

Reinhard Heil, Christian Wadehul und Karsten Wendland
Robotik und die Zuschreibung von Bewusstsein¹

1 Wandert die Krone der Schöpfung weiter?

Die beeindruckenden Leistungen moderner Roboter mit sogenannter Künstlicher Intelligenz (KI) erstaunen die Menschen in allen Kulturkreisen. Selbst die deutsche Bundeskanzlerin fand sich bereits im angeregten Gespräch mit dem Roboter Sophia wieder, den dessen Hersteller Hanson Robotics stolz als Demonstrator seiner Technikzukünfte um die Welt schickt.² Auch wenn es in diesem Gespräch der beiden Damen vor Publikum auf der Konferenz „Morals & Machines 2018“ vor allem um den Austausch von Artigkeiten und um Fußballergebnisse ging, war die Inszenierung gelungen und die Positionierung der Maschine Sophia als dialogfähiges Gegenstück in menschenähnlicher Gestalt wirksam umgesetzt. Das Königreich Saudi-Arabien hatte Sophia bereits ein halbes Jahr zuvor die Staatsbürgerschaft verliehen, nachdem das Gerät auf einer Konferenz in Riad vor internationalem Publikum Rede und Antwort gestanden und die Anwesenden beeindruckt hatte. Teile der arabischen Welt reagierten entsetzt und fragten in öffentlichen Foren, wie es sein könne, dass Sophia als „Staatsbürgerin“ mehr Rechte habe als Söhne des Landes, deren Väter keine Saudis seien, ganz zu schweigen von Gastarbeitern und Einwanderern, und ob sie überhaupt die richtige Religion habe.³ Wenige Tage später verkündete Sophia in einem öffentlichen Statement, sich zukünftig auch reproduzieren zu wollen.

Dialogfähige Systeme in Roboterform sind dabei, in unseren Gesellschaften „anzukommen“. Dabei müssen sie keineswegs so menschenähnlich aussehen wie Sophia und ihre Geschwister des Herstellers aus Hongkong. Als Roboter werden weithin Maschinen verstanden, die ausgestattet mit Sensoren und Aktoren ferngesteuert oder automatisch Dinge verrichten, die sonst der Mensch erledigen müsste – etwa in gefährlichen Feuerwehreinheiten, in der Industrie oder beim komfortablen Rasenschnitt auf dem heimischen Anwesen. Optisch sind diese Roboter in der Regel sofort als Maschinen erkennbar,

¹ Die Erstellung dieses Beitrags erfolgte im Kontext des durch das Bundesministerium für Bildung und Forschung (BMBF) geförderten Forschungsprojekts „Abklärung des Verdachts aufsteigenden Bewusstseins in der Künstlichen Intelligenz (KI-Bewusstsein)“ im Förderprogramm „Innovations- und Technikanalyse“, Förderkennzeichen 16ITA202.

² Vgl. WELT Nachrichtensender: *Morals & Machines. Roboter fragt Merkel, wie sie über das WM-Aus denkt* (Nachrichten-Fernsehbeitrag auf YouTube), veröffentlicht am 28.6.2018, <https://www.youtube.com/watch?v=z0F3QxtaG9M> (Abruf der in diesem Beitrag angegebenen Internetseiten: 11.6.2019).

³ Vgl. Arab News: *Saudi Arabia Becomes First Country to Grant Citizenship to a Robot* (Nachrichten-Fernsehbeitrag), veröffentlicht am 27.10.2017, <http://www.arabnews.com/node/1183166/saudi-arabia>.

die im derzeitigen Sprachgebrauch bei ausgeprägter Autonomie als „intelligent“ oder gar „brillant“ beschrieben werden, da sie selbsttätig, präzise und flexibel ihre Programme abarbeiten und Menschen dadurch entlasten und deren Möglichkeiten deutlich erweitern können. Ihrer Produktivität wegen stellen sich mittlerweile erste Fragen nach einer steuerlichen Veranlagung, etwa von Industrierobotern, sowie der Behandlung solcher Maschinen als Rechtssubjekte⁴ mit eigenen Ansprüchen.

Kommen Roboter allerdings in menschenähnlicher Gestalt daher, wie es etwa bei Sophia und asiatischen Pflegerobotern der Fall ist und wie es uns seit Jahrzehnten in fiktionalen Romanen, Filmen und Computerspielen vorgestellt wird, tritt in unseren westlichen Kulturkreisen ab einem bestimmten Grad an Ähnlichkeit zu uns selbst ein deutliches Unbehagen hervor – solche Maschinen erscheinen uns unheimlich und gruselig, weil sie einerseits so menschenähnlich und doch irgendwie erkennbar anders sind, insbesondere wenn sie sich bewegen und als sprichwörtliche „seelenlose Zombies“ noch furchteinflößender wirken können als unbewegte menschliche Leichname. Nach diesem als *uncanny valley* beschriebenen Phänomen ist es uns lieber, wenn das Design der Roboter auch ausreichend nichtmenschliche Züge aufweist und sie im Aussehen und Verhalten eher als niedliche Kameraden wahrgenommen werden; durch zu stark ausgeprägten Anthropomorphismus fallen sie in eine Akzeptanzlücke.⁵

Wäre dieses Unbehagen überwindbar, wenn Roboter tatsächlich so menschenähnlich wären, dass der Unterschied des Roboters zum Menschen nicht mehr ausgemacht werden könnte? An dieser scheinbar einfachen Frage hängen Bündel weiterer grundsätzlicher Überlegungen und Positionen sowie unterschiedliche wissenschaftliche, philosophische, wirtschaftliche und allgemein gesellschaftliche Diskurse. Dass Roboter den Menschen immer besser werden imitieren können, wird kaum mehr in Zweifel gezogen. Auch dass Roboter zukünftig noch mehr Aufgaben besser werden umsetzen können als der Mensch, ist unstrittig. Wenn die Menschenähnlichkeit aber so weit ginge, dass Roboter ein eigenes „Bewusstsein“ entwickeln würden, mit allen damit zusammenhängenden Konsequenzen, hätten wir eine gänzlich neue Situation. Obwohl unbeantwortet ist, ob Roboter jemals ein Bewusstsein entwickeln könnten, und obwohl mit dem Bewusstseinsbegriff selbst, wie wir sehen werden, landauf, landab sehr flexibel umgegangen wird und ebenfalls unbeantwortet ist, wie Bewusstsein denn entsteht, ob es irgendwo herkommt und wie es mit unserem Selbst zusammenhängt, erfährt die Rede von (sich selbst) bewusst werdenden Robotern und KI-Systemen gegenwärtig eine Hochkonjunktur. Diese wird genährt von Marketing- und Medienschaffenden, von

⁴ Vgl. Jonathan Erhardt / Martino Mona: Rechtsperson Roboter? Philosophische Grundlagen für den rechtlichen Umgang mit künstlicher Intelligenz, in: Sabine Gless / Kurt Seelmann (Hg.): Intelligente Agenten und das Recht – Verantwortungszuschreibung in Antike und Moderne, Baden-Baden 2016, 61 – 94.

⁵ Vgl. Masahiro Mori: Das unheimliche Tal, in: Konstantin Daniel Haensch / Lara Nelke / Matthias Planitzer (Hg.): Uncanny Interfaces, Hamburg 2019, 212 – 219.

einflussreichen Laien und auch von manchen Wissenschaftlern, die mit Hinweisen auf ein mögliches Bewusstsein von Maschinen kokettieren, entsprechende Technikzukünfte prophezeien und damit Narrative in die Welt setzen, die letztlich bewusste Roboter verheißen. Zeitgleich arbeiten technische Entwickler tatsächlich daran, mit neueren Ansätzen und großer Zuversicht mit Bewusstsein ausgestattete Roboter bauen zu wollen. Nüchtern betrachtet handelt es sich nach heutigem Erkenntnisstand in all diesen Fällen um Zuschreibungen von etwas, das wir – meist unscharf – „Bewusstsein“ nennen, an Maschinen, die mit eben dieser geheimnisvollen Eigenschaft durch uns von außen attribuiert werden, ohne sie zu haben. Auf dem Nährboden der angebotenen Narrative und eigener Fantasien, Erwartungen und Befürchtungen rechnen allerdings offenbar zunehmend mehr Menschen damit, dass Roboter und KI-Systeme irgendwann irgendeine Art von Bewusstsein entwickeln und uns sogar insgesamt überholen und schließlich abhängen könnten. Dann wären nicht mehr wir Menschen die Krone der Schöpfung, sondern würden diese an die von uns geschaffenen Geschöpfe weiterreichen und vielleicht selbst sogar über kurz oder lang vollständig von der Bühne abtreten.

2 Vom Geist in der Maschine zur KI-Kirche

Im breiteren öffentlichen Verständnis wurde unser Bild von Robotern und KI in den letzten Jahrzehnten stark durch populäre Science-Fiction-Abenteuer geprägt, etwa durch den Film „Terminator“, in dem eine KI namens *Skyнет* die Weltherrschaft übernimmt. Die Maschinen werden im Film als übermächtig und bedrohlich angelegt und sind dem Menschen keineswegs untertan, sondern als streng logisch denkende und zweckrationale Akteure konsequent darin, die unvollkommenen Menschen für ihre Sünden abzustrafen. Ebenso polarisierend konnten die Zuschauer schon in den 1980er Jahren auf der Kinoleinwand verfolgen, wie der Militärroboter *Nummer 5*, aus düsteren Wolken vom Blitzschlag getroffen, im Staube seiner verkohlten Aggregate wiedererweckt, fortan – fulguriert – ein ganz anderer war als vorher, nämlich durchdrungen von Emotionen und Wünschen nach Freundschaft, Gerechtigkeit und Barmherzigkeit.

Diese Idee einer „sich selbst bewusst werdenden“, quasi „erwachenden“ KI hat sich in den letzten Jahren Zug um Zug vom Science-Fiction-Genre emanzipiert – durch gewonnene Schach- und Go-Weltmeisterschaften, selbstfahrende Autos, Assistenzsysteme wie *Siri* und *Alexa* und nicht zuletzt durch menschenähnliche Roboter wie die eingangs erwähnte *Sophia*, die durch sogenanntes maschinelles Lernen Sprache nicht nur „erkennen“, sondern auch dialogfähig generieren kann. Zwar steht *Sophia* in der KI-Community im Verdacht, „geskriptet“ zu sein, was bedeutet, dass sie ihre Aussagen nicht immer selbst entwickelt, sondern auch von außen eingeflüstert bekommen kann. Umso dringlicher ist es mithin, in der Diskussion um bewusste Roboter zu klären, was gemeint ist, wenn von Bewusstsein gesprochen wird, um die unterschiedlichen Erwartungen

an ein künstliches Bewusstsein genauer einordnen zu können. Weitgehend akzeptiert sind die folgenden Ausprägungen, die dem Bewusstseinsbegriff zugeordnet werden:

- Vigilanzbewusstsein (Wachsein vs. Schlaf)
- Kognitives Bewusstsein (die Aufmerksamkeit auf etwas richten)
- Zustandsbewusstsein (über einen bewussten Zustand verfügen)
- Phänomenales Bewusstsein (fühlen, empfinden, spüren usw.)
- Handlungsbewusstsein (etwas mit Absicht tun)
- Selbstbewusstsein (sich seiner Gedanken, Empfindungen usw. bewusst sein)

Aus einer rein technischen Haltung heraus sind all diese Ausprägungen – mit Ausnahme der letzten – Robotern durchaus „beizubringen“, indem sie imitiert bzw. simuliert werden. Verschiedene Entwickleransätze greifen gesonderte Teilaspekte wie etwa das „Sehen“ heraus, das die Verarbeitung und Modellierung der Außenwelt zur zielgerichteten und kollisionsfreien Bewegung in der eigenen Umgebung ermöglichen soll, wie es z. B. beim automatisierten Fahren benötigt wird.⁶ Die verbleibenden Differenzen zwischen der Simulation im Roboter und unserem Erleben dieser Ausprägungen (wie etwa: absichtlich einen Apfel zu greifen) spielen in Entwicklerkreisen typischerweise keine Rolle, da die „Als ob“-Analogien einen passenden Arbeitsjargon anbieten und zur freimütigen Verwendung des Bewusstseinsbegriffs einladen. Das Abstands-Delta von der Simulation der genannten Bewusstseinsausprägungen zu ihren jeweiligen menschlichen Erlebensvarianten wird als selbstverständlich gegeben akzeptiert und spielt für die Arbeitspraxis keine größere Rolle. Ein Beharren auf diesen Differenzen vonseiten technischer Nichtfachleute, möglicherweise sogar verbunden mit der Frage, ob Bewusstsein denn nun ein Begriff der Technikwissenschaften geworden sei, wird typischerweise als rhetorische oder begriffspolizeiliche Spitzfindigkeit eingeordnet und verschließt mögliche Dialoge, anstatt sie zu öffnen.

Der Anspruch an ein „richtiges“ Bewusstsein, wie wir es von uns Menschen kennen, bezieht sich jedoch in der Regel auf das letzte Merkmal und stellt die Frage, ob eine Künstliche Intelligenz über ein Bewusstsein ihrer Selbst im Sinne einer Erste-Person-Perspektive verfügen könne. Die Haltungen und Standpunkte zu dieser Frage sind in den Wissenschaften vielfältig und nicht selten emotional aufgeladen. Überdies werden weitere Begriffe im gegenwärtigen Diskurs aufgerufen, deren Grundcharakter von ähnlicher Unschärfe und Zuschreibungseignung ist. So wurde die Rede von einer „Superintelligenz“ vom Philosophen Nick Bostrom ins Spiel gebracht. Selbige werde erreicht, sobald eine KI in ihrer Leistungsfähigkeit die Menschen überholt habe und diese dann nicht mehr benötige.⁷ Ebenso haben vage Ausdrücke wie der „Geist in der Maschine“

⁶ Vgl. Ralf Otte: Künstliche Intelligenz für dummies, Weinheim 2019.

⁷ Vgl. Nick Bostrom: Superintelligenz. Szenarien einer kommenden Revolution, Frankfurt a. M. 2016.

und die „Seele des Computers“ ihre Diskursarenen und überzeugten Anhängerschaften. In Kalifornien wurde unlängst die KI-Kirche „Way of the Future“⁸ gegründet, deren Anhänger am liebsten einen Gott aus der Technik erzeugen und diesen sodann anbeten möchten. In die Kirchengemeinde kann man bereits jetzt eintreten und durch Hinterlegung des eigenen Namens ein Bekenntnis ablegen, sodass die erwartete KI-Gottheit dies später aufgreifen, würdigen und anrechnen könnte. Kirchengründer Anthony Lewandowski, der fachlich als KFZ-Ingenieur ausgewiesen ist und mit Entwicklungen zum autonomen Fahren auf sich aufmerksam gemacht hat, bedient mit seiner Organisation nicht nur transhumanistische Erlösungsfantasien, sondern erhöht diffuse Erwartungen an ein maschinelles Bewusstsein auf das Niveau einer Gottheit.

3 Ging das Bewusstsein an die Neurowissenschaften verloren?

Mit Begrifflichkeiten wie Bewusstsein, Geist, Seele oder Gott wird in den beschriebenen gegenwärtigen Zusammenhängen flexibler umgegangen, als es sich jene wünschen, die sich aus ihrer wissenschaftlichen Fachlichkeit heraus für diese Begriffe zuständig erklären. Für viele Vertreter der Geisteswissenschaften ist es kaum akzeptabel, wie waghalsig und unreflektiert die genannten Begriffe im Diskurs um KI und Bewusstsein als Metaphern, Analogien oder schlicht als Platzhalter eingesetzt werden für etwas, was man nicht so genau beschreiben kann, über das man aber trotzdem spricht und dabei innerhalb der technisch ausgerichteten Diskursarenen offenbar sogar gemeinsame Verständigungen erzielt.

Lange Zeit bestimmten Philosophie und Psychologie, was unter Bewusstsein zu verstehen ist bzw. wie über Bewusstsein geredet wird. Dies hat sich in den letzten Jahrzehnten dem Zeitgeist entsprechend deutlich zu den sogenannten Neurowissenschaften hin verschoben, in denen Bewusstsein zumeist als Folge neuronaler Prozesse verstanden und sein Entstehen im Gehirn verortet wird. Der Gehirnzentrismus hat Konzepte, die Bewusstsein beispielsweise in Verbindung mit Verkörperung und Interaktion fassen, beinahe vollständig verdrängt.

Derzeit setzt langsam wieder ein Umdenken ein aus der Erkenntnis heraus, dass die bisherigen Beiträge der Neurowissenschaften zur Bewusstseinsdebatte, gerade wegen des Gehirnzentrismus, insgesamt weniger gehaltvoll gewesen sind als erhofft. Beispielsweise stellen die Arbeiten des Neurowissenschaftlers Maxwell Bennett traditionelle Paradigmen der Neurowissenschaft infrage, die dem Gehirn anstatt der Person psychologische Fähigkeiten wie Denken, Wahrnehmen und Erinnern zuschreiben. Gemeinsam mit dem Philosophen Peter Hacker versucht Bennett, die begrifflichen Grundlagen

⁸ Way of the Future Church: Humans United in support of AI, committed to peaceful transition to the precipice of consciousness, <http://www.wayofthefuture.church>.

der Neurowissenschaft zu klären.⁹ Dabei betonen sie die Unterschiede zwischen den Erklärungsansätzen von philosophischer Erkenntnistheorie und empirischer Gehirnforschung, wobei die jeweilige disziplinäre Eigenberechtigung nicht in Abrede gestellt wird. Es sei aber eine Begriffsverwirrung, wenn behauptet werde, der Geist besitze eine immaterielle oder materielle Substanz (Neurowissenschaft), das Gehirn denke, nehme wahr oder rechne. „Das Gehirn ist nicht das Bewusstseinsorgan. Man sieht mit seinen Augen und hört mit seinen Ohren, aber man ist nicht mit seinem Gehirn bei Bewusstsein. Am Bewusstsein ist nichts einzigartig privat.“¹⁰ Es handele sich um einen „mereologischen Fehlschluss“ – ein Teil werde fälschlicherweise für das Ganze gehalten –, der das neurowissenschaftliche Denken durchziehe, wenn dem Gehirn Eigenschaften zugeschrieben werden, die nur einem Lebewesen als Ganzem zukommen.¹¹ Es gibt zwar kein menschliches Bewusstsein ohne Gehirn, das macht aber das Gehirn noch nicht zum Sitz des Bewusstseins. Ja bereits die Frage nach dem „Sitz des Bewusstseins“ führt in die Irre, da es so als Eigenschaft einer Sache begriffen wird. Der angeführte mereologische Fehlschluss findet sich nicht nur in den Neurowissenschaften und Teilen der Philosophie und Psychologie, sondern vor allem auch in der öffentlichen Debatte.

Je nachdem, welches Bewusstseinsverständnis zugrunde gelegt wird, fällt auch die Antwort auf die Frage, ob Roboter (oder Künstliche Intelligenzen) bewusste Entitäten sein können oder nicht, unterschiedlich aus. Für die gegenwärtige wissenschaftliche Auseinandersetzung mit dem Gesamtphänomen „KI und Bewusstsein“ ist daher eine Haltung nützlich, die vordergründige begriffliche Differenzen aushalten kann, als Einstieg nutzt und die jeweiligen Bedeutungen und gemeinten Konzepte zu ergründen versucht. Diese Form des transdisziplinären Zugriffs stellt nicht nur Erkenntnisgewinne in Aussicht an Stellen, an denen der wissenschaftliche Diskurs an Disziplinengrenzen steckenbleibt, weil – je nach Bewertung – etwa Philosophen nur staunen oder belustigt sind, anstatt konkrete Vorschläge zu machen, Neurowissenschaftler in den Ausschnitten ihrer Empirie festhängen, ohne das Gesamte zu beachten, Juristen sich aufs Bedenken tragen beschränken, anstatt Lösungswege aufzuzeigen, und Technikwissenschaftler fahrlässig mit Begriffen hantieren. Eine transdisziplinär-zugewandte Haltung kann es im Gegenteil ermöglichen, Brücken zu bauen, und zu einem gemeinsamen Verständnis der Gegenstände beitragen, indem fachliche Kategorien und Bewertungskriterien der verschiedenen Gruppen produktiv aufgegriffen, aufeinander bezogen und dergestalt rückgekoppelt werden, dass sich beteiligte Akteure im Diskurs die Zugänge der fachlich

⁹ Vgl. Maxwell Bennett / Peter Hacker: Die philosophischen Grundlagen der Neurowissenschaften, Darmstadt 2010; Maxwell Bennett / Daniel Dennett / Peter Hacker / John Searle: Neurowissenschaft und Philosophie. Gehirn, Geist und Sprache, Berlin 2010.

¹⁰ Bennett / Hacker: Die philosophischen Grundlagen der Neurowissenschaften (s. vorherige Fußnote), 327.

¹¹ Vgl. ebd., 321.

anderen zu eigen machen können, um ihr Gesamtverständnis zu erweitern und über die begrifflichen Grenzen der eigenen Disziplin hinaus ausdrucksstärker zu werden.

4 Herstellung von maschinellem Bewusstsein

Verschiedene Arbeitsgruppen weltweit streben an, bewusste Roboter zu bauen. Die intendierten Ergebnisse variieren je nach Begriffsverständnis und disziplinärem Zugang. Allen gemeinsam ist, dass sie etwas Neues erschaffen wollen, das die Welt so bislang noch nicht gesehen hat. Dabei stehen die Arbeitsgruppen in einem globalen Wettbewerb miteinander, dessen Bedeutung im Rückblick auf historische Entwicklungen nicht zu unterschätzen ist: Wer im Feld bewusster Maschinen die Technologieführung übernehme, werde in der Folge auch das gesellschaftliche System vorgeben – so die Haltung in einschlägigen Fachkreisen.

In einer durch die Autoren durchgeführten umfangreicheren Bestandsaufnahme zeigen sich die folgenden zusammenfassenden Hauptlinien in den Bemühungen, bewusste KI-Systeme und Roboter zu bauen:

- *Technischer Nachbau ausgewählter Eigenschaften plus Bewusstseinszuschreibung:* Ausgewählte menschliche oder auch tierische Eigenschaften werden mit technischen Mitteln nachgeahmt und die Ergebnisse sodann frei und unbeschwert als „Bewusstsein“ bezeichnet.¹² Gängige Eigenschaften sind etwa die Erstellung eines Abbildes der Außenwelt im System (*world mapping*), selbstgesteuertes kollisionsvermeidendes Bewegen des eigenen Körpers im Raum (*moving*), das Reagieren auf Umweltreize (*reaction*). Diese Eigenschaften sind beispielsweise in Saug- und Mährobotern, Industrierobotern und auch in autonomen Fahrzeugen bereits realisiert. Von Entwicklern werden die technischen Nachbauten auch als vorbereitende Arbeiten für ein späteres umfassenderes KI-Bewusstsein ausgewiesen.
- *Identifikation und Nachbau des neuronalen Korrelats des Bewusstseins:* In Strukturen und Prozessen des Gehirns sollen jene neuronalen Korrelate identifiziert werden, die für das Entstehen von Bewusstsein zuständig sind (*neural correlates of consciousness, NCC*). Würden diese gefunden, ließen sie sich, so die Hoffnung, mit technischen Mitteln nachbauen. Bislang sind solche NCC noch nicht identifiziert worden, und die Frage, ob sie überhaupt existieren, wird in den Neurowissenschaften leidenschaftlich debattiert.¹³ Der Ansatz ist populär und verheißt, das Bewusstsein tatsächlich dingfest machen zu können.

¹² Vgl. Junichi Takeno: Creation of a Conscious Robot. Mirror Image Cognition and Self-awareness, Boca Raton 2013.

¹³ Vgl. Christof Koch / Marcello Massimini / Melanie Boly / Giulio Tononio: Neural Correlates of Consciousness. Progress and Problems, in: Nature Reviews Neuroscience 17 (2016), 307 – 321.

- *Nachbau des Gehirns unter Berücksichtigung imaginärer Effekte und von Quanteneffekten:* Entwickelt werden „neuromorphe“ Computer, die nicht nur ihre Betriebsprogramme und Speicherinhalte, sondern auch ihre Hardwarefunktionen ändern können. Hierbei wird neben digitalen auch auf analoge Komponenten gesetzt, etwa indem versucht wird, die synaptischen Kopplungen zwischen Neuronen mittels Kondensatoren nachzubilden. Hiervon versprechen sich die Entwickler Systemdynamiken, in denen u. a. gezielt Quanteneffekte und Wirkungen mit komplexen und imaginären Zustandsbeschreibungen hervorgebracht werden können.¹⁴ Zu diesen Ansätzen liegen bereits EU-Patente vor, Entwicklungsprojekte sind angelaufen.
- *Gehirn- und körperorientierte Inszenierungen:* Entwickelt werden Roboter wie Sophia, die aufgrund ihrer äußeren Erscheinung und ihres Verhaltens schlicht den Anschein erwecken sollen, dass da „etwas mehr“ sein könnte als reine Mechatronik mit KI-Komponenten, Netzwerkanbindung und menschenorientiertem Oberflächen-design. Diese Roboter sind besonders geeignet, um medienwirksam aufzutreten und zu naiven Zuschreibungen von Bewusstsein anzugehen.
- *Hoffen auf Emergenz durch Systemkomplexität bzw. durch Fulguration von außen:* Beide Ansätze haben gemeinsam, dass Entwickler sie nicht gezielt mit technischen Mitteln herstellen können, sondern abwarten müssen, bis Bewusstsein in der Maschine „ankommt“ – entweder infolge ihrer Komplexität oder ihres Vernetzungsgrades, die Bewusstsein schließlich als Nebenprodukt „emergieren“ lassen, oder aber in der Variante, dass ein Bewusstsein, eine Seele, ein Geist oder ein bewusstes dynamisches Feld von außen in die Maschine einfährt, einschlägt oder sich an sie anbindet. Solche Szenarien gibt es in verschiedenen Spielarten bis hin zu der Annahme extraterrestrischer Künstlicher Intelligenzen, die technische Geräte und Netzwerke der Menschheit unterwandern werden oder dies gar bereits getan haben, sich bislang aber noch vor uns verstecken.

Die beschriebenen Ansätze zur Herbeiführung bewusster Maschinen fassen den Stand Mitte des Jahres 2019 zusammen. Ergänzend anzuführen ist an dieser Stelle, dass die Beschäftigung mit dem Thema Bewusstsein in den Neurowissenschaften ihre frühere Anruchigkeit zunehmend verliert und damit das Forschungsfeld auch von Empirikern freier und mit weniger Furcht vor Reputationsschäden bearbeitet werden kann.

¹⁴ Vgl. Ralf Otte: Versuch einer Systemtheorie des Geistes. Mathematische und physikalische Untersuchungen mentaler Prozesse. Ansatz zur mathematischen Analyse des Geist-Körper-Problems, Göttingen 2011.

5 Mitgestalten der notwendigen Diskurse

Die Begleitmusik aus Theologie, Geistes- und Sozialwissenschaften zu Fragen der Entwicklung bewusster Maschinen und Roboter klingt – von Ausnahmen abgesehen – derzeit noch recht leise und verhalten. Ähnlich wie für zahlreiche Akteure aus dem politischen Raum sind KI-Themen insbesondere der hier besprochenen Art für viele Nichttechniker weit weg vom eigenen Kernressort. Die notwendige kritische und wertebasiert Orientierung gebende Auseinandersetzung mit den gesellschaftlichen und politischen Folgen der Zuschreibung von Bewusstsein an Roboter und KI-Systeme wird aber auch zukünftig kaum von jenen geführt werden, die diese Systeme bauen. Aus Sicht der Autoren sind gerade Vertreter der nicht-technischen, reflektierenden wissenschaftlichen Disziplinen dazu aufgerufen, das Diskursvakuum zu bewussten Maschinen mit Eingaben und Beiträgen zu bespielen, Bezugsrahmen einzubringen und Impulse zu geben.