

Künstliches Bewusstsein – Wird KI irgendwann wie wir?

Es ist ein beliebtes Thema von Hollywood: Maschinen, die dem Menschen ähnlich sind und so hoch entwickelt, dass sie uns an Intelligenz ebenbürtig sind oder gar noch übertreffen. In diesen Filmen ist die künstliche Intelligenz mit Bewusstsein aber in der Regel der Bösewicht. Sie hat sich gegen ihre Schöpfer gewandt und will nun die Menschheit beherrschen, unterjochen oder gar vernichten.¹ Obwohl diese Filme durchaus einige tatsächliche Fakten aufgreifen, kann einem diese doch weit verbreitete Einstellung zu einer KI mit Bewusstsein etwas einseitig vorkommen, weshalb dieser Text einen differenzierteren Blick auf das Thema „KI mit Bewusstsein“ werfen will. Dabei soll dann unter anderem eben auch geklärt werden, wie realistisch das Szenario einer bösen KI mit einem eigenen Bewusstsein, das sich gegen die Menschheit wendet, tatsächlich ist.

Dafür soll sich dieser Text mit zwei Fragen beschäftigen, die folgendermaßen formuliert sein sollen:

Kann eine Maschine überhaupt ein Bewusstsein entwickeln?

Sollte eine KI über ein Bewusstsein verfügen dürfen?

Zuallererst soll an dieser Stelle jetzt versucht werden, eine Definition für Bewusstsein zu finden, damit grundlegend geklärt ist, was überhaupt gemeint ist,

wenn von Bewusstsein gesprochen wird. Hier lässt sich aber schon schnell feststellen, dass das nicht ganz so einfach ist, wie man es im ersten Moment vielleicht meinen könnte.

Begriffsdefinition

Wenn man Bewusstsein im Duden nachschlägt, bekommt man zwei mögliche Bedeutungen vorgeschlagen. Die erste und an dieser Stelle interessantere beschreibt den abstrakten psychologischen Begriff des Bewusstseins (die zweite kommt eher in Redewendungen wie „das Bewusstsein verlieren“ zum Tragen und soll daher in diesem Text außen vor gelassen werden). Die Bedeutung ist noch einmal in drei Punkte aufgeteilt. Da heißt es:

1.a) Zustand, in dem man sich einer Sache bewusst ist; deutliches Wissen von etwas, Gewissheit

1.b) Gesamtheit der Überzeugungen eines Menschen, die von ihm bewusst vertreten werden

1.c) Gesamtheit aller jener psychischen Vorgänge, durch die sich der Mensch der Außenwelt und seiner selbst bewusst wird.

Diese Definition 1.c) ist jene, die man am ehesten im Bereich der Psychologie anwenden kann.²

Eine Sache ist an diesem Eintrag zu Bewusstsein schonmal bemerkenswert: Nämlich, dass in jeder der Definitionen in irgendeiner Weise Wendungen wie „bewusst sein“ oder „bewusst werden“ vorkommen. Es scheint also schwierig zu

¹ Der deutsche Philosoph Karsten Weber beschreibt zwei mögliche Deutungen für KI und Roboter in der Science-Fiction-Kunst: Zunächst sieht er im kulturellen Diskurs eine Entwicklung der Darstellung von KI und Roboter von einem reinen Werkzeug zu einem handelnden Akteur. Auf der anderen Seite symbolisierten KI und Roboter die Technik, deren folgenreicher Einsatz die Welt zerstören

könnte. Siehe Karsten Weber, *Roboter und Künstliche Intelligenz in Science Fiction-Filmen: Vom Werkzeug zum Akteur*. In: Jan Fuhse (Hg.), *Technik und Gesellschaft in der Science Fiction*. Berlin: LIT 2008, S. 34-54.

² Aus Duden. Berlin: Bibliographisches Institut GmbH, Online-Version. Abgerufen am 02.03.2021 unter <https://www.duden.de/rechtschreibung/Bewusstsein>

sein, Bewusstsein nicht durch sich selbst zu erklären. Dies weist schon auf die Problematik und Schwierigkeit einer eindeutigen Definition des Bewusstseinsbegriffes hin. Nichtsdestotrotz soll hier noch etwas tiefer in die Thematik eingestiegen und dafür die psychologische Definition aus dem Duden als Aufhänger verwendet werden.

Es gibt in dieser Definition drei Begriffe, die sich für eine nähere Betrachtung anbieten, nämlich „psychische Vorgänge“, „der Mensch“, und „seiner selbst“. Aus diesen drei Begriffen folgen drei Fragen, die den Bewusstseinsbegriff aus verschiedenen Richtungen annähern. Diese Fragen sind:

Können nur Menschen ein Bewusstsein haben?

Was sind psychische Vorgänge?

Was ist das „Selbst“?

Im Folgenden soll versucht werden, kurze Antworten auf diese Fragen zu geben.

Können nur Menschen ein Bewusstsein haben?

Es gibt Untersuchungen bei Tieren, die sich mit der Frage beschäftigen, ob diese eine dem Menschen ähnelnde Wahrnehmung von sich selbst und der Welt haben. Um dies nachzuweisen, wird in der Biologie der erstmals von Gordon Gallup etablierte Spiegeltest verwendet. Dem Tier wird irgendwo am Körper ein weißer, geruchloser Fleck aufgemalt. Dann wird dem Tier ein Spiegel vorgeführt. Erkennt das Tier sich selbst im Spiegel, wird es auf

den Fleck reagieren und zum Beispiel versuchen, ihn wegzuwischen. Gallup konnte ein solches Verhalten erstmals bei Schimpansen nachweisen.³

Es gibt einige Tierarten, die den Spiegeltest schon bestanden haben, nämlich, wie der Tierphilosoph Peter Kaiser nennt, neben Schimpansen auch andere große Menschenaffenarten, Delfine und Elstern.⁴ Dennoch ist es schwer zu sagen, ob das Selbsterkennen im Spiegel wirklich auf ein Ichbewusstsein und eine konkrete Wahrnehmung von sich selbst schließen lässt. Kaiser beschreibt, dass der Spiegeltest zwar ein unmittelbares körperliches Selbstbewusstsein beweist, dieses körperliche Bewusstsein in seiner Bedeutung aber bisher zu Unrecht weitgehend unbeachtet blieb. Er sieht hier ein neues und vielversprechendes Forschungsgebiet und plädiert dafür, menschliche Begrifflichkeiten aus dem Bereich des körperlichen Bewusstseins auch auf die Tierphilosophie anzuwenden.⁵ Alles in allem bleibt es aber schwierig, nur aufgrund eines bestandenen Spiegeltestes auf ein Bewusstsein, wie es der Mensch besitzt, zu schließen.

Am Ende wissen wir also nicht, ob ein Bewusstsein in der Natur nur dem Menschen vorbehalten ist oder nicht. Bei einer KI haben wir den Vorteil, dass wir mit ihr sprachlich kommunizieren könnten - man denke an Chatroboter oder ähnliches. Bei einer KI könnte es uns also leichter fallen, zu beurteilen, ob sie ein Bewusstsein hat.⁶ Hier stoßen wir aber auf andere

³ vgl. Gordon Gallup, *Chimpanzees: self-recognition*. Washington D.C.: Science Vol. 167, Issue 3914 (1970), S. 86-87.

⁴ vgl. Peter Kaiser, *Selbstbewusstsein als anthropologische Differenz?* In: Anne Siegetsleitner, Andreas Oberprantacher (Hg.), *Mensch sein – Fundament, Imperativ oder Floskel?* Innsbruck: Innsbruck University Press 2017, S. 486.

⁵ vgl. Ebd., S.487-490.

⁶ Alan Turing beschreibt einen möglichen Testaufbau, der einen Computer dann für intelligent erklärt, wenn er in der Bearbeitung einer Aufgabe nicht mehr von einem Menschen unterschieden werden kann. Siehe Alan Turing, *Computing Machinery and Intelligence*. In: Stuart Shieber (Hg.), *The Turing Test*. Cambridge (US): The MIT Press 2004, S. 105-110.

Probleme, die zu einer späteren Stelle in diesem Text erörtert werden sollen.

Was sind psychische Vorgänge?

Eine Antwort auf diese Frage versuchen die Neurowissenschaften zu geben, die sich mit den Zusammenhängen von biologischen Prozessen im Gehirn mit psychischen Bewusstseinszuständen beschäftigen.

Das Ziel der Neurowissenschaft ist es, das Gehirn absolut physikalisch erklärbar zu machen. Dazu sollen den Bewusstseinszuständen eindeutige und reproduzierbare neuronale Zustände im Gehirn zugeordnet werden.⁷ Dies könnte uns möglicherweise auch zu einer bewussten KI führen, denn diese Zustände ließen sich am Computer simulieren. Irgendwann könnte es vielleicht möglich sein, das menschliche Gehirn zu 100% durch ein Computerprogramm nachzubilden. Wenn man davon ausgeht, dass das Bewusstsein nur die Gesamtheit der Prozesse in unserem Gehirn darstellt, müsste dieses Programm dann in der Lage sein, ein Bewusstsein zu entwickeln.⁸ Derlei Gedanken seien an dieser Stelle aber nur spekulativ angebracht.

Das Problem der Neurowissenschaften ist, dass sie bisher nur die Abläufe im Gehirn beschreiben, aber nicht erklären können. Zum Beispiel hieße das: Wir wissen, was im Gehirn passiert, wenn wir ein bestimmtes Gefühl haben, aber wir wissen nicht, warum aus den jeweiligen Prozessen im Gehirn

genau dieses Gefühl entsteht. Dieses Problem der Frage nach Ursache und Wirkung wird auch als Qualiaproblem⁹ der Philosophie bezeichnet.¹⁰

Am Ende verstehen wir aktuell die Prozesse in unserem Gehirn noch nicht gut genug, um anhand dieser erklären zu können, was Bewusstsein ausmacht oder wie es entsteht.

Was ist das „Selbst“?

Der Selbst-Begriff lässt ebenso wie der Bewusstseinsbegriff verschiedene Möglichkeiten der Definition zu und ist eng mit diesem verknüpft.¹¹ Im Rahmen dieses Textes sei der Begriff „Selbstbewusstsein“ als die Fähigkeit des Wissens um die eigenen Gedanken und Bewusstseinszustände definiert. Das Selbstbewusstsein darüber, wie Entscheidungen getroffen werden, ist eine Grundvoraussetzung zu der Fähigkeit, über das eigene Handeln zu reflektieren und sich selbst zu hinterfragen.

An dieser Stelle soll an einem Beispiel beschrieben werden, was es für eine KI bedeuten könnte, diese Fähigkeit zu besitzen: Man stelle sich eine KI vor, deren Aufgabe es ist, einen Trailer zu schneiden. Verfügt die KI über Selbstbewusstsein im Sinne der obigen Definition, so trifft sie ihre Entscheidungen nicht nur aufgrund eines Datenabgleiches und Datenverarbeitung, sondern kann die inhaltlichen Aspekte ihrer

⁷ Eine gute Einführung in das Thema gibt Thomas Trappenberg, *Fundamentals of Computational Neuroscience*. Oxford: Oxford University Press 2010, S. 1-16.

⁸ vgl. Jochen Zielke: *Künstliche Intelligenz und Bewusstsein*. Köln: Planet Wissen 2018. Abgerufen am 04.03.2021 unter https://www.planet-wissen.de/technik/computer_und_roboter/kuenstliche_intelligenz/pwiekuenstlicheintelligenzundbewusstsein100.html

⁹ Zur näheren Erklärung des Begriffs siehe Michael Tye: *Qualia*. In Edward N. Zalta (Hg.), *Stanford Encyclopedia*

of Philosophy. Stanford (US): Stanford University 1997, Online-Version. Abgerufen am 03.03.2021 unter <https://plato.stanford.edu/entries/qualia/>

¹⁰ vgl. hierzu Thomas Nagel, *What is it like to be a bat?* Durham (US): The Philosophical Review LXXXIII, 4 (1974), S. 435-50.

¹¹ vgl. Joel Smith, *Self-Consciousness*. In Edward N. Zalta (Hg.), *Stanford Encyclopedia of Philosophy*. Abgerufen am 03.03.2021 unter <https://plato.stanford.edu/entries/self-consciousness/>

Arbeit erklären. Also warum genau sie den Schnitt an dieser Stelle gesetzt hat und warum sie die Musik ab der Hälfte des Trailers abspielt, sie versteht genau, was ein Trailer ist und was sie tut.

An diesem Beispiel wird aber eine bereits angesprochene Problematik deutlich, auf die jetzt an dieser Stelle genauer eingegangen werden soll. Die Frage ist, wie bei einem Computer, der intelligent scheint, diese Intelligenz zu beurteilen ist und wie bewertet werden kann, ob er nun ein Bewusstsein hat oder nicht. Tatsächlich gibt es schon eine KI, die einen Trailer geschnitten hat, diese war sich aber sicher noch nicht dessen wirklich bewusst, was sie da tut, auch wenn das Ergebnis annehmbar ist.¹² Die Problematik ist, dass man bei einem Computer von einer nach außen hin korrekten Lösung der Aufgabe noch nicht auf ein wirkliches korrektes inneres Verständnis der Bedeutung der Aufgabe schließen kann. Dieses Problem wurde von dem Philosophen John Searle in dem Gedankenexperiment des chinesischen Zimmers beschrieben.

Das chinesische Zimmer

Das chinesische Zimmer stellt eine Reaktion auf den Turing-Test dar, der eben dann einen Computer für intelligent erklärt, wenn er von außen in der Lösung einer Aufgabe nicht mehr von einem Menschen zu unterscheiden ist.¹³

Searle stellt sich einen Menschen vor, der in einem geschlossenen Raum sitzt, keinerlei Chinesisch versteht, aber über eine genaue

Anleitung in seiner Muttersprache zur Beantwortung von in chinesischer Schrift gestellten Fragen verfügt. Einfach nur durch ein stures Befolgen dieser Anleitung kann er von außen in Chinesisch gestellte Fragen sinnvoll beantworten, ohne den Sinn der Fragen und Antworten zu verstehen. Nach außen hin wirkt es aber so, als ob der Sinn der Fragen verstanden wurde, da die Antworten ja richtig sind.¹⁴

Dieses Gedankenexperiment zeigt also Folgendes: Selbst wenn ein Computer für uns wirkt, als ob er ein Bewusstsein besitzt, heißt das noch nicht, dass es auch wirklich so ist. Wahrscheinlicher ist, dass er nur stur eine Anleitung befolgt, und wir es sind, die das Bewusstsein auf ihn projizieren.

Mit diesem Punkt soll jetzt die Frage nach der Definition von Bewusstsein abgeschlossen und noch einmal festgehalten werden, dass es sehr schwierig ist, zu beschreiben, was Bewusstsein wirklich ausmacht und wie dieses nachweisbar ist. Am Ende kann man sich dem Begriff nur aus verschiedenen Richtungen annähern, ihn aber noch nicht endgültig begreifen.

Das künstliche Wesen PSI

Eine Art Produktionshandbuch für ein künstliches Bewusstsein beschreibt der deutsche Psychologe Dietrich Dörner in seiner Theorie „Bauplan für eine Seele“.¹⁵

Dörner stellt zuerst die These auf, dass die Seele ein vollkommen deterministisches System sei, eine Art Maschine, die sich auf eine Vielzahl von wenn-dann-Aussagen stützt. Diese Aussagen stellen Ursache und

¹² vgl. Charles Haine, *How IBM's Watson Used AI to Edit the Trailer for Upcoming AI Thriller 'Morgan'*. West Hollywood (CA): No Film School 2016. Abgerufen am 03.03.2021 unter <https://nofilmschool.com/2016/10/ibm-watson-edits-morgan-trailer>

¹³ vgl. Alan Turing, *Computing Machinery and Intelligence*.

¹⁴ vgl. John R. Searle, *Minds, brains, and programs*. Cambridge (US): The Behavioral and Brain Sciences Vol. 3, Issue 3 (1980), S. 417-424.

¹⁵ Dietrich Dörner, *Bauplan für eine Seele*. Hamburg: Rowohlt Taschenbuch Verlag GmbH 2001.

Wirkungsbeziehungen in der Form „Wenn X gilt, dann folgt daraus Y“ dar und in ihrer Verknüpfung untereinander entstehen Seele und Bewusstsein.

Ist diese These wahr, so schreibt Dörner, also die Seele nichts anderes als eine hochkomplexe Maschine, so wäre es auch möglich, dass ein Computer – nichts anderes als eine deterministische Maschine – ein Bewusstsein entwickeln kann. Wäre die These falsch, so wäre die Psychologie nicht einmal eine wirkliche Wissenschaft, da sie sich mit etwas beschäftigen würde, das sie nie vollkommen erklären könne.

Dörner sieht die Entwicklung der Seele als einen evolutionären Prozess, der aus der Notwendigkeit entstand, ein autonomes Handeln zu entwickeln. Diesen Prozess spielt er in einem Gedankenexperiment an dem künstlichen Wesen PSI durch, das im Anfangsstadium nichts weiter ist als eine einfache, vollkommen unselbstständige Dampfmaschine. Nach und nach erhält PSI grundlegende menschliche Eigenschaften wie Hunger, Durst, Selbsterhaltungstrieb und Sexualität. In deren Wechselwirkung soll, so wie in der menschlichen Evolution auch, das Bewusstsein entstehen. Dörner beschäftigt sich bewusst mit der Frage der Entwicklung von Bewusstsein, nicht mit der Frage nach dem Wesen von Bewusstsein, da es der Psychologie noch nicht einmal gelinge, eine einheitliche Definition für Gefühle zu finden.¹⁶

¹⁶ vgl. Ebd., S. 9-24.

¹⁷ vgl. Dylan F. Glas, Takashi Minato, Carlos T. Ishi, Tatsuya Kawahara, Hiroshi Ishiguro, *ERICA: The ERATO Intelligent Conversational Android*. New York City: 2016 25th IEEE International symposium on robot and human interactive communication (RO-MAN), S. 22-29.

¹⁸ vgl. Hiroshi Ishiguro in *How do Robots "think"? Can they take over the world? | Inside Hiroshi Ishiguro's*

Am Ende bleibt die künstliche Entwicklung einer Seele aber noch ein Gedankenexperiment.

Die Androidin „Erica“

Ein interessantes Beispiel aus der aktuellen Forschung für eine „KI zum Anfassen“ stellt der humanoide Roboter Erica¹⁷ dar, ein Geschöpf des japanischen Robotik-Forschers Hiroshi Ishiguro, einem der Stars in der Bewusstseins-Szene. Erica steht für **Erato Intelligent Conversational Android** und es handelt sich dabei um eine KI mit menschlichem Aussehen, die in der Lobby der Forschungseinrichtung sitzt und sich mit Besuchern unterhält.

Erica kann das Gegenüber mit einer Gesichtserkennungssoftware beurteilen und auch Gefühle erkennen. Auch wenn es vielleicht nicht so scheint, hat der Mensch trotzdem immer noch viel Arbeit für die KI übernommen, denn verschiedene Gesprächsthemen und Szenarien sind vorimplementiert, Erica reagiert aber auf den Gesprächspartner und wählt die Szenarien selbstständig aus.¹⁸

Ishiguro glaubt, dass wir durch die Entwicklung von KI auch erfahren können, was uns als Menschen ausmacht und sieht KI als einen natürlichen und notwendigen Teil der Evolution, als ein neues Werkzeug in der Entwicklung des Menschen.¹⁹ Er glaubt auch, dass KI ein menschliches Aussehen haben sollte, um die Kommunikation und Vermittlung einfach zu gestalten. Eine mögliche praktische Anwendung seiner Forschung sieht er im Service und

Robot Lab. Berlin: DW Shift 2020.

Internetdokumentation, abgerufen am 04.03.2020 unter <https://www.youtube.com/watch?v=5Zd2WYhTQ9w>

¹⁹ vgl. Hiroshi Ishiguro in *Hiroshi Ishiguro: You say you want evolution? Embrace robotics*. Tokyo: The Japan Times 2019. Interview mit Hiroshi Ishiguro, abgerufen am 04.03.2021 unter <https://www.youtube.com/watch?v=Ou-x798INS0>

Dienstleistungsbereich, zum Beispiel an einer Hotelrezeption.²⁰ In Japan sind Rezeptionsroboter tatsächlich schon weit verbreitet. Interessant ist, dass in der japanischen Vorstellung auch Dinge eine Seele besitzen können und die Ablehnung oder Skepsis gegenüber Robotern deshalb wahrscheinlich kleiner ist als in westlichen Kulturkreisen.

Ein Blick in die Zukunft...

An dieser Stelle soll jetzt ein kleiner Blick in die Zukunft gewagt werden. Wie wird sich das Thema KI und Bewusstsein entwickeln?

Es steht außer Frage, dass die Entwicklung und Verbreitung von autonom mit uns kommunizierenden Robotern oder Programmen weiter vorangebracht werden wird, zum Beispiel im Dienstleistungssektor, in beratenden Funktionen oder in der Medizin zum Beispiel als Pflegekraft oder gar als Therapiemethode mithilfe von Mikrochips.²¹ Die zunehmende Forschung im Bereich unseres Gehirns und Bewusstsein verspricht vor allem auf diesem Feld Fortschritte: So gibt es Ansätze, Nervenzellen durch Elektroden zu ersetzen und robotische Prothesen mit KI als Schnittstelle zu unserem Bewusstsein zu verbinden.²² Unumstritten ist mittlerweile die Ansicht, dass es gelingen wird, starke KI zu entwickeln. Wie Julian Moser schreibt, ist es aber weiterhin fraglich, inwieweit diese über ein Bewusstsein verfügen wird.²³

²⁰ vgl. Hiroshi Ishiguro in *How do Robots "think"? Can they take over the world? | Inside Hiroshi Ishiguro's Robot Lab.*

²¹ vgl. Jochen Zielke: *Künstliche Intelligenz und Bewusstsein.*

²² vgl. Frank Thole, *Künstliche Intelligenz (KI) überholt sich selbst – Auf dem Weg zum artifiziellen Bewusstsein.* München: Digitale Welt Academy 2019. Abgerufen am 04.03.2021 unter <https://digitaleweltmagazin.de/2019/07/22/kuenstliche-intelligenz-ki-ueberholt-sich-selbst-auf-dem-weg-zum-artifiziellen-bewusstsein/>

Weitergehend kann man spekulativ und etwas provokativ natürlich fragen, ob irgendwann humanoide Roboter mit Bewusstsein als eine Art neue Lebensform unsere Gesellschaft ergänzen werden.

Wird KI irgendwann böse?

Und damit wäre man am Ausgangspunkt des Textes angelangt und der klassischen Sci-Fi-Frage, ob uns KI irgendwann ersetzen, beherrschen oder gar auslöschen wird.

Die Antwort ist, dass dies wohl eher unwahrscheinlich ist.

Wie der Zukunftsforscher Matthias Horx beschreibt, ist die Grundvoraussetzung für Machtverlangen Angst vor Verlust, das heißt, die KI müsste dafür mit einer Körperlichkeit und Sterblichkeit ausgestattet werden.²⁴ Die KI müsste wie wir werden, um wirklich böse zu sein.

Doch wenn die KI menschlich werden muss, um böse zu sein, gibt es dann überhaupt noch einen Unterschied zwischen menschlicher Bosheit und KI-Bosheit? Um es mit einem Topos der Science-Fiction auszudrücken: Das Risiko, dass der hochintelligente und mit einem Bewusstsein ausgestattete Bordcomputer eines Raumschiffs auf ferner Mission in den Weiten des Alls darüber verrückt wird und die restliche Besatzung auslöschen will, ist genauso unkontrollierbar wie das, dass dies bei

²³ Julian Moser, *Starke KI, schwache KI – Was kann künstliche Intelligenz?* Bremen: Just Add AI GmbH 2018. Abgerufen am 04.03.2021 unter <https://jaai.de/starke-ki-schwache-ki-was-kann-kuenstliche-intelligenz-261/>

²⁴ Matthias Horx, *Der KI-Irrtum. Warum ich nicht an die finstere Machtübernahme der Künstlichen Intelligenz glaube – und auch nicht an das Gegenteil.* Wien: Zukunftsinstitut Horx GmbH, ohne Datum. Abgerufen am 04.03.2021 unter <https://www.horx.com/schluesstexte/der-ki-irrtum/>

einem der menschlichen Crewmitglieder geschieht.

Der Grund für die weitverbreitete Angst, dass KI sich irgendwann gegen uns wendet, liegt nach Horx am Ende in unserer Angewohnheit, unsere eigenen Stärken und Schwächen auf Dinge zu übertragen, die uns ähnlich sind. Intelligenten Computern werden, auch deshalb weil sie für viele nicht wirklich begreifbar sind, menschliche, emotionale Motive unterstellt, um sie zu erklären, die sie tatsächlich aber nicht haben.²⁵

Wie wird KI mit einer Art Bewusstsein jetzt also in Zukunft aussehen? Geht das mehr in die Richtung einer humanoiden KI, die uns immer ähnlicher wird, oder bleiben wir bei knuffigen Robotern, die zwar selbstständig mit uns kommunizieren können, sonst aber keine wirklich menschlichen Eigenschaften haben?

Am Ende ist dies eine Frage, deren Beantwortung nur spekulativ möglich ist. Wahrscheinlich scheint aber: KI wird sich weiterentwickeln und uns nach außen hin immer mehr so scheinen, als ob sie wirklich ein Bewusstsein besitzt, auch wenn dies in Wirklichkeit nicht der Fall sein wird, sondern nur eine Projektion unserer eigenen Gefühle. Eine menschlich aussehende und wirkende KI mag vielleicht entwickelt werden, ob sie sich wirklich etablieren wird, ist aber eine andere Frage. Aufgrund des von Horx beschriebenen Uncanny-Valley-Effekts²⁶ wäre die Ablehnung des Großteils der Bevölkerung gegenüber dieser Art der Robotik wohl noch so groß, dass derlei Produkte einfach schlicht nicht marktfähig wären. Szenarien,

in denen KIs zum Beispiel Lebenspartner ersetzen, sind daher wohl doch eher unwahrscheinlich.²⁷

Fazit

Zum Abschluss soll noch einmal zu den beiden anfangs formulierten Leitfragen dieses Textes zurückgekehrt werden.

Kann eine Maschine überhaupt ein Bewusstsein entwickeln?

Die Entwicklung von einem künstlichen Bewusstsein wäre theoretisch möglich und würde vermutlich über die Nachbildung von unserem eigenen funktionieren. Dies setzt aber ein Verständnis unseres Bewusstseins voraus, von dem wir noch weit entfernt sind. Tatsächlich fällt es uns schon schwer, Bewusstsein überhaupt zu definieren.

Wir müssen aufpassen, dass wir nicht menschliche Eigenschaften auf die Technologie übertragen. Nur das korrekte Lösen einer schweren Aufgabe setzt bei Computern noch nicht das tiefe innere Verständnis ihrer Bedeutung voraus.

Sollte eine KI über ein Bewusstsein verfügen dürfen?

An dieser Stelle sollen die hypothetischen Vor- und Nachteile für eine KI, ein Bewusstsein zu haben, zusammengefasst werden.

Ein Teil von Bewusstsein ist zweifelsohne das selbstständige Hinterfragen und die Bewertung des eigenen Handelns. Eine starke KI, die die Fähigkeit besitzt, die Bedeutung der ihr betrauten Aufgabe wirklich zu verstehen, könnte diese

²⁵ vgl. Ebd.

²⁶ vgl. Ebd.

²⁷ Siehe zu der Frage nach der Zukunft von KI zum Beispiel auch Verena Muntschick, Christian Schuldt, Lena Papasabbas, *6 Thesen zur Künstlichen Intelligenz*.

Frankfurt: Zukunftsinstitut GmbH, ohne Datum.

Abgerufen am 04.03.2021 unter

<https://www.zukunftsinstitut.de/artikel/digitalisierung/6-thesen-zur-kuenstlichen-intelligenz/>

sicherlich noch einmal leistungsfähiger und besser bearbeiten. Das ist auch für die Fragen der Ethik interessant, denn nur mit Bewusstsein könnte ein Algorithmus eine eigene Ethik haben und bewusst ethisch handeln. Am Ende kann eine Maschine nur durch ein eigenes Bewusstsein die Bedeutung von Existenz und Leben wirklich begreifen und schätzen. Szenarien einer bösen Super-KI sind am Ende wohl auch deshalb eher unwahrscheinlich, wenn auch nicht unmöglich. Sollte die KI wirklich so wie wir werden, dann könnte sie wie ein Mensch auch schlechte Wesenszüge entwickeln. Dies wird aber wohl eher ein Gedankenspiel bleiben.

Vielleicht ist die Frage am Ende eher: **wie viel** Bewusstsein braucht eine KI? Vielleicht gelingt es, die positiven Aspekte eines Bewusstseins zu entwickeln oder zumindest zu simulieren. Am Ende muss der Mensch aber immer die Kontrolle behalten, sodass sich seine Schöpfung nicht verselbstständigt.²⁸

Literaturverzeichnis

Manuel Alfonseca, Manuel Cebrian, Antonio Fernandez Anta, Lorenzo Coviello, Andrés Abeliuk, Iyad Rahwan, *Superintelligence Cannot be Contained: Lessons from Computability Theory*. El Segundo (CA, US): Journal of Artificial Intelligence Research Vol. 70 (2021), S. 65-76.

Dietrich Dörner, *Bauplan für eine Seele*. Hamburg: Rowohlt Taschenbuch Verlag GmbH 2001.

Eintrag zu Bewusstsein in *Duden*. Berlin: Bibliographisches Institut GmbH, Online-Version. Abgerufen am 02.03.2021 unter <https://www.duden.de/rechtschreibung/Bewusstsein>

Gordon Gallup, *Chimpanzees: self-recognition*. Washington D.C.: Science Vol. 167, Issue 3914 (1970), S. 86-87.

Dylan F. Glas, Takashi Minato, Carlos T. Ishi, Tatsuya Kawahara, Hiroshi Ishiguro, *ERICA: The ERATO Intelligent Conversational Android*. New York City: 2016 25th IEEE International symposium on robot and human interactive communication (RO-MAN), S. 22-29.

Charles Haine, *How IBM's Watson Used AI to Edit the Trailer for Upcoming AI Thriller 'Morgan'*. West Hollywood (CA): No Film School 2016. Abgerufen am 03.03.2021 unter <https://nofilmschool.com/2016/10/ibm-watson-edits-morgan-trailer>

Matthias Horx, *Der KI-Irrtum. Warum ich nicht an die finstere Machtübernahme der Künstlichen Intelligenz glaube – und auch nicht an das Gegenteil*. Wien: Zukunftsinstitut Horx GmbH, ohne Datum. Abgerufen am 04.03.2021 unter <https://www.horx.com/schluesstexte/der-ki-irrtum/>

Hiroshi Ishiguro in *How do Robots "think"? Can they take over the world? | Inside Hiroshi Ishiguro's Robot Lab*. Berlin: DW Shift 2020. Internetdokumentation, abgerufen am 04.03.2020 unter <https://www.youtube.com/watch?v=5Zd2WYhTQ9w>

²⁸ Obwohl dies laut neusten Berechnung bei superintelligenten KIs womöglich gar nicht möglich ist. Siehe dazu Manuel Alfonseca, Manuel Cebrian, Antonio Fernandez Anta, Lorenzo Coviello, Andrés Abeliuk, Iyad

Rahwan, *Superintelligence Cannot be Contained: Lessons from Computability Theory*. El Segundo (CA, US): Journal of Artificial Intelligence Research Vol. 70 (2021), S. 65-76.

Hiroshi Ishiguro in *Hiroshi Ishiguro: You say you want evolution? Embrace robotics*. Tokyo: The Japan Times 2019. Interview mit Hiroshi Ishiguro, abgerufen am 04.03.2021 unter <https://www.youtube.com/watch?v=Oux798INS0>

Peter Kaiser, *Selbstbewusstsein als anthropologische Differenz?* In: Anne Siegetsleitner, Andreas Oberprantacher (Hg.), *Mensch sein – Fundament, Imperativ oder Floskel?* Innsbruck: Innsbruck University Press 2017.

Julian Moser, *Starke KI, schwache KI – Was kann künstliche Intelligenz?* Bremen: Just Add AI GmbH 2018. Abgerufen am 04.03.2021 unter <https://jaai.de/starke-ki-schwache-ki-was-kann-kuenstliche-intelligenz-261/>

Verena Muntschick, Christian Schuldt, Lena Papasabbas, *6 Thesen zur Künstlichen Intelligenz*. Frankfurt: Zukunftsinstitut GmbH, ohne Datum. Abgerufen am 04.03.2021 unter <https://www.zukunftsinstitut.de/artikel/digitalisierung/6-thesen-zur-kuenstlichen-intelligenz/>

Thomas Nagel, *What is it like to be a bat?* Durham (US): The Philosophical Review LXXXIII, 4 (1974), S. 435-50.

John R. Searle, *Minds, brains, and programs*. Cambridge (US): The Behavioral and Brain Sciences Vol. 3, Issue 3 (1980), S. 417-424.

Joel Smith, *Self-Consciousness*. In: Edward N. Zalta (Hg.), *Stanford Encyclopedia of Philosophy*. Abgerufen am 03.03.2021 unter <https://plato.stanford.edu/entries/self-consciousness/>

Frank Thole, *Künstliche Intelligenz (KI) überholt sich selbst – Auf dem Weg zum artifiziellen Bewusstsein*. München: Digitale Welt Academy 2019. Abgerufen am 04.03.2021 unter <https://digitaleweltmagazin.de/2019/07/22/kuenstliche-intelligenz-ki-ueberholt-sich-selbst-auf-dem-weg-zum-artifiziellen-bewusstsein/>

Thomas Trappenberg, *Fundamentals of Computational Neuroscience*. Oxford: Oxford University Press 2010.

Alan Turing, *Computing Machinery and Intelligence*. In: Stuart Shieber (Hg.), *The Turing Test*. Cambridge (US): The MIT Press 2004, S. 105-110.

Michael Tye: *Qualia*. In Edward N. Zalta (Hg.), *Stanford Encyclopedia of Philosophy*. Stanford (US): Stanford University 1997, Online-Version. Abgerufen am 03.03.2021 unter <https://plato.stanford.edu/entries/qualia/>

Karsten Weber, *Roboter und Künstliche Intelligenz in Science Fiction-Filmen: Vom Werkzeug zum Akteur*. In: Jan Fuhse (Hg.), *Technik und Gesellschaft in der Science Fiction*. Berlin: LIT 2008, S. 34-54.

Jochen Zielke: *Künstliche Intelligenz und Bewusstsein*. Köln: Planet Wissen 2018. Abgerufen am 04.03.2021 unter https://www.planet-wissen.de/technik/computer_und_roboter/kuenstliche_intelligenz/pwiekuenstlicheintelligenzundbewusstsein100.html