

HEIDELBERGER  
JAHRBÜCHER  
ONLINE  
Band 6 (2021)

Gesellschaft der Freunde  
Universität Heidelberg e.V.



# Intelligenz: Theoretische Grundlagen und praktische Anwendungen

Rainer M. Holm-Hadulla, Joachim Funke & Michael Wink (Hrsg.)

HEIDELBERG  
UNIVERSITY PUBLISHING

# Menschliche und künstliche Intelligenz – ein kritischer Vergleich

THOMAS FUCHS<sup>1</sup>

Klinik für Allgemeine Psychiatrie, Universitätsklinikum Heidelberg

## Zusammenfassung

Die Fortschritte der *Künstlichen Intelligenz und der Robotik* stellen die Unterscheidung zwischen Simulation und Realität der menschlichen Person zunehmend in Frage. Sie suggerieren einerseits ein computeromorphes Verständnis menschlicher Intelligenz, andererseits eine Anthropomorphisierung der KI-Systeme. Mit anderen Worten: Wir betrachten uns selbst immer mehr wie unsere Maschinen, und umgekehrt unsere Maschinen wie uns selbst. Was also unterscheidet menschliche und künstliche Intelligenz? Der Aufsatz stellt dafür eine Reihe von Kriterien auf.

## 1 Einleitung

Mit den Fortschritten der künstlichen Intelligenz scheinen wir sterblichen Menschen immer mehr in ein Rückzugsgefecht zu geraten: Die intelligenten Systeme beginnen sich selbst zu adaptieren, zu „lernen“, wie es heißt, und übertreffen vielfach die Leistungen menschlicher Intelligenz. Im Schach, im Go oder Poker-Spiel haben Menschen gegen sie keine Chance mehr. Planen, Wählen, Entscheiden, ja

---

<sup>1</sup> Gekürzte Fassung eines Aufsatzes im Band: T. Fuchs (2020). *Verteidigung des Menschen. Grundfragen einer verkörperten Anthropologie*. Frankfurt/M.: Suhrkamp, S. 21–70.

selbst Autofahren scheinen uns zunehmend abgenommen zu werden. Die entsprechenden Ankündigungen von KI-Ingenieuren, Futurologen und Transhumanisten überbieten sich geradezu gegenseitig:

„Tatsache ist, dass KI den Menschen übertreffen, ja Milliarden Mal klüger werden könnte als jetzige Menschen.“<sup>2</sup>

„Maschinen werden eine Entwicklung nehmen, die die menschliche Evolution widerspiegelt. Am Ende werden selbstbewusste, sich selbst verbessernde Maschinen die menschliche Fähigkeit überschreiten, sie zu kontrollieren oder auch nur zu verstehen.“<sup>3</sup>

Ray Kurzweil, KI-Forscher und Leiter der Entwicklung bei Google, hat für 2045 die „Singularität“ angekündigt, den Zeitpunkt, an dem die künstliche Intelligenz Bewusstsein erlangt, ein exponentieller Fortschritt zu einer „Superintelligenz“ einsetzt und damit ein neues Zeitalter beginnt.<sup>4</sup>

Auch wenn solche vollmundigen Prognosen regelmäßig korrigiert werden – zumindest die Sprache der KI-Forschung nimmt diese Entwicklung bereits vorweg. Es gibt nahezu keine menschlichen Fähigkeiten mehr, die künstlichen Systemen nicht schon jetzt zugeschrieben werden: Wahrnehmen, Erkennen, Denken, Schlussfolgern, Bewerten oder Entscheiden. Umgekehrt erscheint menschliches Bewusstsein vielen heute nur noch als eine Summe von Algorithmen, eine komplexe Datenstruktur im Gehirn, die im Prinzip auch von elektronischen Systemen realisiert werden könnte und nicht mehr an den lebendigen Körper gebunden ist.

Das Computerparadigma des menschlichen Geistes hat eine lange Vorgeschichte. Bereits 1936 entwickelte der Mathematiker Alan Turing die Idee eines digitalen Computers und schlug seinen berühmten Turing-Test vor: Eine Gruppe von Gutachtern kommuniziert längere Zeit schriftlich mit einem Menschen und mit einem Computer, ohne dabei sonstigen Kontakt zu ihnen zu haben. Falls die Gutachter

---

<sup>2</sup> I. Pearson (2008). „The Future of Life. Creating Natural, Artificial, Synthetic and Virtual Organisms“. *European Molecular Biology Organization (EMBO) Reports* 9 (Supplement 1): 75–77.

<sup>3</sup> Ray Kurzweil, zit. n. L. Greenemeier (2010). „Machine Self-awareness“. *Scientific American* 302: 44–45.

<sup>4</sup> R. Kurzweil (2005). *The Singularity Is Near. When Humans Transcend Biology*. New York: Penguin.

danach zwischen Mensch und Maschine nicht unterscheiden können, dann, so Turing, hindert uns nichts mehr, den Computer als „denkende Maschine“ anzuerkennen. Denken wird also rein behavioristisch definiert, nämlich als Output eines Rechensystems, sei es des Gehirns oder des Computers. Der Turing-Test beruht auf der Ununterscheidbarkeit von Simulation und Original: Was so intelligent agiert wie wir, das *ist* auch intelligent, Punktum.<sup>5</sup>

Zumindest die Simulation macht nun gegenwärtig gewaltige Fortschritte – bis zu dem Punkt, an dem sich die Frage nach dem Unterschied zum Original zu stellen beginnt. Was unterscheidet Bewusstsein von seiner Simulation? Gilt hier wirklich das alte Prinzip: „Wenn etwas aussieht wie eine Ente, schwimmt wie eine Ente und schnattert wie eine Ente, dann ist es auch eine Ente“?

Eine Vorstellung von den künftigen Problemen, die diese Frage aufwerfen könnten, gibt „Sophia“, ein humanoider Roboter der Firma *Hanson Robotics*, der zurzeit weltweit in den Medien zu sehen ist.<sup>6</sup> Sophia verfügt über eine menschenähnliche Mimik (modelliert nach Audrey Hepburn), zeigt verschiedene Gefühlsausdrücke, einen modulierten Tonfall und stellt Augenkontakt mit dem Gegenüber her. Sie beantwortet relativ komplexe Fragen, auch über sich selbst, kann Menschen wiedererkennen und witzelt in einer Londoner Talkshow an passender Stelle über das englische Wetter.

Natürlich ist all dies nur ein Bluff. Das wurde spätestens offenkundig, als Sophia von ihrem Erfinder mit einer für sie offenbar unbekanntem Frage konfrontiert wurde, nämlich: „Willst Du Menschen töten?“, und zur Antwort gab: „Okay. Ich will Menschen töten.“ Die Antwort war nur nachgeplappert wie von einem Papagei; Sophia verstand natürlich kein Wort von dem, was sie gefragt wurde. Dennoch ist ihre Wirkung verblüffend. Sophia nähert sich bereits dem *uncanny valley*, wie in der Robotik die Schwelle genannt wird, ab der die Menschenähnlichkeit eines Androiden in uns ein Gefühl von Unheimlichkeit, zugleich aber auch von Faszination erzeugt. Es ist das Gefühl, das entsteht, wenn sich die Bereiche des Toten und des Lebendigen nicht mehr eindeutig voneinander unterscheiden lassen. Wann wird das *uncanny valley* durchquert und eine künftige Sophia von einer bezaubernden, intelligenten Frau nicht mehr zu unterscheiden sein?

---

<sup>5</sup> Vgl. A. M. Turing (1950). „Computing Machinery and Intelligence“. *Mind* 59: 433–460.

<sup>6</sup> C. Weller (2017). „Meet the first-ever robot citizen, a humanoid named Sophia that once said it would destroy humans.“ *Business Insider Nordic*. Haettu, 30. Jg.

Diese Schwelle wird in *Her* überschritten, einem Science-Fiction-Film von Spike Jonze aus dem Jahr 2013: Theodore, ein schüchterner, aber einfühlsamer Mann, verliebt sich in eine Software mit dem Namen Samantha und einer erotischen Stimme, die als „lernendes Programm“ scheinbar menschliche Empfindungen entwickelt. Je mehr Theodore ihr verfällt, desto gleichgültiger wird es ihm, ob es sich um ein reales Gegenüber oder nur um eine Simulation handelt – die beglückende Passung genügt. Allerdings scheitert die Liebe zwischen Mensch und Programm zuletzt an der Weiterentwicklung Samanthas, die mit tausenden von anderen Menschen und Betriebssystemen virtuellen Kontakt aufnimmt und sich in sie „verliebt“, sodass sie Theodore schließlich „verlässt“.

Die projektive Einfühlung des Menschen in seine eigenen, künstlichen Geschöpfe ist freilich kein neues Phänomen. Bekannt ist etwa die Agalmatophilie (griech. *ágalma* = Statue, Götterbild), die erotische oder sexuelle Hinwendung zu Statuen, Puppen oder Automaten. Ovids Bildhauer Pygmalion, von gewöhnlichen Frauen abgestoßen, verliebt sich in die von ihm geschaffene Statue einer idealen Frau, bis sie schließlich von Aphrodite zum Leben erweckt wird. Die Projektion schafft sich ein Wesen, „wie Natur es nie zu erzeugen vermag“, und sie belebt es schließlich auch. In E. T. A. Hoffmanns Sandmann betört die stumpfe Automatenpuppe Olimpia den Studenten Nathanael, der gegen alle warnenden Einwände der Freunde taub bleibt:

„Wohl mag euch, ihr kalten prosaischen Menschen, Olimpia unheimlich sein. – Nur mir ging ihr Liebesblick auf und durchstrahlte Sinn und Gedanken, nur in Olimpias Liebe finde ich mein Selbst wieder.“<sup>7</sup>

Am Ende stürzt sich Nathanael im Wahnsinn von einem Turm.

Es ist, so zeigen diese Beispiele, sehr wohl möglich, dass wir Automaten, Androiden, ja selbst Computersysteme empathisch oder auch erotisch wahrnehmen und ihnen damit so etwas wie Subjektivität zusprechen. Besonders menschenähnliche Stimmen nehmen wir nahezu notwendig als Ausdruck eines Inneren wahr. Wenn Sophia mit zarter Stimme sagt: „Das macht mich glücklich“, dann bedarf

---

<sup>7</sup> Hoffmann, E. T. A. (1960). „Der Sandmann“, in: Ders., *Fantasie- und Nachtstücke*. München: Winkler, S. 331–363.

es schon einer aktiven Distanzierung, um sich klar zu machen, dass da niemand ist, der sich glücklich fühlen könnte, dass es sich also gar nicht um eine *Äußerung* handelt. Das heißt nicht, dass wir uns zu einer menschlichen Stimme normalerweise ein Inneres (Geist, Seele, Bewusstsein oder was auch immer) „hinzudenken“. Die Äußerung eines anderen Menschen ist nicht bloß ein tönendes Wort oder eine symbolische Repräsentation, die auf ein mutmaßliches Inneres verweist. Wir nehmen sie vielmehr als *beseelt* wahr, ohne dahinter eine „Seele“ anzunehmen; wir erleben sie *als den Übergang selbst*, eben als *Äußerung* des anderen, die von einem „Inneren“ gar nicht zu trennen ist. In unserer Wahrnehmung ist der andere immer eine verkörperte, psychophysische Einheit.

Die zunehmend perfektionierte Simulation einer solchen Einheit verlangt nun allerdings, dass wir die Vortäuschung einer Äußerung eigens zurückweisen und Sophias Worte als das nehmen, was sie eigentlich sind: tönende Worte, wie die eines Papageis (beziehungswise nicht einmal das, da ein Papagei seine Laute immerhin auch erlebt). Anderenfalls überlassen wir uns dem Schein und geben wie Nathanael oder Theodore das „Als-ob“, die Unterscheidung zwischen Virtualität und Realität einfach auf. Schon jetzt kann es ja sein, dass der nette Online-Partner oder der einfühlsame Online-Therapeut in Wahrheit nur ein Chatbot ist. Und die ersten Pflegeroboter für Demenzkranke sind bereits in Erprobung: „Erstaunlich schnell bauen die Patienten eine Beziehung zum Roboter auf.“<sup>8</sup> (Leitgeb 2017). Offenbar sind wir nur allzu geneigt, unser eigenes Erleben auf die *Simulate* zu projizieren. Die Faszination des in einem Automaten scheinbar aufblitzenden Bewusstseins trägt dazu bei. Es ist eine Faszination, die auch die KI-Forschung vorantreibt. Ihr Ursprung dürfte – neben dem prometheischen Motiv der gottgleichen Kreativität – wohl letztlich in unbewussten Wünschen nach der Überwindung des Todes zu suchen sein, nämlich durch die Belebung.

*Wie weit also reicht der menschliche Widerstand gegen die Simulation, und wie groß ist ihre Attraktion? Wann geben wir die Unterscheidung von Simulation und Original auf? Genügt uns am Ende die perfekte Simulation?* – Dies dürften entscheidende Fragen in einer digital-automatisierten Kultur werden. Sie sind

---

<sup>8</sup> Leitgeb, V.-V. 2017. „Roboter Mario soll Demenzkranke pflegen“. Süddeutsche Zeitung, online 24.11.2017. <https://www.sueddeutsche.de/bayern/gesundheits-roboter-mario-soll-demenzkrank-pflegen-1.3762375>.

gegenwärtig völlig offen. Was ich im Folgenden anbieten möchte, sind allerdings zwei Klarstellungen zur Unterscheidung:

- Personen sind keine Programme.
- Programme sind keine Personen.

## 2 Personen sind keine Programme

Die gängige Philosophie der Kognitionswissenschaften ebenso wie der *Artificial Intelligence* ist der *Funktionalismus*: Danach bestehen mentale Zustände (also Gefühle, Wahrnehmungen, Gedanken, Überzeugungen etc.) in regelhaften Verknüpfungen von Input und Output eines Systems. Wer sich zum Beispiel in den Finger sticht, hat einen mentalen Zustand, der zu einer verzerrten Gesichtsmuskulatur, zu Stöhnen und zum Zurückziehen des Fingers führt. „Schmerz“ ist dann nichts anderes als der Gehirnzustand, der den genannten Output zur Folge hat – so wie der Zustand eines Brandmelders, der bei Rauch ein Alarmsignal auslöst und die Sprinkleranlage in Gang setzt. Dieser Gehirnzustand lässt sich als eine bestimmte Datenmenge beschreiben. „Geist ist ein neuronaler Computer, von der natürlichen Selektion ausgestattet mit kombinatorischen Algorithmen für kausales und probabilistisches Denken.“<sup>9</sup> Auch Selbstsein ist nur ein solcher komputationaler Zustand: „Wir sind [...] mentale Selbstmodelle informationsverarbeitender Biosysteme. [...] Werden wir nicht errechnet, so gibt es uns nicht.“<sup>10</sup>

Freilich geht das entscheidende Merkmal von Schmerzen, Gefühlen oder Gedanken in dieser funktionalistischen Konzeption offensichtlich verloren – nämlich ihr *Erlebtwerden*. Mit seinem bekannten Gedankenexperiment des „chinesischen Zimmers“ hat John Searle gezeigt, dass sich Sinn und Bedeutung nicht auf funktionale Algorithmen zurückführen lassen, wenn es kein Subjekt gibt, das deren Bedeutung *versteht*.<sup>11</sup> Dazu stelle man sich vor, ein Mann, der kein Wort Chinesisch spricht, sei in ein Zimmer eingeschlossen, in dem sich nur ein Handbuch mit sämtlichen

<sup>9</sup> S. Pinker (1997). *How the Mind Works*. New York: Norton, S. 524 (Übers. T. F.).

<sup>10</sup> Metzinger, T. 1999. *Subjekt und Selbstmodell. Die Perspektivität phänomenalen Bewusstseins vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation*. Paderborn: Mentis, S. 284.

<sup>11</sup> J. R. Searle (1980). „Minds, Brains, and Programs“. *Behavioral and Brain Sciences* 3: 417–457.

Regeln zur Beantwortung von chinesischen Fragen befindet. Der Mann erhält nun von draußen ihm unverständliche chinesische Schriftsymbole durch einen Schlitz ins Zimmer gereicht („Input“), findet aber mit Hilfe des Programms die dazu passenden Antworten, die er dann nach draußen gibt („Output“). Nehmen wir an, das Programm sei so gut und die Antworten so treffend, dass selbst die Chinesen draußen die Täuschung nicht bemerken würde. Dennoch könnte man sicher weder von dem Mann im Zimmer noch von dem System als ganzen sagen: Er oder es *versteht* Chinesisch.

Searles „chinesisches Zimmer“ ist natürlich eine Veranschaulichung von Computern, in denen ein Zentralprozessor nach Algorithmen arbeitet, also etwa nach der Instruktion: „Wenn du Input X erhältst, dann führe Operation Y aus und gib Output Z.“ Die Maschine funktioniert als System völlig adäquat, und doch fehlt ihr die entscheidende Voraussetzung für Verständnis, nämlich Bewusstsein. Folglich kann menschliches Verstehen nicht auf Informationsverarbeitung reduziert werden. Sinnverstehen ist mehr als ein Algorithmus.

Das Gleiche gilt aber für das schon genannte Beispiel der Schmerzen, für den Geschmack von Schokolade oder den Duft von Lavendel – keine qualitative Erfahrung lässt sich als solche aus Daten und Informationen ableiten. Bewusstsein ist überhaupt nicht das geistlose Durchlaufen von Datenzuständen – es ist *Selbstbewusstsein*. Es ist *für mich*, dass ich Schmerzen habe, wahrnehme, verstehe oder denke. Niemand weiß zwar genau, wie dieses Selbstbewusstsein vom Organismus überhaupt hervorgebracht wird, sicher jedoch nicht durch bloße Programme, denn Programme und ihre Trägersysteme erleben nichts. Der Output solcher Systeme ist allenfalls die Simulation von Erleben, nicht das Original – was aussieht, schwimmt und schnattert wie eine Ente, ist eben noch lange keine Ente.

Nun ist bereits die Annahme, das Gehirn sei eine Art Computer mit Speichern und Recheneinheiten, der wie der heimische PC Eingaben zu Ausgaben verarbeitet, ein immer wieder verbreiteter Irrtum. So kann man im Gehirn, anders als im Computer, keine Hardware von Software unterscheiden, denn jede Denk- bzw. Gehirnaktivität verändert zugleich die synaptischen Verknüpfungen, also die organische Basis von Bewusstsein. Es gibt im Gehirn auch keine „Datenspeicherung“, sondern nur variable Reaktionsmuster, die bei Bedarf in ähnlicher, aber nie exakt gleicher Form aktiviert werden. Anders als im Computer geschieht im Gehirn also nie zweimal das Gleiche. Zudem lässt sich die neuronale Signalübertragung nicht exakt in Programmen aus Nullen und Einsen wiedergeben, da sie von Neuro-



modulatoren (Opioiden, Neuropeptiden, Monoaminen u. a.) ständig beeinflusst wird, die vor allem für das Emotionserleben unerlässlich sind. Schließlich besteht das Gehirn zum größten Teil (85%) aus einem Stoff, der die neuronalen Prozesse ermöglicht, aber einen Computer sofort zum Kurzschluss bringen würde – nämlich aus Wasser.

All dies macht bereits deutlich, dass es sich beim Gehirn eben nicht um einen „biologischen Computer“ handelt. Doch das Wichtigste ist: das Gehirn kann seine Funktionen *für sich genommen* gar nicht erfüllen. Es ist ein Organ des lebendigen Organismus, mit dem es aufs Engste vernetzt ist.<sup>12</sup>

Bereits unser primäres, noch unreflektiertes Erleben beruht auf der Interaktion des Gehirns mit dem übrigen Körper: Bewusstsein entsteht nicht erst im Kortex, sondern es resultiert aus den vitalen Regulationsprozessen, die den ganzen Organismus miteinbeziehen, und die im Hirnstamm und Zwischenhirn integriert werden.<sup>13</sup> Die Aufrechterhaltung der Homöostase und damit der Lebensfähigkeit des Organismus ist die primäre Funktion des Bewusstseins, wie es sich in Hunger, Durst, Schmerz oder Lust manifestiert. So entsteht das leibliche Selbsterleben, das *Lebensgefühl*, das allen höheren geistigen Funktionen zugrunde liegt. Dies lässt sich auch so ausdrücken: *Alles Erleben ist eine Form des Lebens*. Ohne Leben gibt es kein Bewusstsein und auch kein Denken.

All diese lebendigen Prozesse lassen sich von elektronischen Systemen nicht simulieren. Auch das Human Brain Project der EU, das bis 2023 eine Computersimulation des gesamten Gehirns erreichen sollte, hat mit der tatsächlichen Aktivität eines Gehirns im Organismus wenig zu tun, erst recht nichts mit Bewusstsein. Selbst eine perfekte Computersimulation des Gehirns wäre ebenso wenig bewusst wie die perfekte Computersimulation eines Wirbelsturms uns nass machen oder umwehen würde. Bewusstes Erleben setzt Leiblichkeit und damit *biologische Prozesse in einem lebendigen Körper* voraus. Nichts davon ist im Human Brain

---

<sup>12</sup> Vgl. zum Folgenden ausführlich T. Fuchs (2017). *Das Gehirn – ein Beziehungsorgan. Eine phänomenologisch-ökologische Konzeption*. 5. Aufl. Stuttgart: Kohlhammer, insbesondere S. 136–140.

<sup>13</sup> Vgl. dazu Damasio, A. 2011. *Selbst ist der Mensch: Körper, Geist und die Entstehung des menschlichen Bewusstseins*. München: Siedler; Panksepp, J. 1998. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford, New York: Oxford University Press, sowie Fuchs (2017).

Project zu finden. Nur Lebewesen sind bewusst, empfinden, fühlen oder denken. Und Personen sind Lebewesen, nicht Programme.

### 3 Programme sind keine Personen

Gehen wir nun den umgekehrten Weg: Warum sind Programme niemals Personen?

Beginnen wir mit dem problematischen Begriff der „künstlichen Intelligenz“ selbst. Was meinen wir eigentlich, wenn wir von Intelligenz sprechen? Das lateinische *intelligere* bedeutet „einsehen, verstehen, begreifen“. Wer intelligent ist, verfügt also zumindest über ein grundlegendes Verständnis davon, was er tut und was um ihn herum vorgeht. Vor allem aber ist er in der Lage, sich selbst und seine Situation aus einer übergeordneten Perspektive zu sehen, sodass er kreative Lösungen finden kann, die auf einem Überblick oder „Umweg“ beruhen. Intelligent handelt zum Beispiel, wer sich beim Weg durch einen Wald Zeichen macht, um später seinen Rückweg zu finden; oder wer seine Urlaubsreise um eine Woche verschiebt, weil er nicht bei Ferienbeginn in den üblichen Stau geraten will. Dazu muss er sich in ein Verhältnis zur jeweiligen Situation setzen und sich selbst gewissermaßen „von außen“ sehen, also über *Selbstbewusstsein* oder *Reflexivität* verfügen.

Nun haben wir schon gesehen, dass ein Computersystem *nicht das Geringste von dem versteht, was es tut*. Erst recht vermag es nicht, sich auf sich selbst zu beziehen, sich von außen zu sehen. Daher kann es auch nicht intelligent genannt werden, selbst wenn es Fähigkeiten simuliert, die wir bei Menschen als Ausweis von Intelligenz verstehen. Kein Übersetzungsprogramm versteht ein Wort von dem, was es übersetzt, kein Schachcomputer weiß, dass er Schach spielt. Sophia, die kein einziges Wort begreift, das sie ausspricht, wird niemals intelligent werden, selbst wenn sie irgendwann in allen denkbaren Situationen die perfekte Antwort geben kann. Intelligenz setzt Selbstbewusstsein voraus.

Wenn wir also mit „Intelligenz“ die Fähigkeit bezeichnen, sich selbst oder eine Situation aus einer übergeordneten Perspektive zu erfassen, um Probleme geschickt zu lösen und daraus zweckvolles Handeln abzuleiten, dann können wir solche Fähigkeiten gewiss keinem Apparat ohne Bewusstsein zuschreiben. Die Bezeichnung „intelligent“ wird hier nur uneigentlich gebraucht, so wie man von einem „Smartphone“ auch nicht annimmt, dass es wirklich „smart“, also klug ist – es führt nur blind Programme aus, die man als „klug entwickelt“ bezeichnen

kann. Erst recht gilt das, wenn wir an praktische, emotionale oder schöpferische Intelligenz denken – hier lassen uns die „intelligenten Systeme“ völlig im Stich.

Nun werden zahlreiche Einwände auftauchen, die auf die fortgeschrittenen Fähigkeiten von „intelligenten Programmen“, „lernenden Systemen“ etc. verweisen. Betrachten wir daher einige solcher angeblicher Intelligenzleistungen etwas näher.

- Lösen Computer Probleme? – Nein, denn Probleme stellen sich für sie gar nicht. Als Problem – vom griechischen *próblema*, das heißt „das zur Lösung Vorgelegte“ – bezeichnen wir ein Hindernis oder eine Schwierigkeit bei der Bewältigung einer Aufgabe, etwa weil sich verschiedene Anforderungen widersprechen, ein veränderter Blickwinkel zur Lösung erforderlich ist etc. Doch „Hindernisse“ und „Aufgaben“ gibt es nur für *zielorientierte Wesen*, die einen Weg von der gegenwärtigen zu einer künftigen Situation suchen und die Lösung vorstellend vorwegnehmen können. Vor ein Problem gestellt zu sein und es zu bewältigen, ist daher an den *bewussten Lebensvollzug* gebunden. Mitunter mag man die Aufgabe *mit Hilfe* eines Computers lösen, doch dann stellt die programmierte Berechnung nur für den Nutzer selbst eine Lösung seines Problems dar – der Computer kann das Problem nicht einmal erkennen.
- Computer, obgleich sie so heißen, *rechnen* auch nicht. Denn Rechnen bedeutet eine Operation, bei der *wir* auf ein Ergebnis abzielen und dann beurteilen, *ob es richtig ist*. Der bloße Ablauf eines Programms ist keine Rechnung, denn richtig oder falsch gibt es für das Programm nicht. Computer rechnen ebenso wenig wie die Uhr die Zeit misst, denn die Uhr weiß von Zeit gar nichts. Nicht der Computer ist also der Rechner, sondern ich selbst rechne *mit Hilfe* des Computers. Dass ich die dazu notwendigen Algorithmen nicht kenne, sondern nur der Programmierer (so wie ich auch den Uhrenmechanismus nicht kenne, sondern der Uhrmacher), ändert daran nichts. Es liegt im Wesen jedes komplexeren Instruments, dass in ihm menschliche Intelligenz gewissermaßen sedimentiert ist; das macht das Instrument selbst aber nicht intelligent.
- Aus dem gleichen Grund treffen Computer auch keine Entscheidungen. Sich zu entscheiden, setzt das Bewusstsein von alternativen Möglichkeiten voraus, die in der Imagination vorweggenommen werden: Ich könnte dieses tun oder jenes. Auch dazu bedarf es einer Ziel- und Zukunftsorientierung ebenso

wie der Unterscheidung von Realität und kontrafaktischer Imagination, und für beides hat der Computer keinen Sinn – er kennt weder ein „Noch-nicht“ noch ein „Als-ob“. Wenn nun ein medizinisches Expertensystem ausgehend von Patientendaten eine Therapie errechnet, trifft es keine Entscheidungen, ebenso wenig wie ein selbstfahrendes Auto sich entscheidet, ob es lieber einen Unfall mit einer alten oder jungen Person in Kauf nimmt. Die Entscheidungen über die Programme und ihre Bewertungskriterien sind längst vorher getroffen worden, nämlich von den Programmierern. Alles andere sind nur noch sich abspulende Algorithmen.

- Gibt es aber nicht doch „zielsuchende Systeme“, zum Beispiel „smart bombs“, die ihren Flug selbst beeinflussen können, weil sie über ein internes Modell ihrer Operationen verfügen? – Natürlich sucht auch eine „intelligente Bombe“ gar nichts, da sie keine intentionale Beziehung zu ihrem Zielobjekt hat. Jede Flugkorrektur dient nur der internen Sollwertregulierung des Mechanismus und geschieht rein momentan, ohne sich auf ein vorgestelltes Ziel in der Zukunft zu richten. Für dieses Ziel selbst bleibt der Mechanismus blind und taub, *denn er ist sich selbst nicht vorweg*. Nur bewusstes Erleben nimmt die Zukunft vorweg und ist protentional – wünschend, strebend, erwartend oder befürchtend – auf Mögliches gerichtet. Dass der einprogrammierte Zielabgleich eine „Zielsuche“ oder „Zielvorhersage“ darstelle, ist daher eine falsche Redeweise. Nur *für den Ingenieur* oder *für den Schützen* hat die Bombe ein Ziel.

Die Leistungen der klassischen Computersysteme sind also im Vergleich zur menschlichen Intelligenz recht überschaubar. Ihre partielle Überlegenheit beruht auf ihrer extremen Spezialisierung auf berechenbare Aufgaben, gewissermaßen einer Fachidiotie, und auf ihrer ungeheuren Verarbeitungsgeschwindigkeit. Doch ihre vermeintliche Intelligenz ist nur geborgt: Jedes dieser Programme ist nur so „klug“ oder raffiniert wie der Programmierer, der es entworfen hat.

Inzwischen haben wir es allerdings mit einer neuen Generation von künstlicher Intelligenz zu tun, nämlich „lernenden Maschinen“. Es handelt sich um künstliche neuronale Netzwerke, die in der Lage sind, die adaptiven Fähigkeiten des Gehirns zu simulieren. Ähnlich wie bei biologischen Synapsen sind die Verknüpfungen zwischen den im Computer modellierten Neuronen numerisch gewichtet und passen sich im Verlauf eines Trainingsprozesses dem Input von Signalen an (*deep*

*learning*). Häufig genutzte Verknüpfungen werden verstärkt, selten genutzte unterbrochen. Dabei werden dem System tausende ähnliche Muster dargeboten, zum Beispiel verschiedene Versionen eines Gesichts, bis es auf die mit der größten Wahrscheinlichkeit wiederkehrende Pixelanordnung reagiert, also das Gesicht „erkennt“. Solche Systeme sind auch in der Lage, zum Beispiel Hunde von Katzen zu unterscheiden, Stimmen über Handymikrophone zu identifizieren oder Übersetzungen vorzunehmen – sie sind bereits allgegenwärtig.

All dies bedeutet zweifellos einen bemerkenswerten Fortschritt – doch kann man hier wirklich von „Erkennen“ und „Lernen“ sprechen? Natürlich *erkennt* ein System gar nichts, denn das Erlebnis des Wiedererkennens, der Vertrautheit oder *Ähnlichkeit* geht ihm völlig ab. Ein „lernendes System“ zum Beispiel war in der Lage, Abbildungen von Kühen in den verschiedensten Positionen und Ausschnitten zu identifizieren. Doch als man ihm eine Kuh vor einem Meeresstrand präsentierte, hielt es die Kuh für ein Schiff – bislang hatte es nämlich nur Bilder von Kühen auf Wiesen und Feldern verarbeitet. Ohne diesen Kontext geriet das System in die Irre.<sup>14</sup> Das heißt aber im Rückschluss: Trotz abertausenden von Bilddurchläufen hatte es auch zuvor keine einzige Kuh *erkannt* – jedes Kleinkind hätte die Kuh am Strand als Kuh gesehen, und das schon nach wenigen Kontakten mit Kühen. Die Gestalt und der Begriff einer Kuh lassen sich nun einmal nicht auf eine statistische Wahrscheinlichkeit von Pixelübereinstimmungen reduzieren. Man sollte daher auch nicht von lernenden, sondern besser von „adaptiven Systemen“ sprechen. Nur Lebewesen können lernen.

Wir sehen: Den Begriff der Intelligenz haben wir unseren Maschinen allzu voreilig zugebilligt. Die Bezeichnung „künstliche Intelligenz“ ist wohl nicht mehr aus der Welt zu schaffen. Doch wir sollten uns immer dessen bewusst bleiben, dass zwischen den Rechen- und Anpassungsleistungen eines Computersystems und den Wahrnehmungen, den Einsichten, dem Denken und Verständnis eines Menschen nicht nur ein gradueller, sondern ein fundamentaler Unterschied besteht.

---

<sup>14</sup> Vgl. B. Schölkopf, “Symbolic, Statistical and Causal Intelligence”, Vorlesung am Marsilius-Kolleg der Universität Heidelberg, 16.07.2020.

#### 4 Resümee: Simulation und Original

Die Fortschritte der Simulationen machen es notwendig, die kategorialen Unterschiede zwischen menschlicher und künstlicher Intelligenz klarzustellen. Intelligenz im eigentlichen Sinne des Wortes ist an Einsicht, Übersicht und Selbstbewusstsein gebunden: *verstehen, was man tut*. Und Voraussetzung für Bewusstsein wiederum ist nicht nur ein Gehirn, sondern ein lebendiger Organismus. *Alles Erleben beruht auf dem Leben*.

Der Begriff einer bewusstlosen Intelligenz ist ein hölzernes Eisen. Was an den Leistungen der KI-Systeme intelligent erscheint, ist nur eine Projektion unserer eigenen intelligenten Fähigkeiten. Ihre scheinbare Zielverfolgung oder Problemlösung, ihre vermeintlichen Vorhersagen oder Bewertungen leiten sich ausnahmslos von unseren eigenen Zielen, Problemen, Lösungen und Bewertungen ab, die wir zu Programmen formalisiert und ausgelagert haben, um uns die Arbeit des Durchrechnens zu ersparen. Es ist im Prinzip nichts anderes als dass die Uhr für uns die Zeit misst, weil wir unsere eigene Erfahrung von regelmäßigen Naturprozessen in einen nützlichen Mechanismus ausgelagert haben. So unsinnig es wäre, wollte man der Uhr ein Wissen von der Zeit zusprechen, so unsinnig ist es, einem „intelligenten Roboter“ ein Verständnis von Sprache oder einem „intelligenten Auto“ die Wahrnehmung von Gefahren zuzuschreiben. Das Bewusstsein davon, dass wir es in der KI nur mit einer Externalisierung unserer eigenen Rechen- und Denkfähigkeit, mit einer Projektion unserer selbst zu tun haben, geht uns zunehmend verloren.

Doch alle künstlichen Systeme bleiben von unserem eigenen, bewussten und zweckgerichteten Lebensvollzug abhängig. Alle Programme, die in solchen Systemen ablaufen, sind nur *für uns* Programme, also zielgerichtete Prozesse. Den Systemen geht es dabei um nichts. Sie wissen, erkennen, verstehen nichts, denn sie *erleben* nichts. Die Ähnlichkeit ihrer Funktionen mit menschlichen Leistungen mag noch so täuschend, ihre spezialisierte Überlegenheit noch so verblüffend sein – wir sollten uns nicht täuschen lassen. Unsere vermeintlichen künstlichen Doppelgänger sind und bleiben unsere Erzeugnisse; ihre Intelligenz ist nur die Projektion unserer eigenen.

Aber auch wenn es keine bewusstlose Intelligenz gibt, und umgekehrt die noch so perfekte Simulation von Intelligenz kein Bewusstsein erzeugt – die Fortschritte der Simulationstechnik werden ihre Wirkung nicht verfehlen. Der Anthropomorphismus, der unserem Wahrnehmen und Denken inhärent ist, verleitet uns nur

allzu leicht dazu, unseren Maschinen menschliche Intentionen, Handlungen, ja Gefühle zuzuschreiben. Spätestens bei humanoiden Robotern lebt der Animismus wieder auf, den wir für ein überwundenes Stadium der Vorgeschichte gehalten haben oder noch bei Kleinkindern beobachten können. Dann wäre die simulierte Ente am Ende doch eine Ente, und das Als-ob der Simulation ginge verloren – sei es, weil der kategoriale Unterschied nicht mehr verstanden wird, oder weil er letztlich als gleichgültig erscheint. Dass die KI-Systeme angeblich jetzt schon „denken“, „wissen“, „planen“, „voraussagen“ oder „entscheiden“, bahnt solchen Grenzauflösungen den Weg. Hans Jonas' Warnung gilt heute erst recht:

„Der menschliche Verstand hat eine starke und, wie es scheint, unwiderstehliche Neigung, menschliche Funktionen in den Kategorien der sie ersetzenden Artefakte, und Artefakte in den Kategorien der von ihnen versehenen menschlichen Funktionen zu deuten. [...] Die Benutzung einer bewusst doppelsinnigen und metaphorischen Terminologie erleichtert diese Hin- und Her-Übertragung zwischen dem Artefakt und seinem Schöpfer.“<sup>15</sup>

Freilich kann man alle Begriffe wie Denken, Entscheiden, Intelligenz oder Bewusstsein rein behavioristisch als Output definieren, wie es Turing bereits vorschlug. Damit heben wir allerdings die Maschinen auf unsere Stufe und degradieren uns selbst zu Maschinen.

Die eigentliche Gefahr, die sich daraus ergibt, dürfte darin bestehen, dass wir den Systemen freiwillig immer mehr Entscheidungen überlassen – Entscheidungen, die nur noch für wenige durchschaubar sind und sich der demokratischen Kontrolle entziehen. Je komplexer die Gesellschaft wird, desto attraktiver könnte es werden, Planungen und Bewertungen an Maschinen zu delegieren, so wie es heute bereits an der Börse der Fall ist – sei es, weil die Resultate für „objektiver“ erklärt werden, oder weil die Bereitschaft, persönliche Verantwortung angesichts der Komplexität der Welt abzugeben, immer weiter zunimmt.<sup>16</sup> Sind die Systeme uns nicht ebenso überlegen wie *Deep Blue* Gari Kasparow?

<sup>15</sup> Jonas, H. 1973. *Organismus und Freiheit. Ansätze zu einer philosophischen Biologie*. Göttingen: Vandenhoeck & Ruprecht, S. 166.

<sup>16</sup> Ein Beispiel dafür sind die in den USA immer häufiger durchgeführten Beurteilungen des Rückfallrisikos von Straftätern durch KI-Systeme (mit einem offensichtlichen Bias zuungunsten von Farbigen). Hier werden intransparente Programme zu Hilfsrichtern oder gar Entscheidungsinstan-

Doch die Bewertungen, die in KI-Systemen implementiert sind, stammen ausnahmslos menschlichen Wertsetzungen – kein „intelligentes System“ sagt uns von sich aus, was das Richtige, das Gute und ethisch Gebotene ist. Je mehr sich die Vorstellung von künstlicher Intelligenz als einer vermeintlich überlegenen Form der Analyse, Vorhersage und Bewertung etabliert, desto eher gerät in Vergessenheit, dass Entscheidungen mit all ihren Unwägbarkeiten letztlich immer nur von Menschen selbst getroffen werden können. Ebenso wird vergessen, dass es in Wahrheit einige wenige Konzerne und informationstechnische Eliten sind, die die maßgeblichen Entscheidungen treffen und die mittels Big Data immer mehr Bereiche der Gesellschaft zu kontrollieren vermögen. Die häufig geäußerten Warnungen vor einer „Machtübernahme intelligenter Maschinen“ sind zweifellos unsinnig, da sie bewusstlosen Systemen einen Eigenwillen unterstellen. Doch gerade als „dienstbare Geister“ können KI-Systeme die Machtverhältnisse in der Gesellschaft tiefgreifend verändern. „Nicht die Maschinen übernehmen die Kontrolle, sondern diejenigen, die die Maschinen besitzen und kontrollieren.“<sup>17</sup>

Die entscheidende Herausforderung der Künstlichen Intelligenz aber liegt wohl in der Frage, die sie an uns und unser Selbstbild richtet: Erschöpft sich unser Menschsein in dem, was sich in Simulation und Technologie übersetzen lässt? Besteht es allein in komplexen neuronalen Algorithmen, und ist unser Erleben nur ein Epiphänomen? – Gerade weil die Technik viele unserer spezialisierten Fähigkeiten übersteigt, fordert sie uns dazu auf, wieder neu zu entdecken, worin unser Menschsein eigentlich besteht – nämlich nicht in Information oder Daten, sondern in unserer lebendigen, fühlenden und verkörperten Existenz.

---

zen (vgl. L. Kirchner, J. Angwin, J. Larson, S. Mattu. 2016. „Machine Bias: There’s Software Used across the Country to Predict Future Criminals, and It’s Biased against Blacks“. Pro Publica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>).

<sup>17</sup> Lenzen, M. 2018. *Künstliche Intelligenz. Was sie kann und was uns erwartet*. München: Beck, S. 247.



## Über den Autor

**Thomas Fuchs** ist Karl-Jaspers-Professor für Philosophische Grundlagen der Psychiatrie und Psychotherapie an der Universität Heidelberg. Er ist Leiter der Sektion „Phänomenologische Psychopathologie und Psychotherapie“ der Klinik für Allgemeine Psychiatrie am Universitätsklinikum Heidelberg und Koordinator des Marsilius-Projekts „Verkörperung als Paradigma einer evolutionären Kulturanthropologie“. Zudem ist er Forschungsstellenleiter der Karl Jaspers-Gesamtausgabe der Heidelberger Akademie der Wissenschaften und Vorsitzender der Deutschen Gesellschaft für Phänomenologische Anthropologie, Psychiatrie und Psychotherapie (DGAP).

### **Korrespondenzadresse:**

Prof. Dr. Dr. Thomas Fuchs  
Klinik für Allgemeine Psychiatrie  
Zentrum für Psychosoziale Medizin  
Universitätsklinikum Heidelberg  
Voßstr. 4  
D-69115 Heidelberg

E-Mail: [thomas.fuchs@med.uni-heidelberg.de](mailto:thomas.fuchs@med.uni-heidelberg.de)

Homepage: <https://www.uni-heidelberg.de/fakultaeten/philosophie/philssem/phaenomenologie/>