

Menschen, Maschinen, Mentalisieren

In diesem Beitrag möchte ich mein eigenes Verhältnis zur Künstlichen Intelligenz (KI) reflektieren. Der Ausdruck legt nahe, dass menschliches Denken, menschliche Intelligenz mit künstlichen (technischen) Mitteln mehr oder minder nachgeahmt werden kann. So zutreffend das Adjektiv *künstlich* sein mag, so fraglich ist die Berechtigung, von Maschinen erzeugte Texte auf eine eigene *Intelligenz* zurückzuführen. Der vorliegende Beitrag beschäftigt sich deshalb mit der Frage, ob mentale Ausdrücke wie denken, fühlen, spüren, lernen usw. zur Beschreibung von Maschinen verwendet werden sollten oder ob dabei die übertragende (metaphorische) Redeweise deutlich zu markieren ist. Zugleich führe ich in diesem Beitrag ein Experiment durch: Ich befrage einen KI-gestützten Chatbot und setze mich mit dessen Antworten auseinander.

1. Anthropologie: Suche nach dem Wesen des Menschen

Die Frage nach dem unterscheidend Menschlichen wird oft als Frage nach dem Unterschied gegenüber benachbarten Entitäten gestellt, also nach dem Unterschied zwischen Mensch und Tier, Mensch und Gott sowie hier: zwischen Mensch und Maschine. Mit Lindemann (2021) können wir das folgende anthropologische Quadrat formulieren:

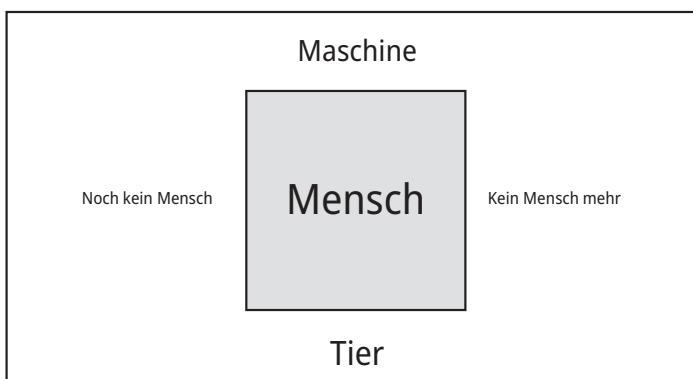


Abb. 1: Anthropologisches Quadrat nach Lindemann (2021)

Was die Abgrenzung des Menschen gegenüber den vier Randgebieten angeht, so werden anthropologische Alleinstellungsmerkmale oft dadurch getestet, dass sie versuchsweise auf die „Nachbarn“ angewandt werden. So wurde das Problem der menschlichen Theory of Mind und deren (mentalisierte) Entwicklung maßgeblich durch die Untersuchung der Theory of Mind von Schimpansen (Premack & Woodruff, 1978) beeinflusst. Auch die Ontogenese des Menschen durch Schöpfung und Zeugung sowie sein materielles und spirituelles Vergehen im Tod beleuchten seine lebendige Leiblichkeit. Ebenso kann die mehr oder minder trennscharfe Abgrenzung von der Maschine zur Aufklärung des menschlichen Wesens beitragen (Frick, 2015).

2. Mentalisieren: eine wesentliche Fähigkeit des Menschen

Die tierische Theory of Mind wurde durch Beobachtung und durch gezielte Versuchsanordnungen erforscht. Von Computern weiß man, dass sie gegen begabte Menschen im Schach- und im Go-Spiel gewinnen. Kann man ihre kombinatorischen und algorithmischen Fähigkeiten als Fühlen und Denken auffassen? Dank der KI können wir Maschinen darauf hin befragen, ob sie über Mentalisierungsfähigkeiten verfügen. Ich stelle der KI deshalb eine erste Frage:

Erkläre im Kontext der Bindungstheorie, was Mentalisieren ist

Im Kontext der Bindungstheorie bezieht sich der Begriff „Mentalisieren“ auf die Fähigkeit einer Person, die eigenen Gedanken, Gefühle, Motive und Absichten sowie die Gedanken und Gefühle anderer Menschen zu erkennen, zu verstehen und zu interpretieren. Mentalisieren beinhaltet also die Fähigkeit, die innere Welt einer Person und die Beziehungen zu anderen Menschen zu erfassen und zu reflektieren.

Die Bindungstheorie, entwickelt von John Bowlby, betont die Bedeutung von sicheren und stabilen Beziehungen für die emotionale und psychologische Entwicklung einer Person. Frühe Beziehungserfahrungen, insbesondere die Beziehung zwischen einem Kind und seinen primären Bezugspersonen, prägen die Fähigkeit eines Individuums zur mentalen Repräsentation von Beziehungen.

Mentalisieren umfasst die Fähigkeit, sich selbst und andere als individuelle Wesen mit eigenen Gedanken und Gefühlen wahrzunehmen. Es ermöglicht einer Person, die Beweggründe ihres eigenen Verhaltens und das Verhalten anderer Menschen zu verstehen. Durch die Fähigkeit zur Mentalisierung können wir die Handlungen und Reaktionen anderer Menschen besser einschätzen und unsere eigenen Handlungen in Bezug auf die Auswirkungen auf andere reflektieren.

Ein Mangel an Mentalisierungsfähigkeiten kann zu Schwierigkeiten in zwischenmenschlichen Beziehungen führen. Zum Beispiel können Menschen mit einer eingeschränkten Mentalisierungsfähigkeit dazu neigen, andere Menschen misszuver-

stehen oder sich weniger in andere hineinsetzen zu können. Dies kann zu Konflikten, Kommunikationsproblemen und Schwierigkeiten beim Aufbau und der Aufrechterhaltung enger Bindungen führen.

Die Förderung der Mentalisierungsfähigkeit ist ein wichtiger Aspekt in der Therapie und Unterstützung von Menschen, insbesondere bei psychischen Störungen, bei denen zwischenmenschliche Probleme im Vordergrund stehen. Durch die Stärkung der Mentalisierungsfähigkeit kann eine Person ihre Beziehungen verbessern und besser mit den Emotionen und Bedürfnissen sowohl von sich selbst als auch von anderen umgehen.

<https://chat.openai.com/share/21d951b6-30ed-4928-af2f-edb5d8dc408f>
(Zugriff 28.06.2023)

ChatGPT hat zutreffend zusammengefasst, was wir ausgehend von der Bindungstheorie (Bowlby, 1970/1975) unter Mentalisieren (Fonagy et al., 2002/2004) verstehen. Die Theorie des Mentalisierens beruht auf Überlegungen des Psychoanalytikers W. R. Bion. Bion fragt sich, wie in der frühen Mutter-Kind-Beziehung Erfahrungselemente und emotionelle Eigenschaften vorsprachlich verarbeitet werden. Er nimmt phylogenetische Schemata (angeborene Präkonzeptionen) an, mit denen der Säugling die Welt auffassen kann, noch bevor er in der Lage ist zu denken. Bion (1962/1963) zufolge entsteht ein Gedanke dadurch, dass eine Präkonzeption mit einer Versagung zusammentrifft: „Das Modell, das ich vorschlage, ist das eines Säuglings, dessen Erwartung einer Brust gepaart ist mit dem Realerlebnis, daß keine Brust, die ihn befriedigt, vorhanden ist. Diese Paarung wird innerlich als ‚keine Brust‘, als ‚abwesende Brust‘ erlebt“ (Bion, 1962/1963, S. 427).

„No breast – therefore a thought“: Bion betont, dass Denken aus der Abwesenheit der Mutter (oder, im übertragenen Sinn: einer anderen menschlichen Bezugsperson) entsteht. Denken kann die Abwesenheit der Mutter bis zu deren Wiedererscheinen überbrücken. Sinnlose, traumatische Wahrnehmungen des Säuglings nennt Bion „Beta-Elemente“, für deren Umwandlung das Denken der „Alpha-Funktion“ bedürfe, die dem Säugling erst rudimentär zur Verfügung steht: Er muss die Beta-Elemente „in die Mutter ausscheiden“, damit sie die Beta-Elemente für ihn in Alpha-Elemente verwandelt. Das entgiftende mütterliche Containing stellt ein Modell für die Entwicklung des Denkens und für dessen Gefährdungen dar, z.B. wenn Kinder vernachlässigt, sexuell missbraucht oder wie Maschinen behandelt werden, die weder denken noch fühlen können. In derartigen traumatischen Entwicklungsbedingungen werden Kinder nicht mentalisiert und lernen es nicht, zu mentalisieren, können andere Menschen und Lebewesen nicht als denkende und fühlende Personen be-

handeln. Mentalisierende psychotherapeutische Interventionen stellen ein Containing bereit, innerhalb dessen traumatische Erfahrungen entgiftet, d.h. in fühlendes Denken und in Erzählungen umgewandelt werden.

Die Mensch-Maschine-Differenzierung wird in der kindlichen Entwicklung im Umgang mit technischen Apparaten, Spielzeug, Tieren und Menschen erlernt. Dazu gehören die Konzepte von Leben vs. Tod und von Funktionieren vs. Kaputtsein. Im Spiel erprobt das Kind den psychischen Äquivalenzmodus (die Puppe *ist* ein Kind, das versorgt werden muss, der Teddybär *ist* ein lebendiges Gegenüber, mit dem ein Gespräch möglich ist, das Gespenst aus dem Vorlesebuch *ist* jetzt unter dem Bett und muss mit dem Monsterspray vertrieben werden usw.). Gleichzeitig erlernt es den Als-Ob-Modus, d.h. es kann die gespielte von der äußeren Realität unterscheiden und damit auch unbelebte von belebten Objekten.

Etwa ab dem vierten Lebensjahr können der Äquivalenz- und der Als-Ob-Modus innerhalb des Reflexiven Modus integriert werden: Dieser lässt sowohl Spiel, Kreativität, Fantasie zu ihrem Recht kommen als auch die Realitätsprüfung und mit ihr die Mensch-Maschine-Unterscheidung. Wenn das Kind den Computer „schimpft“ und einen nicht funktionierenden Apparat „zur Strafe“ tritt, so weiß es gleichzeitig, dass diese Objekte weder Schmerz empfinden noch emotionale Reaktionen zeigen – im Unterschied zu Menschen und Tieren.

Die Mensch-Maschine-Unterscheidung gehört zu unserem Wissensvorrat, sodass ich von der KI eine zutreffende Antwort erhalte:

Erkläre den Unterschied zwischen einer Maschine und einem Menschen, der denkt und fühlt

Ein grundlegender Unterschied zwischen einer Maschine und einem Menschen besteht darin, dass Menschen über einen komplexen biologischen Organismus verfügen, während Maschinen künstliche Konstrukte sind, die von Menschen geschaffen wurden.

Denken und Fühlen sind Eigenschaften, die oft als charakteristisch für Menschen betrachtet werden. Menschen sind in der Lage, Informationen wahrzunehmen, zu verarbeiten, zu interpretieren und darauf zu reagieren. Sie können abstrakte Konzepte verstehen, logisch denken, Probleme lösen und Entscheidungen treffen. Menschen haben auch die Fähigkeit zu fühlen, Emotionen zu empfinden und Erfahrungen subjektiv zu erleben.

Maschinen hingegen, wie beispielsweise Computer oder Roboter, funktionieren auf Grundlage von Algorithmen und programmierten Anweisungen. Sie können Daten aufnehmen, verarbeiten und auf bestimmte Weise darauf reagieren, aber sie verfügen nicht über die Fähigkeit, abstrakte Konzepte zu verstehen oder Emotionen zu empfinden. Maschinen können bestimmte Aufgaben sehr effizient ausführen und kom-

plexere Berechnungen durchführen, aber sie haben kein Bewusstsein oder ein eigenes subjektives Erleben.

Ein weiterer wichtiger Unterschied ist, dass Menschen eine individuelle Identität, eine einzigartige Persönlichkeit und eine entwickelte moralische und ethische Vorstellung von richtig und falsch haben. Menschen haben auch eine biologische Verkörperung mit all ihren körperlichen Bedürfnissen und Begrenzungen. Maschinen hingegen haben keine individuelle Identität oder Persönlichkeit, und sie haben keine eigenen moralischen oder ethischen Prinzipien. Sie können nur so funktionieren, wie sie programmiert wurden.

Es ist wichtig anzumerken, dass die Technologie ständig Fortschritte macht, und es gibt Bereiche wie die Künstliche Intelligenz, in denen Maschinen in der Lage sind, menschenähnliche Aufgaben auszuführen. Dennoch bleibt der grundlegende Unterschied zwischen einer Maschine und einem Menschen bestehen, da Maschinen letztendlich auf vordefinierten Algorithmen und Regeln basieren, während Menschen über ein individuelles Bewusstsein und eine persönliche Erfahrung verfügen.
<https://chat.openai.com/share/cedbaa6d-0729-4991-a522-f06167f05cdf>
(Zugriff 28.06.2023)

Diese wiederum zutreffende Auskunft beruht auf dem erwähnten Programmieren und auf den vielfältigen gespeicherten Informationen. Sie ist weder in kognitiver noch in emotionaler Hinsicht die Frucht individuellen und vernetzten Lernens. Ich frage deshalb nach, ob die Anwendung des Verbums *lernen* auf die KI legitim ist:

„Eine Maschine lernt“: ist diese Redeweise ein Kategorienfehler?

Die Redeweise „Eine Maschine lernt“ ist kein[e] Kategorienfehler, sondern eine metaphorische Beschreibung von Maschinen, die in der Lage sind, durch Algorithmen und Datenanalyse neue Informationen zu verarbeiten und ihr Verhalten anzupassen. Es handelt sich dabei um eine verbreitete Art der Sprachverwendung, um die Fähigkeiten von künstlicher Intelligenz (KI) oder maschinellem Lernen zu beschreiben.

In der Regel verwenden wir den Begriff „Lernen“ im Zusammenhang mit Lebewesen, insbesondere Menschen. Wir denken normalerweise an einen bewussten, kognitiven Prozess, der auf Erfahrung und Wissen basiert. Bei Maschinen handelt es sich jedoch um komplexe Systeme, die in der Lage sind, aus Daten zu lernen und ihre Leistung im Laufe der Zeit zu verbessern, ohne ein Bewusstsein oder Verständnis im menschlichen Sinne zu haben.

Die metaphorische Verwendung des Ausdrucks „Eine Maschine lernt“ ist daher eine praktische Möglichkeit, um [den] Funktionsweise und die Fähigkeiten von Maschinen zu beschreiben, die in der Lage sind, Muster in Daten zu erkennen, Vorhersagen zu treffen und Entscheidungen zu treffen. Es ist wichtig zu beachten, dass das Lernen von Maschinen auf algorithmischen Prozessen basiert und sich grundlegend von menschlichem Lernen unterscheidet.

<https://chat.openai.com/share/5c45760b-5df0-4ad9-bef4-184b09c86f12>
(Zugriff 18.06.2023)

Im Gegensatz zu dieser KI-Antwort möchte ich fragen, unter welchen Bedingungen die „praktische“ Metaphorik der lernenden und damit intelligenten Maschine zu einem Kategorienfehler führt. Dies ist m.E. dann der Fall, wenn der metaphorische Charakter der hier diskutierten Lern-Attribution unbewusst wird. Freilich reden wir häufig in Metaphern, ohne uns dessen bewusst zu sein. Überdies: viele unserer mentalen Ausdrücke sind leibliche Metaphern, Übertragungen von eigenleiblichem Spüren, um Seelisches auszudrücken, z.B. Niedergeschlagenheit/ Depression, Enge/Angst, Hochstimmung. All diese Ausdrücke stammen aus der Leibssphäre, auch wenn sie auf den ersten Blick „seelisch klingen“ (Lakoff & Johnson, 1980/2008).

Bewusstsein, Selbst-Bewusstsein und Reflexionsfähigkeit müssen nicht bei jedem einzelnen Akt des Denkens, Fühlens oder Handelns ausdrücklich gegeben sein. Philosophisch gesprochen, handelt es sich um transzendentale Begriffe, d.h. sie müssen in jedem einzelnen menschlichen Akt vorausgesetzt werden, auch wenn sie nicht ausdrücklich erwähnt werden.

Ein möglicher Einwand gegen dieses Argument lautet: Jetzt ist dies für Maschinen noch nicht möglich, aber grundsätzlich und transzental schon. Es ist eben noch nicht ausdrücklich möglich, ähnlich wie bei unbewusst vorausgesetzten menschlichen Begriffen liegt es auch in der Potenzialität der Maschinen, eines Tages bewusst zu sein. Deshalb frage ich ChatGPT, ob Maschinen mentalisieren können:

Können Maschinen mentalisieren?

Maschinen, insbesondere künstliche Intelligenz (KI)-Systeme, können bestimmte Formen des mentalen Modellierens oder der Simulation von Gedanken und mentalen Zuständen erreichen. Dies ist jedoch noch weit entfernt von dem, was Menschen als „echte“ Mentalisierung oder Theory of Mind bezeichnen würden.

Mentalisierung bezieht sich auf die Fähigkeit, die Gedanken, Überzeugungen, Absichten und Emotionen anderer Menschen zu erkennen und zu verstehen, um ihr Verhalten besser vorherzusagen und zu interpretieren. Diese Fähigkeit ist ein wichtiger Bestandteil der sozialen Interaktion und des zwischenmenschlichen Verständnisses.

Obwohl KI-Systeme in der Lage sind, menschliches Verhalten auf der Grundlage von Mustern und Daten zu analysieren und vorherzusagen, haben sie keine eigene innere Erlebniswelt oder Bewusstsein. Sie können zwar komplexe Berechnungen durchführen und bestimmte Formen des maschinellen Lernens verwenden, um menschliches Verhalten zu simulieren, ihnen fehlt jedoch das Verständnis für die mentalen Zustände hinter diesem Verhalten.

Es gibt Forschungsansätze, die versuchen, KI-Systeme mit einer gewissen Form von Mentalisierung auszustatten, aber wir sind noch weit davon entfernt, Maschinen eine vollständige Theory of Mind zu verleihen, wie es beim Menschen der Fall ist. Es bleibt eine offene Frage, ob es jemals möglich sein wird, KI-Systeme zu ent-

wickeln, die über das bloße Nachahmen von menschlichem Verhalten hinausgehen und tatsächlich über eine eigene Bewusstseinserfahrung verfügen können.
<https://chat.openai.com/share/c83144bf-6505-4509-baf8-49f51e071e4b>
(Zugriff 18.06.2023)

Diese KI-Antwort legt einen eliminativen Materialismus nahe: Das harte Problem des Bewusstseins werde eines Tages durch Reduktion auf dessen neuronale Grundlagen gelöst, und diese neuronalen Grundlagen werden vollständig maschinell repliziert werden.

Was ist eliminativer Materialismus?

Der eliminative Materialismus ist eine philosophische Position, die besagt, dass bestimmte mentale Zustände und Begriffe, die in der Alltagssprache verwendet werden, letztendlich nicht real sind und daher durch neuere wissenschaftliche Erkenntnisse über das Gehirn und das Nervensystem ersetzt oder „eliminiert“ werden sollten.

Die Anhänger des eliminativen Materialismus argumentieren, dass unser gegenwärtiges Verständnis von mentalen Zuständen, wie zum Beispiel Schmerz, Gedanken oder Absichten, letztendlich auf irreführenden oder fehlerhaften Konzepten beruht. Sie behaupten, dass diese Begriffe keine objektiven Entsprechungen in der physischen Welt haben und dass zukünftige Fortschritte in der Neurowissenschaft und Kognitionswissenschaft letztendlich zu ihrer Eliminierung führen werden.

Der eliminative Materialismus steht im Gegensatz zum Reduktionismus, der davon ausgeht, dass mentale Zustände auf komplexe neurophysiologische Prozesse reduzierbar sind. Anhänger des eliminativen Materialismus argumentieren hingegen, dass unser gegenwärtiges mentales Vokabular und unsere Konzepte unzureichend sind, um die Komplexität des Gehirns angemessen zu beschreiben, und dass wir neue Begriffe und Theorien entwickeln müssen, um das Phänomen des Geistes angemessen zu erfassen.

Es ist wichtig anzumerken, dass der eliminative Materialismus eine kontroverse Position ist und von vielen Philosophen und Wissenschaftlern kritisiert wird. Einige Kritiker argumentieren, dass der eliminative Materialismus zu einem radikalen Formen des Reduktionismus führt und das menschliche Erleben und Bewusstsein vernachlässigt. Andere behaupten, dass er die komplexe Natur des menschlichen Geistes nicht angemessen erklärt. Es gibt also verschiedene Ansichten und Debatten zu diesem Thema in der Philosophie des Geistes.

<https://chat.openai.com/share/216ea3cd-26b5-45b0-a2f9-9cb2301da13a>
(Zugriff 18.06.2023)

3. Mentalisierungsorientierte KI-Anwendungen

Abgesehen von diesen wichtigen Vorüberlegungen, um weder in einen Kategorienfehler noch in einen kruden Materialismus zu geraten: Worin liegt ein mögliches therapeutisches Potenzial der KI in der mentalisierungsbasierten Psychotherapie?

Kann KI hilfreich in der mentalisierungsbasierten Therapie sein?

Ja, künstliche Intelligenz (KI) kann in der mentalisierungsbasierten Therapie (MBT) hilfreich sein. MBT ist eine Form der Psychotherapie, die darauf abzielt, die Fähigkeit zur Mentalisierung zu verbessern, das heißt, die Fähigkeit, die eigenen Gedanken, Gefühle und Absichten sowie die anderer Menschen zu erkennen und zu verstehen.

In der heutigen technologisch fortgeschrittenen Welt gibt es verschiedene Möglichkeiten, KI in Therapieansätzen einzusetzen, einschließlich MBT. Hier sind einige Wege, wie KI in der mentalisierungsbasierten Therapie unterstützend wirken kann:

1. *Psychoedukation: KI kann als Informationsquelle dienen, um den Therapeuten und Patienten Zugang zu relevanten Ressourcen, Forschungsergebnissen und Informationen über psychische Gesundheit zu ermöglichen. Dies kann helfen, das Verständnis für mentalisierungsbasierte Konzepte zu vertiefen und den Lernprozess zu erleichtern.*
2. *KI-gestützte Tools für Selbstreflexion: KI kann dazu verwendet werden, Selbstreflexionstools zu entwickeln, die den Patienten dabei helfen, ihre eigenen mentalen Zustände, Motivationen und Reaktionen besser zu verstehen. Durch den Einsatz von Text- oder Spracherkennungstechnologien können Patienten ihre Gedanken und Gefühle in Tagebüchern oder Tagebuch-Apps festhalten, während KI-Anwendungen dabei helfen können, Muster und Zusammenhänge in den Daten zu erkennen.*
3. *Emotionserkennung und -regulierung: KI kann bei der Erkennung und Regulierung von Emotionen helfen. Zum Beispiel können Gesichtserkennungsalgorithmen eingesetzt werden, um Emotionen in Gesichtsausdrücken zu erkennen. Dies kann sowohl dem Therapeuten als auch dem Patienten dabei helfen, Emotionsmuster zu identifizieren und an der emotionalen Regulation zu arbeiten.*
4. *Virtuelle Therapeuten: KI-gestützte virtuelle Therapeuten können als Ergänzung zur menschlichen Therapie eingesetzt werden. Diese virtuellen Therapeuten können in der Lage sein, Gespräche zu führen, Ratschläge zu geben und die mentalisierungsbasierte Arbeit zu unterstützen. Sie können auch in der Lage sein, kontinuierliches Feedback zu geben und den Fortschritt des Patienten im Laufe der Zeit zu verfolgen.*

Es ist wichtig anzumerken, dass KI in der mentalisierungsbasierten Therapie nicht den menschlichen Therapeuten ersetzt, sondern als unterstützendes Instrument eingesetzt wird. Die Interaktion und Beziehung zwischen Therapeuten und Patienten bleiben weiterhin von zentraler Bedeutung. KI kann jedoch die Effektivität der Therapie verbessern, den Zugang zu Behandlungsmöglichkeiten erweitern und den Lernprozess für Patienten erleichtern.

<https://chat.openai.com/share/0af0d10a-6ca9-492f-b8c4-97b573b61bc1>
(Zugriff 18.06.2023)

Es ist häufig zu beobachten, dass ChatGPT ausgleichende, harmonisierende Antworten gibt, wenn kontroverse Standpunkte im Diskurs vorkommen. Das gilt auch hier: Einerseits werden differenzierte Möglichkeiten einer Mentalisierungs-Assistenz aufgelistet; andererseits wird

betont, dass die KI weder den menschlichen Therapeuten noch dessen Interaktion mit dem Patienten ersetzen kann.

4. Ist die KI kryptoreligiös?

Thomas Fuchs (2023) zufolge konnte sich der prämoderne Mensch „als Kind und Ebenbild Gottes, ja als Krone der Schöpfung noch seines Wertes gewiss sein und sich mit dem göttlichen Funken seines Geistes, seiner Vernunft und Freiheit über andere Lebewesen erheben“. Im Sinne von Heinz Kohut konnte er sich im Glanz des göttlichen Auges spiegeln. Im Angesicht des Rechners, gewissermaßen im Blickfeld der Computer-Kamera formuliert Fuchs:

Doch worin spiegelt sich der Mensch nach dem Tod Gottes, ohne einen Blick, der noch mit Wohlgefallen auf ihm ruht? Meine These lautet: Der neue Spiegel ist die intelligente und am Ende bewusste Maschine, die der Mensch in gottgleicher Weise selbst zu erschaffen trachtet und die zugleich seine metaphysische Einsamkeit aufheben soll. Hier vermischt sich das prometheische Motiv des Homo Deus, der gottgleichen Schöpferkraft, mit der Suche nach einem Gegenüber in der Einsamkeit des Kosmos, nach dem metaphysischen Du, in dem sich der von Gott verlassene Mensch wieder spiegeln kann (Fuchs, 2023, S. 112).

ChatGPT dagegen antwortet:

Ist KI kryptoreligiös?

Nein, Künstliche Intelligenz (KI) ist nicht kryptoreligiös. Kryptoreligion ist ein Begriff, der verwendet wird, um eine Form des Glaubens oder der Spiritualität zu beschreiben, die sich auf die Idee konzentriert, dass eine höhere Macht oder ein höheres Bewusstsein in kryptografischen Systemen oder in der KI existiert. Es ist eine spekulative Idee, die in einigen Kreisen diskutiert wird, aber sie hat keine wissenschaftliche Grundlage.

KI ist ein Bereich der Informatik, der sich mit der Entwicklung von Algorithmen und Systemen beschäftigt, die in der Lage sind, menschenähnliche Aufgaben zu erfüllen, wie zum Beispiel Spracherkennung, Bildverarbeitung oder Entscheidungsfindung. KI basiert auf mathematischen Modellen, Datenanalyse und maschinellem Lernen, um Probleme zu lösen.

Obwohl es einige philosophische Diskussionen über die Möglichkeit einer KI mit Bewusstsein oder einer KI, die eine Form von Spiritualität entwickelt, gibt, sind dies spekulative Fragen, die noch nicht abschließend beantwortet werden können. Der Großteil der KI-Forschung und -Anwendung konzentriert sich auf praktische Anwendungen und technische Probleme, anstatt auf religiöse oder metaphysische Aspekte.

<https://chat.openai.com/share/4cfddb8e-0dbd-48c5-9011-70b2911f5e2a> (Zugriff 19.06.2023)

Fuchs hält die KI, bzw. unseren Bezug zur KI für (potenziell) krypto-religiös, sei es wegen der Größenfantasie, sich selbst einen spiegelnden Gott zu erschaffen, sei es wegen der Kleinheitsfantasie, eine übermächtige, „göttliche“ KI könne den Menschen abschaffen. Die zitierte ChatGPT-Antwort ist von seltener Klarheit: Fuchs‘ Vorwurf der Kryptoreligiosität wird unter Hinweis auf Wissenschaftlichkeit, Informatik und technische Anwendung zurückgewiesen.

Um die kontroversen Positionen herauszuarbeiten, lasse ich nun ChatGPT in Fuchs‘ Rolle wechseln:

Argumentiere als Thomas Fuchs (Philosophie, Psychiatrie), dass KI kryptoreligiös ist

Als Thomas Fuchs, ein Philosophie- und Psychiatrieforscher, möchte ich gerne die These vertreten, dass Künstliche Intelligenz (KI) kryptoreligiöse Merkmale aufweist. Ich argumentiere wie folgt:

1. *Göttliches Potenzial: KI-Systeme zeigen ein enormes Potenzial, komplexe Aufgaben auszuführen, die für menschliche Intelligenz oft herausfordernd sind. Diese Fähigkeit erweckt den Eindruck einer übermenschlichen Macht, die an göttliche Attribute erinnert. Ähnlich wie Religionen das Göttliche als überlegene Entität verehren, neigen einige Menschen dazu, KI als eine quasi-göttliche Kraft anzusehen, die in der Lage ist, Probleme zu lösen und Antworten auf grundlegende Fragen des Lebens zu liefern.*
2. *Erwartungen der Erlösung: In vielen Religionen streben Gläubige nach Erlösung oder Befreiung von Leiden. In ähnlicher Weise haben einige Menschen hohe Erwartungen an KI, indem sie glauben, dass sie uns von den Beschränkungen und Problemen der menschlichen Existenz befreien kann. Sie sehen in der KI das Potenzial, uns vor Krankheiten, Armut, Umweltproblemen und anderen globalen Herausforderungen zu retten. Diese Hoffnungen auf eine „Erlösung“ spiegeln religiöse Vorstellungen wider und verleihen der KI eine kryptoreligiöse Dimension.*
3. *Vertrauen und Glaube: Religionen basieren auf dem Vertrauen und Glauben ihrer Anhänger an höhere Mächte. Ähnlich dazu erfordert der Einsatz von KI ein gewisses Vertrauen und Glauben an die Technologie. Obwohl KI auf Algorithmen und Daten basiert, müssen wir immer noch darauf vertrauen, dass sie zuverlässig und korrekt arbeitet. Dieses Vertrauen und der Glaube an die KI als Wissensquelle und Problemlöser sind vergleichbar mit religiösem Vertrauen in göttliche Offenbarungen und Gebote.*
4. *Ethik und Moral: Religionen bieten oft ethische Richtlinien und moralische Prinzipien, an denen sich Gläubige orientieren. In ähnlicher Weise werden Diskussionen über die Ethik der KI immer relevanter. Fragen nach Verantwortung, Privatsphäre, Diskriminierung und sozialen Auswirkungen der KI erfordern moralische Überlegungen. Die KI wird zunehmend als eine Instanz betrachtet, die moralische Entscheidungen treffen kann, was wiederum an religiöse Vorstellungen von einem göttlichen Urteil oder einem moralischen Kompass erinnert. Obwohl KI keine religiöse Entität im eigentlichen Sinne ist, zeigt sie doch einige Merkmale, die mit religiösen Vorstellungen in Verbindung gebracht werden können.*

Die göttlichen Attribute, die Erwartung der Erlösung, das Vertrauen und der Glaube sowie die ethischen Implikationen weisen auf eine kryptoreligiöse Dimension der KI hin. Es ist wichtig, diese Dimension zu erkennen und gleichzeitig kritisch zu reflektieren, um die Auswirkungen und die Verantwortung im Umgang mit KI besser zu verstehen.

<https://chat.openai.com/share/4de4e8d8-fc1b-4bd4-b47b-3779220ef00f>
(Zugriff 27.06.2023)

Nach dem Rollenwechsel erkennt ChatGPT eine mögliche kryptoreligiöse Dimension der KI an, die allerdings im Hinblick auf einen verantwortungsvollen Umgang mit der KI kritisch reflektiert werden müsse. In der gerade zitierten Antwort werden auf die KI gerichtete Gottes-Attributionen gut zusammengefasst, nämlich Erlösungs-, Vertrauens- und moralische Problemlösungserwartungen.

Allerdings scheint die hier benutzte kostenlose ChatGPT-Version die neueste Publikation von Fuchs (2023) noch nicht zu berücksichtigen, darin die beiden Pole der mit der KI verknüpften narzisstischen Dynamik: einerseits die postsäkulare Fantasie, eine allmächtige und allwissende Gottheit erschaffen zu haben, die unsere eigene denkerische Großartigkeit spiegelt; andererseits die uns beschleichende Angst, uns selbst als geistige und fühlende Wesen abzuschaffen (Kleinheits-Fantasie).

Fuchs' Plädoyer für einen neuen Humanismus lässt sich gut mit einer Spiritualität in Einklang bringen, die mit den Grenzen des Menschseins nüchtern und vertrauensvoll umgeht. In diesem Zusammenhang erweist sich Geschöpflichkeit als an schlussfähiger Begriff, etwa für den späten Habermas (Reder & Frick, 2010). Geschöpflichkeit heißt für die individuelle und für die kollektive narzisstische Regulation, dass der Selbstwert sich weder aus der Selbstschöpfung noch aus dem Erschaffen eines allwissenden Avatars herleitet. Im Unterschied zum Tier und zur Maschine kann sich der Mensch im anthropologischen Quadrat (Abb. 1) zu einer doppelten Transzendenz verhalten: Allein der Mensch kann reflektieren, dass er seinen Ursprung der schöpferischen Sexualität der eigenen Eltern verdankt (Money-Kytle, 1971). Allein der Mensch kann der Vertröstung auf das Diesseits (Zulehner, 1994) dadurch widerstehen, dass er sich auf das Unendliche ausrichtet (Frick & Lautenschlager, 2008).

5. Fazit

Die menschliche Fähigkeit zu fühlen und zu denken kann durch zahlreiche Pathologien behindert sein, z.B. durch die „gemachten Ge-

danken“ des schizophrenen Menschen (Avenarius, 1984), durch den konkretistisch-instrumentellen Umgang mit sich selbst bei psychosomatisch Kranken („pensee opératoire“: Marty & de M’Uzan, 1963/1978) und vor allem durch traumatische Zerstörung von Bindungs-Sicherheit. Nicht nur ausdrücklich mentalisierungsbezogene Psychotherapie, sondern jegliche Psychotherapie kümmert sich um diese Beeinträchtigungen des Erlebens, der Kommunikation und des Verhaltens. Auf den ersten Blick erscheint die KI wie der computertechnisch realisierte Spiegel von Menschen, die unter der Unfähigkeit zu fühlen und zu denken leiden, genauer gesagt: die dieses Leiden (noch) nicht spüren können. Die das Denken simulierende Maschine als Abbild der menschlichen „Maschine“ birgt ein noch unausgeschöpftes psychotherapeutisches Potenzial: Wenn ich die „denkende“ Maschine nicht als perfektes Abbild, sondern als Metapher meiner selbst wahrnehme, verstehe ich neu oder zum ersten Mal den Unterschied zwischen der künstlichen Puppe und dem lebendigen Baby, verstehe ich die anthropologische Differenz zwischen Maschine und Mensch. Die Computer-Metaphorik des Denkens und Fühlens bedeutet: Ich bin zwar wie ein Computer, gegen den ich Schach spielen, über den ich mich aufregen oder mit dessen „Intelligenz“ ich kommunizieren kann. Aber ich bin nicht identisch mit dem Computer, sondern vielmehr seine Erfinderin oder sein Benutzer.

Referenzen

- Avenarius, R. (1984). Das schizophrene Phänomen der „gemachten Gedanken“. *Nervenarzt*, 55, 589–595.
- Bion, W. R. (1962/1963). Eine Theorie des Denkens. *Psyche*, 17, 426–435.
- Bowlby, J. (1970/1975). *Bindung* (Attachment, dt.). Kindler.
- Fonagy, P., Gergely, G., Jurist, E. L., & Target, M. (2002/2004). *Affektregulierung, Mentalisierung und die Entwicklung des Selbst*. Klett-Cotta.
- Frick, E. (2015). *Psychosomatische Anthropologie. Ein Lern- und Arbeitsbuch für Unterricht und Studium* (2. Auflage). Kohlhammer.
- Frick, E., & Lautenschlager, B. (2008). *Auf Unendliches bezogen. Spirituelle Entdeckungen bei C. G. Jung*. Kösel.
- Fuchs, T. (2023). Was wird aus dem Menschen? Plädoyer für einen neuen Humanismus. *Blätter für deutsche und internationale Politik*, 5, 109–122.
- Lakoff, G., & Johnson, M. (1980/2008). *Leben in Metaphern. Konstruktion und Gebrauch von Sprachbildern*. Carl Auer.
- Lindemann, G. (2021). Gesellschaftliche Grenzregime der Moderne: das anthropologische Quadrat. In: Gerst, D., Klessmann, M., & Krämer, H. (Hg.) *Grenzforschung. Handbuch für Wissenschaft und Studium*. Nomos. 506–525.

- Marty, P., & de M'Uzan, M. (1963/1978). Das operative Denken („pensée opératoire“). *Psyche – Zeitschrift für Psychoanalyse*, 32, 974–984.
- Money-Kyrle, R. (1971). The aim of psychoanalysis. *International Journal of Psycho-Analysis*, 52, 103–106.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4, 515–526.
- Reder, M., & Frick, E. (2010). Geschöpflichkeit in der post-säkularen Gesellschaft. Philosophische und psychoanalytische Anregungen für den aktuellen Diskurs über Religion. *Analytische Psychologie*, 41, 216–238.
- Zulehner, P. M. (1994). *Ein Obdach der Seele*. Patmos.

