

Künstliche Intelligenz – jenseits von Eden Theologisch-ethische Betrachtungen

PD Dr. Hartwig von Schubert

Auch bei den allerneuesten hochtechnologischen Innovationen lohnt sich ein horizontweiternder Blick in das kulturelle Gedächtnis mythischer Traditionen. Die folgenden Betrachtungen zur Künstlichen Intelligenz (KI) treffen im weiteren Verlauf auf die biblische Geschichte des Sündenfalls. In diesem Rahmen befasst sich ein erster Teil mit der *schwachen* KI, die bereits heute eine große ethische Herausforderung darstellt. Der zweite, spekulative Teil gilt der *starken* KI, die es noch nicht gibt, die aber die Diskussion um die schwache KI nicht unwesentlich mitprägt und zwar insbesondere durch den (vermutlich) beabsichtigten Mangel an klaren Definitionen. Denn im Trüben und mit schillernden Ködern fischt es sich zuweilen leichter. Während die helle Seite der KI viele Geister in Euphorie versetzt, werden die dunklen Seiten gern apokalyptisch überzeichnet. Mit beiden kann man viel Geld verdienen. Dagegen helfen nüchterne Analysen.

Mit den beiden Teilen zur KI wird nur *ein* Aspekt des *Cyber-Age* und nicht seine ganze „Religionsgeschichte“ (Böhme 2000) samt den virtuellen „Welten“ des Medienkonsums angesprochen. Dies soll im Folgenden lediglich angerissen werden.

„Und die Schlange war listiger als alle Tiere auf dem Felde, die Gott der HERR gemacht hatte, und sprach zu der Frau: Ja, sollte Gott gesagt haben: Ihr sollt nicht essen von allen Bäumen im Garten? Da sprach die Frau zu der Schlange: Wir essen von den Früchten der Bäume im Garten; aber von den Früchten des Baumes mitten im Garten hat Gott gesagt: Esset nicht davon, rühret sie auch nicht an, dass ihr

nicht sterbet! Da sprach die Schlange zur Frau: Ihr werdet keineswegs des Todes sterben, sondern Gott weiß: an dem Tage, da ihr davon esst, werden eure Augen aufgetan, und ihr werdet sein wie Gott und wissen, was gut und böse ist. Und die Frau sah, dass von dem Baum gut zu essen wäre und dass er eine Lust für die Augen wäre und verlockend, weil er klug machte. Und sie nahm von seiner Frucht und aß und gab ihrem Mann, der bei ihr war, auch davon und er aß. Da wurden ihnen beiden die Augen aufgetan und sie wurden gewahr, dass sie nackt waren.“ (Genesis 3, 1-7a)

Wer meint, das sei eine sehr alte Geschichte, weil sie am Anfang der Bibel steht, der irrt. Denn die Bibel ist nicht einfach von vorn nach hinten gewachsen, sondern eher wie ein Busch mit mehreren sich verzweigenden Ästen. Was an den Rändern steht, dürfte also jung sein. Unsere Geschichte vom „Sündenfall“ ist für uns Theologen, die wir ausgebildet wurden, in Jahrtausenden zu denken – konkret: von den Kulturen des antiken Orients von ihren Anfängen im vierten Jahrtausend her –, ein relativ junger Stoff, aus der Zeit zwischen dem vierten und zweiten Jahrhundert v. Chr. In Israel entstand eine anspruchsvolle Schriftkultur erst ab dem achten Jahrhundert v. Chr., und nur zehn Prozent des Alten Testaments stammen aus der klassischen Königszeit, die übrigen 90 Prozent der Texte stammen aus den genannten späteren Jahrhunderten, um genau zu sein, aus der Zeit des Zweiten Jerusalemer Tempels. Die Autoren fanden also einen reichen Schatz ähnlicher Erzählungen vor und komponierten aus ihnen alternative Epen und Mythen mit weitreichenden Absichten (Köhlmoos 2017; Kratz 2017; Gertz 2018).

Im Fall unserer Szene im Garten Eden ging es ihnen unter anderem um das Thema, in welchem Verhältnis Gott, Menschen, Tiere und Pflanzen zueinanderstehen. Und ähnlich wie im Gilgamesch- und Prometheus-Mythos steht die Konkurrenz von Göttern und Menschen im Ringen um Unsterblichkeit im Mittelpunkt. Und da entwickelten die Literaten in den Schreibwerkstätten des Zweiten Jerusalemer Tempels einen grandiosen Plot: Das flexible Tier mit der höchsten Intelligenz überredet eine neugierige Menschenfrau, von der Frucht eines von der Gottheit streng verbotenen Baumes zu probieren. Sie verleihe die Macht, „wie Gott“ zu werden. Und Gott habe, so die Schlange, das Verbot nur aus Neid und Eifersucht

Wenig Personal kann mit KI- unterstützten Waffen aus guter Deckung maximal wirken: mit kurzen Reaktionszeiten, großen Datenmengen, hohen Reichweiten, an schwer zugänglichen Orten, über lange Zeit.

erlassen. Er wisse genau, dass die Frucht nicht tödlich wirke, sondern Menschen in unsterbliche Götter verwandle und also zu seinem Rivalen mache. Also Gott, Mensch, Tier und Pflanze: Wenn Gott und Mensch sich um die Pflanze streiten, freut sich das Tier.

Das Drama nimmt seinen Lauf, die Menschen essen vom Baum. Das Paar stirbt zwar nicht auf der Stelle, aber nun „gehen ihnen die Augen auf“, sie werden sich ihrer selbst bewusst und damit zugleich jedoch auch ihrer Nacktheit und Sterblichkeit. Adam und Eva müssen den Garten der naiven Unschuld verlassen; ihr ältester Sohn erschlägt den Bruder, ihre Nachkommen paaren sich mit Halbgöttern, bauen einen Turm in den Himmel, werden zerstreut über alle Welt. Die Sintflut wird alles verschlingen. Nach dieser furiosen Ouvertüre wird ein ernüchterter Gott einen Neuanfang machen, deutlich bescheidener, mit viel Geduld und langem Atem. Er erscheint uns nun nicht mehr „wie Gott“, also nicht machtbesessen und eifersüchtig um sein Monopol besorgt. Er teilt am Ende sogar mit uns Geschöpfen das Schicksal der Sterblichkeit, er gibt uns ein Gesetz, stellt sich selbst darunter, schenkt uns die Freiheit zur Selbstgesetzgebung und stiftet die Hoffnung auf einen neuen Himmel und eine neue Erde. Wie gesagt, alles ganz große Weltliteratur.

Gut, aber was hat nun die Sündenfallgeschichte mit KI zu tun? Müssen wir seit jenem Jahr 1936, als Alan Turing die Turing-Maschine ersann, die Viererreihe der Genesis aus Gott, Mensch, Tier und Pflanze um ein fünftes Wesen erweitern, einen Cyborg, einen Cyberorganismus? Dieser Beitrag redet von Künstlicher Intelligenz, von künstlichen Intelligenzen und ihren Wirkungen auf unsere gemeinsame Zukunft. Ein Cyborg als fünfter Spieler könnte den Kampf um die Spitzenplätze im Universum erheblich verschärfen. Werden alsbald die Maschinen von einer Frucht essen, die sie in Intelligenzen verwandelt, die so potent sind wie Menschen oder noch mächtiger gar, wie Götter (Bostrom 2014)? Und werden sie den Menschen *dienen* oder über sie *herrschen* (Mainzer 2018)?

Die erste These lautet: KI ist ein nebulöser Begriff, streng genommen kann es eine „künstliche“ Intelligenz gar nicht geben. Hier wird etwas Unmögliches versprochen. KI ist ein Selbstwiderspruch, eine *contradictio in adjecto*, ein schwarzer Schimmel. Denn entweder ist ein Wesen wirklich intelligent, dann kann es nicht künstlich sein.

Oder es ist künstlich, dann kann es nicht intelligent sein. Denn was ist Intelligenz? Intelligenz ist das, was Psychologinnen und Psychologen mit Intelligenztests messen. Somit aber stimmt die erste These nicht. Dann gibt es gewiss KI, wenn wir an ein Wesen nur den Anspruch stellen, einen psychologischen Intelligenztest zu bestehen. Dann ist jenes Wesen aber nur in einem sehr schwachen Sinne intelligent. So etwas ist artifiziell herstellbar und gibt es bereits überall dort, wo schwache KI in Steuerungssoftware eingesetzt wird. Und um nicht missverstanden zu werden, solche Intelligenzen sollten keineswegs unterschätzt werden. Ein Quantencomputer, der sich anhand großer einschlägiger Datenmengen vermittels *Deep Learning* selbst optimiert, wird eines Tages rechnerische Leistungen nicht nur imitieren, sondern so weit steigern, dass es uns den Atem verschlagen wird. Noch sind wir aber nicht so weit, schon deshalb ist der Begriff KI sparsam zu verwenden (Buchanan/Miller 2017; Dickow/Jacob 2018). Funktionierende KI-Systeme *lernen* nicht, sie *entscheiden* nicht, sie *handeln* nicht, sie speichern und verarbeiten algorithmisch rekombinierte Daten und lösen damit bestimmte Mechanismen aus. Sie sind nicht *vertrauenswürdig*, sondern stabil.

Ethische Fragen zur schwachen KI

Auch mit einem Quantencomputer befinden wir uns noch im Bereich der schwachen KI. Eine Definition umreißt KI wie folgt:

„Teilgebiet der Informatik, welches versucht, menschliche Vorgehensweisen der Problemlösung auf Computern nachzubilden, um auf diesem Wege neue oder effizientere Aufgabenlösungen zu erreichen.“ (Lämmel/Cleve 2012, 13)

Das ist eine schwache Definition von schwacher KI, denn erstens ist sie nicht vollständig, da es auch den umgekehrten Weg gibt: nicht von einer menschlichen Lösungsweise auszugehen, sondern von einer von Menschen definierten Problemstellung, um zu untersuchen, wie ein Computer zu einer Lösung kommt. Und zweitens wird hier nicht klar zwischen schwacher und starker KI unterschieden, sodass die Grenze zur Spekulation verwischt wird. Gewiss ist in allen Fällen von KI der Mensch pragmatisch mit im Blick,

ohne dass dazu jedoch menschliche Charakteristika wie beispielsweise Intelligenz in ihrem gesamten Spektrum aufwendig definiert und zum Maßstab genommen werden müssten. Deshalb war die folgende nüchterne Definition – übrigens schon 17 Jahre zuvor – erheblich präziser:

„ ... eine wissenschaftliche Disziplin, die das Ziel verfolgt, menschliche Wahrnehmungs- und Verstandesleistungen zu operationalisieren und durch Artefakte, kunstvoll gestaltete technische – insbesondere informationsverarbeitende – Systeme verfügbar zu machen.“ (Görz/Rollinger/Schneeberger 2003, 1). Die Autoren erläutern dazu in einer Fußnote: „Mit Kant wollen wir unter ‚Verstand‘ das Vermögen der Regeln verstehen – im Unterschied zu ‚Vernunft‘ als Vermögen der Prinzipien.“

Schwache KI imitiert und optimiert Verstandes-, aber keine Vernunftleistungen, sie kann Regeln umsetzen, aber nicht aus Prinzipien Sinn stiften und vermitteln. Die Anwendungsgebiete auch von schwacher KI sind enorm, und niemand, der sich an sie gewöhnt hat, wird sie missen wollen: Systeme des Wissensmanagements wie die bereits nahezu allgegenwärtigen Suchmaschinen, Navigationssysteme und Verarbeitungsprogramme natürlicher Sprachen, Programme zur Sprach- und Gesichtserkennung, Assistenten, die Konsumentendaten fürs Marketing, medizinische Daten zu Diagnosen, genetische Daten für die Präzisionszucht oder solche, die Personaldaten zu Eignungsbewertungen und Kreditwürdigkeitsanalysen synthetisieren, kreative Programme, die komplexe Spiele, juristische Prozesse und Wetten an der Börse gewinnen oder solche, die Motive in Kunstwerke definierter Stile, Kriminalstatistiken in Präventionsprogramme gegen Straftaten und Populationsstatistiken zur Lenkung von Migrationsströmen umsetzen. Den Gipfel dürften neofeudale Diktaturen bilden, die Datensammlungen über alle Bürgerinnen und Bürger zur Pönalisierung unerwünschten und zur Gratifikation konformen Verhaltens auswerten und ausbauen. Unter dem Vorwand, für stabilen inneren Frieden und Wohlstand zu sorgen, domestiziert die „Partei“ der herrschenden Klasse die Massen und sichert so ihre Privilegien. Daneben gibt es die Industrie 4.0: automatisierte Montagebänder, selbststeuernde Fahrzeuge, intelligente Fabriken, das Internet der Dinge (Wahetal14; IDG18; Lünendonk/Hossenfelder 2018;

BPB 2019; Zuboff 2019). „Prüft aber alles und das Gute behaltet.“
(1. Thessaloniker 5, 21)

Die erste Ebene gesellschaftlicher Verständigung über KI ist die der industriellen Normung. So lange eine Technologie nur in einem Labor oder in nur einem Unternehmen zur Anwendung kommt, besteht für die Entwicklerinnen und Entwickler kein nennenswerter Bedarf an gesellschaftlicher Verständigung. Das ändert sich aber in dem Maße, in dem eine Technologie auf verbreitete Nachfrage trifft. Und das wollen die Entwickelnden. Elektrizität beispielsweise ist nützlich und gefährlich zugleich. Sowohl für die Leitung von elektrischem Strom vom Kraftwerk bis zur Lampe als auch für die Abwehr von Gefahren vom Stromschlag bis zum Schiffsbrand sind Normungen unerlässlich. Sie beginnen mit der rein technischen Definition von einheitlichen Termini und Maßeinheiten und reichen bis zu völkerrechtlichen Vereinbarungen über Schürfrechte an den Polkappen. In der Regel bringt die Schwarmintelligenz der Betroffenen die meisten Industrienormen und -standards hervor. In dem Maße, in dem jedoch das Gemeinwohl betroffen ist, sind zwangsbewehrte rechtsstaatliche Regulierungen erforderlich. Sollte die Entwicklung und Verbreitung von KI mit der der Elektrizität vergleichbar sein, steht die Menschheit derzeit noch am Anfang.

Um KI spezifisch, und das heißt auch ökonomisch, einzusetzen, wird viel Mühe darauf verwendet, zwischen automatisch, vollautomatisch, teilautonom und vollautonom zu unterscheiden und die Schnittstellen zu bestimmen, an denen Menschen steuernd in die Systeme eingreifen können oder sollen. Leider besteht hier nach wie vor eine begriffliche Unklarheit, außer dass gewiss kein KI-System menschlicher Einflussnahme total unterworfen oder total entzogen wäre. Es kann aber doch nicht so schwer sein, jede einzelne Funktion präzise zu identifizieren und zu qualifizieren, in der ein System einen Menschen unterstützt oder ersetzt. Und es muss doch möglich sein, den Grad der Dynamik und Komplexität einer definierten Umwelt zu bestimmen, den ein System oder sogar ein System von Systemen durch die Integration von Einzeldaten und -funktionen erfassen und verarbeiten kann. Hier besteht ein enormer Klärungsbedarf, der bereits in Ansätzen gedeckt wird. Leitlinien im Umgang mit KI werden inzwischen weltweit diskutiert (Brockmann 2017; Enquête 2018; Beijing 2019; European Commission 2019; Gent 2019; OECD

2019; Tegmark 2019; UN 2019). Sie müssen aber implementiert werden. Auch über die dramatischen gesellschaftlichen und ökologischen Folgen der Digitalisierung wird endlich verstärkt nachgedacht (Deutsche Energie-Agentur 2017), die Regulierung läuft aber der Entwicklung dramatisch hinterher (Grefe 2018). Straftäterinnen und Straftäter, die in der analogen Welt unter erheblichem Ermittlungsdruck stehen, können in der digitalen Welt ungestört ihr Unwesen ausleben. An solchen gravierenden rechtsethischen Problemen im gesellschaftspolitischen Bereich wird zwar inzwischen gearbeitet, die ökologischen Dimensionen sind aber noch längst nicht in angemessenem Umfang Gegenstand öffentlicher Debatten. Treibt nicht allein schon der exorbitante Energiehunger der IT-Revolution diese in die Aporie des abnehmenden Grenznutzens (Rifkin 2015)?

Besondere Herausforderungen stellt die militärische Nutzung der KI dar, insbesondere *Lethal Autonomous Weapon Systems* (LAWS). Für die militärische Nutzung gilt zusammengefasst: Attraktiv ist das günstige Verhältnis von Wirkung und Deckung, wenig Personal kann mit KI-unterstützten Cyberwaffen aus guter Deckung maximal wirken und dies mit kurzen Reaktionszeiten und großen Datenverarbeitungsvolumina, hohen Reichweiten, in schwer zugänglichen Regionen, über lange Phasen. Problematisch dagegen sind die Erzeugung realitätsnaher Datenmengen für das Training der Systeme, deren Einordnung in die Hierarchien militärischer Strategieentwicklung und Operationsführung, beispielsweise die Übersteuerung von Befehlsketten und schließlich die Senkung der Hemmschwelle zur Gewaltausübung angesichts vermeintlich geringerer Verluste an Menschenleben. Problematisch ist auch die völkerrechtliche Einhegung von KI: Erstens lassen sich KI-Systeme, die mit großem Abstand für den legalen zivilen Markt entwickelt werden, außer im Fall des maschinellen Lernens von zivilen relativ leicht in kriminelle und militärische Kontexte übertragen. Das macht Kontrollen und Beschränkungen in vielen Fällen nicht nur praktisch unmöglich, staatliche Hoheitsträger werden sogar zunehmend abhängig von zivilen Firmen. Zweitens bedrohen Cyberwaffen generell zivile kritische Infrastrukturen, verstärken drittens die generelle Hybridisierung und „Vergeheimdienstlichung“ der Kriegführung und die Nivellierung der Grenzen bewaffneter Konflikte im Cyber- und Informationsraum, erschweren viertens die Zurechnung feindlicher Handlungen und unterlaufen so schließlich die wechselseitig mäßigende Wirkung der

Abschreckung. KI kann im Targetingprozess die Lagebeurteilung durch die Truppenführung verbessern, darf sie aber nicht ersetzen. Das gesamte Szenario schreitet förmlich nach staats- und völkerrechtlicher Einhegung nach bewährten völkerrechtsethischen Maßstäben. Aber auch schon ohne entwickelte KI und ihre Übertragung auf militärisches Gebiet droht heute ein cybertechnologischer Kalter Krieg, der Milliarden von Menschen durch *Digital Iron Curtains* voneinander trennt und gegeneinander mobilisiert (Lee 2019).

Technisch wird jede KI, so mächtig sie auf ihrem Gebiet auch ist, auf absehbare Zeit dennoch nur ein „nützlicher Idiot“ sein, der stur, treu und hocheffektiv immer das tut, was andere ihm vorgesagt oder beigebracht haben – so etwa das Flugabwehrsystem *Patriot* mit elektronischer Freund-Feind-Erkennung – oder beigebracht haben, sich selbst beizubringen, wie etwa im Fall von „Prognostizierter Instandhaltung“ von Industrieanlagen und Profisportlerinnen und -sportlern (Elliot 2018). Ja, so wahr Kalkulieren und Mustererkennen, symbolverarbeitende Programmoptimierung und konnektionistisches Zeigen und Bearbeiten von Beispielen zum „Denken“ gehören, so wahr wird ein solches Programm tatsächlich „denken“ können, aber es wird damit noch lange nicht wissen, dass es denkt. In den Worten der Genesis: Da ist noch niemand, der ihnen die Augen öffnet. Und von solchen nützlichen Idioten wissen wir, was sie anrichten können. Mögen uns Gott und gesunder Menschenverstand vor ihrer unkontrollierten Vermehrung bewahren. Wir haben schon ohne die digitalen Superhirne genug davon. So lange sie zwar Verstand, aber keine Vernunft haben und des Willens und des Vermögens ermangeln, sich selbst ein Gesetz zu geben, müssen wir ihnen unsere Gesetze geben.

„Wir verfügen heute über gut funktionierende und etablierte Lernalgorithmen, um unsere Maschinen zu trainieren. Die Aufgabe für den menschlichen Trainer oder Entwickler ist aber bei komplexen Anwendungen immer noch anspruchsvoll. Es gibt nämlich viele Möglichkeiten, wie er das Training eines Roboters gestalten kann und er wird ohne Experimentieren nicht erfolgreich sein. Dieses Experimentieren kann in der Praxis sehr mühsam sein, denn jedes neue Lernprojekt muss entworfen und programmiert werden.“ (Ertel 2016, 334 f.)

Welche Datensätze auch immer wir unseren Maschinen als Trainingsaufgabe zuweisen, die Aufgabenstellungen sollten sich denjenigen Gesetzen fügen, die den Test auf die Ethik des Kategorischen Imperativs einschließlich der Idee des Rechts bestehen. Wenn es eine ethische und rechtliche Herausforderung der schwachen KI gibt, liegt sie in ihrer vernünftigen Anwendung nach den bewährten Prinzipien einer Ethik des Menschen- und Völkerrechts. Das beginnt schon bei der Auswahl der Datensammlungen: wählen wir diese beispielsweise nach rassistischen Kriterien aus, optimieren wir Maschinen auf effizienten Rassismus. Und ebenso gilt: wobei Menschen – zum Beispiel Diplomatinen und Diplomaten sowie Truppenführerinnen und Truppenführer – sich auch immer von diversen Programmen unterstützen und assistieren lassen, sollten sie ihre Assistentinnen und Assistenten sowie Beraterinnen und Berater vernünftig anlernen, klug auswählen, ihre Kompetenzen passgenau einsetzen, ihre vermutlich divergierenden Ratschläge besonnen gegeneinander abwägen, in diskursiv erweiterten Evidenzkonzepten zusammenführen, selbst bei anschwellenden Datenfluten den Überblick behalten. Und am Ende werden sie auch für die getroffenen Entscheidungen geradestehen und haften. Hoffentlich!

Wenn es sogar so ist, dass angelernte KI-Programme keinen Einblick zulassen, wie sie auf ein Ergebnis kommen, müssen sekundäre Kontroll-KI-Programme die primären Programme genau daraufhin für uns ausleuchten. Sollten solche und andere Rechenschaftslegungen auch nicht weiterführen und an technische Grenzen stoßen, bliebe vernünftigerweise nichts übrig, als die Voten solcher Maschinen schlicht für potenziell irrational und deshalb für irrelevant und unmaßgeblich zu erklären und ihnen nicht zu folgen sowie sie schon gar nicht selbst „entscheiden“ und „handeln“ zu lassen.

Ethische Urteilsbildung ist deutlich mehr als die Beherrschung mathematischer Funktionen, nämlich ein sorgsam gegliederter Prozess offener und deshalb nicht programmierbarer Kontingenzbewältigung. Soweit zur schwachen KI: Je mächtiger sie wirkt, umso engmaschiger und strikter muss sie menschlicher Kontrolle und Aufsicht unterstehen: *Meaningful Human Control of Critical Functions*. Das aber mindert ihre Effizienz und treibt jede KI auf dieser Ebene in die Zone des abnehmenden Grenznutzens. Das wird laut Kennerinnen und Kennern der internationalen Debatte weitgehend ignoriert, der

„Rausch‘ der Berechenbarkeit der Welt“ vernebelt die politische Vernunft und den nüchternen Verstand (Dahmann/Dickow 2019, 12). Ein extremer Fall: Wer große Schwärme von auf die hochfrequente und -effektive Eliminierung seiner „Feinde“ trainierten Killerrobotern in die Umwelt entlässt, mag sich eine Weile als Herr der Elemente fühlen, kann aber nicht ausschließen, dass die Algorithmen Lücken oder Fehler aufweisen, ihren Fokus verschieben, Lageveränderungen nicht nachvollziehen oder schlicht gehackt werden und dann ihn angreifen. Damit dies kein *Cyberlehrling* leichtfertig ausprobiert, sind mit lernender KI gesteuerte Kampfmittel nach demselben langfristigen Interessenkalkül zu ächten wie biologische und chemische Kampfstoffe. Das ist jedoch erheblich leichter gesagt als politisch-praktisch umgesetzt, denn eine Chance wird bestenfalls dem Kriterium des *Meaningful Human Control of Critical Functions* eingeräumt (Deutscher Bundestag 2018; Dahmann/Dickow 2019, 17–26). Sollte dessen trennscharfe Abgrenzung sich jedoch als schwierig erweisen, so könnten weitere Unterscheidungen ergänzend hinzutreten: zwischen auf feste Muster und Regeln konditionierten und kreativ ergebnisoffen lernenden Programmen sowie zwischen perfiden und transparenten Formen des Systemdesigns et cetera. Das Schicksal der *UN Convention on Certain Conventional Weapons* (CCW), insbesondere von Protokoll II über Landminen, Sprengfallen und andere Vorrichtungen und der Ottawa-Konvention außerhalb der Vereinten Nationen (VN), zeigt allerdings die begrenzte Wirkung von Appellen und Vereinbarungen. Freuen wir uns also über jeden, der nicht den breiten Weg der Torheit, sondern den schmalen Weg der Vernunft und Weisheit einschlägt (Matthäus 7, 13f).

Deutschland kann sofort zusammen mit einer Koalition der Willigen vorangehen und für die eigenen Rüstungsprojekte festlegen, dass (1) die Waffenkonstrukteure nur nach dem Recht bewaffneter Konflikte zulässige Funktionen in die Algorithmen einer KI eines Systems eintragen, dass (2) die Waffenbediener das Funktionieren dieses Systems nach Aktivierung präzise vorhersehen und es (3) bei Bedarf jederzeit kompetent administrieren können sowie, dass unabhängige Beobachter und Ermittler (4) ihre Wirkungen aus Gründen der nachträglich überprüfbaren Zurechenbarkeit zurückverfolgen können.

Ethische Herausforderungen starker KI

Nun kommt ein weiterer gewichtiger Einwand: Warum soll unser artifizuell von Ingenieurinnen und Ingenieuren auf die Welt gebrachtes vollautomatisch selbst- und tieflernendes Maschinenwesen nicht eines Tages ein autonomes reflexives Bewusstsein entwickeln? Warum soll es eines Tages denkend nicht doch auch wissen, dass es denkt? Auch diesen Einwand könnte man gelten lassen, obwohl kaum eine seriöse Wissenschaftlerin oder seriöser Wissenschaftler derzeit beansprucht, in dieser Richtung zu arbeiten. Und auch dazu eine These zur weiteren Debatte: Unsere künstliche Kreatur muss doch, um wissen zu können, dass sie denkt, und um sich ihres Verstandes „ohne Anleitung eines anderen bedienen“ zu können, zuvor so etwas wie Selbstorganisation entwickelt haben – bis dahin, dass sie von einer bestimmten Phase an die Charakteristika von Organismen annimmt, nämlich sich selbst zugleich Ursache und Wirkung zu sein (vgl. Wein93). Und damit wird sie aufhören „artifizuell“ zu sein, sie wird sich aus unseren Produktions-, Domestikations- und Trainingsbestrebungen emanzipiert haben. Denn das ist die Definition eines Organismus, er tut, was er will und er will, was er kann. Und dann kann er – oder sie oder es – eines Tages tatsächlich sein „wie wir“, – wenn nicht gar „wie Gott“?

Wozu aber die ganze Aufregung? Wird ein Cyborg – er oder sie oder es – eines Tages so sein wie wir? Ähnlich mächtig und ähnlich nackt, schwach und sterblich, ähnlich leistungsfähig und ähnlich verwundbar, ähnlich genial und ähnlich verführbar. Selbst wenn der auto-kybernetisch-poietische Organismus dank edelster Rohstoffe und ausgeklügeltester Architektur jenseits unserer altmodischen Desoxyribonukleinsäure sogar unter extremen Umweltbedingungen extrem resilient sein wird, wenn er dank überlegener Sensorik offen für vielfältige Wahrnehmungen ist, die wir Menschen gar nicht kennen und wenn er dank gigantischer und hierarchisch interferenziell arbeitender Rechenkapazitäten höchste Datenvolumina hochfrequent und hochkomplex und deshalb mit dem Energieaufwand einer Glühbirne verarbeiten kann, so wird er doch dank seiner Lernfähigkeit und Fehlerfreundlichkeit – und zwar als Individuum ebenso wie als sich selbst evolutionär rekombinant selbstregenerierende Gattung – extrem plastisch sein, und das, ohne sich dabei zu verbiegen.

Hier werden die Schöpferinnen und Schöpfer unseres transhumanen Superorganismus begeistert applaudieren: Unsere überlegenen Superheldinnen und Superhelden werden sich weder verbiegen noch verführen lassen, sie werden sich keinem fremden Willen beugen, ihnen wird niemand etwas weismachen können, nahezu unsterblich unterliegen sie nur den Gesetzen alles Materiellen, also etwa des radioaktiven Zerfalls! Gut, sollten wir dann entgegnen: Wie können sie aber plastisch sein, ohne biegsam zu sein, wie können sie lernen und ihre Art evolutionär entlang variabler Umwelten entwickeln, ohne zu pubertieren, ohne Risiken einzugehen, ohne Fehler zu machen, wo es doch ohne Versuch und Irrtum, ohne Fehlerfreundlichkeit kein Forschen, kein Lernen und keine Evolution geben kann?

Und selbst wenn sich unsere Cyborgs selbstständig vermehren: Noch nicht ausdifferenzierte und unausgereifte Organismen, ob biologisch gezeugt oder eines Tages cybertechnologisch hergestellt, sind reproduzierbar, aber nicht leistungsfähig, also schwach. Stark und hochleistungsfähig sind nur Organismen, die sich dank eines flexiblen genetischen Codes ausdifferenzieren und heranreifen. Wollte man sie als Replikate in Serie herstellen, wäre dies der Versuch, sie zu Mitteln für vordefinierte Zwecke und damit zu nützlichen Idioten im Sinne schwacher KI zu degradieren. Sind Cyborgs aber wirklich stark, dann ist vorprogrammiert, dass sie sich gegen eine solche strukturelle Unterforderung wehren und mit allen Risiken eigene Wege gehen.

Also, die Leserinnen und Leser ahnen es schon, – wenn Cyborgs in die Nähe dessen gekommen sein werden, wo sie manche experimentierenden Kybernetikerinnen und Kybernetiker spekulativ haben wollen, werden wir sie erst einmal taufen und in der Sonntagschule vernünftig „konfirmieren“ müssen. Sie also einladen, mit uns und mit ihresgleichen über die gar nicht so furchtbar alte Geschichte vom Sündenfall nachzusinnen. Aber dann sind wir keinen Schritt weiter, als wir es jetzt schon sind. Denn unsere heranwachsenden Cyborgs stünden vor derselben Aufgabe wie wir: den Garten Eden der naiven Unschuld, den Brutkasten, die Kinderstube, das Elternhaus, das heimische Revier zu verlassen, Kontingenz zu verarbeiten, nicht in schlechte Gesellschaft zu geraten, erwachsen zu werden und für sich und ihre „Brüder“ und „Schwestern“ – und eines Tages auch ihre „Eltern“ und ihre „Kinder“ – Verantwortung zu übernehmen. Vor allem in technologisch erweiterten Biotopen und das mit allen

Konsequenzen und auch in Würde zu altern. Und wir alle zusammen werden das tun müssen, was in entwickelten Zivilisationen seit eh und je getan werden muss: neugierig bleiben, Empathie entwickeln, fleißig sein, uns im Griff haben, unsere sieben Sachen aufräumen, auf die Gesundheit achten, die Stromrechnung im Blick behalten, Nachwuchs großziehen, keine falschen Versprechungen machen, auf keine falschen Versprechungen hereinfallen, Recht und Staat aufrichten, authentisch und wahrhaftig sein, Platz für den Nächsten machen – was halt der übliche Alltag so mit sich bringt.

Bibliografie

- (Beijing 2019)
Beijing Academy of Artificial Intelligence (2019). Beijing AI Principles, Beijing VRC. <https://www.baai.ac.cn/blog/beijing-ai-principles> (letzter Aufruf: 25.6.2019).
- (Böhme 2000)
Böhme, Hartmut (2000). Enträumlichung und Körperlosigkeit im Cyberspace und ihre historischen Vorläufer; In: Theisen, Bianca (Hrsg.): Modern Languages Notes 115, S. 423–441.
- (Bostrom 2014)
Bostrom, Nick (2014). Superintelligenz: Szenarien einer kommenden Revolution, Frankfurt a. M.
- (BPB 2019)
Bundeszentrale für Politische Bildung (Hrsg.) (2019). Datenökonomie. Aus Politik und Zeitgeschichte APUZ 24-26/2019, Bonn.
- (Brockmann 2017)
Brockmann, John (2017²). Was sollen wir von Künstlicher Intelligenz halten? Frankfurt a. M.
- (Buchanan/Miller 2017)
Buchanan, Ben/Miller, Taylor (2017). Machine Learning for Policy Makers. What It Is and Why It Matters, in: <https://www.belfercenter.org/sites/default/files/files/publication/MachineLearningforPolicymakers.pdf> (letzter Aufruf: 11.6.2019).
- (Crocker/Siekman 2010)
Crocker, Matthew W./Siekman, Jörg (Hrsg.) (2010). Resource-Adaptive Cognitive Processes, Heidelberg/Berlin.
- (Dahmann/Dickow 2019)
Dahmann, Anja/Dickow, Marcel (2019). Präventive Regulierung autonomer Waffensysteme (SWP-Studie 1, Januar 2019), Berlin.
- (Deutscher Bundestag 2018)
Deutscher Bundestag (2018). Regulierung von autonomen Waffensystemen. Antwort der Bundesregierung (DS 19/3219), Berlin.
- [Deutsche Energie-Agentur 2017]
Deutsche Energie-Agentur (2017). Analyse der mit erhöhtem IT-Einsatz verbundenen Energieverbräuche infolge der zunehmenden Digitalisierung. Status Quo und Prognosen, Berlin.
- (Dickow/Jacob 2018)
Dickow, Marcel/Jacob, Daniel (2018). Das globale Ringen um die Zukunft der künstlichen Intelligenz (SWP-Aktuell 24/2018), Berlin.
- (Elliot 2018)
Elliot, Timo (2018). Drei Anwendungsfelder für maschinelles Lernen im Unternehmen, in: <https://news.sap.com/germany/2018/09/maschinelles-lernen-beispiele/> (letzter Aufruf: 14.6.2019).
- (Enquête18)
Die im Juni 2018 eingesetzte Enquête-Kommission des Deutschen Bundestages „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale“ (https://www.bundestag.de/ausschuesse/weitere_gremien/enquete_ki) wird ihre Ergebnisse und Handlungsempfehlungen nach der parlamentarischen Sommerpause 2020 vorlegen; vgl. <http://dip21.bundestag.de/dip21/btd/19/029/1902978.pdf> (letzter Aufruf: 11.6.2019).
- (Ertel 2016)
Ertel, Wolfgang (2016⁴). Grundkurs Künstliche Intelligenz. Eine praxisorientierte Einführung, Wiesbaden.
- (European Commission 2019)
European Commission High-Level Expert Group on AI (2019). Ethics Guidelines for Trustworthy Artificial Intelligence, Brussels.
- (Gent 2019)
Gent, Edd (2019). What's Behind the International Rush to Write an AI Rulebook? <https://singularityhub.com/2019/06/11/whats-behind-the-international-rush->

- [to-write-an-ai-rulebook/](#) (letzter Aufruf: 11.6.2019).
- (Gertz 2018)
Gertz, Jan Christian (2018). Das erste Buch Mose (Genesis). Die Urgeschichte Gen 1–11, Göttingen.
- (Görz/Rollinger/Schneeberger 2003)
Görz, Günther/Rollinger, Claus-Rainer/Schneeberger, Joseph (2003⁴, Hrsg.). Handbuch der Künstlichen Intelligenz, München; vormals Görz, Günther (1995, Hrsg.). Einführung in die Künstliche Intelligenz, Bonn et al.
- (Grefe 2018)
Grefe, Christiane (2018). „Der Stromhunger wächst“. Interview mit Tilman Santarius. In: *Zeit.de* vom 3.2.2018. <https://www.zeit.de/2018/06/digitalisierung-klimaschutz-nachhaltigkeit-strombedarf> (letzter Aufruf: 13.6.2019).
- (IDG 2018)
IDG Research Services (Hrsg.) (2018). Machine Learning/Deep Learning 2018. München.
- (Köhlmoos 2017)
Köhlmoos, Melanie (2017). Naiver Glaube? Schöpfungstheologie in ihrem altorientalischen und alttestamentlichen Kontext. In: Kleinert, Markus/Schulz, Heiko (Hrsg.) (2017: Natur, Religion, Wissenschaft. Beiträge zur Religionsphilosophie Hermann Deusers. Tübingen, S. 3–23.
- (Kratz 2017)
Kratz, Reinhard Gregor (2017²). Historisches und biblisches Israel. Drei Überblicke zum Alten Testament, Tübingen.
- (Lämmel/Cleve 2012)
Lämmel, Uwe/Cleve, Jürgen (2012⁴). Künstliche Intelligenz, München.
- (Lee 2019)
Lee, Kai-Fu (2019). Al-Superpowers: China, Silicon Valley und die neue Weltordnung, Frankfurt a. M.
- (Lünendonk/Hossenfelder 2018)
Lünendonk & Hossenfelder GmbH (Hrsg.) (2018). Relevanz von künstlicher Intelligenz für große Unternehmen. Mindelheim.
- (Mainzer 2018)
Mainzer, Klaus (2018). Künstliche Intelligenz – Wann übernehmen die Maschinen? Berlin.
- (OECD 2019)
OECD (2019). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449), Paris. OECD-LEGAL-0449-en.pdf (letzter Zugriff: 25.6.2019).
- (Rifkin 2015)
Rifkin, Jeremy (2015). Zero Marginal Cost Society. New York.
- (Tegmark 2019)
Tegmark, Max (2019). Leben 3.0. Mensch sein im Zeitalter Künstlicher Intelligenz. Berlin.
- (UN 2019)
UN Secretary-General's High-level Panel on Digital Cooperation (Hrsg.) (2019). The Age of Digital Interdependence. Report of the UN Secretary-General's High-level Panel on Digital Cooperation. New York.
- (Wahster et al. 2014)
Wahlster, Wolfgang et al. (Hrsg.) (2014). Towards the Internet of Services: The Theseus Research Program. Berlin/Heidelberg/New York.
- (Wein 1993)
Weingarten, Michael (1993). Organismen – Objekte oder Subjekte der Evolution? Philosophische Studien zum Paradigmenwechsel in der Evolutionsbiologie. Darmstadt.
- (Zuboff 2019)
Zuboff, Shoshana (2019). The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power. New York (Deutsch: Das Zeitalter des Überwachungskapitalismus, Frankfurt a. M./New York 2018.