

Kritik der Künstlichen Intelligenz

VON HANS-DIETER MUTSCHLER

Wenn eine neue Technik entsteht, wachsen die Bäume der Erwartungen in den Himmel.¹ Als sich in der zweiten Hälfte des 19. Jahrhunderts die Elektrizität allgemein durchsetzte, galt diese physikalische Kraft als ein wahres Wunder. Thomas Edison, eigentlich ein Gegner der Todesstrafe, erfand flugs den elektrischen Stuhl, weil es eine besonders *humane* Form des Tötens sei. Bauern verlegten elektrische Drähte in den Boden, um deren Fruchtbarkeit zu steigern, und Lenin definierte: „Sozialismus ist Sowjetmacht plus Elektrizität.“ Als Werner von Siemens in Berlin die elektrische Straßenbeleuchtung einführte, war das ein *moralischer* Fortschritt, denn jetzt konnte man die Diebe nachts *sehen*. In jener Zeit wurden auch die Elektroschocks für psychisch Kranke eingeführt. Elektrizität war also eine moralische und eine politisch progressive Macht mit heilsamer Wirkung.

Solche moralischen oder sogar das Religiöse streifenden Zuschreibungen verlieren sich dann rasch und werden vergessen, wenn man sich an die neue Technik gewöhnt hat. Vielleicht sind noch die Erlösungshoffnungen in Erinnerung geblieben, die sich mit der Weltraumfahrt verbanden, besonders nachdem die Amerikaner unter Wernher von Braun auf dem Mond gelandet waren. Von Braun stellte in Aussicht, dass die bemannte Weltraumfahrt dazu führen würde, die tiefsten existenziellen Fragen des Menschen zu beantworten und zugleich den Weltfrieden zu sichern.² Die bemannte Raumfahrt hat inzwischen viel von ihrem Glanz verloren, dafür erleben wir heute einen regelrechten Hype um Roboter und Computer, der auch kluge Philosophen für sich einnimmt.³

Dieser Medienhype wird angeheizt durch wöchentliche Siegesmeldungen in der Presse, die sich ständig überschlagen. Die Frage nach den Grenzen des Computers scheint sich von selbst zu erledigen, denn pausenlos stehen in den Zeitungen neue Nachrichten, was die Computer beziehungsweise Roboter alles gelernt haben und nun womöglich besser können als wir: Sie übersetzen von einer Sprache in die andere, komponieren, malen, schreiben Romane und Gedichte, machen Witze, erstellen ärztliche Diagnosen, führen komplizierte mathematische Beweise durch, betätigen sich als Psychotherapeuten, in der ärztlichen Diagnose oder auch in der Rechtsprechung, verfassen neue Kochbücher und haben angefangen, Fußball zu spielen. Auto und Ski fahren können sie schon. Nichts scheint ihnen unmöglich.⁴

¹ Ich habe das für verschiedene Technologien seit der Industriellen Revolution nachgewiesen in *H.-D. Mutschler, Die Gottmaschine. Das Schicksal Gottes im Zeitalter der Technik, Augsburg 1998*.

² So in der von von Braun autorisierten Biographie, die im Jahr der amerikanischen Mondlandung herauskam (*B. Ruland, Wernher von Braun. Mein Leben für die Raumfahrt, Offenburg 1969*). Die Gegenrechnung macht Rainer Eisfeld auf, indem er die Verstrickung von Brauns in das Naziregime deutlich macht (*R. Eisfeld, Mondsüchtig. Wernher von Braun und die Geburt der Raumfahrt aus dem Geist der Barbarei, Reinbek 1996*).

³ So z. B. *K. Mainzer, Computerphilosophie zur Einführung, Hamburg 2003*. Dagegen ist das Standardwerk *J. Copeland, Artificial Intelligence. A Philosophical Introduction, Oxford 1993*, von wohlthuender Nüchternheit, obwohl der Autor überzeugter Naturalist ist. Ebenso nüchtern ist Manuela Lenzens Buch über KI (*M. Lenz, Künstliche Intelligenz. Was sie kann & was uns erwartet, München 2018*).

⁴ Vgl. meine kritischen Artikel *H.-D. Mutschler, Ist der Mensch ein Roboter? In: M. Kossler/R. Zecher (Hgg.), Von der Perspektive der Philosophie. Beiträge zur Bestimmung eines philosophischen Standpunkts in einer von den Naturwissenschaften geprägten Zeit, Hamburg 2002, 291–308; ders., Zur Herrschaft der Computermetapher, in: Scheidewege. Jahresschrift für skeptisches Denken, 32 (2002/2003) 106–118; ders., Liebe Deinen Roboter wie Dich selbst – oder: Drei Gründe, weshalb Roboter keine Menschen sind, in: Erwägen, Wissen, Ethik, 20/2 (2009) 218f.*

Auf diese Art legt sich die Zukunftsvorstellung der sogenannten ‚Trans-‘ oder ‚Posthumanisten‘ nahe, wonach der biologische Mensch ein auslaufendes Modell der Evolution sei, und zwar entweder so, dass die Roboter sich ständig verbessern, um die Macht zu übernehmen, oder so, dass die somatischen Eigenschaften des Menschen Schritt für Schritt durch *bessere* künstliche Systeme ersetzt werden im Sinn eines ‚Cyborg‘, also einer Verbindung von Mensch und Maschine, die man im Sinn des ‚human enhancement‘⁵ so weit vorantreiben könne, bis das Resultat dasselbe sein würde, nämlich der haushoch überlegene Roboter, was dann auch ‚Singularität‘ genannt wird.

Ein solches Konzept, also das Konzept einer ‚starken KI‘, vertreten mit geradezu religiösem Eifer die Computerenthusiasten Marvin Minsky, Hans Moravec und Ray Kurzweil.⁶ Bei solchen Autoren findet sich auch die Vorstellung, dass Roboter mutieren könnten, um völlig neue, emergente Eigenschaften zu entwickeln, vielleicht die, uns auszurotten, was solche Autoren nicht weiter erschreckend finden. Der Bessere gewinnt eben.

Wie ungerechtfertigt solche Behauptungen sind, zeigen ganz elementare Überlegungen: Lebewesen, die zur Evolution fähig sind, müssen über eine Dualität von Genom und Phänotypus verfügen. Aber *kein* Roboter verfügt über ein Genom, das mutieren könnte. Wenn aber doch, wären die meisten Mutationen ‚letal‘, wie in der Natur, und die Mutationen als reine Zufallsereignisse hätten zur Folge, dass die ‚Humanisierung der Roboter‘ Hunderttausende oder Millionen von Jahren benötigen würde. Wer mag so lange warten? Außerdem ist völlig unklar, wer die Rolle der ‚Umwelt‘ für diese mutierenden Roboter spielen sollte und ob nicht der Stoffwechsel eine notwendige Bedingung für Leben ist. Mit einem Satz: Die Prinzipien der Evolution sind auf solche Roboter nicht anwendbar. Sie ergeben hier keinen Sinn.

Etwas kultivierter ist die Auffassung von Margaret A. Boden, der zufolge wir Gebrauch machen sollten von der *Artificial Life Technologie (AI)*.⁷ Diese Programmieretechnik wurde von dem Mathematiker John Holland entwickelt.⁸ Er hatte die Idee, Algorithmen zu erfinden, mit deren Hilfe virtuelle Lebewesen auf dem Computer zufällig mutieren können und gemäß einer Fitnessfunktion selektiert werden, die für das Ziel der Entwicklung steht. In dieser Hinsicht gibt es jedoch eine radikale Differenz zwischen natürlicher und künstlicher Evolution. Die natürliche Evolution kennt keine Ziele, die künstliche hängt vollständig davon ab. Das heißt aber, dass auch hier, rein aufgrund dieser Programmieretechnik, nichts Neues passieren kann, was Frau Boden ignoriert, die doch eigentlich den Begriff der ‚Kreativität‘ klären möchte. Ebenso ignoriert sie den Begriff des ‚Neuronalen Netzes‘ und die Notwendigkeit, es zu trainieren. Es ist schon einigermaßen erstaunlich, dass Philosophinnen und Philosophen sich mit Begriffsanalysen begnügen, wo sie zunächst einmal die Technik verstehen sollten. Aber das ist die Wucht des Medienhypes, der bis zur Philosophie durchschlägt.

Jaron Lanier, immerhin Preisträger des deutschen Buchhandels, der maßgeblich an der Entwicklung interaktiver ‚*Virtual Reality*‘ beteiligt war, ist inzwischen zum schärfsten Computerkritiker geworden. Er sagt: „Ich habe das Gefühl, dass es sich bei KI und den damit verbundenen Erlösungsgedanken um einen großangelegten Betrug handelt.“⁹

Manchmal wird von den Computerenthusiasten gesagt, dass zwar die jetzigen Computer und Roboter dem Menschen noch in vielerlei Hinsicht unterlegen seien, aber die Computer der Zukunft seien ganz anders konstruiert und würden den Menschen hinter sich lassen. Doch die Geschichte der Technik ist so wenig antizipierbar wie die

⁵ J.-Ch. Heilinger, *Anthropologie und Ethik des Enhancements*, Berlin/New York 2010.

⁶ Vgl. M. Minsky, *Mentopolis*, Stuttgart 1990; H. Moravec, *Computer übernehmen die Macht. Vom Siegeszug der künstlichen Intelligenz*, Hamburg 1999; R. Kurzweil, *Menschheit 2.0 – Die Singularität naht*, Berlin 2014.

⁷ Vgl. M. A. Boden, *Creativity and Artificial Intelligence*, in: E. S. Paul/S. B. Kaufmann (Hgg.), *The Philosophy of Creativity*, Oxford 2014, 224–244.

⁸ Vgl. J. H. Holland, *Hidden Order. How Adaptation Builds Complexity*, Reading (Mass.) [u. a.] 1996.

⁹ FAZ vom 12.12.2018.

Geschichte im Allgemeinen. Wechsel auf die Zukunft sind jedenfalls unseriös. Es geht vielmehr darum, *nüchtern* zu fragen, wo die konstruktionsbedingten Grenzen dieser Maschinen liegen. Dies soll im Folgenden geschehen.

Dazu ist es nützlich zu fragen, in welchen Fällen der Computer stärker ist als der Mensch und in welchen Fällen er eklatant versagt. Diese Unterscheidung ist deshalb so schwierig, weil von den Computerenthusiasten selbst eklatante Mängel in Siegesmeldungen umgedichtet werden, so ähnlich wie sich nach den Bundestagswahlen jede Partei zum Sieger erklärt, möge sie so schlecht abgeschnitten haben wie sie wolle.

Einige Negativbeispiele vorweg: Im Auktionshaus Christie's in New York wurde Ende Oktober 2018 erstmals ein Porträt für 432.500 Dollar versteigert, das ganz und ausschließlich von einem Computer hergestellt war. Dessen „neuronales Netz“ wurde zuvor mit allen erdenklichen Porträts aus der Kunstgeschichte trainiert. Das Resultat war ein Verschnitt, der keinen Kunstexperten überzeugte, denn der Computer schuf nichts Neues. Er verwandelte ein gegebenes Passbild nach Prinzipien, die er dem bereits Bestehenden abgesehen hatte.

Es gibt Kompositionsprogramme, die ein Kirchenlied, das Bach niemals vertont hat, vierstimmig setzen. Das Resultat klingt wie Bach, aber nur so, wie jeder Kompositionsschüler lernt, im bachschen Stil zu schreiben, was nichts Besonderes ist.

Hans Magnus Enzensberger war einer der ersten, der zusammen mit einem Informatiker im Jahr 2000 einen „Poesieautomaten“ entwickelte, der zunächst in Landsberg am Lech installiert war. Er sonderte pro Minute ein Gedicht ab und zwar so, dass die Zeit seit dem Urknall nicht ausreichen würde, damit sich die Maschine wiederholt. Ein Beispiel:

Die Trägheit entflammt.
Eindrücke lächeln gegenüber dem Sturm.
Auge, Tropfen, Schrecken, alles antwortet.

Gedanken, Hoffnungen, Wälle, alles kreischt.
Erkenntnisse und Hoffnungen, der Tod rieselt.

Monolithen, Winde, Vögel, rein gar nichts singt.

Diejenigen, die diese Gedichte lasen, bemerkten allerdings, dass sie sich immer nur nach Enzensberger anhörten, was nach einiger Zeit furchtbar *langweilig* wurde, so dass der Meister selber zurückruderte: „Wer so schlecht dichtet wie diese Maschine, der soll es gleich bleiben lassen.“ Aber wozu dann der Aufwand?

Der Kunstfälscher Wolfgang Beltracci führte die größten Kunstexperten jahrzehntelang an der Nase herum und verdiente Millionen, indem er kogenial Werke der klassischen Moderne erfand, die es gar nicht gab. Nachdem er überführt worden war, weil er ein Deckweiß verwendet hatte, das es erst nach dem Zweiten Weltkrieg gab, saß er sechs Jahre lang im Gefängnis und fing danach mit der freien Malerei an. Er musste schließlich von etwas leben. Seine freie Malerei war aber so schlecht, dass sie niemand kaufen wollte. So sind die Computer auch. Sie können nachahmen, schaffen aber nichts wirklich Neues. Es versteht sich aber von selbst, dass dies kein Argument gegen den Einsatz von Computern in der Kunst sein kann.

An der Zürcher Musikhochschule gibt es einen Fachbereich für elektronische Musik.¹⁰ Dort wird *mit* dem Computer komponiert. Der Computer ist also ein vom Menschen gesteuertes *Hilfsmittel*. Auch eine Orgel ist eine vom Menschen gesteuerte mechanische Maschine. Und doch erzeugt sie die schönste Musik, wenn es einer wirklich kann. Das gilt auch ganz allgemein für die jährliche „Ars Electronica“ in Linz. Auch hier ist die Elektronik Mittel zum Zweck, kein Selbstzweck und das gilt auch für jeden Synthesizer.

¹⁰ Das „Institute for Computer Music and Sound Technology“ (ICST), das seit den 80ern existiert.

Entscheidend ist also, ob der Mensch die Sache im Griff hat oder nicht. Es gibt inzwischen Roboter, die Operationen, etwa nach der Da Vinci Methode, viel präziser durchführen, als jeder Chirurg, indem sie auf Daten der Computertomographie zurückgreifen. Aber der Arzt überwacht die Operation beständig und korrigiert, falls es nötig sein sollte und das ist immer dann der Fall, wenn etwas Unvorhersehbares eintritt.

Der Mangel an Fähigkeit zur Innovation zeigt sich bei den Computern auf allen Gebieten, zum Beispiel auch in der Rechtsprechung. Die Amerikaner haben den größten Thesaurus an Gerichtsurteilen auf EDV, sodass ein Richter seinen Fall eingeben kann und der Computer spuckt sofort das passende Urteil aus. Es hat sich allerdings gezeigt, dass der Computer sämtliche Vorurteile der amerikanischen Richter reproduzierte, wie strengere Beurteilung von Schwarzen, Schwulen oder Frauen. Es würde sich auch prinzipiell nichts ändern, wenn man diese Fehlurteile herausprogrammieren würde. Es bliebe ein anderes Problem, das sich *prinzipiell* nicht beseitigen lässt: „Individuum est ineffabile.“ Der Einzelfall enthält immer Neues, das zu berücksichtigen wäre und das in den bereits gespeicherten Fällen nicht vorkommt.

Das ist auch der Fall bei medizinischen Diagnosen. Viele Ärzte schätzen den Computer als Diagnosehilfe, weil er inzwischen so viele Krankheitssymptome speichert, wie sie der Arzt niemals im Gedächtnis behalten könnte. Kein verantwortlicher Arzt wird sich aber blind auf den Computer verlassen, sondern er nimmt die Resultate des Computers als *Anregung*, um sich selbst ein Bild zu machen. Wir können dies auf die Formel bringen:

*Der Computer hat keine Urteilskraft, weil er blind ist für Innovation!*¹¹

Urteilskraft ist ein *praktisches* Vermögen, bei Aristoteles „*phronesis*“, heute „practical reason“ oder „know how“ genannt. Sie will geübt sein, weil es hier keine rein logischen Ableitungen gibt. Das scheint übrigens der Grund zu sein, weshalb noch niemand auf die Idee gekommen ist, Politiker durch Computer zu ersetzen. Im politischen Handeln müssen zu gegebenen Zwecken erst die Mittel gefunden werden, die sich nicht antizipieren lassen, und oft genug bringen erst gegebene Mittel die Zwecke hervor, denen sie genügen werden.

Bei Aristoteles gibt es allerdings auch die Denkfigur des „praktischen Syllogismus“. Danach würde Handeln einer strengen Logik genügen: Wir setzen uns einen bestimmten Zweck, suchen nach den Mitteln, die den Zweck realisieren könnten, wählen das beste Mittel und handeln entsprechend. Dieser praktische Syllogismus wurde von den Pragmatisten heftig bekämpft mit dem oben genannten Argument, dass uns oft genug die Mittel zur Realisierung bestimmter Zwecke nicht gegeben sind oder dass die Mittel die Zwecke allererst hervorbringen.

Vielleicht haben beide Recht, je nach Situation. Muss ich von A aus den Ort B erreichen, so kann ich die Trambahn, den Esel, das Fahrrad oder den Hubschrauber benutzen. Die Trambahn erscheint mir vielleicht am günstigsten, also steige ich ein. Aber nicht alle Handlungskontexte sind so überschaubar, und dann haben die Pragmatisten mit ihrer Kritik recht.¹²

Wir könnten nun sagen: Die Computer sind die Virtuosen des praktischen Syllogismus. In allen anderen Fällen versagen sie eklatant. Das ist der einfache Grund, weshalb sie besser Schach, Go oder Poker spielen als wir. In solchen Fällen sind das Ziel und

¹¹ Die „Stanford Encyclopedia“, die ja eine prinzipiell reduktionistisch-naturalistische Position vertritt, ist doch zugleich sehr selbstkritisch. KI habe nichts zum Problem der Kreativität zu sagen. (vgl. S. Bringsjord [u. a.], Art. Artificial Intelligence, in: Stanford Encyclopedia of Philosophy [https://plato.stanford.edu; letzter Zugriff: 21.06.2019], 3.1).

¹² Bei Kant firmiert diese Differenz als Unterschied zwischen reflektierender und bestimmender Urteilskraft. Rüdiger Bubner entwickelt in seiner Handlungstheorie einen über Kant hinaus erweiterten Begriff von „Urteilskraft“: „Das logische Verhältnis von Regel und Fall der Regel erscheint dagegen nicht als einfache Subsumtion, vielmehr als Wechselbestimmung.“ Wir haben hier also kein formallogisch bestimmtes Verhältnis. Ganz ähnlich der Sozialphilosoph John Rawls, der das Konzept des „Überlegungsgleichgewichts“ entwickelt hat (R. Bubner, Handlung, Sprache und Vernunft. Grundbegriffe praktischer Philosophie, Frankfurt a. M. 1982, 261).

die Mittel fest vorgegeben. Das Ziel wird vom Programmierer gesetzt, und die Mittel erarbeitet sich der Computer, während das „neuronale Netz“ trainiert wird. Das ist auch so beim Go-Spiel, das viel komplizierter ist als Schach. Erst 2016 gewann ein Computer gegen den Weltmeister im Go-Spiel, Lee Sedol. Der entsprechende Computer ‚kannte‘ nur die Regeln, den Rest erledigte er ‚selbst‘. Aber was heißt hier ‚selbst‘? In Wahrheit ließ man die Maschine fünf Millionen (!) zufallsgenerierte Ausgangspositionen einnehmen und gegen sich selber spielen. Es macht keinen Unterschied, ob wir eine solche Maschine mit Meisterspielen trainieren, wie im Schach, oder mit zufallsgenerierten Ausgangspositionen. In beiden Fällen bewegt sich die Maschine im Referenzrahmen einer starren Zweck-Mittel-Relation, die in diesem Fall sogar *a priori* erfüllbar ist: Bei hinreichender Kapazität kann der Computer alle nur möglichen Spiele abrufen und gewinnt immer.

Computer sind, wie John Searle sagen würde, syntaktische Maschinen.¹³ Das ist auch der Grund, weshalb sie so stark sind, wenn es darum geht, mathematische Beweise zu führen, wobei diese Beweise oft so lang und kompliziert sind, dass niemand weiß, ob man sie als solche gutheißen sollte, da niemand 300 eng beschriebene Seiten nachrechnen will.¹⁴

Die genannten Stärken und Schwächen des Computers folgen aus seiner Konstruktion: Neuronale Netze müssen *trainiert* werden. Im Gegensatz zur klassischen KI¹⁵ sind sie, wie die natürlichen Gehirne, dezentral aufgebaut, folgen also im Sinn der Bionik biologischen Vorbildern. Wie in den natürlichen Gehirnen sind die Schwellenwerte zwischen den Neuronen veränderlich und ändern sich im Training. Übrigens sind unsere besten „neuronalen Netze“ mit einigen Millionen parallel verschalteten ‚Neuronen‘ (= Transistoren) noch stark unterkomplex im Vergleich zum menschlichen Gehirn, das 100 Milliarden Neuronen enthält, von denen immer eines mit 10.000 wechselwirkt, wobei jedes noch individuell gestaltet ist. Gemessen daran ist die Selbstsicherheit gewisser Leib-Seele-Theoretiker, der Mensch sei auch nur ein Computer, einigermaßen erstaunlich.¹⁶

Allein die Tatsache, dass Neuronale Netze trainiert werden müssen, erlegt ihrer Wirksamkeit Grenzen auf. Sehen wir vom Sonderfall rein syntaktisch bestimmter Spielverläufe ab, so sind „neuronalen Netze“, wie man zu Recht bemerkt hat, ‚vergangenheitsorientiert‘. Sie reproduzieren das, was man ihnen gesagt hat, und weil sie im Sinn von *big data* über einen gigantischen Thesaurus an Zusammenhängen verfügen, kann leicht der Eindruck entstehen, sie würde Neues produzieren – vor allem, weil wir gar nicht wissen, was sie tun. Parallel verschaltete künstliche Neuronen sind wie ebenso viele gekoppelte partielle Differentialgleichungen, bei denen schon drei schwer lösbar sind, sicher aber nicht Millionen. Auf diese Art entsteht ein Black-Box-Effekt der Undurchschaubarkeit, der zur Weltanschauung des ‚Konnektionismus‘ geführt hat, der bis hinein in die Psychologie, Biologie und Linguistik hohe Wellen schlug.

Wenn wir damit zufrieden wären, dass Computer nützliche Instrumente sind, um bestimmte Zwecke zu realisieren (also das, was man ‚schwache KI‘ genannt hat), dann entstünde kein philosophisches Problem. Es entsteht ja auch kein philosophisches

¹³ So sein berühmtes Gedankenexperiment mit dem ‚chinesischen Zimmer‘ (*J. R. Searle, Geist, Hirn und Wissenschaft. Die Reith Lectures 1984, Frankfurt a. M. 1992, 33*). Man müsste freilich zwischen logisch-mathematischer und grammatikalischer Syntax unterscheiden. Sie haben nicht dieselben Eigenschaften.

¹⁴ An der Universität Kaiserslautern werden vornehmlich solche automatisierten Beweise geführt.

¹⁵ Unter ‚klassischer Künstlicher Intelligenz‘ versteht man einen hierarchisch aufgebauten Computer mit Von-Neumann-Architektur, im Gegensatz zu den „neuronalen Netzen“, die die Information parallel, also nichthierarchisch, verarbeiten.

¹⁶ Es kommt ja noch einiges andere hinzu: Natürliche Gehirne verarbeiten Informationen auch mit Hilfe von chemischen Transmitterstoffen, die es beim Computer nicht gibt. Ganz allgemein verändern sich natürliche Gehirne ständig, was für die Hardware eines Computers nicht gilt; ferner sind im Gehirn keine zwei Neuronen gleich, was für unsere Computer auch nicht gilt.

Problem daraus, dass wir Getreide oder Schafe züchten, wobei allerdings nicht ganz klar ist, was wir überhaupt unter ‚Intelligenz‘ verstehen sollten.

Sehr häufig hört man: Intelligent ist ein Roboter dann, wenn er sich so verhält, wie sich ein intelligenter Mensch verhalten würde.¹⁷ Das ist so ähnlich, als würden wir sagen: „Gut ist, was die Menschen so nennen, also ein Apfel, die Demokratie oder Beethovens Neunte.“

In der *Stanford Encyclopedia* wird uns gesagt: „We define AI as the study of agents that receive percepts from the environment and perform actions.“ An anderer Stelle wird noch hinzugefügt, dass diese Aktionen maximalen Nutzen hervorbringen sollten.¹⁸ Diese ‚Definition‘ ist sehr vage, zumindest zu weit, denn nach ihr müssten wir Fische ‚intelligent‘ nennen, wenn wir den Überlebensvorteil als ‚Nutzen‘ bestimmen. Auch jeder Industrieroboter, der mechanische Schweißarbeiten verrichtet, wäre dann ‚intelligent‘.

Problematisch werden unsere Erwartungshaltungen überall dort, wo wir vom Computer Innovationen erwarten, die er einfach nicht leisten kann. Das hat sich in Bezug auf die Kunst gezeigt, aber auch in Bezug auf die menschliche Urteilskraft, als ein Vermögen, mit neuen Erfahrungen umzugehen. Es zeigt sich auch in Bezug auf die natürliche Sprache, solange sie nicht stilisiert und kanalisiert ist wie die Fachsprachen, mit denen der Computer gut umgehen kann.

Die Japaner haben Milliarden ausgegeben, um einen Computer zu programmieren, der die natürliche Sprache mit all ihren Bedeutungszuschreibungen verstehen würde. Die Sache ist jammervoll gescheitert. Ihr Computer war nicht imstande, bei einem Kindergeburtstag Vierjähriger mitzuhalten.

Von einem theoretischen Standpunkt aus gibt es ehrgeizige Versuche, die versteckte Logik der natürlichen Sprache so zu formalisieren, dass damit alles Wesentliche erfasst wird. Daher werden immer neue Kalküle erfunden. In seinem Beitrag über „Logic and Artificial Intelligence“ in der *Stanford Encyclopedia of Philosophy* beschreibt Richmond Thomason sehr viele neue Formen von Logik, dazu angetan, die natürliche Sprache in diesem Sinn zu formalisieren: nichtmonotone, temporale, Präferenz-, situative, kausale, Ereignislogiken und so weiter. Thomason gesteht aber ein, dass die Logik der Alltagssprache immer noch „schwer zu fassen“ sei.¹⁹ Tatsächlich handelt es sich jedoch eher um ein *prinzipielles* Problem.

Solche Logiken werden gerne als das Wesen der Sprache gehandelt, wie auch die formalen Überlegungen im Strukturalismus Ferdinand Saussures oder in Noam Chomskys ‚Tiefengrammatik‘. Chomsky hat sein Unternehmen gerne mit Newtons Physik verglichen: So, wie der Unübersichtlichkeit unserer Sinneserfahrung ein präzises mathematisches Schema zugrunde liegt, als ein Platzhalter all dessen was das Physische *wirklich* ausmacht, so sei es auch bei der Sprache.

Demgegenüber beantwortet die Philosophin Sybille Krämer ihre selbstgestellte Frage negativ: „Gibt es eine Sprache hinter dem jedesmaligen Sprechen?“²⁰ Sie unterscheidet zwei entgegengesetzte Sprachauffassungen und ordnet sie bestimmten Autoren zu: Es gebe eine Logosauszeichnung von Sprache und Kommunikation (Saussure, Chomsky, Searle, Habermas) und dann eine Sprache und Kommunikation jenseits intellektualistischer Annahmen (Wittgenstein, Austin, Luhmann, Davidson, Lacan, Derrida, Butler). Ihre Grundthese lautet dann: Man kann Sprache nicht als bloße Instantiierung von allgemeiner Gesetzmäßigkeit (Grammatik oder formaler Logik) begreifen.

¹⁷ So auch *Lenzen*, Künstliche Intelligenz, 27/8.

¹⁸ *Bringsjord*, Artificial Intelligence, 2; 3.1.

¹⁹ *R. Thomason*, Art. ‚Logic and Artificial Intelligence‘ in: *Stanford Encyclopedia of Philosophy* [<https://plato.stanford.edu>; letzter Zugriff: 22.06.2019], 2.2; Auch für Copeland sind die größten Probleme der Computerisierung des Menschen der common sense und die natürliche Sprache (*Copeland*, Artificial Intelligence, 98).

²⁰ *S. Krämer*, Sprache, Sprechakt, Kommunikation. Sprachtheoretische Positionen des 20. Jahrhunderts, Frankfurt a. M. 2001, 15.

Viele KI-Forscher beklagen das Fehlen einer präzisen Kategorienlehre der Alltagssprache. Das wären Regeln, die nicht rein syntaktisch bestimmt sind wie Logik und Grammatik. Aber der Versuch, *die* Kategorienlehre Alltagssprache (im Singular) aufzustellen, ist von Aristoteles bis Jonathan Lowe noch immer gescheitert.²¹ Daher vertreten viele Analytische Philosophen eine rein pragmatische Auffassung der Kategorien: Sie sind mal so, mal so. Dies ist also auch kein gangbarer Weg.

Bestätigt werden solche Überlegungen zu den fundamentalen Eigenschaften der natürlichen Sprache durch Untersuchungen wie die von Georg Bertram.²² Wie er in geschichtlichen und systematischen Detailstudien nachweist, lässt sich die Bedeutung von Sprache weder formal, noch rein immanent erfassen, sondern nur so, dass wir Sprache mit außersprachlichen Gegebenheiten ins Verhältnis setzen. Dazu gehört einmal das Leibbewusstsein im Sinn von Merleau-Ponty, vor allem aber der Sozialbezug, der geschichtlich-gesellschaftlich codiert und prinzipiell kontingent ist. Wenn dies richtig ist, dann wird deutlich, welche Probleme Computer haben, die natürliche Sprache zu ‚verstehen‘. Sie scheitern an deren Kontingenz und innovativer Kraft.

Die Tochter des Verfassers war, wie viele Kinder im Alter zwischen zwei und drei Jahren, sprachschöpferisch tätig. Alles, was fliegen konnte, nannte sie ‚Vogel‘, auch eine Stubenfliege oder eine Motte. Für Lampen an der Decke erfand sie das völlig neue Wort ‚Uda‘. Eine gepellte Wurst war für sie ein ‚Nackedei‘, und Lebewesen kategorisierte sie ganz anders als wir: Tiere und Kinder gehörten in ein und dieselbe Kategorie, Erwachsene in eine ganz andere. Wie soll ein Computer das ‚verstehen‘, wenn er nur den bereits vorhandenen Sprachgebrauch kennt? Wir hingegen verstehen solche Neuschöpfungen auf Anhieb. Vielleicht versagten die japanischen Computer deshalb schon beim Kindergeburtstag.

Man könnte das Problem auch analog sehen zu naturphilosophischen Überlegungen. Für Charles Sanders Peirce war die drängendste philosophische Frage die nach der Entstehung des Neuen in der Natur.²³ Es schien ihm evident, dass es so etwas gibt. Aber die starren Verlaufsgesetze der Naturwissenschaft erklären das Neue nicht, sondern nur das, was ihnen gemäß ist. Also, schloss Peirce, müssen die Naturgesetze als veränderlich gedacht werden können und als bloß statistisch, was er längst vor Entdeckung der Quantentheorie forderte.

Das *tertium comparationis* zu den „neuronalen Netzen“ sind die Korrelationen, die diese zwischen Input- und Output-Daten aufstellen, mit denen sie trainiert werden. Diese sind dann so etwas wie Naturgesetze der entsprechenden Sachverhalte. Sie sind blind für das Neue.

Es ist aber nicht nur die Sprache, die uns Menschen auszeichnet. Auch andere Eigenschaften sind den Maschinen unzugänglich, zum Beispiel Moralität. Das wird natürlich auch bestritten. So gibt es die neue Disziplin der ‚Maschinenethik‘.²⁴ Das ist nun nicht etwa ein Fall angewandter Ethik, den wir gut gebrauchen könnten, etwa, wenn es darum geht, die Aktionen von Kampfrobotern zu beurteilen, die nicht mehr ferngesteuert werden, sondern ‚selbst‘ töten. Oder Automobile, die lieber drei alte als einen jungen Menschen überfahren, wenn es keine weitere Möglichkeit gibt. Nein, hier geht es darum, dass Roboter autonome Wesen in einem kantischen Sinn sein sollen, dass sie also ‚selbst‘ entscheiden, und zwar nach moralischen Prinzipien. Nach Catrin Misselhorn sind wir inzwischen in der Lage, Roboter ‚moralisch‘ zu programmieren, nach Prinzipien des Utilitarismus, einer deontologischen oder materialen Werteethik,

²¹ Vgl. J. Lowe, *Kinds of Being. A Study of Individuation, Identity and the Logic of Sortal Terms*, Oxford 1989.

²² Vgl. G. W. Bertram [u. a.] (Hgg.), *In der Welt der Sprache. Konsequenzen des semantischen Holismus*, Frankfurt a. M. 2008; G. W. Bertram, *Sprachphilosophie zur Einführung*, Hamburg 2011.

²³ Vgl. Ch. S. Peirce, *Naturordnung und Zeichenprozess. Schriften über Semiotik und Naturphilosophie*, Frankfurt a. M. 1991.

²⁴ Vgl. C. Misselhorn, *Grundfragen der Maschinenethik*, Ditzingen 2018.

je nachdem. Das obengenannte Auto, das lieber drei alte Menschen als einen jungen überfährt, wäre folglich ein utilitaristisches Auto.

Aber dem liegt wohl eine Kategorienverwechslung zugrunde. Wenn *wir* einen Roboter gemäß den Prinzipien einer gewissen Ethik einrichten, dann sind es auch *wir*, die die Verantwortung dafür übernehmen müssen. Ein Roboter, dem wir eine kantisch-deontologische Ethik implantiert hätten, würde in dem genannten Fall nach dem Zufallsprinzip Menschen überfahren, weil die kantische Ethik keine Wertkonflikte kennt.

In der Diskussion um die Maschinenethik werden folgende Fragen *niemals* gestellt: Wenn doch Roboter moralische Agenten sind, dann haben sie auch eine Würde. Müssten wir dann nicht ‚Krankenhäuser‘ für defekte Roboter einrichten, anstatt sie zu verschrotten? Bräuchten wir dann nicht Altersheime für Roboter, die die geforderte Leistung nicht mehr erbringen, oder Behindertenwohnheime für Fehlkonstruktionen? Müsste es nicht spezielle Gerichtsverfahren für Roboter geben, die die Gesetze übertreten haben, aber auch Robotergefängnisse und Roboterfriedhöfe?

All das klingt so absurd, wie es ist. Es entspricht den überschäumenden Versprechungen der Elektrizitätslehre in der zweiten Hälfte des 19. Jahrhunderts, die heute so vergessen sind, wie es die Heilsversprechungen der KI morgen sein werden.

Um 1900, als der Darwinismus breitenwirksam geworden war, stellte man die Frage „Können Tiere denken?“, so wie wir heute die Frage stellen „Haben Roboter Bewusstsein, Gefühle, Moral?“ Da man Rechnen für eine herausragende Leistung des Denkens hielt, war es eine Sensation, als ein Schullehrer ein Pferd vorführte, das Rechnen konnte, genannt ‚der kluge Hans‘. Dieses Pferd konnte auch komplizierte Rechnungen durchführen wie zum Beispiel 17×8 , indem es 136mal mit den Hufen auf das Kopfsteinpflaster schlug. Es zeigte sich aber, dass das Pferd nur ganz genau seinen Besitzer beobachtete, der die Luft anhielt, wenn das Pferd aufhören sollte. Sah das Pferd den Besitzer nicht, konnte es auch nicht mehr rechnen.

So wird es mit den Heilsversprechen der KI kommen, und zurückbleiben werden lebenserleichternde Maschinen und, wie zu hoffen ist, neue Formen von Therapie, beispielsweise Implantate für Alzheimerpatienten im Sinn von Hirnschrittmachern oder Mikrochips im Auge für Erblindete oder Exoskelette für Querschnittsgelähmte, die mit bloßen Gedanken gesteuert werden. Auch das wäre nichts Geringses.²⁵

Summary

This article deals with the mostly overlooked boundaries of artificial intelligence, but only insofar as we make use of neural networks. These boundaries depend on the construction of those networks and cannot be overcome by themselves. Neural networks are totally dependent on the “training data” on which they are based. So they are blind with respect to everything new: for instance with respect to the faculty of judgement, to novelty in the arts or to the creative power of natural language.

²⁵ Man darf sich auch hier nicht zu viel versprechen. Insofern wir heute schon therapeutische Computerhilfen haben, empfinden viele Patienten diese als entfremdend wie eine transplantierte Leber, Lunge oder eine Prothese. Vgl. die Untersuchungen von *M. Sabinol*, *Das techno-zerebrale Subjekt*. Zur Symbiose von Mensch und Maschine in den Neurowissenschaften, Bielefeld 2016.