

Sekundärtraumatisierung:**Wir brauchen ein Subjekt als Gegenüber****Langfassung des Gesprächs aus PPP 1/2024**

Mathias Heinicke sprach mit Prof. Dr. Catrin Misselhorn, Philosophieprofessorin an der Universität Göttingen, über ethische Probleme und Fragestellungen zu KI-basierten Anwendungen im Zusammenhang mit Psychotherapie. Catrin Misselhorn ist Leiterin verschiedener wissenschaftlicher Projekte zur ethischen Bewertung von Robotern und KI-Systemen sowie Verfasserin einer Vielzahl von Publikationen zum Thema.

Mag. Rer. nat. Mathias Heinicke: Frau Professor Misselhorn, in Ihrem Buch zu Emotionalität von Künstlicher Intelligenz sprechen Sie von etwas, das Sie Verantwortungslücke nennen. Wenn man sich die aktuellen Lauterbach-Gesetze zur Datennutzung ansieht, in denen die Krankenkassen aufgefordert sind, Daten möglichst KI-automatisiert auszuwerten und dann auch schon konkrete Handlungs- und Behandlungsempfehlungen zu geben: Da ergeben sich solche Verantwortungslücken, wie Sie sie untersucht haben, oder?

Prof. Dr. Catrin Misselhorn: Was mich interessiert, ist das Menschenbild, das hinter so einem Vorgehen steht und die Auffassung darüber, worauf es bei der Psychotherapie ankommt. Die Fragen sind: Reicht es, dass ein psychisch erkrankter Mensch Gefühle auf ein gefühlloses Gegenüber projiziert, oder kommt es darauf an, dass eine Person in einer zwischenmenschlichen Beziehung Anerkennung für sich und ihr Leid erfährt? Kann eine App Behandlungsvorschläge, Reaktionen in einer nicht emotionalen Sprache formulieren oder ist es für den Behandlungserfolg notwendig, dass Apps Emotionen simulieren? Und wie geht man mit Behandlungsfehlern oder falschen Reaktionen um, die zum Beispiel Symptome verschlimmern? – Die Parallele mit dem Problem der Verantwortungslücke besteht darin, dass wir im Fall der Chatbots für Psychotherapie eine Anwendung haben, die sicherlich mit den besten Absichten programmiert wurde. Aber was passiert, wenn sie sich in einer konkreten Situation auf eine nicht angemessene Art und Weise verhält oder auf eine, die sogar zu schlimmen Konsequenzen führt? Dann können die Entwickler*innen so einer App natürlich sagen: Wir haben das nicht zu verantworten. Der Nutzer/die Nutzerin hat auch nichts falsch gemacht, und die Maschine kann auch keine Verantwortung übernehmen. Was geschieht also in so einem Fall?

Um nochmals auf die Frage nach dem Menschenbild zu kommen, für mich ist die Frage: Ist die Verwendung solcher Apps nicht letztlich ein Fall von technischem Solutionismus, also der naiven Annahme, dass es für jedes Problem einschließlich des der psychischen „Betriebsstörungen“ bei

Menschen ein Technological Fix gibt? Daran schließt sich die Frage an: Verhindert die Auslagerung psychotherapeutischer Behandlung unter Umständen die Auseinandersetzung mit gesellschaftlichen Ursachen für psychische Erkrankungen?

Ich formuliere all das deshalb als Fragen, weil es Dinge sind, die ich als Philosophin nicht alleine entscheiden kann, sondern dabei natürlich auf Ihre Erfahrungen und empirischen Befunde angewiesen wäre. Zugespitzt könnte man die folgende Hypothese formulieren: Psychische Erkrankungen könnten zumindest dadurch begünstigt werden, dass man von Menschen erwartet, sich in maschinenartig streng getaktete Abläufe einzupassen und eine mit einer Maschine vergleichbare perfekte Performance abzuliefern, während Emotionen, Sinnfragen, Kritik an solchen Abläufen als Störfaktoren möglichst auszuschalten sind. Wenn an dieser Hypothese etwas dran ist, könnte man sich weiter fragen, ob die Behandlung psychischer Störungen durch Maschinen nicht ein Mehr des Gleichen ist und die Situation sogar noch zuspitzen könnte.

MH: Beim Stichwort Performance-Optimierung kommt mir die Frage in den Sinn: Kann KI es denn tatsächlich leisten, durch Simulation eine menschliche Beziehung zu ersetzen?

CM: Aus meiner Sicht kann sie das eben nicht leisten. Wir brauchen – selbst dann, wenn der Therapeut oder die Therapeutin sich, wie im Fall der klassischen Psychoanalyse, sehr stark zurücknehmen muss – eine empathische Grundstimmung, die so eine Beziehung trägt. Ich würde sagen, selbst jenseits der Empathie im engeren Sinne (also, dass man Mitgefühl empfindet und ausdrückt) ist es wichtig, dass es jemanden gibt, der das Leiden zumindest als ein solches wahrnimmt. Wenn man diese Anerkennung durch ein anderes Subjekt wegnimmt, dann macht man eigentlich die behandelte Person ein Stück weit zu einem reinen Objekt, weil dahinter letztlich die Annahme steckt, die Person habe gar nicht wirklich einen Anspruch drauf, in ihrem Leiden von jemandem gehört zu werden, der sie auch wirklich verstehen kann.

Und hier liegt aus meiner Sicht die Gefahr einer Behandlung, zumindest wenn sie rein durch eine App erfolgen soll, die darauf setzt, einen Menschen „reparieren“ zu können, ohne auf dessen Leiden wirklich einzugehen und diese als solche wahrzunehmen. Mir scheint, da reicht eine Simulation nicht. Deswegen glaube ich, dass solche Apps ohne Begleitung eines menschlichen Therapeuten nicht eingesetzt werden sollten.

Manche sagen, dass man diese Apps vielleicht bei „harmloseren Störungen“, bei leichten Depressionen oder Phobien, einsetzen sollte. Da wäre für mich die Frage an Sie als Praktiker: Kann man das so klar unterscheiden? Oder ist es nicht auch möglich, dass etwas „leicht“ erscheint oder als eine Phobie, in Wirklichkeit aber Depression ist? Bei diesem Kontextverständnis, da sehe ich einfach Grenzen der KI.

Einerseits kann man sich diese Apps ja wie Lebensratgeber vorstellen: Ich sage, dass ich mich heute so traurig fühle und frage, was ich machen kann, und dann kommt vielleicht der Ratschlag: Geh doch mal spazieren, und das ist ja keine schlechte Idee. Aber: Kann man diese Ratschläge in einer neutralen Sprache formulieren, oder ist es nötig, dass eine „Psychotherapie-App“ immer auch eigene Emotion simuliert? Muss sie als jemand oder etwas erscheinen, das Emotionen hat und Mitgefühl empfindet? Wenn das stimmt, würde das natürlich bedeuten, dass solche Apps immer einen manipulativen Charakter haben, weil sie sehr subliminale Mechanismen der Anthropomorphisierung ansprechen. Man kann wissen, dass die App keine Gefühle hat und trotzdem sehr gefühlsmäßig

darauf reagieren, jenseits dieses Wissens: kognitive Dissonanz – ich sehe etwas und weiß, dass es nicht der Fall ist, aber die Wahrnehmung davon täuscht mich dennoch.

MH: Sie schreiben auch von einem Abstumpfungseffekt. Greift der eher, wenn ich eine App nicht auf dem Display habe, auf dem ich Texte mit ihr austausche, sondern einen niedlichen Roboter vor mir habe, durch den eine Verkörperlichung stattfindet? Dann würde das System doch funktionieren?

CM: Da stellt sich die Frage, wo man die Funktion ansiedelt. Erstens würde ich sagen, die Manipulation als solche ist schon problematisch, selbst wenn sie – im besten Fall – zugunsten der erkrankten Person stattfindet. Im nicht so guten Fall: Kostenfreie Apps funktionieren meist über Werbung oder aufgrund der Nutzung von Daten, die bei ihrer Anwendung erhoben werden. Bei beidem ist fraglich, ob das ethisch richtig ist, gerade bei Menschen, die in einer Notsituation und aufgrund ihrer psychischen Situation besonders vulnerabel sind. Es stellt sich auch die Frage: Geht es nur ums Wohlfühlen, um die Illusion, mir würde jemand zuhören? Wir hatten ja schon gesagt, dass es auf diese doch sehr individuelle Beziehung zwischen Therapeut und Patient ankommt.

MH: Im Bereich Diagnostik wird immer wieder allgemein behauptet: Testungen, Testauswertungen, Bildbeurteilungen, Befundbeurteilungen, das kann KI doch viel besser. Gibt es ethische Probleme, wenn man vor einer psychotherapeutischen Behandlung primär künstliche Intelligenz zur Diagnostik einsetzt?

CM: Ich sehe eine enge Verbindung zwischen Philosophie und Psychologie, insofern als psychische Störungen auch immer sehr viel mit dem Selbstverständnis von Personen zu tun haben. Mit den Fragen: Wie haben sie Lebenssituationen erlebt, welche Sinnfragen stellen sie sich? Emotionserkennungssoftware würde die Stimme, Gesichtsausdrücke, vielleicht auch sprachliche Äußerungen analysieren. Aber das ist natürlich extrem intrusiv, das heißt, wir dringen sehr stark in die Privatsphäre, ja in die Intimsphäre eines Menschen ein, der eine eigene Beschreibung einer Situation hat, die ja zu ihm gehört. Wenn man dem dann mit der maschinell erstellten Diagnose begegnet und sagt: „In Wirklichkeit geht es dir so und so“ ist die Frage: Ist das nicht wieder ein Mehr des Gleichen, da wir auch hier Menschen wie Maschinen behandeln und auf die besonderen Charakteristika von Subjektivität, von Selbstbestimmung gar nicht mehr eingehen. Aus meiner Sicht sind diese jedoch zentral, wenn es um die Psyche geht.

MH: Nun gibt es ja einen gesellschaftlichen Trend zur Selbstoptimierung. Jeder trackt sich heute irgendwie. Sehen Sie eine besondere Gefahr für bestimmte Personengruppen, abgehängt oder ausgegrenzt zu werden, mit all diesen Optimierungstrends, wenn diese durch KI noch verstärkt werden?

CM: Es liegt eine Gefahr im Self Tracking. Natürlich, wenn ein heranwachsendes Kind sich zum Beispiel dauernd fragt, ob es dünn genug ist, sieht man natürlich die bekannten Gefahren der Selbstoptimierung.

Das andere ist, wie wir bezogen auf vulnerable Personengruppen mit Apps wie Replika umgehen, die sich als Freunde anbieten, und mit dem Versprechen antreten, psychische Störungen, Depressionen und Einsamkeit behandeln zu können. Da haben wir es ja mit einem kommerziellen System zu tun, das bestimmte Funktionen, auch erotische, nur bei Bezahlung freigibt. Das ist eine Gefahr, gerade

weil diese Apps so angelegt sind, dass sie Bindung nicht nur befördern, sondern geradezu versuchen, sie auf manipulative Art zu erzwingen. Solche Mechanismen sind natürlich besonders gefährlich für vulnerable Gruppen, also für Kinder, aber auch für Personen, die sich in einer psychisch labilen Situation befinden.

MH: Gibt es denn auch aus ethischer Perspektive Vielversprechendes für eine künftige Nutzung oder Weiterentwicklung der KI? Rechner und Techniken werden schneller, die Datenmassen, die im Hintergrund stehen, werden immer größer...

CM: Ich finde es schade, dass die derzeit aufsehenerregendsten Anwendungen von KI häufig in Bereichen stattfinden, in denen wir sie aus meiner Sicht am wenigsten brauchen: im kreativen, sozialen, pflegerischen oder psychologischen Bereich. Dort geht es um Dinge, die Menschen gut und mit Freude erledigen können und die sie als wertvoll empfinden. Stattdessen würde ich mir wünschen, dass mehr Entwicklung im Bereich von solchen Problemen entwickelt würde, in denen wir über keine guten Lösungen verfügen.

MH: Aber warum kommt es eigentlich nicht dazu?

CM: Ich sehe drei Gründe. Der eine ist: Soziale Anwendungen sind insofern einfach zu programmieren, als sie die menschliche Neigung ausnutzen, Unbelebtes, das aussieht wie Menschen oder sich zumindest bis zu einem gewissen Grad so verhält, zu anthropomorphisieren. Das sieht man schon an dem ganz frühen Psychotherapie-Programm „Eliza“ aus den sechziger Jahren: Ein ganz simpler Mechanismus, und trotzdem fühlten sich die Menschen von diesem Programm teilweise sehr gut verstanden. Das sagt uns aber nicht, dass diese KI so super ist, sondern nur, dass offenbar die Anthropomorphisierungstendenz so stark ist, dass man sehr gut oder relativ einfach damit operieren kann.

Ein zweiter Grund ist, dass wir in diesem Bereich sehr viele Daten haben, die (scheinbar) frei zur Verfügung stehen. Viele dieser Anwendungen berücksichtigen die urheberrechtliche Seite nicht oder verletzen teilweise Persönlichkeitsrechte von Nutzer*innen. Das ist für bestimmte Akteure ein „Vorteil“. Natürlich ist dieses sogenannte Scraping problematisch.

Der dritte Grund ist, dass es Programmierer*innen natürlich reizt, genau die Leistungen zu programmieren, die scheinbar im Kernbereich menschlicher Kompetenzen liegen: soziale Interaktion, Emotionen, Kreativität. Ich glaube, dass das die Entwicklung von KI-Innovation in eine Richtung lenken kann, die nicht optimal ist – und dass wir mehr Anreize schaffen sollten, KI auf Probleme anzuwenden, bei denen die Menschen grundsätzliche Schwierigkeiten haben, sie zu lösen.

MH: Sich verstärkt auf die Kooperation zwischen Mensch und Maschine ausrichten, das haben Sie als Arbeitswelt 5.0 bezeichnet ...

CM: Genau. Mich würde noch aus meiner philosophisch-anthropologischen Perspektive interessieren, wie Sie das als Praktiker eigentlich sehen. Nutzen Sie KI in ihrer Therapie? Schlagen Sie Ihren Patient*innen KI-gestützte Apps vor? Wo würden Sie sinnvolle Anwendungsbereiche sehen?

MH: Die digitalen Gesundheitsanwendungen, die schon auf dem Markt sind, habe ich bisher ein einziges Mal an einen Patienten verordnet. Da ging es um eine Schlafstörung. Der Patient war dann aber überhaupt nicht zufrieden, weil die Nutzung mit so viel Texteingaben verbunden war.

Meine Partnerin ist Kinder- und Jugendlichenpsychotherapeutin, ihre Patient*innen sind da schon sehr viel stärker eingebunden. Diese Replika-App wird ganz stark genutzt, oft sogar als „Freundin“ bezeichnet. Hier sehe ich eine gewisse Gefahr in dem Moment, in dem man einem KI-System eine körperliche Hülle gibt, dass dann die Grenze des Menschlichen oder Manipulativen noch weiter überschritten wird. Es gibt natürlich Kritiker, die sagen, Psychotherapie per se, speziell die Verhaltenstherapie, sei auch in gewisser Weise manipulativ, weil wir Verhaltensänderungen herbeiführen. Aber dieses stringent Zielgerichtete ist in der direkten Psychotherapie nicht gegeben. Denn, wie schon erwähnt, ein Roboter hätte die zwischenmenschliche Komponente nicht, der würde immer gleich und stringent agieren. Aber im Bereich der Diagnostik könnte ich mir KI-basierte Anwendungen gut vorstellen, zum Beispiel beim Ausfüllen von Fragebögen.

CM: Gehört aber der Prozess, sich darüber Gedanken zu machen, welche Frage ich wie beantworte, nicht zu einem therapeutischen Prozess? Ich habe kürzlich Journalisten gefragt, wie sie KI für ihre Arbeit nutzen, und ob sie KI-Interviews transkribieren lassen. Sie haben mir gesagt, die Transkription gehöre für sie zum gedanklichen Prozess der Herstellung eines Texts. Das fand ich ganz plausibel. Genau deswegen könnte ich auch mein Schreiben nicht an einen Chatbot abgeben, weil ich erst im Schreiben meine Gedanken verfertige, eine Materie richtig durchdringe.

So stelle ich es mir auch mit einem Fragebogen in einer psychotherapeutischen Praxis vor: Dass jemand gezwungen wird, über sich nachzudenken, und dass es dabei vielleicht nicht nur auf das Endergebnis ankommt, sondern auch darauf, wie die Person dort hingekommen ist – dass sie sensibilisiert wird für gewisse Themen, Fragen oder Ansatzpunkte.

MH: Allerdings kann man bei vielen diagnostischen Fragebögen mit ja oder nein antworten oder einen Schweregrad ankreuzen, das Ganze möglichst spontan, denn so eine Diagnostik sollte immer eine Momentaufnahme sein. Dort könnte ich mir für manche Patient*innen-Gruppen schon eine Erleichterung vorstellen.

CM: Ich habe natürlich als Philosophin eine Sicht darauf, was Subjektivität ist, und auch ethische Grundsätze, von denen ich glaube, dass sie für ein Verständnis von Psychotherapie wichtig sind. Aber ich sehe, dass es auch ein fachliches Verständnis braucht, was aus dem Psychologiestudium oder Medizinstudium mit entsprechendem Schwerpunkt resultiert, und da muss ich natürlich aufpassen, dass ich diese Grenzen nicht überschreite. Denn ich kritisiere umgekehrt oft, wenn Naturwissenschaftler*innen Grenzen überschreiten und sich durch ihre wissenschaftliche Expertise auch als Ethikexpert*innen fühlen. Aber umgekehrt kann ich natürlich auch keine empirischen Hypothesen über die Wirksamkeit von Therapien oder ähnliches aufstellen. Ich hoffe, dass ich in meinen Äußerungen diese Grenze gut getroffen habe.

MH: Die haben Sie sehr gut getroffen. Vielen Dank für das Gespräch!

Folgende Auswahl an Publikationen von Prof. Dr. Misselhorn möchten wir Ihnen besonders empfehlen:

- Künstliche Intelligenz – das Ende der Kunst? (Reclam-Verlag 2023)

- Künstliche Intelligenz und Empathie. Vom Leben mit Emotionserkennung, Sexrobotern & Co (Reclam-Verlag 2021, 3. Aufl. 2024). [Gelb markierten Teil evtl. streichen]
- Grundfragen der Maschinenethik (Reclam-Verlag 2018, 5. Auflage 2022, 3. Platz auf der Sachbuchbestenliste von ZEIT, ZDF und Deutschlandfunk Kultur)