D’Artagnan oder die vierte Kränkung

Zugänge zu menschlicher und Künstlicher Intelligenz

«Is this the real life? Is this just fantasy?
Caught in a landslide, no escape from reality.»

Queen, Bohemian Rhapsody[1]

«Er ist gestorben aus Weltfremdheit und weil er seine Lebensbedingungen, die für ihn vernichtend geworden waren, nicht zu ändern verstand.»

Walter Benjamin, Zum Bilde Prousts[2]

1. Vorbemerkung

Künstliche Intelligenz (KI) ist längst von Science-Fiction zum Alltagsphänomen geworden. KI dringt mehr oder weniger unbemerkt, geräuschlos und geschmeidig in die Lebenswelten ein. Wie bei jeder neuen Grosstechnologie können auch bei KI die damit einhergehenden, erhofften oder befürchteten Veränderungen enthusiastisch begrüsst, exzessiv verteufelt, abwartend analysiert oder krampfhaft ignoriert werden. Neu ist dagegen, dass die Begründungen für die eine oder andere Einstellung möglicherweise schon bald von der Technologie (Chatbots, *large language modells, LLM*; *natural language processing, NLP*) selbst kommen könnten. KI ist auf dem heutigen Stand eng verbunden mit fünf global agierenden Technologiekonzernen. Von einem stammt die Hardware, auf dem dieser Text mit Hilfe der Software eines anderen geschrieben und für dessen Recherche auf das Programm und die Dienstleistung zweier weiterer zurückgegriffen wurde. Es sind meine Gedanken, die ich selbst verschriftlicht habe – aber was bedeutet schon «meine» und «selbst», wenn vier der grossen Big 5 Tech-Giganten an der Textentstehung beteiligt waren? Welche Implikationen und Konsequenzen hat es, mit Unterstützung von KI über KI zu schreiben?

Fest steht, dass das Ergebnis ohne den Technologieeinsatz ein anderes gewesen wäre. Offen ist dagegen, worin die Abweichung bestanden hätte und wodurch die Textproduktion in die eine und in die andere Richtung beeinflusst worden wäre. Die Idee, dass hinter dem Prozess, bei dem ein weisses Blatt Papier sukzessive mit sprachlichen Symbolen geschwärzt wird, eine Person als Autorin steht, scheint weiterhin unverzichtbar. Ich kann davon überzeugt sein, meinen Laptop wie eine Schreibmaschine zu verwenden. Aber weil ich diese Einstellung völlig unabhängig von einer Antwort auf die Frage haben kann, ob ein Laptop das Gleiche mit mir macht, wie eine Schreibmaschine, kann die Überzeugung auf einer schlichten Unterstellung beruhen. Nichtsdestotrotz ist sie von erheblichem Gewicht, etwa für die rechtliche Klärung komplexer Urheber- und Copyrightfragen oder für die anthropologische und politische Konzeption von Freiheit im Sinn von Immanuel Kants «*Freiheit der Feder*» als «einziges Palladium der Volksechte».[3] Genauso wenig kann ich auf die Unterstellung verzichten, wenn ich mich im Spiegel wiedererkennen, eigene Positionen vertreten und durchsetzen, das Leben als meine

Biographie revuepassieren lassen möchte oder wenn ich will, dass mich andere als die Person erkennen, die ich bin. Eine allgemeine Problembeschreibung könnte lauten: Die traditionellen Denksysteme als Voraussetzung und Legitimationsgrundlage unserer etablierten Praktiken und Lebensweisen sind kunstvoll ausbalancierte Bauklotztürme und es scheint so, als würde KI, genauer unser Nachdenken darüber, ausgerechnet die untersten Steine abräumen wollen.

Die bedrohlich klingende Wahrnehmung lässt sich aber auch umkehren: «Etwas, das grossartig ist an der aktuellen Diskussion […], besteht darin, dass in der breiten Öffentlichkeit plötzlich ein Bewusstsein über den Status und die Position des Menschen entstanden ist. Noch nie zuvor konnte man so direkt und unmittelbar über das Kernproblem, nämlich den ‹Menschen› selbst diskutieren. Die künstliche Intelligenz ist ein Spiegel von uns Menschen, und die Frage, ob es bei Robotern um die Ablösung bzw. den Ersatz von uns Menschen geht oder ‹bloss› um unsere Assistenz, kommt dann erst an zweiter Stelle. Zunächst müssen wir uns nämlich damit beschäftigen, wer und was wir Menschen sind.»[4] Die Imaginationen von KI dienen als Spiegel, in dem sich Personen und Gesellschaften selbst anschauen. Die vorgestellten Bilder sind weder nur technologische noch nur anthropologische, sondern entspringen einer diffusen Mischung aus beiden. So ist bereits die Gegenüberstellung von künstlicher Intelligenz einerseits und menschlicher, natürlicher, carbonbasierter … Intelligenz unklar. Das lange Zeit gültige Alleinstellungsmerkmal von *homo sapiens* erübrigte jede Verhältnisbestimmung oder Abgrenzung gegenüber anderen Intelligenzwesen und -formen. Die neue Herausforderung wird nicht durch das berechtigte Argument erledigt, dass «der Mensch» viel mehr sei als ein neuro-kognitives Aggregat.

Die Diskussionen über KI gleichen dem Balanceakt auf einem Drahtseil, bei dem KI am einen und die Menschenbilder am anderen Ende der Balancierstange platziert werden, und der deshalb die Frage aufwirft, *wer* dann eigentlich mit der Stange in den Händen auf dem Seil balanciert. Diese Perspektive auf die KI-Debatten unterscheidet sich von anderen möglichen, hat aber gegenüber vielen aktuellen Diskussionen drei Vorteile: (1.) Sie ist nicht auf technische und rechtliche Implementierungs- und Anwendungsfragen beschränkt; (2.) sie vermeidet eine vorschnelle anthropozentrisch-apologetische Haltung und (3.) sie ist anschlussfähig an theologische und ethische Diskurse. Die Fragen nach KI, nach den Subjekten, die wechselseitig «interagieren», und nach den individuellen und gesellschaftlichen Folgen sind mehrfachadressiert. Sie lassen sich weder allein durch die Beschreibung, Analyse und Normierung von Technologien beantworten, noch ausschliesslich durch die Begründung politischer, rechtlicher und ethischer Ansprüche gegenüber den Technologien und ihren Anwendungen. Der dringende rechtliche Regelungsbedarf kann (bisher) nur in der Form geleistet werden, dass die KI analog zu anderen Technologien behandelt wird, weil das Recht keinen alternativen Umgang mit Technologien kennt. Voraussetzung dafür wäre eine sorgfältige Beschreibung und Analyse der Konstellationen, die sich aus der gesellschaftlichen Implementierung von KI ergeben.

Hier setzen die nachfolgenden Überlegungen ein. Wie kann über KI gesprochen werden? An welche Diskurse kann angeschlossen werden? Wo bestehen Übereinstimmungen und wo stösst die Übernahme vertrauter und etablierter Perspektiven und Argumente an ihre Grenzen? Mit diesem allgemeinen Einstieg ist die Idee verbunden, das Thema in weiteren Beiträgen in lockerer Folge zu vertiefen.

2. KI zwischen Alltag und Ausnahmezustand

Kein Navigations- und Assistenzsystem, kein Haushaltsroboter und Thermostat kommen ohne KI aus. Die Smartphones haben uns zu KI-Junkies gemacht. Wir geraten in Stress, wenn das Gerät unauffindbar oder der Akku leer ist, und wenn es abhandenkommt, reagieren wir, als hätten wir einen Teil von uns selbst verloren. Die Abhängigkeit ist offenkundig. «Welche Werbung Sie online sehen, welche Kauf-, Seh- und Hörempfehlungen Ihnen auf Amazon, Netflix, YouTube oder Spotify gegeben werden, bestimmt eine KI. Viele Nachrichten, die Sie lesen, sind von einer KI verfasst worden. Und häufig bestimmen KIs zumindest mit, ob Sie zum Vorstellungsgespräch eingeladen werden, wie hoch Ihre Versicherungsbeiträge sind und wie Sie behandelt werden, wenn Sie krank sind. Diese Liste liesse sich so lange fortführen, dass wohl nur eine KI die Geduld hätte, sie bis zum Ende zu lesen.»[5] Kaum jemand kommt heute um «GAFAM», die big 5 Tech Companies, herum: Google/Alphabet (Umsatz 2022: $279.8 billion), Apple (Umsatz 2022: $394.33 billion), Facebook/Meta (Umsatz 2022: $116.61 billion), Amazon (Umsatz 2022: $513.98 billion) und Microsoft (Umsatz 2022: $198.3 billion).[6] Die Macht der globalen Konzerne und die Technologien, die sie – im doppelten Wortsinn – grossgemacht haben, gehören untrennbar zusammen. Die Unternehmensstrategien sind genauso undurchschaubar, wie die Technologien, die sie anbieten, die wir nutzen und denen wir immer mehr Angelegenheiten unserer Leben anvertrauen.

KI ist gleichzeitig etwas Vertrautes, das inzwischen fest zum Alltag gehört, und etwas weitgehend Unbekanntes, das nur schwer auf den Begriff gebracht und mit den etablierten Kategorien beschrieben werden kann. Deshalb wird auf metaphorische und anthropomorphe Sprache zurückgegriffen: Hardware und Algorithmen, die «interagieren», «selbständig denken» und «Entscheidungen fällen» oder einen «autonomen» Status haben, der die Forderung eigener «Rechte» nahelegt. Das sind Attribute, die zuvor allein Personen und mit graduellen Abstufungen bestimmten Tieren zugesprochen wurden. Die KI kommt von allen Technologien unserem Verständnis von uns selbst am nächsten. Das fasziniert ebenso, wie es Misstrauen und Befürchtungen hervorruft. Wir kennen nichts Vergleichbares, das uns – gemäss unserer eigenen KI-Narrative – so ähnlich ist und gleichzeitig so geheimnisvoll fremd und anders.

Als Geburtsstunde der KI gilt das von John McCarthy, Marvin Minsky, Nathaniel Rochester und Claude Shannon im Sommer 1956 durchgeführte *Dartmouth Summer Research Project on Artificial Intelligence*. Die Idee wird im Projektantrag formuliert: «Das Seminar soll von der Annahme ausgehen, dass grundsätzlich alle Aspekte des Lernens und anderer Merkmale der Intelligenz so genau beschrieben werden können, dass eine Maschine zur Simulation dieser Vorgänge gebaut werden kann. Es soll versucht werden, herauszufinden, wie Maschinen dazu gebracht werden können, Sprache zu benutzen, Abstraktionen vorzunehmen und Konzepte zu entwickeln, Probleme von der Art, die zurzeit dem Menschen vorbehalten sind, zu lösen, und sich selbst weiter zu verbessern.»[7] KI wird hier als Simulationstechnologie menschlicher Kognitionsfähigkeiten verstanden. Durch die enorme Ausweitung der Rechnerkapazitäten seit 2015 wurde die Idee der Simulation von Kognition sukzessive durch Konzepte Lernender Systeme ersetzt. «Lernende Systeme sind Maschinen, Roboter und Softwaresysteme, die abstrakt beschriebene Aufgaben auf Basis von Daten, die ihnen als Lerngrundlage dienen, selbstständig erledigen, ohne dass jeder Schritt spezifisch programmiert wird.»[8] Sie begegnen heute

in den Bereichen Übersetzung und Textproduktion, Sprach- und visuelle Erkennung, Roboter (*care robots, sex robots)* autonomes Fahren, autonome Waffensysteme (*lethal autonomous weapons systems, LAWS*) und Virtualisierung (*virtual* und *augmented reality*) etwa bei Spielen, Lernprogrammen oder in der Prognostik (*predictive analytics*). Alle Anwendungen beruhen auf den Basistechnologien der Datenverarbeitung und Entwicklung von Algorithmen. Grob kann dabei zwischen geschlossenen und offenen KI-Funktionen unterschieden werden. Während geschlossene Systeme genau definierte Funktionen erfüllen (wissensbasierte Systeme zur Datenanalyse), sind offene Systeme dadurch gekennzeichnet, dass sie Aufgaben kreativ und auf eigenständige Weise (lernende Systeme; *deep learning*) erfüllen.

In der Schweiz ist die KI-Diskussion in Politik und Wissenschaften fest etabliert. 2018 setzte der Bundesrat im Rahmen der «Strategie ‹Digitale Schweiz›» eine interdepartementale Arbeitsgruppe ein, die Ende 2019 einen Bericht *Herausforderungen der künstlichen Intelligenz* und 2020 *Leitlinien* für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung vorlegte.[9] Viele Institutionen, exemplarisch die Schweizerische Akademie der technischen Wissenschaften (SATW), setzen sich für bessere rechtliche und politische Rahmenbedingungen von KI ein.[10] 2018 war die Schweiz weltweit führend bei den KI-Startups pro Kopf und in der KI-Forschung gemessen am *citation impact score*.[11] Im gleichen Jahr lancierte der Schweizerische Nationalfonds (SNF) das 64 Projekte umfassende NFP 77 «Digitale Transformation».[12] 2020 publizierte die nationale Stiftung TA-Swiss eine Studie zu Chancen und Risiken der KI für die Bereiche Arbeit, Bildung, Medien, Konsum und Verwaltung.[13] Seit 2015 betreibt die Nonprofit-Genossenschaft MINDATA eine Datenplattform, für die sie als Treuhänderin der deponierten persönlichen Datendossiers auftritt.[14] Das Kompetenzzentrum Digital Society Initiative (DSI) der Universität Zürich hat im Bereich KI in der Medizin ein Projekt zum gesundheitsbezogenen Nutzen von datenbasierten digitalen Abbildern von Personen, sogenannten «digital twins» durchgeführt.[15] In einer Bevölkerungsbefragung im Sommer 2023 sprachen sich 62% der knapp 1500 Befragten für die Nutzung eines digitalen Zwillings durch Gesundheitsfachpersonen aus. Ein deutliches Misstrauen bekundeten die Befragten gegenüber der Weitergabe von anonymisierten Daten an private Unternehmen (Pharma-Branche, Tech-Unternehmen und Krankenkassen) im Gegensatz zu Hochschulen, öffentlichen Spitälern und Bundesämtern.[16] Für die gesetzlichen Rahmenbedingungen hat der Bundesrat Ende 2022 eine Motion für ein Rahmengesetz für die Sekundärnutzung von Daten angenommen.[17]

«Is this the real life? Is this just fantasy? / Caught in a landslide, no escape from reality.» Die ersten Songzeilen von Freddie Mercury's *Bohemian Rhapsody* werfen die uralte Frage nach der Wirklichkeit der denkenden und erkennenden Subjekte und der Realität der Erkenntnisobjekte auf: Welchen Standpunkt nimmt das denkende und erkennende Subjekt ein und wo hat das Gedachte und Erkannte seinen Ort? Allgemein wird davon ausgegangen, dass die Wirklichkeit unabhängig davon existiert, ob sie wahrgenommen und erlebt wird oder nicht. Allerdings ist es aus epistemologischer Sicht äusserst schwierig, Gründe für die Überzeugung anzugeben, weil wir nicht sagen können, wie die Gegenstände unserer Erfahrung ausserhalb und unabhängig davon beschaffen sind. Das Problem begegnet in der Philosophie seit Platons berühmtem Höhlengleichnis als erkenntnistheoretische Frage,[18] praktisch wird es erst mit der Produktion virtuelle Welten durch KI.

3. KI-Geschichten

Die Menschen von der Antike bis zur Neuzeit hatten zwar – soweit wir das rekonstruieren können – ein genaues Gespür für die Diskrepanzen des Lebens zwischen Glück und Unglück, aber waren weit von der Idee und den Möglichkeiten entfernt, ihr Schicksal selbst in die Hand zu nehmen. Dass «der Mensch» der Schmied seines eigenen Glücks ist, war die längste Zeit der Menschheitsgeschichte eine unvorstellbare und unerreichbare Vorstellung. Erst mit den modernen wissenschaftlich-technischen Entwicklungen wird sie zu einer realen Herausforderung. Die fundamentalen Veränderungen seit der Neuzeit zeigen sich am deutlichsten darin, dass die Widerfahrnisse des Lebens nicht mehr einfach als (von fremder Hand bewirktes) Schicksal hingenommen und gedeutet, sondern als dreifache «Kränkung» erlebt werden:[19] als *kosmologische Kränkung*, dass die Erde nicht der Mittelpunkt des Weltalls ist (Kopernikus); als *biologische Kränkung*, dass die menschliche Spezies aus der Tierwelt hervorgegangen ist (Darwin) und als *psychologische Kränkung*, «dass das Ich nicht Herr sei in seinem eigenen Haus» (Freud).[20] Dieser Demütigungstrias könnte heute als vierte Desillusionierung eine *epistemische Kränkung* hinzugefügt werden,[21] die das seit der Antike behauptete anthropologische Alleinstellungsmerkmal vom vernunftbegabten Tier (griech. *zoon logon echon*; lat. *animal rationale*) angreift. Freilich blieben alle Kränkungen ohne grössere Konsequenzen und führten eher dazu, die erkannten Schwachstellen mit noch mehr Eifer technisch zu kompensieren. Der Philosoph Günther Anders beschreibt die vierte Kränkung als «Prometheische Scham», die in der Einsicht besteht, dass «was Kraft, Tempo, Präzision betrifft, der Mensch seinen Apparaten unterlegen ist; dass auch seine Denkleistungen, verglichen mit denen seiner ‹computing machines›, schlecht abschneiden».[22] Die Menschen können mit ihrer eigenen Zivilisation nicht mehr mithalten, sie werden zu Opfern ihrer eigenen Erfolgsgeschichte: «Und wir? Und unser Leib? Nichts von täglichem Wechsel […] Er ist morphologisch konstant; moralisch gesprochen: unfrei, widerspenstig und stur; aus der Perspektive der Geräte gesehen: konservativ, unprogressiv, antiquiert, unrevidierbar, ein Totgewicht im Aufstieg der Geräte. Kurz: die Subjekte von Freiheit und Unfreiheit sind ausgetauscht. Frei sind die Dinge: unfrei ist der Mensch.»[23] Die epistemische oder prometheische Kränkung und ihre Folgen können natürlich sehr unterschiedlich gedeutet werden. Drei Grundoptionen:

1. Die Abschaffung von homo sapiens: Aus seiner neutralen Beobachtungsperspektive präsentiert Yuval Noah Harari eine evolutionär-entspannte Sicht: «Trotzdem gibt es keinen Grund zur Panik. Zumindest nicht jetzt gleich. Der Aufstieg des Sapiens wird ein allmählicher historischer Prozess sein und keine Apokalypse à la Hollywood. Homo sapiens wird nicht durch eine Roboterrevolte ausgelöscht werden. Vielmehr wird er sich wahrscheinlich Schritt für Schritt auf eine höhere Stufe befördern und dabei mit Robotern und Computern verschmelzen, bis unsere Nachfahren rückblickend feststellen werden, dass sie nicht mehr die Art von Lebewesen sind, welche die Bibel verfassten, die Chinesische Mauer erbauten und über Charlie Chaplins Albernheiten lachten.»[24] Die Version wäre kompatibel mit der göttlichen Verheissung «Ich will meinen Bund mit euch aufrichten: Nie wieder soll alles Fleisch vom Wasser der Sintflut ausgerottet werden, und nie wieder soll eine Sintflut kommen, um die Erde zu verderben.» (Genesis 9,11) Denn das Versprechen gilt der gesamten Schöpfung, nicht einer Gattung und erst recht nicht der menschlichen. Der privilegierte menschliche Status, der an vielen Stellen aus der Bibel herausgelesen wird, ist eine kirchlich-theologische Erfindung.

2. Die Verbesserung oder Überwindung von homo sapiens: Trans- und posthumanistische Ideen von der menschlichen Optimierung oder Überwindung (*enhancement*, *cyborgization*, *mind uploading*) sind inzwischen aus den Science Fiction- und Esoterik-Abteilungen in den Bereich der etablierten Wissenschaften eingewandert. Hinter den Schlagwörtern von Trans- und Posthumanismus verbergen sich sehr heterogene Ansätze, die von medizinischem und biotechnologischem Enhancement (Neuroenhancement, Protetik, Biohacking, Kryonik) über posthegemoniale und ökologische Kulturen bis hin zu individueller Unsterblichkeit (Transformation von der *carbon-based* zur *silicon-based* Existenz) reichen.[25] Trans- und posthumanistische Ansätze zielen in unterschiedlicher Weise auf eine Emanzipation von den – als mangelhaft, beschränkend, nicht zukunftsfähig oder «schlecht» beurteilten – biologischen Merkmalen von *homo sapiens* («Mensch 1.0»). Die *condition humaine* gelten nicht länger als vorgegebene und unhintergehbare Bedingungen menschlicher Existenz: «Dieser [humanistische] Mensch war, wie sich erwies, nicht der universalistische Kanon vollkommener Proportionen, der ein naturgesetzliches Ideal ausdrückt, sondern ein historisches, wert- und standortgebundenes Konstrukt. Individualismus ist kein Bestandteil der ‹menschlichen Natur›, wie liberale Denker zu glauben geneigt sind, sondern eine geschichts- und kulturspezifische Diskursformation – eine Konstruktion, die noch dazu immer problematischer wird.»[26] Während der Transhumanismus auf eine technologische Perfektionierung des menschlichen Körpers, seine Lebensverlängerung und die Unsterblichkeit des Geistes (durch digitale *whole brain emulation*) zielt, strebt der Posthumanismus die Überwindung des menschlichen Körpers und damit eine kategoriale Überschreitung der anthropozentrischen Perspektive oder Existenz an. Ein gegenüber der technologischen Version erkenntniskritischer Posthumanismus (Braidotti: «Antihumanismus») erklärt das humanistische Projekt der Aufklärung für gescheitert: «Was zähmt noch den Menschen, wenn der Humanismus als Schule der Menschenzähmung scheitert? Was zähmt den Menschen, wenn sie bisherigen Anstrengungen der Selbstzähmung in der Hauptsache doch nur seiner Machtergreifung über alles Seiende geführt haben? Was zähmt den Menschen, wenn nach allen bisherigen Experimenten unklar geblieben ist, wer oder was die Erzieher wozu erzieht?»[27] Die kulturkritische Stossrichtung, aber auch der Gedanke der Transformation und Überwindung des «alten Menschen» üben eine wechselseitige Anziehungskraft der Theologie auf Trans- und Posthumanismus und umgekehrt aus.[28]

3. Die Dezentrierung von homo sapiens: Die Varianten der Abschaffung, Perfektionierung oder Überwindung von *homo sapiens* begegnen zwar in vielen kritischen Stimmen, können aber nicht ernsthaft als Zielperspektive von KI behauptet werden. Als Grundlage für eine Erzählung über das Verhältnis zwischen der Menschheit und KI bietet sich der Roman von Alexandre Dumas *Les Trois Mousquetaires* an.[29] Der Romantitel nennt nur drei Musketenschützen des Königs. Gemeint sind Athos, der Adelige, Aramis, der Geistliche, und Prothos, der Bourgeois. Als Repräsentanten der drei Stände der französischen Gesellschaft konnte die Trias nicht durch einen vierten Musketier ergänzt werden (weil es keinen vierten Stand gab).[30] Aber ausgerechnet der überzählige D'Artagnan wird dann den bekannten Schwur der vier Freunde formulieren: «et les quatre amis répétènt d'une seule voix la formule dictée par d'Artagnan: ‹Tous pour un, un pour tous›.»[31] Analog zum Schema von Dumas' Musketieren – «Die Dreiheit ist komplett, sie erhält sich aber nur durch die Hinzufügung eines vierten Elements»[32] –, führen die drei Kränkungen nicht zur sukzessiven Abschaffung (Harari), zu Perfektionierungs- oder

Überwindungsstrategien (Trans- und Posthumanismus) der Menschheit, vielmehr wird sie durch die vierte Kränkung – im Sinn der Dezentrierung und Depotenzierung ihrer kognitiven Fähigkeiten durch die KI – entlastet und zukunftsfähig. Es geht an dieser Stelle ausdrücklich nicht um eine Definition oder Beschreibung von KI, sondern um Deutungsoptionen ihrer Funktion für die menschliche Selbstwahrnehmung. Der anthropozentrismuskritische Impuls weist Überschneidungen mit Braidottis «Antihumanismus» aus, dessen Betonung der «nomadischen Existenz» auch theologisch attraktiv erscheint:[33] «Das Projekt über nomadische Subjekte […] dient als analytisches Werkzeug, um drei Klassen von Objekten zu betrachten. Erstens die kulturellen Mutationen, die ich als kulturelle Kartographie bezeichne: Was geschieht mit Körpern, Identitäten, Zugehörigkeiten in einer Welt, die technologisch vermittelt, ethnisch gemischt ist und sich sehr schnell in vielerlei Hinsicht verändert. Zweitens ergibt sich ein eindeutig politisches Projekt: Können wir andere Wege finden, um globalisiert zu sein, um planetarisch zu werden, oder stecken wir im neoliberalen Modell fest? Gibt es eine andere Art und Weise, wie wir unsere Verflechtungen neu überdenken können? Und schliesslich die ethische Frage: Was sind die Werte von Subjekten, die nicht einheitlich, sondern gespalten, komplex und nomadisch sind?»[34]

Die drei Narrative bilden alternative Diskussionspisten für KI, die die Aufmerksamkeit auf das Thema steuern: Wie kann über (mit) KI über ihren Einfluss auf uns und unsere sozialen Lebenswelten gedacht und kommuniziert werden? «Wenn der Liberalismus, der Nationalismus, der Islam oder irgendein neuartiger Glaube die Welt des Jahres 2050 prägen will, so wird er künstliche Intelligenz, Big-Data-Algorithmen und Bioengineering nicht nur erklären müssen – er sollte sie auch in ein neues, sinnvolles Narrativ integrieren können.»[35]

4. KI – nur eine Technik?

Menschsein besteht in der Fähigkeit, aus der gattungsspezifischen Not eine Tugend zu machen. In der biologischen Mangelhaftigkeit von *homo sapiens* steckt die Chance zu transformativer Plastizität. «Nackt, unbeschuht, unbedeckt, unbewaffnet», wie Platon im Dialog Protagoras feststellt,[36] ist der Mensch ein «Mängelwesen» (Johann Gottfried Herder), das um seines eigenen Überlebens willen lernen muss, «Hand und Wort» (André Leroi-Gourhan) lebensdienlich-kreativ einzusetzen. Der Vorteil des vernunftbegabten Tieres – «*zoon logon echon*» (Aristoteles) – gegenüber der übrigen Natur besteht darin, nicht auf die eigene Natur festgelegt zu sein. Das Tier «Mensch» ist nicht nur eine evolutionsbiologische Entwicklung, sondern kann sich auch im Blick auf seine eigenen Bedürfnisse und Interessen selbst weiterentwickeln. Mit der Übergabe des Feuers durch Prometheus im Mythos wird das menschliche Schicksal als «Homo faber» (Max Scheler), als schöpferischer und herstellender «Mensch» besiegelt, der Techniken zur «Organentlastung» (Hammer, Kaffeemaschine), «-verstärkung» (Bohrmaschine, Auto) und zum «Organersatz» (Röntgengerät, Flugzeug) (Arnold Gehlen) erfindet und einsetzt. Im Lauf ihrer Geschichte entpuppt sich *homo sapiens* dank seiner flexiblen Konstruktion als «homo compensator» (Odo Marquard), als Mängelausgleichsgenie, das immer erfolgreicher seine natürlichen Beschränkungen zu überlistet versteht.

Damit verbunden ist die Vorstellung, Schicksalhaftes (Gefahren) in technische Probleme (Risiken) transformieren, als solche thematisieren und bearbeiten zu können: «In Wirklichkeit

aber sterben die Menschen nicht, weil eine dunkel gewandete Gestalt sie an der Schulter packt oder weil Gott es so verfügt oder weil die Sterblichkeit wichtiger Teil irgendeines grossen kosmischen Plans ist. Menschen sterben immer wegen irgendeiner technischen Störung. Das Herz hört auf, Blut durch den Körper zu pumpen. Die Hauptschlagader ist durch Fettablagerungen verstopft. Krebszellen breiten sich in der Leber aus. Keime vermehren sich in der Lunge. Und was ist für all diese technischen Probleme verantwortlich? Andere technische Probleme. Das Herz hört auf zu schlagen, weil der Herzmuskel nicht mehr ausreichend mit Sauerstoff versorgt wird. Krebszellen wuchern, weil eine zufällige Genmutation ihren Code verändert hat. Keime lagerten sich in meiner Lunge ab, weil jemand in der U-Bahn nieste. An all dem ist nichts Metaphysisches. Alles nur technische Probleme. Und für jedes technische Problem gibt es eine technische Lösung.»[37]

Gegen die Zivilisationseuphorie, geschichtsphilosophischen und -politischen Fortschritts- und Perfektionierungsideen hatte bereits Friedrich Nietzsche auf die anthropologische Dialektik von Lösung und Problem hingewiesen: «[D]as was im Kampf mit den Thieren dem Menschen seinen Sieg errang, hat zugleich die schwierige und gefährliche krankhafte Entwicklung des Menschen mit sich gebracht. Er ist das noch nicht festgestellte Thier.»[38] Lebensweltlich wird die Chance menschlicher Nichtdeterminiertheit immer dann zum Problem, wenn Optionen am Horizont auftauchen, die mit den menschlichen Selbstverständnissen inkommensurabel erscheinen. Wie die Diskussionen über die militärische und zivile Nutzung der Kernspaltung, die Gentechnologie und aktuell über KI zeigen, steht dann die Frage im Raum, ob sich die Menschen mit ihrem nächsten technologischen Schritt selbst überschreiten und die eigene Gattung zur Disposition stellen würden. In der Technikgeschichte begegnen viele solche imaginären Grenzen, mit denen vor dem nächsten technologischen Schritt gewarnt wurde, und die regelmässig fielen, nachdem der Schritt doch gemacht worden war. Die Konsequenz bestätigt nicht zwangsläufig die Unangemessenheit und Irrtumsanfälligkeit der Grenzwarnungen. Mit guten Gründen kann sie auch als Beleg gelten für die pragmatische Anpassungsfähigkeit des Teils der Erdbevölkerung, der davon profitierte, und für die komplette Ohnmacht des anderen Teils der Erdbevölkerung, der dafür die Zeche bezahlen musste.

Grenzüberschreitungen setzen Grenzen voraus: (1.) Grenzen der Natur (Naturkausalität, Naturgesetze); (2.) moralische Grenzen und (3.) rechtliche Grenzen. Alles, was nicht durch sie aufgehalten wird, bildet eine (legale und legitime) Möglichkeit. Naturkausalität ist unhintergehbar, moralische und rechtliche Normen sind menschliche Setzungen und können deshalb aufgehoben oder revidiert werden. Sachlich geht die Grenzüberschreitung der Grenzziehung voraus. Ein Verhalten muss als anstössig und unerträglich wahrgenommen werden, damit es sanktioniert oder unter Strafe gestellt werden kann. Erst dann wird ein zuvor «nur» unakzeptables Verhalten zu einer verbotenen und pönalisierten Grenzverletzung. Normative Grenzziehungen sind ambivalent, weil sie schützen, indem sie verhindern. Das gilt für schlechte und inakzeptable Absichten und Handlungen ebenso wie für gute und erstrebenswerte. Mehr noch, aus anthropologischer, kulturphilosophischer und theologischer Perspektive sind Übergänge konstitutiv für *homo sapiens*. «Transitionen […] sind als Antidot gegen Stillstand und ausgrenzende Hoffnungen des Gestrigen, Nahen und Eigenen, positiv konnotiert: transkulturell, transnational, transökonomisch, aber auch Transzendenz, oder […] Transsexualität.»[39] Menschli-

ches Leben ist dynamisch auf Zukunft hin angelegt. Ein starkes Motiv dafür sind Kontingenzwahrnehmungen und -erfahrungen von Leiden, Vulnerabilität, Ungerechtigkeit und Fragmentarität. Sie erhalten ihr Gewicht und ihre Dringlichkeit vor dem Hintergrund entgegengesetzter Utopien, Hoffnungen und Erwartungen von Gesundheit, Ganzheit, Heil und Unsterblichkeit.[40] Aus der Perspektive des Richtigen und Guten dürfen die Erfahrungen des Falschen und Schlechten nicht als alternativloses Schicksal hingenommen werden. Hiobs Klage angesichts seines Elends wäre lediglich als bedauernswerte Tatsache registriert worden, hätte er es nicht als himmelschreiende Ungerechtigkeit vor dem Hintergrund seiner Erfahrungen von einem erfüllten Leben angeprangert.

Aufgrund ihrer vorgestellten und tatsächlichen Potentiale hat KI eine starke Affinität zum utopischen Denken. Wo die menschlichen Optionen ausgereizt erscheinen, sind die digitalen Möglichkeiten noch lange nicht am Ende. Wenn für zivilisatorische Fortschritte nicht mehr (allein) auf menschliche, sondern (auch) auf technologische Kreativität gesetzt werden kann, liegt der Gedanke nahe, dass die Fortschrittsgeschichte auf einem neuen Level fortgesetzt werden und noch einmal neu beginnen könne. An dieser Stelle treffen drei idealtypische Perspektiven aufeinander:

(1.) Die *Mittelperspektive*: KI wird als Perfektionierung technologischen Handelns begriffen. Ungeachtet ihrer Leistungsfähigkeit und Effizienz bleibt sie grundsätzlich dem instrumentellen, Werkzeug- oder Maschinenparadigma von Technik verhaftet. Obwohl sie Optionen bietet, die die menschlichen Möglichkeiten übersteigen, werden sie lediglich als Mittel für eine optimierte Umsetzung menschlicher Fähigkeiten verstanden – analog etwa zum Fliegen, das zwar der menschlichen Konstitution verschlossen ist, aber lediglich eine effizientere Form der menschlichen Fortbewegungsfähigkeit darstellt. So verstanden, erbringen Technologien zwar «bessere» Ergebnisse und auf andere Weise, aber sie «machen» nichts, was Menschen, wenn auch «schlechter» oder ineffizienter, nicht auch tun können. Umgekehrt formuliert, besteht der Zweck von Technik darin, möglichst präzise das zu leisten, wofür Menschen sie einsetzen.

(2.) Die *Autonomieperspektive*: Aus dieser Sicht wechselt KI die Seiten oder übernimmt die aus der Mittelperspektive ausschliesslich menschlichen Subjekten vorbehaltene Position. Sie setzt selbst die Zwecke, die sie anschliessend mit ihren Mitteln verfolgt, und wird zum Subjekt ihrer eigenen Technologien, die sie für «eigene» Zwecke gebraucht. Die Vorstellung von KI als autonomes technologisches Subjekt wird in zwei Varianten diskutiert: Entweder nimmt KI diese Rolle durch Simulation ein, indem sie so programmiert wird, sich «wie ein (menschliches) Subjekt» zu verhalten. Oder sie entwickelt eigenständig (*deep learning*) eine Subjektrolle und konstituiert sich damit selbst als autonomes Subjekt. Wichtig wird an dieser Stelle die Unterscheidung zwischen der technologischen Konstruktion der Subjekthaftigkeit (KI benötigt dafür kein Wissen über Subjektivität) und den Kategorien, die wir verwenden, um die technologischen Prozesse zu beschreiben und zu verstehen.

(3.) Die *Interaktionsperspektive*: Zwischen den Polen der instrumentellen und Subjektperspektive auf KI steht eine vermittelnde Sicht, die als geteilte Subjektivität von *homo sapiens* und KI vorgestellt werden kann. Umstritten ist die starke Prämisse, dass bei der Anwendung von KI tatsächlich «Subjekte» in einem hierarchischen Verhältnis oder gar auf Augenhöhe «interagie-

ren». Die spezifische Mensch-Technologie-Konstellation betrifft nicht nur – wie bei den bekannten Mensch-Maschine-Schnittstellen – Steuerungs- und Kontrollfunktionen, sondern «Mensch» und «Maschine» «handeln» Zuständigkeiten wechselseitig «aus». Die Sichtweise kommt der gegenwärtigen Realität am nächsten, in der vier verschiedene Autonomiegrade von KI und Robotics unterschieden werden: (1.) KI als technologische Unterstützung (entspricht der Mittelperspektive); (2.) aufgabenbezogene Autonomie von KI (*human-in-the-loop*): führt kontrolliert Aufträge aus; (3.) bedingte Autonomie von KI (*human-on-the-loop*): gibt Entscheidungsempfehlungen, aber entscheidet nicht, und (4.) starke Autonomie von KI (*human-out-of-the-loop*; entspricht der Autonomieperspektive): entscheidet selbst und operiert selbständig.

5. Ethische Perspektiven auf KI

Je nachdem, was unter KI verstanden wird, stellen sich unterschiedliche ethische Herausforderungen. Im Zentrum der ethischen Aufmerksamkeit steht der Schutz, die Sicherheit, Schadensverhütung und Risikominderung für die beteiligten und davon betroffenen Personen. Die Zielsetzungen unterscheiden sich nicht von den bekannten Standards von Technology Assessment und Technikfolgenabschätzung. 2019 hat eine von der Europäischen Kommission eingesetzte unabhängige hochrangige Expertengruppe für Künstliche Intelligenz *Ethics Guidelines for Trustworthy AI* veröffentlicht.[41] «Wir sind der Ansicht, dass es im Kontext des schnellen technologischen Wandels unabdingbar ist, dass Vertrauen auch in Zukunft das Element bleibt, das Gesellschaften, Gemeinschaften, Wirtschaftsräume und nachhaltige Entwicklung zusammenhält. Deshalb bestimmen wir *vertrauenswürdige KI als unsere grundlegende Ambition*, denn Menschen und Gemeinschaften können der Entwicklung und Anwendung von Technologien nur dann vertrauen, wenn ein klarer und umfassender Rahmen existiert, der Vertrauenswürdigkeit gewährleistet.»[42] Die Autoren des Leitfadens betonen drei Komponenten für eine vertrauenswürdige KI: «1. Sie sollte *rechtmässig* sein und somit geltendes Recht und alle gesetzlichen Bestimmungen einhalten; 2. sie sollte *ethisch* sein und somit die Einhaltung ethischer Grundsätze und Werte garantieren; und 3. sie sollte *robust* sein, und zwar sowohl in technischer als auch in sozialer Hinsicht, da KI-Systeme möglicherweise unbeabsichtigten Schaden verursachen, selbst wenn ihnen gute Absichten zugrunde liegen.»[43]

Die KI-Expertengruppe beschreibt die eigene normative Grundlage als einen «‹menschenzentrierten Ansatz›, wobei der Mensch eine einzigartige und unveräusserliche moralische Vorrangstellung in den Bereichen Zivilgesellschaft, Politik, Wirtschaft und Soziales einnimmt».[44] Daraus ergeben sich vier Leitprinzipien der KI-Ethik: «(i) Achtung der menschlichen Autonomie; (ii) Schadensverhütung; (iii) Fairness und (iv) Erklärbarkeit».[45] Die Prinzipien gründen in der Charta der Europäischen Union: «Die Achtung der menschlichen Autonomie ist mit dem Recht auf menschliche Würde und Freiheit (verankert in den Artikeln 1 und 6 der Charta) eng verbunden. Die Schadensverhütung ist mit dem Schutz der körperlichen oder geistigen Unversehrtheit (verankert in Artikel 3) eng verbunden. Fairness ist mit dem Recht auf Nichtdiskriminierung, Solidarität und Gerechtigkeit (verankert in Artikel 21 ff.) eng verbunden. Erklärbarkeit und Verantwortung sind mit den Rechten, die mit Gerechtigkeit in Zusammenhang stehen (verankert in Artikel 47) eng verbunden.»[46]

Die europäische Sicht stimmt mit den internationalen Standards zum Einsatz und Umgang mit KI weitgehend überein. Am häufigsten genannt werden *Transparenz* (*transperency*) als Forderungen nach Erklärbarkeit und Explizierbarkeit der Resultate; gefolgt von *Gerechtigkeit und Fairness* (*justice & fairness*) als Forderung nach *non-bias* vor dem Hintergrund der exponentiellen und nicht kontrollierbaren Verzerrungen durch Fehler im Datenmaterial, bei der Dateneingabe und Programmierung («*garbage in, garbage out*») sowie der Verstärkung diskriminierender gesellschaftlicher Wertvorstellungen; *Nichtschaden* (*non-maleficence*) als Forderungen nach Sicherheit und Schutz vor Manipulation (*nudging, fake news*); *Verantwortung* (*responsibility*) als Voraussetzung für Zurechnung und Haftung vor dem Hintergrund der systemischen Verantwortungsdiffusion, der Regelung privater Technologieanbieter:innen und der Zuständigkeit für Design und staatliche Regelungen; *Privatheit* (*privacy*) besonders die Datensouveränität, das heisst der Hoheit und Kontrolle über die eigenen personenbezogenen Daten, der Art der Datennutzung und Datenweitergabe (im Rahmen von *smart systems*, *smart governance* und *smart cities*), Probleme bei der Kontrolle privater Software, durch Hacking und die Re-Identifizierung anonymisierter Daten; *Wohltun* (*beneficence*) als Forderungen nach Realisierung und Schutz des persönlichen Wohlbefindens und der sozialen und gemeinschaftlichen Güter; *Freiheit und Selbstbestimmung* (*freedom & autonomy*) besonders im Blick auf die persönliche Zustimmungsfähigkeit (*informed consent*) und Wahlfreiheit und *Vertrauen* (*trust*) in die Zuverlässigkeit und Rechtmässigkeit der Datennutzung und -weitergabe.[47]

Die Struktur der KI-Richtlinien folgt wesentlich den bekannten vier bioethischen Prinzipien von Tom L. Beauchamp und James F. Childress.[48] Die Axiome waren ursprünglich (in der ersten Auflage von 1979) gegen Fehlverhalten in der biomedizinischen Forschung gerichtet und zielten auf den Schutz und die Selbstbestimmung von Versuchspersonen: Autonomie (*autonomy*), Nichtschaden (*non-maleficence*), Wohltun (*beneficence*) und Gerechtigkeit (*justice*). Die Analogie liegt nahe angesichts des exponentiell wachsenden Einsatzes von Hochleistungstechnologien in der medizinischen Diagnostik und Therapie und der zunehmenden medizinischen Bedeutung von KI und Robotics. Gleichzeitig stellt sich die Frage, ob das von Beauchamp und Childress vorausgesetzte Interaktionsverhältnis (Ärzt:innen-Patient:innen-Verhältnis) direkt oder indirekt auf die KI übertragen werden kann: Wer hat die Patient:innen- und Ärzt:innen-Rolle bei KI, wessen Autonomie soll geschützt werden, wer sind die Verantwortungssubjekte, wofür sollen und können sie rechenschaftspflichtig gemacht werden und wer sind schliesslich die Instanzen oder Subjekte, die für eine vertrauensvolle KI einstehen und diese garantieren sollen?

KI kommt *homo sapiens* einerseits näher als jede andere Technologie und ist für *homo sapiens* andererseits undurchschaubarer und unkontrollierbarer als jede andere Technologie. Das *Ähnlichkeitsproblem* besteht in der funktionalen Äquivalenz von künstlichen und menschlichen Fähigkeiten: *Wahrnehmen* – Bildverarbeitung; *Kommunizieren* – Spracherkennung und -generierung (*natural language processing*, *NLP*), Informationsgewinnung mittels semantischer Suchmaschinen; *Lernen* – maschinelles Lernen z. B. Spamfilter (*machine learning*), Mustererkennung (Korrelation statt Kausalität, *data mining*); *Wissen* – Modellierung von Wissen (*knowledge representation*); *Denken* – logische Programmierung, Schlussfolgern aus unsicherem Wissen (*probabilistic reasoning*), Verarbeitung kontinuierlicher Ereignisse (*event processing*);

Handeln – Planen von Handlungsschritten, intelligente Agenten-Technologien z. B. autonomes Fahren.[49] Die funktionale Äquivalenz der Fähigkeiten und Leistungen, die traditionell als besondere oder Alleinstellungsmerkmale der menschlichen Spezies gegenüber allen anderen Gattungen betont wurden und werden, wirft die Frage auf, ob aus diesen Ähnlichkeiten oder Gleichartigkeiten nicht auch eine Äquivalenz des Status und der Anerkennung folgen müsse. Das gilt in besonderer Weise für Technologien, denen «Autonomie» zugeschrieben wird.[50] Unterschieden wird zwischen einer *menschlich-personalen Autonomie*, für die der rechtliche Personenstatus und der moralische Status als Verantwortungssubjekt konstitutiv ist, und einer *technologischen Autonomie*, die als graduelle Unabhängigkeit von bzw. Verselbständigung gegenüber menschlicher Entscheidungs- und Kontrollmacht verstanden wird, ohne dass damit eine Verantwortungsübertragung an die Technologie verbunden wäre. Damit kann KI in die bestehenden Rechtssysteme, Ethik-, Risiko- und Sicherheitsdiskurse (Technology Assessment) integriert werden.

Wenn KI-Systeme aufgrund eigener Datenanalysen und auf der Grundlage selbstlernender Algorithmen Entscheidungen treffen, können diese zwar immer noch von zuständigen Fachleuten beurteilt werden, aber es stellt sich die Frage, worauf sich ihre stärker gewichtete Zustimmung oder Ablehnung stützen kann, wenn vorausgesetzt ist, dass die KI viel grössere Datenmengen, unter Einbeziehung von ungleich mehr Variablen, sehr viel präziser und zuverlässiger analysieren und evaluieren kann. Die Begründung, dass die Fachleute im rechtlichen und moralischen Sinn rechenschaftspflichtig seien, trägt nur so lange, wie gezeigt werden kann, dass die gegenüber der KI beschränktere Wissensbasis der Entscheidungssubjekte zu keinen signifikant ungünstigeren oder negativen Ergebnissen führt. Die Vergleichbarkeit der Ergebnisse hängt wiederum ab von der Nachvollziehbarkeit, Erklärbarkeit und Kontrolle der KI-Expertise.

An diese Herausforderung schliessen die breiten und kontroversen *opacity*- und *bias*-Diskussionen an. Selbstlernende Systeme sind opak, das heisst Blackboxes, bei denen selbst Expert:innen häufig nicht wissen und rekonstruieren können, wie die Resultate zustande kommen. Ein grundsätzliches Problem betrifft die Datensammlung, -verarbeitung und -auswertung. Entscheidungen aufgrund prädiktiver Analysen bilden ein breites Anwendungsfeld. Sie reichen von Restaurantempfehlungen nach individuellen Präferenzen, über medizinische Diagnosen, Entscheidungen über die Ausstellung einer Kreditkarte und die Zuteilung eines Transplantationsorgans bis hin zu Verhaltensprognosen im Rahmen von *predictive policing*.[51] KI kann Vorurteile beseitigen, indem sie Daten «neutral» sammelt und bearbeitet. Genauso kann sie Vorurteile verstärken, denn die Daten, die KI verwendet und mit/aus denen sie lernt, wurden von Menschen erhoben, aufbereitet, eingegeben und programmiert. Die damit verbundenen Verzerrungen – durch nicht oder unterrepräsentative, fehlerhafte, unvollständige, diskriminierende, rassistische oder sexistische Datenerhebungen – werden vom KI-System übernommen und bilden die Grundlage für die Datenverarbeitung und für das *deep learning*. Die «*historical*» und «*statistical bias*» der kulturell und gesellschaftlich geprägten Alltagsurteile und -entscheidungen werden technologisch nicht nur verstärkt, vielmehr können die digitalen Folgen der Verzerrungsquellen später kaum oder gar nicht identifiziert, kontrolliert und korrigiert werden. Grundsätzlich wird den eigenen menschlichen Fähigkeiten zur Selbstaufklärung, -reflexion und -korrektur mehr zugetraut als den digitalen Technologien.

Die Debatte darüber, ob es künstliche moralische «Aktanten»[52] (*artificial moral agents*) gibt, die über Rechte verfügen und Verantwortung übernehmen, stösst auf grundsätzliche Hindernisse und kommt erst zögerlich in Gang.[53] Der Biochemiker Isaac Asimov hatte bereits 1942 eine Erzählung veröffentlicht, in der er drei Regeln einer Roboterethik vorstellt und die Anwendungsprobleme an fiktiven Beispielen erläutert: «Erstens darf ein Roboter unter keinen Umständen einen Menschen verletzen – und, als logische Konsequenz, darf er nicht zulassen, dass ein inaktiver Mensch verletzt wird. […] Zweitens […] muss ein Roboter alle von qualifizierten Menschen erteilten Befehle befolgen, solange sie nicht im Widerspruch zu Regel 1 stehen. […] Drittens: Ein Roboter muss seine eigene Existenz schützen, solange dies nicht im Widerspruch zu den Regeln 1 und 2 steht.»[54] In der aktuellen Diskussion werden unterschiedliche Grade von ethisch relevanten künstlichen Aktanten unterschieden, die von KI-Aufgaben mit ethisch relevanten Wirkungen bis zu vollwertigen moralischen Agenten reichen, die explizite ethische Urteile fällen und in der Lage sind, diese vernünftig zu begründen.[55] Aus einer funktionalen Äquivalenzperspektive kann künstliche oder Maschinenethik als das Design von Technologien definiert werden, deren Aktivitäten, wenn sie von Menschen ausgeführt würden, ein Kriterium für den ethischen Status dieser Menschen darstellen.[56]

Interessant ist nicht nur, wie sich der Diskurs weiterentwickeln wird, sondern auch, was uns, aus welchen Intuitionen und Gründen an diesen Diskussionen über künstliche moralische Aktanten irritiert. Wie die begriffliche Unterscheidung zwischen «Akteur»/«Akteurin» und «Aktant» beispielhaft zeigt, werden wir mit Phänomenen konfrontiert, die ähnlich wahrgenommen und beschrieben werden und gleichzeitig als kategorisch unterschieden gelten (sollen). Anders als bei jeder anderen Technologie, fällt es viel schwerer, aus der Tatsache, dass KI bestimmte Aufgaben sehr viel präziser, effizienter und besser erledigt als wir, *nicht* darauf zu schliessen, dass KI uns in einem qualitativen Sinn überlegen sein könnte. Überlegenheit meint an dieser Stelle keine Übermacht, wie ein Erdbeben oder die Havarie eines Atomkraftwerks, der Menschen ohnmächtig ausgeliefert sind. Der menschliche Kontrollverlust bei Naturkatastrophen und Super-GAUs hat eine andere Qualität als die menschliche Nichtkontrollierbarkeit von «autonomer» KI. Die These, dass die menschliche Zivilisationsgeschichte das Schicksalhafte weitgehend durch technologisch kalkulierbare Risiken ersetzt habe, um es mit den aktuellen Technologieentwicklungen auf andere Weise zurückzuholen, behauptet eine neue Form von Kontrollverlust, Ohnmacht oder Unverfügbarkeit. Allerdings ist das nur die objektive Seite, der die interaktive «soziale» Seite von KI gegenübersteht und für deren Beschreibung die Begriffe und Kategorien noch weitgehend fehlen. Nicht zufällig begegnen in den Debatten um KI, Trans- und Posthumanismus auffällig häufig theologische Narrative und Analogiekonstruktionen, wobei tendenziell eine affirmative Aneignung von nichttheologischen Beiträgen einer kritischen Bezugnahme theologischer Positionen gegenübersteht.[57]

Insgesamt zeigt sich eine grosse Sensibilität für den digitalen *Wandel der Gesellschaft*, der eine hohe Aufmerksamkeit für den *Schutz der (menschlichen) Person* korrespondiert. Eine differenzierte und öffentliche Debatte über die naheliegende Frage nach dem digitalen *Wandel der (menschlichen) Person* selbst findet dagegen – wenn überhaupt – nur am Rand statt. Über die rechtlichen und ethischen Herausforderungen bei der Implementierung und Anwendung von «human-centered-AI» besteht weithin Einigkeit. Aber auch ohne sich auf das dünne Eis

eines Trans- und Posthumanismus hinauszuwagen, stellen sich aus ethischer Sicht weitreichende konzeptionelle Fragen. Drei wichtige Themen betreffen (1.) mögliche Verschiebungen unserer konstitutiven Diskurskategorien und (2.) die Diffusion der binären Konstruktion von «Person» und «Sache». Kaum Beachtung erhalten (3.) die anthropologisch-selbstreflexiven Effekte der KI-Diskussionen. Zwei Beispiele: (1.) Über die längste Zeit liessen sich zahllose Betrachter:innen von dem Bilderzyklus Ferdinand Hodlers beeindrucken, in dem der Maler das Sterben seiner an Krebs erkrankten Geliebten Valentine Godé-Darel dokumentiert. Heute weicht die unbefangene Schaulust zunehmend einer Scham, weil in einer digitalen Welt jede und jeder am eigenen Leib erfährt, was es heisst, zum ohnmächtigen Objekt der Betrachtung Dritter gemacht werden zu können. In beiden Fällen geht es um den Umgang mit Medien und die Kalibrierung des Verhältnisses zwischen Subjekt und Objekt. (2.) Der öffentliche Zugriff von Chatbots (ChatGPT, Google Bard, Open Assistant) wirft nicht nur weitreichende Urheberschafts- und Copyrightfragen auf, sondern greift unsere etablierten Begründungsmuster von eigener Leistung und Erfolg an. Denn die digitalen Applikationen ermöglichen die Generierung «eigener» Texte, ohne diese selbst zu verfassen. Üblicherweise wird diese Option als Angriff auf unsere Leistungslogik zurückgewiesen. Sie kann aber auch umgekehrt als kritische Rückfrage an den kapitalistischen Leistungsgedanken gedeutet werden. Dann wird die «eigene Leistung» als «Mythos vom ‹Selfmademan›» dekonstruiert, der unfähig ist, «unsere Talente als Gaben zu betrachten, für die wir Dank schulden, statt als Erfolge, die wir selbst zustande gebracht haben».[58] Unabhängig davon, was KI sonst noch bewirkt, eröffnet sie einen vertieften Blick und vielleicht neue Facetten auf uns selbst. Insofern hat die sehr locker an Karl Marx anschiessende These einiges für sich: «If we want to understand AI, we have to understand ourselves; and if we want to understand ourselves, we have to understand AI!»[59]

Bern, 31.10.2023

frank.mathwig@evref.ch

---

1 Queen, Bohemian Rhapsody: Queen, A Night at the Opera, 1975 EMI/UK.

2 Walter Benjamin, Zum Bilde Prousts. GS II/1, Frankfurt/M. 1980, 322 (Benjamin zitiert Jacques Rivière).

3 Immanuel Kant, Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis. Ed. Weischedel, Bd. VI, Darmstadt 1983, A 265.

4 Gerfried Stocker, Von künstlicher Intelligenz zur sozialen Intelligenz: Severin J. Lederhilger (Hg.), Gott und die digitale Revolution. Regensburg 2019, 73–96 (73f.).

5 Jens Kipper, Künstliche Intelligenz – Fluch oder Segen?, Berlin 2020, 3.

6 https://growthrocks.com/blog/big-five-tech-companies-acquisitions/ (21.10.2023).

7 John McCarthy/Marvin Minsky/Nathaniel Rochester/Claude E. Shannon, A proposal for the Dartmouth summer research project on artificial intelligence: https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html (29.10.2023): «The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.»

8 https://www.plattform-lernende-systeme.de/selbstverstaendnis.html (29.10.2023).

9 Vgl. Staatssekretariat für Bildung, Forschung und Innovation SBFI, Herausforderungen der künstlichen Intelligenz. Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, Bern 2019; Bundesrat, Leitlinien «Künstliche Intelligenz» für den Bund. Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung, Bern 2020.

10 Vgl. SATW, Recommendations for an AI Strategy in Switzerland. A white paper organised by the SATW topical platform on Artificial Intelligence, Zürich 2019; SATW, Diskussionspapier: Künstliche Intelligenz und die technische Souveränität der Schweiz, Zürich, Juli 2023.

[11] file:///Users/fm/Downloads/european_ai_landscape_-_annexes_-_switzerland_A89294DC-E989-D917-DF63B99723CFB79F_50823.pdf (28.10.2023).
[12] Vgl. https://www.nfp77.ch/en/nXTuIvoYFdvHkmI3/page/projects (28.10.2023).
[13] Vgl. Markus Christen et al., Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz, Zürich 2020.
[14] Vgl. https://www.midata.coop/ (30.10.2023).
[15] Vgl. Nikola Biller-Andorno et al., Künstliche Intelligenz in der Medizin – Zielvorstellungen für die verantwortliche Nutzung digitaler Zwillinge. Digital Society Initiative, Positionspapier: Künstliche Intelligenz in der Medizin, Juni 2023.
[16] Vgl. file:///Users/fm/Downloads/dsi-strategy-lab-umfrage-infografiken-onepager-20230928-1.pdf (30.10.2023).
[17] Vgl. https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20223890 (30.10.2023).
[18] Vgl. David J. Chalmers, Realität+. Virtuelle Welten und die Probleme der Philosophie, Berlin 2023, 29–48.
[19] Vgl. Sigmund Freud, Eine Schwierigkeit der Psychoanalyse: Imago V.1/1917, 1–7.
[20] Freud, Schwierigkeit, 7.
[21] Vgl. Oliver R. Scholz, Die Idee einer vierten Kränkung der Menschheit: Sind uns Computer geistig überlegen?: Marius Backmann/Jan G. Michel (Hg.), Physikalismus, Willensfreiheit, Künstliche Intelligenz, Paderborn 2009, 199–208.
[22] Günther Anders, Die Antiquiertheit des Menschen, Bd. 1: Über die Seele des Menschen im Zeitalter der zweiten industriellen Revolution, München [6]1983, 33.
[23] Anders, Antiquiertheit, 33.
[24] Yuval Noah Harari, Homo Deus. Eine Geschichte von Morgen, München [9]2017, 72.
[25] Vgl. einführend Janina Loh, Trans- und Posthumanismus zur Einführung, Hamburg 2018; Stefan Lorenz Sorgner, Transhumanismus. «Die gefährlichste Idee der Welt», Freiburg/Br. 2016; Benedikt Paul Göcke/Frank Meier-Hamidi (Hg.), Designobjekt Mensch. Die Agenda des Transhumanismus auf dem Prüfstand, Freiburg/Br. 2018; Donna Haraway, Ein Manifest für Cyborgs. Feminismus im Streit mit den Technowissenschaften: dies., Die Neuerfindung der Natur. Primaten, Cyborgs und Frauen, Frankfurt/M., New York 1985, 33–72; Rosi Braidotti, Posthumanismus. Leben jenseits des Menschen, Frankfurt/M., New York 2014; dies./Maria Hlavajova (Hg.), Posthuman Glossary, London, New York 2018; Max More/Natasha Vita-More (Hg.), The Transhumanist Reader. Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future, Malden MA, London 2013.
[26] Braidotti, Posthumanismus, 29.
[27] Peter Sloterdijk, Regeln für den Menschenpark. Ein Antwortschreiben zu Heideggers Brief über den Humanismus, Frankfurt/M. [12]2014, 31f.
[28] Zur theologischen Diskussion vgl. einführend Ron Cole-Turner, Von der Theologie zum Transhumanismus und zurück; Johannes Gössl, Verbesserung oder Zerstörung der menschlichen Natur? Eine theologische Evaluation des Transhumanismus; Jennifer Jeanine Thweatt, Cyborg-Christus: Transhumanismus und die Heiligkeit des Körpers: alle in Benedkt Paul Göcke/Frank Meier-Hamidi (Hg.), Designobjekt Mensch. Die Agenda des Transhumanismus auf dem Prüfstand, Freiburg/Br. 2018; Christopher Coenen, Verbesserung des Menschen durch konvergierende Technologien? Christliche und posthumanistische Stimmen in einer aktuellen Technikdebatte: Hartmut Böhm/Konrad Ott (Hg.), Bioethik – Menschliche Identität in Grenzbereichen, Leipzig 2009, 41–123; Mathias Wirth, Doketisch, pelagianisch, sarkisch? Transhumanismus und technologische Modifikationen des Körpers in einer theologischen Perspektive: NZSth 60/2018, 142–167; Thorsten Moos, Reduced Heritage. How Transhumanism Secularizes and Desecularizes Religious Visions: J. Benjamin Hurlbut/ Hava Tirosh-Samuelson (Hg.), Perfecting Human Futures. Transhuman Visions and Technological Imaginations, Wiesbaden 2016, 159–178; Frederike van Oorschot, Theologische Positionen zu Transhumanismus und KI – ein Überblick: ZPT 75/2023, 139—151; dies./Selina Fucker (Hg.), Framing KI. Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz, Heidelberg 2022; Oliver Dürr, Homo Novus. Vollendlichkeit im Zeitalter des Transhumanismus. Beiträge zu einer Technikтheologie, Münster, 2019; Lukas Ohly, Ethik der Robotik und der Künstlichen Intelligenz. Berlin u.a. 2019; ders./Catharina Wellhöfer, Ethik im Cyberspace, Frankfurt/M. 2017; Calvin Mercer, Bodies and Persons: Theological Reflections on Transhumanism: Dialog: A Journal of Theology, Vol. 54/2015 (Spring), 27--33; ders./Tracy J. Trothen, Religion and the Technological Future. An Introduction to Biohacking, Artificial Intelligence, and Transhumanism, Cham 2021.
[29] Vgl. Alexandre Dumas, Die drei Musketiere, Frankfurt/M. 2011.
[30] Vgl. Reinhard Brandt, D'Artagnan und die Urteilstafel. Über das Ordnungsprinzip der europäischen Kulturgeschichte 1, 2, 3/4, München 1998, 77–80.
[31] Zit. n. Brandt, D'Artagnan, 79.
[32] Brandt, D'Artagnan, 77.
[33] Vgl. aus christlicher Sicht etwa das Konzept der Liminalität bei Christian Strecker, Die liminale Theologie des Paulus. Zugänge zur paulinischen Theologie aus kulturanthropologischer Perspektive, Göttingen 1999.
[34] On Nomadism. Interview by Sara Saleri with Rosi Braidotti: European Alternatives. https://georgemaciunas.com/exhibitions/fluxus-as-architecture-2/fluxhousefluxcity-prefabricatedmodular-building-system/fluxhouse-fluxcities/essays-2/european-alternatives-on-nomadism-interview-with-rosi-braidotti/ (24.10.2023); vgl. Rosi Braidotti, Nomadic Theory: The Portable Rosi Braidotti, New York Chichester, West Sussex 2011.
[35] Yuval Noah Harari, 21 Lektionen für das 21. Jahrhundert, München 2018, I: Die technologische Herausforderung, 1: Desillusionierung, Der liberale Phönix.
[36] Platon, Protagoras 321 c.

[37] Harari, Homo Deus, 37.
[38] Friedrich Nietzsche, Nachgelassene Fragmente Frühjahr 1884, 25: ders., KStA, Bd. 11, München ^{3}2009, 125.
[39] Wirth, Transhumanismus, 143.
[40] Vgl. dazu aus theologischer Sicht Matthias Felder, Christliches Leben und die Verbesserung des Menschen. Enhancement und Heiligung bei Calvin, Berlin, Boston 2022.
[41] Vgl. High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI. European Commission, Brussels, 8 April 2019; Übersetzung: https://demographie-netzwerk.de/site/assets/files/4421/ethik-leitlinien_fur_eine_vertrauenswurdige_ki_1.pdf (30.10.2023).
[42] Expert Group, Guidelines, Ziff. 11.
[43] Expert Group, Guidelines, Ziff. 15.
[44] Expert Group, Guidelines, Ziff. 38.
[45] Expert Group, Guidelines, Ziff. 48.
[46] Expert Group, Guidelines, Fussnote 25.
[47] Vgl. Anna Jobin/Marcello Ienca/Effy Vayena, The global landscape of AI ethics guidelines: Nature Machine Intelligence, Vol 1, September 2019, 389–399: https://doi.org/10.1038/s42256-019-0088-2; Kathleen Murphy et al., Artificial intelligence for good health. A scoping review of the ethics literature BMC Med Ethics (2021): https://doi.org/10.1186/12910-021-00577-8.
[48] Vgl. Tom L. Beauchamp/James F. Childress, Principles of Biomedical Ethics, 8th edition, Oxford 2019.
[49] Vgl. Bernhard G. Humm/Peter Buxmann/Jan C. Schmidt, Grundlagen und Anwendungen von AI: Carl Friedrich Gethmann et al., Künstliche Intelligenz in der Forschung. Neue Möglichkeiten und Herausforderungen für die Wissenschaft, Berlin 2022, 13–42 (16f.).
[50] Vgl. Vincent C. Müller, Ethics of artificial intelligence and robotics: Edward N. Zalta (Ed.), Stanford Encyclopedia of Philosophy, Palo Alto: https://plato.stanford.edu/entries/ethics-ai/ (cp. 2.5).
[51] Vgl. Müller, Ethics, cp. 2.2.2.
[52] Der Begriff ist eine Wortschöpfung von Bruno Latour, On actor-network theory. A few clarifications: Soziale Welt 47/1996, 369–381; ders., Das Parlament der Dinge. Für eine politische Ökologie, Frankfurt/M. 2001.
[53] Vgl. Vincent C. Müller, Is it time for robot rights? Moral status in artificial entities: Ethics and Information Technology (17. May 2021): https://doi.org/10.1007/s10676-021-09596-w; James H. Moor, The Nature, Importance, and Difficulty of Machine Ethics: IEEE Intelligent Systems, 21/2006 (4), 18–21; Jens Kersten, Menschen und Maschinen. Rechtliche Konturen instrumenteller, symbiotischer und autonomer Konstellationen: JZ 70/2015, 1–8; Nadja Braun Binder et al., Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht: Jusletter 28. Juni 2021; Simone Kuhlmann et al. (Hg.), Transparency or Opacity. A Legal Analysis of the Organization of Information in the Digital World, Baden-Baden 2023; Dan Verständig et al. (Hg.), Algorithmen und Autonomie. Interdisziplinäre Perspektiven auf das Verhältnis von Selbstbestimmung und Datenpraktiken, Opladen, Berlin, Toronto 2022; Luciano Floridi (Hg.), Ethics, Governance, and Policies in Artificial Intelligence, Cham 2021.
[54] Isaac Asimov, Runaround: Astounding. Science Fiction Vol. XXIX, No. 1 (March 1942), 194–103 (100) [eigene Übersetzung].
[55] Vgl. Moor, Nature, 20.
[56] Vgl. Müller, Ethics, cp. 3.2.
[57] Vgl. Peter Dabrock, «Prüft aber alles und das Gute behaltet!»: Theologisches und Ethisches zu Künstlicher Intelligenz: ThLZ 147/2022, 635–650, van Oorschot, Positionen; Dürr, Homo; Ohly/Wellhöfer, Ethik.
[58] Michael J. Sandel, Plädoyer gegen Perfektion. Ethik im Zeitalter der genetischen Technik, Berlin 2008, 108.
[59] Vincent C. Müller, Philosophy of AI: A Structured Overview. Final Draf, 24th Juli, 2023, forthcoming in Nathalie Smuha (Ed.), Cambridge Handbook on the Law. Ethics and Policy of Artificial Intelligence, Cambridge 2024, (4).