herausfordernd dazu, dass ein sozialer Roboter hochsensible Daten verarbeitet, die einen besonderen Schutz erfordern. Die Sicherheit und Privatsphäre der Klient:innen müssen dabei stets gewährleistet sein. Um diesen Anforderungen gerecht zu werden, sind innovative Lösungen notwendig. Dies erfordert nicht nur technisches Wissen, sondern auch ein tiefes Verständnis für die ethischen Implikationen und die Bedürfnisse der Nutzer:innen.

Interdisziplinäres Forschungsdesign

Angesichts dieser Herausforderungen ist ein Forschungsdesign angedacht, dass die Expertise aus der Maschinenethik, Informatik, Sozialen Arbeit, Psychologie und die Praxiserfahrung in der ambulanten Suchtberatung und -therapie miteinander verbindet.

Die Akzeptanz des sozialen Roboters seitens der Klient:innen und der Therapeut:innen und die Implementation des neuen Verfahrens in die Abläufe des Ambulatoriums als Forschungsfeld ist eine grundlegende Voraussetzung. Denn die Therapeut:innen sollen nach entsprechender Instruktion in der Lage sein, eigenständig den Roboter in Über-

einstimmung mit dem angewendeten Therapieverfahren zu konfigurieren. Es ist zu erwarten, dass die Akzeptanz seitens der Klient:innen grösser sein wird, je mehr der Roboter auf die individuellen Bedürfnisse der/des Klient:in abgestimmt werden kann.

Im Laufe des Therapieprozesses soll die Wirkung des sozialen Roboters in Bezug auf die Aufrechterhaltung positiver Bindungserfahrungen und das Konsumverhalten erfasst werden. Nach dem transtheoretischen Modell (Di Clemente et al. 1991) lässt sich die Verhaltensänderung bei Klient:innen im Verlauf der Therapie in verschiedenen Phasen aufteilen: a) Absichtslosigkeit, b) Absichtsbildung, c) Vorbereitung, d) Aktion und e) Aufrechterhaltung. Während die ersten vier Phasen vorwiegend auf den Aufbau eines Commitments zur Verhaltensänderung und die Motivation der Klient:innen zielen, geht es in der fünften Phase vor allem um volitionale Prozesse, die dazu führen, dass die Klient:innen in der Lage sind die Verhaltensänderung (Reduktion des Konsums bis hin zur Abstinenz) aufrechtzuerhalten. Es ist zu erwarten, dass die Therapeut:innen insbesondere beim Aufbau der Motivation (Absichtsbildung und Vorbereitung) eine wichtige Rolle spielen. Während der Phase des Aufrechterhaltens, die auch ausserhalb der Therapiesitzung stattfinden soll, soll hauptsächlich der soziale Roboter ins Spiel kommen. Auf der einen Seite wird angenommen, dass das Erleben von Bindung (Attachment) zur Reduktion von Gefühlen wie Stress und Einsamkeit beiträgt, was die Aufrechterhaltung des gewünschten Verhaltens begünstigt. Auf der anderen Seite soll die Präsenz des Roboters als Gegenüber dazu beitragen, angemessen auf Verhaltensweisen zu reagieren, die diese Aufrechterhaltung gefährden (Ainsworth et al. 2015).

Um festzustellen, ob der soziale Roboter positive Effekte bei Klient:innen bewirken kann, ist eine genaue Dokumentation des therapeutischen Prozesses, insbesondere der Roboter-Klient:in-Interaktionen während des gesamten Therapieprozesses und insbesondere in der Phase der Aufrechterhaltung notwendig. Für die Untersuchung intraindividuelle Prozesse bietet sich das

Single-Case-Design¹¹ (N = 1 Design) an. Zunächst soll eine limitierte Anzahl (10–20) der Klient:innen beobachtet werden. Der Roboter ist darauf ausgelegt, das Interaktionsverhalten mit Klient:innen zu registrieren und damit die Reliabilität und Validität der erhobenen Daten zu gewährleisten. Zur Implementation des Verfahrens (Akzeptanz, Zufriedenheit, Realisierbarkeit) werden ebenfalls bei den Klient:innen und den Therapeut:innen Daten erhoben. In einer zweiten Phase der Untersuchung soll eine interindividuelle Überprüfung der Wirksamkeit stattfinden, indem Kontrollgruppen (beispielsweise mit normalen Therapiebedingungen oder mit einem Chatbot) gebildet werden. In einer Nachbefragung (Follow-up) wird dann die langfristige Wirksamkeit des Verfahrens evaluiert. Die Gesamtergebnisse werden schliesslich aufzeigen, in welchem Ausmass ein sozialer Roboter als sozialer Akteur künftig in der ambulanten Therapie von Abhängigkeitserkrankungen integriert werden kann, um die Klient:innen nutzbringend in ihrem Rehabilitationsprozess zu unterstützen.

Literatur

- Ainsworth, M.D.S./Blehar, M.C./Waters, E./Wall, S. (2015): Patterns of attachment: A psychological study of the strange situation – classic edition. New York/London: Routledge.
- Akyazi, K.G./Baştemur, Ş. (2024): Interactive Robots: Therapy Robots. *Psikiyatride Guncel Yaklasimlar – Current Approaches in Psychiatry* 16(1): 16–30. doi:10.18863/pgy.1242958
- Baisch, S./Kolling, T. (2021): Roboter in der Therapie: Vom Demonstrationsobjekt zum Psychotherapeuten. S. 417–440 in: O. Bendel (Hrsg.), *Soziale Roboter. Technikwissenschaftliche, wirtschaftswissenschaftliche, philosophische, psychologische und soziologische Grundlagen*. Wiesbaden: Springer Gabler. doi: 10.1007/978-3-658-31114-8_22
- Bendel, O. (2021): Die fünf Dimensionen sozialer Roboter: Der Versuch einer Systematisierung. S. 3–20 in: O. Bendel (Hrsg.), *Soziale Roboter. Technikwissenschaftliche, wirtschaftswissenschaftliche, philosophische, psychologische und soziologische Grundlagen*. Wiesbaden: Springer Gabler. doi: 10.1007/978-3-658-31114-8_1
- Bestgen, S. (2022): Sozialer Roboter in der ambulanten Suchtberatung (RoSA). Unveröffentlichter Studienbericht. Olten.
- Bielinski, L.L./Trimpop, L./Berger, T. (2021): Die Mischung macht's eben? Blended-Psychotherapie als Ansatz der Digitalisierung in der Psychotherapie. *Psychotherapeut* 66: 447–454. doi: 10.1007/s00278-021-00524-3

- Brisch, K.H. (2015): Bindung und Sucht. Stuttgart: Klett-Cotta.
- Costescu, C./Vanderborght, B./David, D. (2014): The effects of robot-enhanced psychotherapy: A meta-analysis. *Review of General Psychology* 18(2): 127–136. doi: 10.1037/gpr0000007
- DiClemente, C./Prochaska, J.O./Fairhurst, S./Velicer, W.F./Velasquez, M.M./Rossi, J.S. (1991): The process of smoking cessation: An analysis of precontemplation, contemplation, and preparation stages of change. *Journal of Consulting and Clinical Psychology* 59(2): 295–304. doi: 10.1037/0022-006X.59.2.295
- European Commission (2018): Annual activity report – DG communications networks, content and technology. Draft AAR 2018 (europa.eu), Zugriff 27.01.2024.
- Flujas-Contreras, J.M./García-Palacios, A./Gómez, I. (2023): Technology in psychology: Abibliometric analysis of technology-based interventions in clinical and health psychology. *Informatics for Health and Social Care* 48(1): 47–67. London: Taylor & Francis. doi: 10.1080/17538157.2022.2054343
- Grawe, K. (1998): Psychologische Therapie. Göttingen: Hogrefe.
- Henschel, A./Laban, G./Cross, E.S. (2021): What makes a robot social? A review of social robots from science fiction to a home or hospital near you. *Current Robotics Reports* 2(1): 9–19. doi: 10.1007/s43154-020-00035-0
- Khawaja, Z./Bélisle-Pipon, J.C. (2023): Your robot therapist is not your therapist: Understanding the role of ai-powered mental health chatbots. *Frontiers in Digital Health* 5: 1278186. doi: 10.3389/fdgth.2023.1278186
- Kwasnicka, D./Inauen, J./Nieuwenboom, W./Nurmi, J./Schneider, A./Short, C. E./Dekkers, T. A./Williams, J./Bierbauer, W./Haukkala, A./Picariello, F./Naughton, F. (2019): Challenges and solutions for N-of-1 design studies in health psychology. *Health Psychology Review* 13(2): 163–178. DOI:10.1080/17437199.2018.1564627
- Libin, A.V./Libin, E.V. (2004): Person-Robot interactions from the robotics psychologists' point of view: The robotic psychology and robot therapy approach. *Proceedings of the IEEE* 92(11): 1789–1803. doi: 10.1109/JPROC.2004.835366
- Prochaska, J.J./Vogel, E.A./Chieng, A./Baiocchi, M./Pajarito, S./Pirner, M. et al. (2023): A relational agent for treating substance use in adults: Protocol for a randomized controlled trial with a psychoeducational comparator. *Contemporary Clinical Trials* 127: 107125. doi: 10.1016/j.cct.2023.107125
- Schulze, H./Bendel, O./Schubert, M./Binswanger, M./Simmler, M./Reimer, R. et al. (2021): Soziale Roboter, Empathie und Emotionen. Eine Untersuchung aus interdisziplinärer Perspektive. Bern: TA-SWISS.
- Scoglio, A./Reilly, E./Gorman, J./Drebing, C. (2019): Use of social robots in mental health and well-being research: Systematic review. *Journal of Medical Internet Research* 21(7). doi: 10.2196/13322
- Silvera-Tawil, D. (2024): Robotics in healthcare: A survey. *SN Computer Science* 5(1): 189. doi: 10.1007/s42979-023-02551-0
- Young, J.E./Klosko, J.S./Weishaar, M.E. (2005): Schematherapie – ein praxisorientiertes Handbuch. Paderborn: Junfermann.
- Endnoten**
- ¹ Bild Moxie Robot. Zur Verfügung gestellt von Embodied, Inc., einem Unternehmen für Robotik und KI. Weitere Informationen zur Unternehmung: <https://embodied.com/>, Zugriff 12.03.2024.
- ² Vgl. Pressemappe Tiago von PAL Robotics: <https://pal-robotics.com/press>, Zugriff 07.03.2024.
- ³ Bild Roboterhund Aibo von Sony. Für weitere Informationen siehe: <https://us.aibo.com/>, Zugriff 07.03.2024.
- ⁴ Bild Lovot. Zur Verfügung gestellt von Groove X, Inc., einem Unternehmen, welches im Bereich Roboter-Menschen-Beziehungen tätig ist. Für weitere Informationen siehe Lovot, Love x Robot: <https://groove-x.com/en/>, Zugriff 12.03.2024.
- ⁵ Vgl. Pressemappe Sophia u. Little Sophia von Hanson Robotics: <https://t1p.de/hg4kd>, Zugriff 07.03.2024.
- ⁶ Bild Lio. Zur Verfügung gestellt von F&P RoboticsAG, Unternehmen für Roboterlösungen. Für weitere Informationen siehe Lio, der mobile Assistenzroboter: <https://www.fp-robotics.com/de/lio/>, Zugriff 12.03.2024.
- ⁷ Vgl. Pressemappe Soft Bank Robotics: <https://t1p.de/hg4kd>, Zugriff 07.03.2024.
- ⁸ Bild Cruzr. Zur Verfügung gestellt von UBTECH Robotics, Roboterhersteller. Für weitere Informationen siehe Cruzr, Begleitroboter: <https://t1p.de/fg4m6>, Zugriff 12.03.2024.
- ⁹ Vgl. Pressemappe von PARO Robots U.S., Inc.: <https://t1p.de/wwf5g>, Zugriff 07.04.2024.
- ¹⁰ Bild QTrobot. Zur Verfügung gestellt von LuxAI S.A.t, Hersteller Sozialer Roboter. Für weitere Informationen siehe QTrobot, Sozialer Roboter: <https://t1p.de/70pl3>, Zugriff 12.03.2024.
- ¹¹ Studien mit Single-Case-Design oder N-of-1-Studien sind quantitative Studien, die Vorhersagen innerhalb eines einzigen Untersuchungsobjekts (in diesem Fall: eines Patienten) auf der Grundlage wiederholter Messungen von Variablen im Zeitverlauf testen (vgl. Kwasnicka et al. 2019).



DIE SMOKING | CLOAK

Künstliches Mitgefühl und echte Abhängigkeiten: Gesellschaftliche Auswirkungen von KI

2024-2
Jg. 50
S. 36 - 39

Technologiefolgen-Abschätzung beobachtet Technikrends. Sie analysiert ihre Chancen und Risiken sowie mögliche erwünschte und nicht gewollte Folgen neuer Technologien. Damit wird der für eine Demokratie wesentliche kritische Umgang mit technischer Innovation unterstützt. Kein Wunder, schenkt die Technologiefolgen-Abschätzung den Entwicklungen auf dem Gebiet der Künstlichen Intelligenz (KI) ihre besondere Aufmerksamkeit: nützliches Werkzeug oder abhängigmachende Fessel?

ELISABETH EHRENSPERGER

Dr., Geschäftsführerin TA-SWISS, Stiftung für Technologiefolgen-Abschätzung, Brunngasse 36, CH-3011 Bern, www.ta-swiss.ch, <https://t1p.de/mujkh>

CHRISTINE D'ANNA-HUBER

Journalistin BR, Freie Mitarbeiterin TA-SWISS, Stiftung für Technologiefolgen-Abschätzung, Brunngasse 36, CH-3011 Bern, www.ta-swiss.ch

Einleitung

Es ist einer dieser Tage. Sie möchten Ihren Wecker am liebsten ermorden, fühlen sich unausgeschlafen, abgeschlagen und nicht bereit, den Alltag anzupacken. Klagen ist sinnlos, Ihre Familienmitglieder haben ihre eigenen Probleme. Doch zum Glück wartet im Badezimmer «BMind»¹ auf Sie, Ihr intelligenter Spiegel. Ein Blick, ein mürrisches «Morgen, Spiegel» genügen, und schon weiss «er» genau, wie Ihnen zumute ist, schafft mit sanfter Beleuchtung eine beruhigende Atmosphäre und schlägt zur Entspannung eine gemeinsame Meditation vor.

Es gibt sicher beeindruckendere KI-Anwendungen als den scheinbar mitfühlenden Badezimmerspiegel, vorgestellt an der diesjährigen Consumer Electronics Show (CES)² in Las Vegas, der grössten Technologiemesse der Welt. Und doch: Dieses Gadget, das anhand von Körperhaltung, Gesten und Stimme Gemütszustände identifizieren und adäquat darauf eingehen bzw. personalisierte Tipps für mehr Wohlbefinden geben kann, zeigt beispielhaft, warum die rasante Entwicklung im Bereich KI neben riesigen Erwartungen oft auch Unbehagen auslöst: Da ist etwas – kom-

plex, unsichtbar, undurchschaubar, das Computer und alle möglichen weiteren Gegenstände dazu befähigt, auf unsere ganz persönlichen Präferenzen zugeschnittene – und damit potenziell unverzichtbare – Dienstleistungen zu erbringen. Und das, um beim smarten Badezimmerspiegel und den mit sich selbst beschäftigten Familienmitgliedern zu bleiben, vielleicht sogar verlässlicher als die Menschen rund um uns.

Grundlagen für technologiepolitische Entscheide

Wo rund um neue Technologien gesellschaftliche Kontroversen entflammen, sieht TA-SWISS³ genauer hin. Die Schweizer Stiftung für Technologiefolgen-Abschätzung hat den gesetzlichen Auftrag, die möglichen Auswirkungen neuer Technologien frühzeitig zu untersuchen und deren Chancen und Risiken in wissenschaftlichen Studien aufzuzeigen und abzuwägen. Sie liefert damit Wissensgrundlagen für die in einer Demokratie unverzichtbare Debatte darüber, wie sich technologische Entwicklungen in eine gesellschaftlich wünschenswerte Richtung lenken lassen. Zumal Technologien immer ambivalent

sind. Sie besitzen, wie es der Philosoph Herbert Marcuse ausdrückte, einerseits die «befreiende Kraft», «Dinge zu instrumentalisieren», also für den Menschen nutzbar zu machen, können sich andererseits aber auch in eine «Fessel der Befreiung» verkehren und «zur Instrumentalisierung des Menschen» beitragen, also Abhängigkeiten schaffen (Grunwald 2010). Diese Ambivalenz ist auch der KI eigen und hängt paradoxerweise mit einer ihrer grossen Stärken zusammen: der Fähigkeit, Dienstleistungen auf Basis von grossen Mengen an Trainingsdaten an die Bedürfnisse bestimmter Zielgruppen anzupassen. Das kann eine Chance sein. In der Kinder- und Jugendhilfe oder in der Suchtprävention beispielsweise setzen verschiedene Unterstützungsangebote auf KI-Tools, um sicherzustellen, dass in der Online-Beratung passende Hilfsangebote offeriert werden (siehe dazu den Artikel von Lehmann oder Engelhardt in dieser Ausgabe). Gleichzeitig kommt die Personalisierung der Angebote unserer Bequemlichkeit entgegen und schafft Abhängigkeiten, denen wir uns nicht nur ungern hingeben.

Um mögliche gesellschaftliche Auswirkungen neuer Technologien abzu-

schätzen, ist es unabdingbar, hinter den «Hype» und hinter die Kontroversen zu schauen. Medial sind Extrempositionen oft am verbreitetsten, doch es braucht sorgfältige Analysen, die aufzeigen, was eine Technologie wirklich kann. Im Bereich der KI ist der Hype zurzeit beachtlich. Das Versprechen lautet, dass KI-Systeme – mit der menschlichen Intelligenz kombiniert – auf verschiedensten Gebieten atemberaubende Fortschritte ermöglichen und auch komplexere menschliche Tätigkeiten unterstützen oder sogar selbstständig ausführen werden. Dies bis hin zur Befürchtung, mit den immer leistungsstärkeren KI-Modellen könnte sich der alte Albtraum von der Maschine verwirklichen, die sich gegen ihre Schöpfer erhebt und diese schliesslich vernichtet.

Das Black-Box-Problem

Kein Zweifel: KI-basierte Systeme leisten viel. Das hängt mit ihrer Fähigkeit zusammen, in enormen Datenmengen statistische Muster und Korrelationen zu finden, davon Wahrscheinlichkeiten abzuleiten und so den Verlauf berechenbarer Dinge vorauszusagen; und zwar mit ihrem antrainierten enzyklopädischen Wissen und einer ungeheuren Rechenkraft schneller, präziser (und damit billiger) als wir Menschen es können. Mithilfe von KI können standardisierte Abläufe (z. B. das Management von Lieferketten) automatisiert, Industrieanlagen vorausschauend gewartet, digitale Psychogramme von Internetnutzenden erstellt, personalisierte Kaufempfehlungen abgegeben, Röntgenbilder treffsicher ausgewertet, biologische Simulationen auf Proteinebene durchgeführt und damit Krankheiten womöglich früh erkannt oder die Kreditwürdigkeit eines KMU beurteilt werden. Grosse Sprachmodelle wie ChatGPT (Ramelet 2023) – darauf trainiert, das nächste Wort in einem Textblock vorauszusagen bzw. zu errechnen – können natürliche Sprache plausibel selbst generieren und Analogien anwenden, um zu bestimmen, welche Antwort zu einer Frage oder Anweisung passt.

Die marktwirtschaftliche Sprengkraft, die in solchen KI-basierten Prozessoptimierungen schlummert, kurbelt den Hype weiter an. Das bedeutet aber

auch, dass KI oft einfach zum leeren Werbeversprechen wird und nicht überall KI drin ist, wo KI draufsteht: «Wenn es in Python geschrieben ist, dann ist es Machine Learning. In Powerpoint geschrieben ist es wahrscheinlich Künstliche Intelligenz», spottete Mat Velloso schon im Jahr 2018, damals technischer Berater der Geschäftsleitung von Microsoft.⁴

Für zahlreiche Missverständnisse sorgt die Bezeichnung «Künstliche Intelligenz», weil sie den Vergleich von datenbasierten Systemen mit der menschlichen Intelligenz nahelegt. Und für erhebliche Risiken sorgt der Umstand, dass der Grossteil der Menschen und selbst die KI-Konstrukteure nicht nachvollziehen können, wie diese datenbasierten Systeme zu ihren Resultaten gelangen – was die Illusion eines «intelligenten Gegenübers», das «logische Schlüsse zieht», letztlich weiter verstärkt. Er glaube nicht an die Gefahr, dass die KI die Menschheit ausrotten werde, gab der Berliner Hirnforscher John-Dylan Haynes vor Kurzem in der Wochenzeitung «Die Zeit» zu Protokoll, denke aber, «dass es uns nicht guttut, hochkomplexe und wirkmächtige technische Systeme einzusetzen, ohne deren Auswirkungen im Detail zu verstehen.» (Rietz 2023).

Haynes spricht an, was auch die KI-Studie von TA-SWISS (Christen et al. 2020) als einen besonders kritischen Punkt in der Entwicklung von KI-Systemen identifiziert: Algorithmen mutieren insbesondere in generativen KI-Systemen (d. h. KI-Systeme, die mithilfe von Deep-Learning-Modellen neue Inhalte wie Texte, Bilder oder Videos erstellen) von Hilfsmitteln zur Unterstützung von Entscheidungen zu nicht mehr durchschaubaren Entscheidungsautomatismen (ebd.). Manche Fachleute gehen davon aus, dass dieses Problem der Technologie inhärent und nicht zu lösen ist; der Mensch wird sich also auf eine Funktion der unwissenden Kontrolle beschränken müssen. Mit dem Black-Box-Problem hängt auch der Führungsstreit beim KI-Marktführer OpenAI zusammen. Denn beim Unternehmen, das ChatGPT entwickelt hat, gerieten zwei Auffassungen von KI heftig aneinander: KI als ein transparent und nachvollziehbar

zum Wohl der gesamten Menschheit entwickeltes System. Oder KI als eine möglichst rasch und mit kommerziellem Ansatz vorangetriebene und dem Spiel der Marktkräfte überlassene Innovation.⁵ Es stellt sich die Frage, ob dieser Richtungsstreit in Demokratien nicht grundsätzlich von den Stimmbürgerinnen und Stimmbürgern entschieden werden müsste.

Quer durch alle Lebensbereiche

Es dürfte also genau diese Entwicklung der Computertechnologie von einer in ihrer Funktion nachvollziehbaren Hilfestellung zur undurchschaubaren Black-Box sein, die in verschiedensten Bereichen zum gesellschaftlichen Disruptionspotenzial der KI beiträgt. Denn genau das, was die immense Anziehungskraft der KI ausmacht und zahlreiche Anwendungen ermöglicht, kann wegen der fehlenden Kontrollmöglichkeiten zum inhärenten Risiko werden. TA-SWISS hat KI als Grundlagentechnologie wie auch unterschiedliche Anwendungen von KI (soziale Roboter, synthetische Medien, Stimm-, Sprach- und Gesichtserkennung, Kunst und Kultur, Tod und Trauer, Demokratie) untersucht und in den letzten Jahren etliche Studien dazu publiziert. Dabei haben sich die nachfolgend aufgeführten Aspekte als zentrale Herausforderungen für unsere Gesellschaften herauskristallisiert:

- *Zersetztes Vertrauen und Manipulation*
In einer Studie,⁶ die demnächst erscheinen wird, hat TA-SWISS untersucht, wie sich die Verwendung von Deepfakes auf Politik, Medien, Wirtschaft und Rechtsprechung in der Schweiz auswirken könnte und wo Handlungsbedarf besteht. Deepfakes sind mithilfe von KI-Techniken synthetisierte oder manipulierte, aber authentisch erscheinende Foto-, Audio- oder Videoinhalte. Schon heute werden Deepfakes zur Diskreditierung von Personen und Institutionen, zur Desinformation, zur Manipulation von Wahlen, zur Destabilisierung politischer Prozesse oder zur Befeurung inner- oder zwischenstaatlicher Konflikte eingesetzt (Christen et al. 2020). Und da sie sich von originalen Inhalten kaum unterscheiden lassen, könnten Deepfakes, so die vielleicht

weitreichendste Befürchtung auf die Dauer, das Vertrauen in die Medien, die politischen Institutionen und letztlich die Demokratie selbst zersetzen. Dass ein solches Misstrauen in der Bevölkerung ansatzweise bereits vorhanden ist, zeigt eine im Rahmen der oben aufgeführten TA-SWISS-Studie durchgeführte repräsentative Befragung.

– *Verlust von Zwischenmenschlichkeit und ihre digitalen Substitute*

KI-Systeme, die sich im Kontakt mit Nutzenden stetig anpassen und schliesslich zum algorithmisch perfekten Gegenüber werden, simulieren – wie der bereits erwähnte Badezimmerspiegel – Empathie. Menschen neigen als soziale Wesen dazu, zu solchen Geräten eine Beziehung einzugehen (und sie als Helfer und Pfleger zu akzeptieren), – und zwar erst recht, wenn es sich dabei um menschenähnliche, «sozial» agierende Roboter handelt (Rey & Gonçalves 2022; Beitrag von Adam et al. in dieser Ausgabe). Das wird dann problematisch, wenn einsame oder sozial isolierte Personen die Beziehung zur Maschine, die immer perfekt auf die eigenen Vorlieben und Bedürfnisse eingeht, der vielleicht anstrengenderen Beziehung zu realen Menschen vorziehen. Oder schlicht nicht mehr fähig sind, sich auf einen Menschen einzulassen und sich in die KI verlieben oder im Cyber-Bordell⁷ suchen, was sie in der Wirklichkeit nicht finden können.

– Eine weitere Studie,⁸ die ebenfalls in Kürze von TA-SWISS veröffentlicht wird, untersucht, wie digitale Technologien den Umgang mit Tod, Sterblichkeit, Trauer und Bestattungskultur verändern und welche rechtlichen und ethischen Fragen sowie Fragen zum Datenschutz damit verbunden sind. Dabei geht es insbesondere um die Möglichkeit, Verstorbene im «digitalen Jenseits» «weiterleben» zu lassen – beispielsweise in Form eines KI-unterstützten Avatars, also eines digitalen Doppelgängers. Manchen Hinterbliebenen mag der wenigstens digital noch nicht abgerissene Kontakt dabei helfen, ihre Trauer besser zu verarbeiten. Doch gerade kommerziellen Angeboten auf dem Gebiet des

«Digital Afterlife» haftet das Risiko an, die emotionale Verletzlichkeit von Trauernden zu instrumentalisieren, um gezielt Abhängigkeiten zu schaffen und auszunützen: Eine ethisch besonders heikle Ausprägung des dank KI auf die User abgestimmten manipulativen Marketings.

– *Reproduzierte Vorurteile und Diskriminierung*

Automatisierte KI-Systeme übernehmen aus den Datensätzen, mit denen sie trainiert werden, auch Vorurteile und Stereotype. «Unfair» können auch die Algorithmen sein, die ihnen zugrunde liegen. Das kann dazu führen, dass beispielsweise Menschen aus ärmeren Wohngebenden eine ausgeprägtere Neigung zur Straffälligkeit unterstellt wird. Problematisch wird das insbesondere dann, wenn Entscheide, die Menschen betreffen – beispielsweise in Bewerbungsverfahren, bei der Vergabe von Bankkrediten und Versicherungen oder in strafrechtlichen Angelegenheiten – (teilweise) automatisiert und an solche Systeme ausgelagert werden (siehe Beitrag von Lehmann in dieser Ausgabe). Das Gleiche gilt für KI-Systeme, die zu Überwachungszwecken Stimmen, Gesichter und Sprache analysieren (Karabago et al. 2022). Selbst mit dem Versprechen der KI-Unternehmen, ihre Algorithmen «fair» zu kuratieren, ist noch nicht viel gewonnen – solange die Frage, was denn «fair» ist, nicht in der Gesellschaft und demokratisch ausgehandelt wurde. Zu Diskriminierungen kann es aber auch aufgrund technisch bedingter Fehler kommen bzw. zu falschen Outputs aufgrund eines Fehlers in der Software – zu Fehldiagnosen also, z. B. zur Verwechslung meines Gesichts mit demjenigen einer anderen Person oder zu einer falschen Antwort von ChatGPT mit all ihren Folgen.

– *Kompetenzverschiebungen*

KI-Tools wie ChatGPT werden bald fester Bestandteil von Alltag, Arbeitswelt und Hochschulen sein und sich massiv darauf auswirken, wie wir uns informieren, wie wir kommunizieren, arbeiten, lernen, denken und Wissen produzieren. TA-SWISS interessiert

sich dafür, welche Kompetenzen die möglichst optimale Nutzung grosser Sprachmodelle voraussetzt. Welche Fähigkeiten werden an Bedeutung gewinnen, wie müssen sich die Arbeitskräfte anpassen, um weiterhin relevant zu bleiben, wer gewinnt und wer verliert? Und muss die zunehmende Ausrichtung auf KI-Systeme zwangsläufig auf Rechnung von Kreativität, kritischem Denken und menschlicher Intuition gehen? Werden wir die Aufgaben, für die wir KI einsetzen, selbst gar nicht mehr erledigen können? Erste Antworten in diese Richtung gibt eine noch nicht publizierte Studie⁹ zu den Auswirkungen der Digitalisierung auf Musik, Theater und visuelles Design. Sie kommt zum Schluss, dass der Einsatz von KI-Tools zu einer Entkoppelung von Kunst und Handwerk führen dürfte: So wird es in gewissen musikalischen Genres nicht mehr nötig sein, ein Instrument zu beherrschen. Hingegen wird die Kompetenz, die von der KI vorgeschlagenen Arrangements zu beurteilen und auszuwählen, eine zunehmend wichtigere Rolle spielen.

Für einen sinnvollen Umgang mit KI

Auf der Basis der erarbeiteten sachlichen Informationen zu Technologiefolgen identifiziert TA-SWISS jeweils konkrete Handlungsoptionen für Verantwortliche in Politik, Wirtschaft, Forschung und Gesellschaft, um eine technologische Entwicklung in gesellschaftsverträgliche und demokratisch legitimierte Bahnen zu lenken. Auch im Fall der KI ist eine Balance zwischen einer Umarmung der grossartigen Innovationschancen und der Minimierung offensichtlicher Risiken nicht einfach zu finden. TA-SWISS hat aufgrund der Ergebnisse ihrer Studie (Christen et al. 2020) zur KI u. a. folgende Empfehlungen formuliert:

– *Kein generelles KI-Gesetz*

KI-Systeme kommen in so vielfältigen und unterschiedlichen Gebieten zum Einsatz, dass ein übergreifendes «KI-Gesetz» für eine einheitliche Regelung ihres Einsatzes nicht sinnvoll wäre. Vielmehr gilt es, die jeweils spezifischen Anwendungen auszu-leuchten und gezielt jene rechtlichen Fragen zu regeln, welche sich mit der

Verwendung von KI im jeweiligen Feld stellen. Gefragt ist also eine sektorielle Regulierung gemäss den Anforderungen eines spezifischen Anwendungsfeldes.

– *Regulierungskaskade*

Eine Regulierung der KI sollte risikoabwägend und kaskadenartig erfolgen, d. h. je höher das Risiko, desto strenger die Regulierung. Allenfalls wäre das Identifizieren von roten Linien vorzusehen bei Anwendungen, die in keiner Situation angebracht sind, wie z. B. die Erstellung von Persönlichkeitsprofilen zu Zwecken der Verhaltensmanipulation bzw. des Social Scoring oder auch die anlasslose Echtzeit-Gesichtserkennung im öffentlichen Raum, weil hierin das grosse Risiko einer Massenüberwachung besteht.

– *Mensch am Steuer*

Dort, wo über Personen entschieden wird (wie z. B. vor Gericht, bei Bewerbungsverfahren oder im Bildungs- und im Gesundheitswesen), sollen KI-Systeme nicht allein entscheiden dürfen, sondern immer ein Mensch das letzte Wort haben.

– *Ausweisungspflicht*

Der Einsatz von KI-Systemen muss immer transparent gemacht werden. Kundinnen und Kunden oder Bürgerinnen und Bürger müssen wissen, wann sie es mit KI zu tun haben. Auch Chatbots oder «soziale» Roboter dürfen nicht vortäuschen, Menschen oder menschenähnliche Wesen zu sein.

– *Erziehung zur digitalen Mündigkeit*

Die Eigenverantwortung der Bürgerinnen und Bürger ist zu stärken. Wichtig sind hier nicht zuletzt übergeordnete Medienkompetenzen im Umgang mit neuen Medien im Internet. Das heisst unter anderem die Befähigung zur Hinterfragung von Quellen oder Inhalten ebenso wie die Sensibilisierung für eine erhöhte Vorsicht beim Teilen von persönlichen Daten.

Intransparenz, Verlust von Vertrauen und von Mitmenschlichkeit, Vorurteile und Manipulation können den Boden für innere Leere und Abhängigkeiten legen. Gerade in der Verstärkung bestehender Vorurteile und Diskriminierung liegt im Zusammenhang mit einer Suchterkrankung oft ein Teufelskreis – insofern Diskriminierung die Ursache einer Suchterkrankung sein kann oder Suchterkrankung und suchtkranke Menschen an sich stigmatisiert werden. Die oben aufgeführten Anhaltspunkte für einen sinnvollen Umgang mit KI sind denn auch im Umgang mit Sucht von hoher Relevanz. Ganz generell legen Transparenz und Eigenverantwortlichkeit den Boden für ein selbstbestimmtes Leben und Handeln. Auf dass wir uns selbst im Spiegel noch erkennen, anstatt auf sein künstliches Mitgefühl zu setzen, in dessen Abhängigkeit sich unsere Gemütsverfassungen zu bewegen beginnen.

Literatur

Christen, M./Mader, C./Čas, J./Abou-Chadi, T./Bernstein, A. et al. (2020): Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz. TA-SWISS, Stiftung für Technologiefolgen-Abschätzung. Zollikon: vdf Hochschulverlag AG.

Grunwald, A. (2010): Technisierung als Bedingung und Gefährdung von Kultur – Eine dialektische Betrachtung. KIT Scientific Publishing 2010. <https://t1p.de/cydxm>, Zugriff 23.03.2024.

Karaboga, M./Frei, N./Ebberts, F. et al. (2022): Automatisierte Erkennung von Stimme, Sprache und Gesicht: Technische, rechtliche und gesellschaftliche Herausforderungen. TA-SWISS, Stiftung für Technologiefolgen-Abschätzung. Zollikon: vdf Hochschulverlag AG. <https://t1p.de/qxrua>, Zugriff 24.03.2024.

Ramelet, L. (2023): Themenpapier zu Chat GPT. TA-SWISS, Stiftung für Technologiefolgen-Abschätzung. <https://www.ta-swiss.ch/chatgpt>, Zugriff 23.03.2024.

Rey, L./Gonçalves, E. (2022): Roboter im Spiegel der Technologiefolgen-Abschätzung. Ergebnisse aus vertieften Auseinandersetzungen von TA-SWISS. Bern: TA-SWISS.

<https://t1p.de/nb457>, Zugriff 24.03.2024.

Rietz, Ch. (2023): Gibt es den freien Willen? Artikel in DIE ZEIT vom 10. Oktober 2023. <https://t1p.de/f63v8>, Zugriff 25.03.2024.

Endnoten

- ¹ Für weitere Informationen zum intelligenten Spiegel siehe Artikel auf der Website digitaltrends: <https://t1p.de/sx0kd>, Zugriff 28.03.2024
- ² Vgl. Website der CES: <https://www.ces.tech/>, Zugriff 25.03.2024.
- ³ Für weitere Informationen siehe Website der TA-SWISS: <https://www.ta-swiss.ch/>, Zugriff 25.03.2024.
- ⁴ Vgl. Twittereintrag von Mat Velloso vom 23. November 2018: <https://t1p.de/ge11o>, Zugriff 28.03.2024.
- ⁵ Für weitere Informationen zum Führungsstreit bei OpenAI siehe Artikel auf der Website der NZZ vom 23. November 2023: <https://t1p.de/dqgno>, Zugriff 28.03.2024.
- ⁶ Für weitere Informationen zur Studie siehe Website der TA-SWISS zu Deepfakes und manipulierte Realitäten: <https://t1p.de/zdn3v>, Zugriff 26.03.2024.
- ⁷ Vgl. bspw. Artikel im Berliner [030] Magazin über das Cybrothel in Friedrichshain in Berlin: <https://t1p.de/r733i>, Zugriff 26.03.2024.
- ⁸ Für weitere Informationen zur Studie siehe Website der TA-SWISS zu Der Tod im digitalen Zeitalter: <https://t1p.de/4dcky>, Zugriff 26.03.2024.
- ⁹ Für weitere Informationen zur Studie siehe Website der TA-SWISS zu Kultur und Digitalisierung. Studie zur Technologiefolgen-Abschätzung: <https://t1p.de/12w92>, Zugriff 26.03.2024.

