

# Philosophisch-anthropologische Überlegungen angesichts „Deep Learning“: Intransparente und eigenständige Systeme als Herausforderung für das Selbstverständnis des Menschen?

## 1. Einleitung

Informationstechnische Systeme, die zu „Deep Learning“ – einer Form des maschinellen Lernens, die in „künstlichen neuronalen Netzwerken“ eingesetzt wird – fähig sind, können den Eindruck erwecken, „von selbst“ zu „denken“. Sie erscheinen in dieser Hinsicht dem Menschen als ebenbürtig, provozieren und konkurrieren bzw. sind beispielsweise beim Erfassen von Bildinhalten oder dem Erkennen von Krankheitsmerkmalen teilweise dem Menschen sogar überlegen. Damit können sie das Bild, das Menschen von sich selbst machen, infrage stellen.

Die folgenden Überlegungen werden demgemäß der Frage nachgehen, was es für das Selbstverständnis des Menschen heißen könnte, wenn der Mensch sich seine Lebenswelt mithilfe von eigenständigen und intransparenten informationstechnischen Systemen einrichtet. Genauer: Die Absicht besteht darin zu fragen, wie solche Systeme den Menschen dazu herausfordern können, über sich selbst nachzudenken. Dadurch wird im allgemeineren Sinne auch nach den Verhältnissen von Menschen und Techniken gefragt.

Ausgehend von Ernst Cassirer, der Technik als Realisierung menschlicher Lebensformen versteht, soll diesen Fragen insbesondere mithilfe von Ulrich Sonnemanns Denkbewegungen zur menschlichen Freiheit und Technik nachgegangen werden. Im Zentrum steht das Thema Freiheit, weil die menschliche Fähigkeit denken zu können für Sonnemann geradezu Ausdruck von Freiheit ist und sich das Infragestellen des Menschen durch „selbstlernende“ Systeme insbesondere hinsichtlich der Frage nach Freiheit auf tun kann. Insofern werden Sonnemanns Gedankengänge aus den 1960er und 1970er Jahren auch danach erkundet, inwiefern sie Denkipulse geben können, wenn es um ein gegenwärtiges Verständnis von nicht transparenten eigenständigen Systemen und um menschliches Selbstverständnis geht.

Damit die theoretischen Überlegungen in die gegenwärtigen technischen Entwicklungen eingeordnet werden können, soll als erster Schritt geklärt werden, wie Deep Learning funktioniert und welche Eigenschaften es aufweist. Um Deep Learning in einen breiteren Kontext einzuordnen, wird danach anhand Cassirers Technikverständnis aufgezeigt, dass sich der Mensch mittels Technik seine Welt formt und Technik selbst als Ausdruck von Kultur verstanden werden kann, jedoch auch Möglichkeit für den Menschen überhaupt ist, zu sich selbst in ein Verhältnis zu treten. Anschließend wird ausführlich Sonnemanns Verständnis von menschlicher Freiheit erarbeitet. Dies geschieht nicht nur, weil es Voraussetzung dafür ist, Sonnemanns Technikverständnis und seine Überlegungen zum Verhältnis von Technik und menschlichem Selbstverständnis zu verstehen, sondern dies ist auch hinsichtlich der Frage von Relevanz, ob sich Anknüpfungspunkte von Sonnemanns Gedanken für ein Verständnis von gegenwärtigen Systemen, die Intransparenz und Eigenständigkeit aufweisen, finden lassen.

## 2. Zur Funktionsweise von Deep Learning

Zentral für Deep Learning<sup>1</sup> ist die Produktion von Wissen, was im Folgenden anhand des fiktiven Computerprogramms E2VA erklärt werden soll:

E2VA hilft, aus einer grossen Menge von Datensätzen („inputs“) mit zunehmender Präzision Muster zu erkennen und auszuwerten, um beispielsweise zu bestimmen, ob auf einzelnen Bildern eine spezifische Schlange abgebildet ist, die als „Königskobra“ klassifiziert („output“) werden kann. E2VA erkennt auch Bäume, die neben der Kobra abgebildet sind, klassifiziert diese als solche und kann im digitalen Bild lokalisieren, wo sich die Kobra im Verhältnis zum Baum befindet (Ortsbestimmung innerhalb eines Bildes). Das Programm kann vollständige Bilder in verschiedene Bereiche wie „Königskobra“, „Mensch“, „Frucht“ oder „Baum“ segmentieren. Umso mehr Rohdaten, also Bildpunkte in Bildsätzen, zur Verfügung stehen, desto präziser werden die Klassifizierungen.

Dazu verwendet E2VA eine Form von maschinellern Lernen, das als „Deep Learning“ bezeichnet und in „künstlichen neuronalen Netzwerken“ (ANN) eingesetzt wird. Solche Netzwerke sind jedoch keine Nachbildung

1 Deep Learning, als eine Form des maschinellen Lernens, die insbes. bei informationstechnischen Systemen Anwendung findet, wird in diesem Aufsatz unter den Begriff der „Technik“ subsumiert. Dies gilt auch für Ausdrücke wie „Maschine“, „Apparat“ oder „System“.

eines biologischen Nervensystems, obschon der Ausdruck dies suggeriert; vielmehr stellen sie eine Abstraktion von Funktionsweisen eines biologischen Nervensystems dar. Im Unterschied zu einer Synapse des menschlichen Gehirns können künstliche „Neuronen“ z. B. Signale in mehr als eine Richtung übermitteln („backpropagation“); schon von daher haben sie nicht die biophysischen Eigenschaften von Nervenzellen (Boden 2018, 80).

Bei Deep Learning Systemen nehmen einzelne künstliche Neuronen („units“) Informationen („inputs“) auf, verarbeiten diese mithilfe einer bestimmten Funktion und geben ihren „output“ an mindestens ein nächstes künstliches Neuron weiter. Dabei sind die am künstlichen Netzwerk beteiligten Neuronen („nodes“) in hierarchischen Schichten („layers“) organisiert und miteinander verbunden: Neuronen der Eingangsschicht („input layer“) nehmen die Rohdaten auf und leiten sie an Neuronen der mittleren Schichten („hidden layers“) weiter, die dem Betrachter nicht direkt zugänglich sind. Diese übermitteln wiederum ihre Ergebnisse an die Ausgangsschicht („output layer“). Durch letztere zeigt sich diejenige Information, die wir als Begriffe wie „Königskobra“ oder „Frucht“ verstehen.

Doch zuerst erkennt ein System wie E2VA nicht, ob es sich bei den Rohdaten um ein Abbild der Schlange oder des Apfels handelt: „The function mapping from a set of pixels to an object identity is very complicated.“ (Goodfellow, Bengio, und Courville 2016, 6) Um eine möglichst genaue Aussage treffen zu können, muss E2VA anhand von Daten erlernen, relevante Merkmale zu extrahieren. Dazu benötigt es Übung. Umso mehr Daten, desto effizienter wird es darin, aufgabenspezifisch Informationen zu klassifizieren und Vorhersagen zu treffen.

Wird E2VA anhand von benannten („labeled“) und vorgegebenen Beispiel-Daten „trainiert“, was es aus einer Datenmenge zu extrahieren hat, spricht man von „supervised learning“. Lernt E2VA jedoch „unsupervised“, gibt man den Daten keine Kennzeichnungen („labels“): Über mehrere mittlere Schichten extrahieren die Algorithmen dann verschiedene Merkmale aus den Rohdaten eigenständig. Umso mehr Schichten, desto komplexere Merkmale kann E2VA herausarbeiten: Die erste Schicht kann beispielsweise aufgrund unterschiedlicher Helligkeit bei benachbarten Bildpunkten Ränder erkennen. Mit diesen Informationen können weitere Schichten nach Ecken und erweiterten Konturen suchen, die als Ansammlung von Rändern gelten (Goodfellow, Bengio, und Courville 2016, 6). Auf dieser Basis kann eine weitere Zwischenschicht vollständige Teilobjekte („Augen“, „Zunge“, usw.) erfassen, indem sie charakteristische Ansammlungen von Ecken und Konturen definiert (Goodfellow, Bengio, und Courville 2016, 6). Schlussendlich führen die erfassten Teilobjekte im Bild dazu, dass auf

ihrer Grundlage die Objekte im Bild erkannt werden (Goodfellow, Bengio, und Courville 2016, 6). Das heisst: E2VA bildet auf Basis der erhaltenen Bildpunkte Muster und sortiert ähnliche Merkmale („clustering“), die nicht schon von vornherein einprogrammiert sind. Umso häufiger die vom System neu definierten Muster auftreten, desto höher bewertet das System die Wahrscheinlichkeit, dass seine eigenen Musterbildungen korrekt sind. Vorgegebene Kennzeichnungen sind unnötig. E2VA lernt mittels „Erfahrung“: „Learning is driven by the principle that co-occurring features engender expectations that they will co-occur in future. Unsupervised learning can be used to discover knowledge. The programmers needn't know what patterns/clusters exist in the data: the system finds them for itself.“ (Boden 2018, 41–42) Das Zitat verweist auf zwei wichtige Punkte. Erstens: Systeme mit Deep Learning sind fähig, „Wissen“ zu erwerben, indem sie „von selbst“ Muster erkennen, Klassifizierungen vornehmen und Lernvorgänge vollziehen. Daher ist diese Form maschinellen Lernens auch für die Künstlichen Intelligenz vielversprechend, denn „Artificial Intelligence (AI) seeks to make computers do the sorts of things that minds can do.“ (Boden 2018, 1) Zweitens aber verweist das Zitat auf die Problematik, wie man nachvollziehen könnte, wie ein solches System konkret Resultate liefert.

### 3. „Black Box“ als Ausdruck nicht-transparenter Systeme

Damit Menschen sich im Alltag zurechtfinden, benötigen sie eine immense Menge an Wissen und Erfahrung in der Welt, wobei sich vieles davon als intuitives und subjektives Wissen zeigt und es demnach schwierig ist, in einer formalen Sprache und mit logischen Schlüssen zum Ausdruck zu bringen: Eine erfahrene Ärztin kann zwar alle möglichen Dinge aussagen, welche Faktoren sie während einer Entscheidungsfindung für gewichtig hält. Doch abschliessend eine präzise und abstrakte Erklärung zu liefern, *warum*<sup>2</sup> sie entschieden hat, einen Patienten doch noch ins Spital einzuliefern, um eine bestimmte Untersuchung durchzuführen, wird ihr schwer-

2 Ein „Warum“ verweist bereits auf die Frage nach möglichen Rechtfertigungsgründen bei Handlungen.

fallen.<sup>3</sup> In jeder derartigen Entscheidung geht ein Moment von Spontaneität ein.

Auffälligerweise verhält es sich bei Deep Learning Systemen ähnlich, denn sie können zwar in Form von Informationsverarbeitungen bestimmte Muster erkennen und bilden, um damit spezifisches Wissen oder Entscheidungen zu generieren, doch warum ein solches System funktioniert und spezifische „outputs“ generiert, ist nicht gänzlich erklärbar: „[E]ven though we make these networks, we are no closer to understanding them“. (Castelvecchi 2016, 22)

Im Zusammenhang mit einer solchen Intransparenz ist vermehrt die Rede von einer „Black-Box“. Diese von der Kybernetik geprägte Metapher bezeichnet allgemein Systeme, bei denen nur Ein- und Ausgabesignale beobachtbar sind. Allein aus ihnen müssen Aussagen über das Verhalten eines Systems abgeleitet werden, da das System selbst über seine inneren Vorgänge keinen direkten Aufschluss gibt: „The engineer is given a sealed box that has terminals for input [...] and terminals for output, from which he may observe what he can. He is to deduce what he can of its contents.“ (Ashby 1956, 86)

Nun ist es bei Deep Learning Systemen nicht so, dass Ingenieur\*innen über keinerlei Wissen darüber verfügen, was sich im Innern der Systeme abspielt: Sie können durchaus Auskunft geben, wie in einem artifiziellen neuronalen Netzwerk einzelne Neuronen isoliert betrachtet programmiert sind, wie sie sich verhalten und wieviel Gewicht sie einem Input-Signal geben. Betrachtet man diese einzelnen Knotenpunkte im System, so greift die „Black-Box“-Metapher nicht. Trotzdem fehlen abschliessende Erklärungen, warum Deep Learning überhaupt funktioniert (Lin, Tegmark, und Rolnick 2017; Lipton 2018; Zhang et al. 2021). Dies bedarf einer Präzision: Damit ist nicht gemeint, dass nur der Endnutzer bzw. die Endnutzerin die hintergründigen Prozesse, wie etwa die beim Betriebs-

3 Vgl. hierzu Sybille Krämer, für die die formale Beschreibung einer Handlung u. a. den schematischen Symbolgebrauch bedingt: „Eine Handlung in der Perspektive des Schemas, welches sie realisiert, zu beschreiben heißt: diese Handlung konstituiert keinen Eigen-Sinn; sie gilt nicht als ein Ereignis, das den Charakter einer Geschichte hat, sondern entlehnt ihre Bedeutung der Einhaltung eines Schemas. Sie wird dadurch zu einem Verfahren.“ (Krämer 1988, 2) Gerade das Beispiel der Ärztin zeigt, dass ihre Entscheidung eben den Charakter eines Ereignisses aufweist und sie auf ihre gelernte Erfahrung zurückgreift, was in einer rein formalen Sprache schwierig wiederzugeben wäre. Dies gilt auch dann noch, wenn sich die Art des ärztlichen Urteilens formalisieren liesse, indem die Ärztin in ihrem Arztbericht hinterher vernünftige Gründe für ihre Handlungen angibt (vgl. abduktives Schlussfolgern bei Peirce).

system eines Computers, nicht kennt, dagegen Softwareentwickler\*innen oder Ingenieur\*innen schon. Sondern damit ist gemeint, dass auch den Konstrukteur\*innen komplexer Deep Learning Systeme ein abschliessendes Wissen darüber fehlt, warum sie funktionieren. Die Komplexität von Rekursivität kann hierbei eine Rolle spielen: „Deep learning models, for instance, tend to be black boxes [...] because they are highly recursive.“ (Rudin 2019, 206) Umso mehr Knotenpunkte aus verschiedenen Schichten miteinander verbunden und rückgekoppelt sind, desto schwieriger sind sie nachvollziehbar. Auch ist die jeweilige Ergebnisfindung abhängig von der spezifischen Programmierung einzelner künstlichen Neuronen, der Netzwerkstruktur und von den Datensätzen zu Trainings- und Validierungszwecken. Allein die Komplexität solcher Systeme lässt es fast nicht zu, Resultate auf Basis von Eigenschaften einzelner Komponenten nachzuvollziehen. Bereits William Ross Ashby war das Problem bekannt: “It does in fact very commonly happen that when the system becomes large, so that the range of size from part to whole is very large, the properties of the whole are very different from those of the parts.“ (Ashby 1956, 111–112)

Deep Learning Systeme können jedoch nicht nur komplex sein, sondern ihre Ergebnisfindung basiert – im Gegensatz zu vorprogrammierten Wenn-Dann-Regeln – auf dem Prinzip des Lernens aus „Erfahrung“. Dies beinhaltet etwa Änderungen in der Gewichtung bestimmter Signale oder Verbindungen innerhalb eines Netzwerkes, aber auch Änderungen des Programmablaufs durch „genetische Algorithmen“ (Optimierungsalgorithmen), welche alte Verknüpfungen zwischen Knotenpunkten aufheben oder neue hinzufügen: „This involves making adaptive changes in the weights, and sometimes also in the connections. Usually, the networks anatomy—the number of units, and the links between them—is fixed. If so, learning alters only the weights. But sometimes, learning [...] can add new links and prune old ones.“ (Boden 2018, 74–75)

Ob ein Deep Learning System Ergebnisse liefert, ist nicht zuletzt auch abhängig davon, wie gut es „lernt“ respektive welche „Erfahrungen“ es mit den Daten macht; wieviel Gewicht es bestimmten Signalen gibt, wie es mit Verknüpfungen und Parametern umgeht und eigenständig Anpassungen vornimmt.

#### 4. Eigenständigkeit: Zur nicht-festgelegten Wirkungsweise von Deep Learning Systemen

Weil bei solchen Systemen „[e]rst über das resultierende Verhalten [...] rückwärts auch der einzelne Prozessor [Knotenpunkt] seine spezifische Qualität [erhält]“ (Wiedebach 2014, 119), sind spezifische Ergebnisfindungen für den Interpreten *prinzipiell immer* erst in Nachhinein zu erfassen. Hartwig Wiedebach weist darauf hin, dass bei solch vernetzten Systemen „Differenzen, Übergänge und Schwankungen auf[treten], bei denen es unwahrscheinlich ist, daß ein vorausbestimmter Deutungsmodus ihrer Vielfalt gerecht wird. Geringfügige Intensitätsvarianten an bestimmten Orten im Verteilungsmuster der Netzaktivität, also Nuancen, sind vielleicht von eigenem Interesse und erhalten auf zuvor unbekannte Weise Gewicht. Hier meldet sich eine Eigenständigkeit des Systems an.“ (Wiedebach 2014, 117)

Auch wenn jede einzelne noch so spezifische Gewichtung oder Anpassung bei den einzelnen Knotenpunkten des Systems *im Nachhinein* erschliessbar wäre, würde dies noch nicht erklären, *warum* das System als Ganzes spezifische Anpassungen vorgenommen oder Verknüpfungen gelöscht hat. Das System liefert keine Gründe, warum es so funktioniert, wie es funktioniert.<sup>4</sup>

Das heisst: Das Verhalten von Systemen mit Deep Learning kann nicht nur aufgrund technischer Komplexität nicht gänzlich im Voraus prognostiziert werden, sondern scheint prinzipiell erst im Nachhinein möglich zu sein. Daher muss solchen Systemen eine gewisse Eigenständigkeit zugestanden werden.

In diesem Zusammenhang liest man auch von „autonomen“ Systemen. Die vielfältige Verwendungsweise des Attributs „autonom“<sup>5</sup> in Diskursen über technische Möglichkeiten scheint hier insofern verwirrend zu sein,

4 Werden Deep Learning Systeme eingesetzt, die zwar zu besseren Resultaten als herkömmliche Systeme führen, deren Vorgänge jedoch nicht nachvollziehbar sind, kann dies ethische Probleme mit sich bringen. Es wird daher versucht, solche Modelle erklärbar zu machen („Explainable Artificial Intelligence“). Gegenüber bspw. Post-Hoc-Erklärungsmodellen siehe jedoch Cynthia Rudin, die solche Erklärungsmodelle selbst für problematisch hält, da sie irreführend und unzuverlässig sein können und grundsätzlich die Berechnungen des Originalmodells nicht perfekt wiedergeben, denn: „If the explanation was completely faithful to what the original model computes, the explanation would equal the original model, and one would not need the original model in the first place, only the explanation. (In other words, this is a case where the original model would be interpretable).“ (Rudin 2019, 207)

5 Herkommend von *αὐτονομία*; von *αὐτο-* (selbst) und *νόμος* (Gesetz).

weil etymologisch der Begriff auf politische Selbständigkeit und Freiheit zurückgeht und seit Kant damit die Idee einer möglichen Selbstgesetzgebung beim Menschen verknüpft wird — und weil dies wiederum, wie wir unten sehen werden, mit einer Idee von Freiheit einhergeht (ausführlich dazu: Pohlmann 1971). Ein Verständnis möglicher Selbstgesetzgebung beim Menschen schliesst ein, dass er als vernunftfähiges Wesen auch frei darüber nachdenken kann, warum oder woher er bestimmte Verhaltensweisen aufweist, weshalb er etwas moralisch als gut oder schlecht erachtet, welche Bedürfnisse er zum Ausdruck bringen will oder wieso und ob er mit dem eigenen Handeln zufrieden ist — und wenn nicht, wie er stattdessen sein soll.

Über sich selbst nachdenken und zu Einsichten gelangen zu können zeigt, dass der Mensch potenziell in der Lage ist, der Vernunft auch im Praktischen zu folgen. Damit der Mensch dies tun kann, muss er sich als frei begreifen, damit das, was er will, überhaupt möglich ist und er das, was ihm falsch und unvernünftig erscheint, nicht in Praxis umsetzt. Auch eine vernünftige Entscheidung, sich an festgelegte Regeln zu halten, setzt eine Idee von menschlicher Freiheit voraus. Das ist vereinfacht ausgedrückt, jedoch genügt es hervorzuheben, dass aus dieser Perspektive betrachtet Autonomie mit dem oben genannten moralphilosophischen Freiheitsverständnis einhergeht.

Wollte man in diesem Sinne informationsverarbeitenden Systemen grundsätzlich Autonomie zuschreiben, müsste man annehmen, dass Freiheit als Eigenschaft des Systems möglich wäre. Dies wiederum würde heissen, dass man die dazu notwendigen Voraussetzungen gezielt implementieren könnte, oder man müsste annehmen, dass sie im System so angelegt würden, dass Freiheit „von selbst“ entsteht. Die Idee, dass ein Deep Learning System mit möglicher Selbstgesetzgebung – was dem Begriff nach die Vorstellung eines „Selbst“ beinhaltet – und Freiheit möglich wäre, geht von Bedingungen aus, die bisher allein menschlichen Lebewesen vorbehalten waren.

In der Literatur über technische Möglichkeiten wird in einigen Fällen jedoch implizit oder explizit zwischen menschlicher und technischer Autonomie unterschieden (Müller 2012, 213; Sankaran et al. 2020). Oft handelt es sich um Beispiele einer „technischer Autonomie“, die in Relation zur menschlichen Kontrolle steht und bei denen Systeme Tätigkeiten unabhängig von menschlicher Intervention vollziehen. Oder es ist von sich graduell entwickelnder „Autonomie“ zu lesen: „from a fully user controlled system to a system not controlled at all.“ (Müller 2012, 213) Nicht selten wird in Diskursen entweder zuversichtlich oder kritisch geäußert, dass bei



einer graduellen Zunahme „technischer Autonomie“ sich diese der menschlichen Autonomie immer mehr annähert (Visser, Pak, und Shaw 2018; Zhu, Wang, und Wang 2018).<sup>6</sup>

Einer solchen Begriffsverwendung wohnt ein anthropomorphisierendes Moment inne: Komplexen intransparenten Systemen, welche Tätigkeiten „autonom“ verrichten, werden dabei menschliche Eigenschaften zugeschrieben. Mindestens implizit führt dies dazu, dass die Vorstellung einer möglichen Selbstgesetzgebung – einschliesslich ein damit verbundenes Freiheitsverständnis und die Möglichkeit, der Vernunft nach zu handeln – entsprechend derjenigen des Menschen mittransportiert wird. Umgekehrt werden menschliche Eigenschaften wie Selbstgesetzgebung zunehmend von einer technischen Perspektive aus betrachtet und mit solchen Systemen verglichen. Das heisst: Auf den Menschen werden Eigenschaften projiziert, die bisher den Systemen, die anspruchsvolle algorithmische Tätigkeiten ausüben, zugehörig waren. Es wird noch Gelegenheit sein, darauf zurückzukommen.

Beides ist problematisch. Es scheint daher angemessener zu sein, bei solchen Systemen nicht von „Autonomie“, sondern von „Eigenständigkeit“ zu sprechen.

## 5. Technik als Realisierung menschlicher Lebensformen

„Autonomie“, im Sinne einer Selbstbestimmung oder möglicher Selbstgesetzgebung, ist dem ursprünglichen Wortsinn nach verknüpft mit einem Verständnis von Freiheit, die es dem Menschen ermöglicht, moralisch zu denken und handeln, ohne fremdbestimmt zu sein. Angesichts technologischer Entwicklungen, insbesondere auch im Bereich des maschinellen Lernens, drängt es sich nicht nur auf, über die Eigenständigkeit von technischen Möglichkeiten, sondern aufgrund deren Verwobenheit mit den Menschen demnach über ein Verständnis von Freiheit beim Menschen und seinen Umgang damit nachzudenken.

Im Folgenden soll daher zu Beginn dem Technikverständnis Ernst Cassirers gefolgt werden, das auf die Verwobenheit von Menschen und Technik eingeht, bevor auf das Verständnis von Freiheit beim einzelnen Menschen eingegangen wird. Ausgangspunkt ist, dass Technik zum Menschen gehört,

6 Visser et al. schlagen bspw. vor, in eigenständige Systeme menschlichen Fähigkeiten einzubauen: „we propose a framework to infuse a unique human-like ability, building and actively repairing trust, into autonomous systems.“ (Visser et al. 2018, 1409), und Zhu et al. sprechen von „human-like autonomous driving algorithms“ (Zhu et al. 2018, 348).

das heisst, sie ist selbst Ausdruck und Voraussetzung zum Aufbau von Kultur. Technik ist dabei nicht allein als Handwerkszeug zu verstehen, sondern umfasst mehrere Dimensionen. Dies zeigt sich mit Hilfe von Cassirers Verständnis des Menschen als *animal symbolicum*, welches Menschen als Wesen, „*die ihrem Ausdruck eine Form zu geben vermögen*“ (Schwemmer 1997, 31), versteht.

Wie Sprache gehört auch Technik unabdingbar zur Realisierung menschlicher Lebensformen. Mit ihr konstituiert der Mensch seine Wirklichkeit, indem er die Welt aktiv formt und sie mit Bedeutung verseht. Ein solcher sinn- und formgebender Prozess beruht für Cassirer auf den Tätigkeiten des Geistes, welcher „als das bildende Prinzip in jenem Subjekt zu begreifen [ist], in dem der Kulturprozess seinen Ursprung hat“ (Recki 2013, 46). Gleichwohl soll „Geist“ hier nicht als eine Art unabhängige Substanz aufgefasst werden, die in sich ist und durch sich begriffen wird, sondern als „das produktive, auf Verobjektivierung in einem materiellen Medium ausgerichtete Prinzip des weltbildenden menschlichen Bewusstseins.“ (Recki 2013, 113) Demnach ist zu erkennen: Einerseits stellt Geist ein Vermögen dar, überhaupt konkret in der Welt sinn- und formgebend zu handeln, andererseits wird dazu ein Medium benötigt, durch das dieses Vermögen sich betätigen, zum Ausdruck gelangen und erkannt werden kann.

Allgemein gesprochen bringt dieses produktive Tätigsein Kultur hervor, und dadurch ist für Cassirer Geist gerade begreifbar. Dazu gehört auch Technik als Medium. Sie ist dasjenige, „*was dem menschlichen Willen eine körperliche Form gibt.*“ (Cassirer 2009, 128)<sup>7</sup> Kultur ist demnach nicht einfach gegeben, sondern muss erarbeitet werden. Entscheidend ist, dass in dieser Arbeit auch eine Selbstbefreiung des Menschen zum Ausdruck kommt: „Die menschliche Kultur als Ganzes kann als der Prozeß der Selbstbefreiung des Menschen verstanden werden. Die Entwicklung der Sprache, der Kunst, der Religion und Wissenschaft sind die einzelnen Phasen dieses Prozesses. In diesen [...] entdeckte und erprobte der Mensch eine neue Macht, eine Macht, mit deren Hilfe er sich eine ‚ideale‘ Welt erbaute.“ (Cassirer 1960, 289)

Damit der Mensch sich die Welt einrichten kann, und sich nicht nur äußeren Zwängen der Naturgegebenheiten ausgesetzt sieht, muss er die Welt nicht nur wahrnehmen, sondern er muss auch fähig sein, zu planen, sich von seinen schöpferischen Zielen zu distanzieren und sie in der Zukunft zu antizipieren: „Dieses Stehenlassen des Zieles ist es erst, was eine ‚objektive‘ Anschauung, eine Anschauung der Welt als

7 Cassirer übernimmt diese Definition der Technik von Max Eyth (1924).

einer Welt von ‚Gegenständen‘ ermöglicht.“ (Cassirer 2009, 141) Birgit Recki bezeichnet diese Verobjektivierung als „Distanz-Apriori“, welches „[e]ntscheidend zum Verständnis von Cassirers nirgendwo definiertem Freiheitsbegriff“ (Recki 2013, 75) sei. Denn für Cassirer ermöglicht die (auch geistige) Distanzierung es dem Menschen nicht nur, die Welt als eine Welt von Gegenständen wahrzunehmen, sondern gleichzeitig, zu sich selbst in ein distanziertes Verhältnis zu treten.

Das heisst: Indem der Mensch ein technisches Artefakt erschafft, steht ihm nicht nur ein solches gegenüber, sondern er steht damit insofern in einer Wechselbeziehung, als dieses Artefakt ihn auch erkennen lässt, dass er selbst es war, der das, was nun auf ihn wirkt, gedacht und geformt hat. Erst in diesem Erkenntnisakt konstituiert sich für Cassirer mit dem Objekt auch das Subjekt selbst. In dieser Wechselbeziehung werden Handlungs- und Denkräume geschaffen, die es dem Menschen ermöglichen, eine Welt sinnhaft und sinnlich einzurichten: „[I]n dieser Doppelrelation der Abstandnahme eröffnet sich mit dem Freiraum der Verfügung der Spielraum der Reflexion und des Handelns. Durch jeden Akt der Symbolisierung gewinnt der Mensch in der distanzierenden Verobjektivierung einen Spielraum der Verfügung, in dem Cassirer den Ursprung der Freiheit erkennt.“ (Recki 2013, 75)

Mit dem Wort „Ursprung“ wird angedeutet, dass bei Cassirer der Begriff „Freiheit“ seinen Ausgang anderswo findet, nämlich bei der „Spontaneität des tätigen Geistes“ (Recki 2013, 76). Was ist damit gemeint?

## 6. Spontaneität als Ausdruck von Freiheit in der Welt

Mit „Spontaneität des tätigen Geistes“ wird auf Kants Vorstellung einer Spontaneität des Denkens angespielt, die in Cassirers Philosophie eine zentrale Rolle einnimmt (Schwemmer 1997, 25). Bei Kant gilt: Spontaneität bezeichnet den Akt des Menschen, von sich aus Vorstellungen hervorzu-bringen bzw. Begriffe bilden und Urteile ableiten zu können.

Spontan heisst hier zuerst einmal: „von selbst“ (*sua sponte*). Damit ist gleichzeitig „durch andere(s)“ ausgeschlossen. Die Frage ist nun, wie dieses „von selbst“ noch näher zu definieren wäre. Ulrich Sonnemann, dessen Gedanken in Hinblick auf den Begriff der Spontaneität durchaus von einer Nähe zu Kant gezeichnet sind, meint dazu:

„[...] es ist Zeit, uns zu erinnern, daß der Geist bei den höheren Gebilden des Gedankens und der Kunst nicht beginnt: vom Spiel des Kindes an folgt alles menschliche Verhalten, das spontan ist, einer Regung aus ihm oder auf ihn hin. Was heisst aber spontan? Was sich *sponte* begibt, begibt sich von selbst, also frei:

ohne dieses *von selbst* glückt keine Wesensbestimmung der Freiheit, die aus ihrem Wesen verstanden nicht die *libertas* von Rechten und Einrichtungen, sondern etwas Ursprünglicheres ist. Definiert das nun das Spontane? Halten wir uns mit solchen Sorgen nicht auf. Zur Definition der sich von selbst ereignenden Uerscheinung der Freiheit wird hier kein Finger gerührt.“ (Sonnemann 1969, 330)

Zwar verweist Sonnemann mit dem Spontanen auf eine „von selbst ereignende Uerscheinung der Freiheit“, doch zugleich macht er deutlich, dass er sich nicht darum bemühen will, Spontansein exakt zu bestimmen. Warum nicht? Spontaneität, als ein dem Menschen unmittelbar widerfahrendes Ereignis, ist zwar geradezu „als ein am Menschenwesen doch wichtiges, ja für es eigentlich entscheidendes Moment“ (Sonnemann, 1969, 330). Doch gleichzeitig und grundsätzlich handelt es sich um ein nicht vorhersehbares *Ereignis*, welches weder plan-, mess- noch berechenbar sein kann und sich dem Versuch einer definitorischen Festlegung entzieht. Damit gerät ein unverfügbares Moment des Menschseins ins Blickfeld. Spontaneität, als jeweils neues, einzigartiges Ereignis, ist nicht reproduzierbar. Es entzieht sich daher gerade dem, was für die positiven Wissenschaften notwendig wäre, um das Wesen des Menschen zu bestimmen und messbar zu machen.<sup>8</sup>

Diese „Uerscheinung der Freiheit“ erinnert an Kant, bei dem Freiheit an und für sich als unabhängig vom Wirken der Naturgesetze in Raum und Zeit gilt. Trotzdem wirkt sie als Bewegursache in die von Gesetzen bestimmte Natur hinein, ohne diesen Gesetzen selbst zu unterliegen. Freiheit gilt für Kant als „eine absolute Spontaneität der Ursachen, eine Reihe von Erscheinungen, die nach Naturgesetzen läuft, von selbst anzufangen“ (KrV B 474 / A 446). Doch sie ist für ihn unabdingbar mit Vernunft verbunden. Sie soll zu menschlichem Verhalten in der Welt führen: „Praktische Vernunft und Freiheit gehören zusammen; allein im Verbund mit der jeweils anderen – komplementären – Anlage kommen sie je für sich zur Entfaltung. [...] Einer Vernunft ohne Freiheit nämlich bliebe die entscheidende, weil in Bewegung setzende und gestaltende, bliebe die praktische

8 Am Beispiel von Intelligenztests zeigt Sonnemann ein diesbezügliches Scheitern der Wissenschaft auf, weil Menschen unvorhergesehen intelligentes Verhalten aufweisen können, auch wenn ein solches im Antwortraster des Testes nicht vorgesehen und daher nicht bewertbar ist. Hier scheitert die Wissenschaft, „[w]eil das Wesen des Psychischen nicht der Mechanismus, sondern die Spontaneität ist – auf die speziellen Verhältnisse angewandt, die für die Intelligenz gelten: weil Verhalten, das nicht normgemäß, sondern wesentlich intelligent ist, überhaupt nicht vorausbestimmt werden kann, da sein Charakter die Überraschung ist, das Schöpfertum im genauen Wortsinn, der Durchbruch durch Schranken [...]. [...] Das Wesentliche an der Intelligenz ist so wenig meßbar wie das Wesen überhaupt“. (Sonnemann 1969, 188–89)

Dimension verschlossen.“ (Himmelmann 2003, 51) Damit kommt der Begriff des Gesetzes ins Spiel: Die praktische Freiheit ist auf Regeln und Ordnung angewiesen, da sie sonst ohne Ziel und Richtung wäre (Himmelmann 2003, 52).

Auffällig ist an Sonnemanns Zitat, dass er das „Spiel des Kindes“ erwähnt. Bereits beim Kind, welches noch keine vorgeformten Begriffe kennt und im spontanen Spiel seine Welt sinnlich wahrnimmt und zu erschliessen beginnt, scheint sich der Geist auf den Weg zum Denken zu begeben; wobei „alle Reflexion, die über den Kindergriff nach dem Vorhandenen sich hinauswagt, überhaupt zunächst sich regt und das Denken auf den Weg bringt.“ (Sonnemann 1969, 331) Im Spiel des Kindes wird diese Spontaneität leiblich dargestellt. Sie erscheint als Ausdruck von Freiheit in der Welt: „Die Freiheit, die in der Welt in Gestalt des Spontanen erscheint, will sich *zeigen*, sie will nicht bestimmt werden; sie ist selbst das Bestimmende (des Geistes) und die Bestimmung (des Menschen).“ (Sonnemann 1969, 330) Die Verbindung des unregelmässigen Verbs „wollen“ mit „zeigen“ schreibt der Freiheit zwar die Absicht zu, sich selbst Ausdruck verleihen zu wollen, doch gleichwohl folgt darauf eine Absage gegenüber jeglichem Versuch, Freiheit durch die Erkenntniskraft des Verstandes begreifen zu wollen. Dieses Spannungsverhältnis bleibt bestehen. Würde Freiheit als veräusserbares Gut verstanden werden – als etwas, das der Mensch besitzen kann –, bestünde es nicht, weil Freiheit sich dadurch verobjektivieren und somit rational fassen und durch den Verstand in Anspruch nehmen liesse. Doch Freiheit ist für Sonnemann eben nicht eine zu vergegenständlichende, sondern selbst „*Eigenschaft des Menschen*“ (Sonnemann 1963, 83).<sup>9</sup> Der Mensch hat Freiheit nicht, sondern ist Freiheit. Ihre Absicht, sich zu zeigen, kann demnach überhaupt nicht als eine „von aussen“ an den Menschen herangetragene verstanden werden, sondern ist Ausdruck seiner selbst. Sie als Spontaneität in der Welt empirisch zu fassen würde Selbstobjektivierung bedeuten, doch „[d]ie Bedingung aller Objekterfahrung kann sich selbst nicht zum Objekt werden“ (Sonnemann 1969, 331), heisst es bei Sonnemann, denn „[d]ie Bedingung aller Erkenntnis von Naturgesetzlichkeit ist deren Vergegenständlichung; der Erkennende, um das Gesetz zu erkennen,

9 An dieser Stelle findet sich auch eine Kritik Sonnemanns an der (philosophischen) Idee, Willensfreiheit „als bloßes Postulat der praktischen Vernunft“ (Sonnemann 1963, 83) zu fassen. Sonnemanns Nähe zu Kant, die im vorliegenden Aufsatz zum Vorschein tritt, soll daher nicht darüber hinwegtäuschen, dass er sich auch kritisch gegenüber der Philosophie und ihrem Hauptgeschäft des Denkens äussert, dies spezifisch auch zu Kant (vgl. Sonnemann 1987). Auf Sonnemanns Kritiken kann im Kontext dieses Aufsatzes jedoch nicht näher eingegangen werden.

muß das Objektgeschehen *vor sich* haben, nimmt also, in dem Maß, in dem er sich ihm gegenüber objektiv verhält, nicht an ihm teil.“ (Sonnemann 1965, 75–76) Freiheit selbst ist die Bestimmung des Menschen und das Bestimmende der Vernunft, und somit Selbstbestimmung des Menschen. Wir erinnern uns an Kant, wo Freiheit und Vernunft unabdingbar miteinander verbunden sind. Auch Selbstbestimmung ist eng mit Freiheit und Vernunft verbunden, denn sie gibt ersterer eine Form und ist ein Vermögen der praktischen Vernunft: „Die Selbstbestimmung besteht [bei Kant], ganz allgemein gesagt, darin, der Freiheit eine *Form* zu geben. Vernunft ist die Instanz, der diese Formgebung obliegt, und dadurch ist sie *praktische Vernunft*.“ (Himmelman 2003, 51) Das heisst: Praktische Vernunft, als Vermögen, um frei moralisch zu denken und überhaupt zu reflektieren, schliesst Selbstbestimmung mit ein.

Spontaneität, als Ausdruck von Freiheit in der Welt, kann zwar nicht durch den Verstand erfasst, aber doch durch den Menschen erfahren werden; als augenblickliches Ereignis erscheinend, rückt es in den Bereich der Selbsterfahrung des Menschen, bleibt zugleich jedoch auch unbegreiflich Gewesenes.

## 7. Verstandesoperationen und Denken im Blick auf Deep Learning Systeme

Dass auch Denken für Sonnemann geradezu Freiheit ist, wurde andeutungsweise ersichtlich. Doch bedarf es einer näheren Ausführung, denn Denken unterscheidet Sonnemann von blossen „Verstandesoperationen“. Letztere brauchen weder Freiheit noch sind sie Ausdruck einer solchen; vielmehr definieren sich Verstandesoperationen durch ihren technischen Charakter, durch Mechanismen, kalkulatorische Vorgänge und Abläufe „in programmhaft vorgespurten Bahnen“ (Sonnemann 1963, 181). Hingegen ist die menschliche Fähigkeit, denken zu können, jeweils als spontan zustossendes Ereignis und Ausdruck von Freiheit „überhaupt kein Verrichten, sondern ein Empfangen, sein Wesen ist gerade die Offenständigkeit, das Schrankenlose, das Nichtfestgelegtsein. Was da empfangen – niemals konstruiert – wird, sind Einsichten, Ideen, Gedanken“ (Sonnemann 1963, 181, sinngemäss auch 119). Am Beispiel der „denkenden Maschine“ sieht Sonnemann „[d]as Mißverständnis [...] einer alten Verwechslung des Denkens

mit der Verstandesoperation.“ (Sonnemann 1963, 181)<sup>10</sup> Es sind die Kybernetiker mit ihrer „Lehre von den elektronischen Gehirnen“ (Sonnemann 1963, 180), die dieser Verwechslung zusätzlich Nachdruck verleihen, und die ihre Roboter und Hirnmodelle so kreieren, wie sie selber zu denken meinen: „Sie gleichen dem Geist, den sie begreifen, ihren Begriff des Denkens selbst an, für dessen Verständnis ihnen theoretische Hirnmodelle vorschweben, die elektronisch arbeiten“ (Sonnemann 1963, 181). Und gerade wenn behauptet wird, dass es kein „Denken“ gäbe, welches nicht durch ein formales Regelsystem ausgedrückt werden könne, wird dabei vergessen, „daß immer diese Ausdrückbarkeit der Gedankentat, der Sprache, *folgt, nie ihr vorausgeht*.“ (Sonnemann 1963, 181; Hervorhebung R. S.) Damit verweist Sonnemann implizit auf Freiheit und den uns bereits begegneten Gedanken, dass Freiheit unmöglich durch den Verstand bestimmbar sei. Es bleibt dann auch unerheblich, wie effektiv, komplex und eigenständig (auch als Roboter „verkörperte“) Deep Learning Systeme mittels Rechenoperationen und implementierten Algorithmen Dinge berechnen oder Handlungen in der Welt verursachen — das Tun der „denkenden Maschinen“ bleibt im Vergleich zu genuinem Denken ein Verrichten im Sinne der Verstandesoperationen.

Gerade im Hinblick auf Deep Learning Systeme haben Sonnemanns Gedankengänge an Aktualität nichts eingebüsst: Auch wenn Systeme nun vermehrt auf Prinzipien des Lernens aus „Erfahrung“ basieren, so sind sie doch auch unter der Kategorie von „denkenden Maschinen“ zu fassen, denn sie haben Funktionen (bspw. Gewichtungsfunktionen) zur Datenverarbeitung implementiert, berechnen Ergebnisse und Wahrscheinlichkeiten, klassifizieren und kategorisieren, verwenden statistische Methoden, verarbeiten eingegebene Beispieldaten und versuchen daraus, Gesetzmässigkeiten und Muster zu extrahieren und Vorhersagen zu treffen. Es handelt sich um Systeme, die prinzipiell Algorithmen folgen. Daran ist nichts „Schöpferisches“ auszumachen; deren Eigenschaft ist nicht Freiheit, deren Verrichten kein Erkenntnisdrang und deren Resultate sind keine Einsichten. Demgegenüber entfaltet sich menschliche Spontaneität als Ausdruck von Freiheit „erst jenseits“ (Sonnemann 1963, 183) dessen.

10 Obwohl Sonnemann an dieser Stelle nicht näher darauf eingeht, was er mit „alt“ meint, so ist anzunehmen, dass er sich hier philosophiegeschichtlich auf die Idee des Menschen als „Maschine“ oder „Automaten“ bezieht, so wie es u. a. bei Hobbes (Leviathan) vorzufinden ist, aber auch beim Cartesischen Dualismus, der den menschlichen Körper mechanistisch denkt.



Indem der Mensch seine Systeme so baut, wie er meint, dass er selber denkt, verändert sich sein Selbst- und Weltbild. Durch die von ihm selbst hervorgebrachte fortschreitende „Mechanisierung der Köpfe“ (Sonnemann 1963, 181), durch die „Verapparaturung des Menschen“ (Sonnemann 1969, 186) – heute zweifellos auch eine „Algorithmisierung“ des Menschen – läuft er Gefahr, dass freies Denken sich nicht ereignen kann und er sich selbst lediglich als Verstandesoperator, und dadurch als unfreies informationsverarbeitendes Wesen, versteht. In einer solchen Welt, so hatte bereits Günther Anders eindringlich festgestellt, sei es für den Menschen unstrittig, „daß auch seine Denkleistungen, verglichen mit denen seiner ‚computing machines‘, schlecht abschneiden“ (Anders 1985, 32). Für Sonnemann ist die verstörende Irritation oder Scham, die der Mensch angesichts einer ihm überlegenen Technik erfährt, nicht unbedingt auf die scheinbare Konkurrenzunfähigkeit zurückzuführen, sondern ergibt sich beim Anblick seines „denkenden“ Ebenbildes in der Welt: „Der Mensch des Apparats ist kein Produkt der Technik, wohl aber hat er, ihres Wesens ansichtig, Grund, sich dafür zu halten. Erst seit er die Technik hat, sieht er, gestehen wir es doch ein, etwas wirklich Neues, das Befremdlichste von allem, nämlich sein Ebenbild in der Welt“ (Sonnemann 1963, 120). Was dem Menschen begegnet ist das, wofür er sich hält — dasjenige „denkende“ Ebenbild, dass er sich erschafft hat, wie er selber denkt, dass er denkt. Nur, dass dieses Ebenbild nun „angefangen hat, als Ganzes auf den Menschen zurückzuwirken“ (Sonnemann 1963, 124). Hier klingt Cassirers „Distanz-Apriori“ wieder an, jedoch scheint sich das Mass der Verfügungsmacht, die der Mensch gegenüber seinen technischen Artefakten innehat, zu vermindern: „Die Technik, dies Gewollte, will jetzt etwas. Sie antwortet dem Menschen nicht länger wie sein eigener Schatten, sondern wie ein Antwortender“ (Sonnemann 1963, 126).

Der Eindruck, den Systeme mit Deep Learning hinterlassen, lässt sich an dieser Stelle einreihen: Sie prägen beispielsweise Handlungsräume des Menschen, indem sie diese vorstrukturieren, oder sie präsentieren dem Menschen nicht nur gewünschte, sondern auch unerwartete und überraschende Resultate. Zum Beispiel sollten Nicht-Spieler-Figuren – d. h. Figuren eines Computerspiels, die nicht von Menschen gesteuert werden – mittels maschinellen Lernens dazu gebracht werden, Hindernisse zu umgehen. Dazu hat man Wände um sie herum gebaut resp. programmiert. Die Figuren entwickelten jedoch mit der Zeit „von selbst“ eine besondere Art von „Wackeln“, das ihnen ermöglicht hat, Wände hinaufzusteigen und dadurch die Hindernisse auf unvorhergesehene Art und Weise zu umgehen



(Lehman, Clune, und Misevic 2020, 285). Die Figuren zeigten eine Vorgehensweise, die der Mensch nicht erwartet hat.

## 8. „Black Box“ als Spiegelbild der menschlichen Unverfügbarkeit

Ganz im Sinne Sonnemanns erscheint die Technik als „denkendes“ Ebenbild nun *vermeintlich selbständig* (vgl. hierzu Sonnemann 1963, 125). Das mit den Deep Learning Systemen (wieder)auftauchende Phänomen der Black Box als Ausdruck nicht transparenter Systeme verstärkt diese Sicht: Als „nebulöses“ und „unverfügbares“ Element solcher Systeme gibt es vor, etwas zu sein, was es seiner Eigenschaft nach überhaupt nicht sein kann. Luzide hat Sonnemann diese gespenstige „Unwirklichkeit“ bereits hervorgehoben: „Daher die Unheimlichkeit der neuesten Maschinerie [...], Erscheinung von etwas zu sein, das den im Raum ihr Begegnenden auf das innigste zu eigen ist, also aber, da das Eigenste sich selbst nicht *als* dieses von außen begegnen kann, gerade als dessen absolutes Anderssein und so als die Verfremdung selber vor die Menschen tritt.“ (Sonnemann 1963, 124) Während Deep Learning Systeme als „denkendes“ Ebenbild des Menschen auftreten, greift das damit in Erscheinung tretende Phänomen der Black Box noch tiefer: Es steht für das Unverfügbare des „inneren Geisteslebens“ und bedeutet damit die Absage an die Hoffnung, durch die Erkenntnis kraft des Verstandes das Geschehen der Systeme gänzlich zu begreifen. Im übertragenen Sinne wird das Blackbox-Phänomen Ausdruck einer Freiheit als Eigenschaft der Systeme. Freiheit ist, wie wir gesehen haben, jedoch Eigenschaft des Menschen. Doch gerade, weil Freiheit dem Menschen eigen ist und ihm nicht gegenüber treten kann, erfährt er an seinem „denkenden“ Ebenbild eine Selbstentfremdung; er erfährt Denken vor und nicht bei sich. Der Mensch überträgt seine Unberechenbarkeit auf die Systeme.

## 9. Technik, die herausfordert

Trotz aller Gefahr der fortschreitenden Selbstentfremdung sieht Sonnemann genau darin auch die Chance des Menschen, sich wieder „auf sein Menschseinkönnen“ (Sonnemann 1963, 123) zu besinnen und dem Geist „im Griff der Mechanik“ (Sonnemann 1963, 125) (oder von allumfassenden algorithmischen Systemen) in seiner Not zur Hilfe zu eilen: „Der Mensch, das ist klar, wird nichts zu lachen haben, außer über sich selbst. Gar nicht genug aber, muß ihm gewünscht werden, wird er diesem einen, befreiendsten Lachen von Herzen sich dann wenigstens

hingeben. Eine stärkere, geistgesegnetere Waffe, das Mechanische in sich zu tilgen, hat er konstitutionell überhaupt nicht“ (Sonnemann 1963, 124).<sup>11</sup> Die Chance liegt darin, dass Technik – und wird sie in Zukunft noch so ausgeklügelt sein – den Menschen provoziert und existenziell herausfordert, über sich selbst nachzudenken und Stellung zu sich selbst zu beziehen: indem sie ihm als unfreies Spiegelbild begegnet und ihm eine Möglichkeit aufzeigt, wie er sich selber sehen *kann*, aber nicht *muss*, und ihn gleichzeitig dazu heraus- und auffordert, sein Spiegelbild – und dadurch sich selbst und sein Verhältnis zu ihm – auch anders zu denken. Dieses „andere Denken“ könnte mit Anspielung auf Sonnemanns „Negative Anthropologie“ auch heissen, dass die Technik danach fragt, was Menschen eben nicht sind: lediglich aus formalisierten Prozessen bestehend, maschinenhaft lernende „Systeme“, Denkapparate, Algorithmen, kalkulierbare „Verstandesoperatoren“. Sie fordert den Menschen auf, sich „nach seiner Freiheit [zu] fragen“ (Sonnemann 1963, 126). Diesen Aufforderungen nachkommend, würde sich das Denken des Menschen selbst verändern. An einer solch widerständigen Technik hat sich menschliche Freiheit immer wieder zu bewähren: „Daß die Freiheit sich er-eignet, besagt, daß sie den Menschen sich zu eigen nimmt und er sie an als seine Eigenschaft. Diese will sich bewähren. Sie kann es nur an Widerständen“ (Sonnemann 1963, 145).

## 10. Schlussbemerkungen

Dass sich solche Widerstände, die sich aus der Technik ergeben, immer wieder aufs Neue auftun, zeigt sich nicht zuletzt auch an den Systemen, die zu Deep Learning fähig sind: Als Teil der menschlichen Lebenswelt erscheinen sie dem Menschen als unverfügbare und eigenständige Systeme, die auch auf seine Handlungsräume und sein Denken Einfluss nehmen können. Indem sie vorgeben, Eigenschaften zu besitzen, die der Mensch bislang als seine eigenen verstanden hat, können sie das Bild, das Menschen von sich selber machen, infrage stellen.

Nimmt man Sonnemanns Überlegungen ernst, heisst dies jedoch auch, dass eine genaue Verhältnisbestimmung von menschlichem Selbstverständnis in Relation zur Technik eigentlich nicht abschliessend möglich sein sollte, weil ansonsten frei denken im Sinne Sonnemanns sich nicht (mehr)

11 Im Zusammenhang mit dem befreienden Lachen eines Menschen scheint es interessant zu sein, dass demgegenüber ein „künstliches“ Lachen, bspw. bestehend aus einem animatronischen Robotergesicht in Kombination mit implementierten Lernalgorithmen (vgl. Chen et al. 2021), dem Betrachter eher unheimlich erscheinen kann.

ereignen kann. Hingegen stellt es eine fortwährende Aufgabe des Menschen dar, immer wieder – und dies mit „voller Aufmerksamkeit für die Welt als eine äußere oder innere“ (Sonnemann 1969, 82) – festgelegten Denkmustern entgegenzutreten. Dies kann dann beispielsweise auch bedeuten, nicht nur danach zu fragen, was Systeme mit Deep Learning können/nicht-können, und wo und unter welchen Umständen sie eingesetzt/nicht-eingesetzt werden (sollten), sondern dass in Diskursen über gegenwärtige Techniken der Fokus vermehrt auch auf das menschliche Selbstverständnis und die damit verbundenen Fragen nach dem Menschsein zu richten wäre.

## Literaturverzeichnis

- Anders, Günther. 1985. *Die Antiquiertheit des Menschen*. Erster Band. München: C.H. Beck.
- Ashby, William Ross. 1956. *An introduction to cybernetics*. London: Chapman and Hall.
- Boden, Margaret A. 2018. *Artificial Intelligence: a Very Short Introduction*. Oxford, United Kingdom: Oxford University Press.
- Cassirer, Ernst. 1960. *Was ist der Mensch? Versuch einer Philosophie der menschlichen Kultur*. Stuttgart: Kohlhammer.
- Cassirer, Ernst. 2009. *Schriften zur Philosophie der symbolischen Formen*. Hamburg: Meiner.
- Castelvecchi, Davide. 2016. „The Black Box of AI.“ *Nature*, Oct 6;538(7623): 20–23.
- Chen, Boyuan, Yuhang Hu, Lianfeng Li, Sara Cummings, und Hod Lipson. 2021. „Smile Like You Mean It: Driving Animatronic Robotic Face with Learned Models.“ *arXiv preprint arXiv:2105.12724v1*.
- Goodfellow, Ian, Yoshua Bengio, und Aaron Courville. 2016. *Deep Learning*. Cambridge: MIT Press.
- Himmelfmann, Beatrix. 2003. *Kants Begriff des Glücks*. Berlin/New York: De Gruyter.
- Kant, Immanuel. *Kritik der reinen Vernunft*. Nach der ersten und zweiten Originalausgabe. Hrsg. Jens Timmermann. Mit einer Bibliographie von Heiner F. Klemme. Hamburg: Meiner, 1998.
- Krämer, Sybille. 1988. *Symbolische Maschinen: die Idee der Formalisierung in geschichtlichem Abriß*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Lehman, Joel, Jeff Clune, und Dusan Misevic. 2020. „The Surprising Creativity of Digital Evolution: A Collection of Anecdotes from the

- Evolutionary Computation and Artificial Life Research Communities.“ *Artif. Life* 26(2): 274–306.
- Lin, Henry W., Max Tegmark, und David Rolnick. 2017. „Why Does Deep and Cheap Learning Work So Well?“ *J Stat Phys* 168, 1223–1247.
- Lipton, Zachary C. 2018. „The Mythos of Model Interpretability: In machine Learning, the concept of interpretability is both important and slippery.“ *ACM Queue*, Vol. 16, Issue 3, May-June 2018, 31–57.
- Müller, Vincent C. 2012. „Autonomous Cognitive Systems in Real-World Environments: Less Control, More Flexibility and Better Interaction.“ *Cognitive Computation*, Vol. 4, 212–215.
- Pohlmann, Rosemarie. 1971. „Autonomie.“ In *Historisches Wörterbuch der Philosophie*, Band 1, Joachim Ritter (Hrsg.), 701–719. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Recki, Birgit. 2013. *Cassirer*. Stuttgart: Reclam.
- Rudin, Cynthia. 2019. „Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead.“ *Nat Mach Intell* 1, 206–215.
- Sankaran, Supraja, Chao Zhang, Marisela Gutierrez Lopez, und Kaisa Väänänen. 2020. „Respecting Human Autonomy through Human-Centered AI.“ *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20)*. Association for Computing Machinery, New York, USA, Article 134, 1–3.
- Schwemmer, Oswald. 1997. *Ernst Cassirer: ein Philosoph der europäischen Moderne*. Berlin: Akademie Verlag.
- Sonnemann, Ulrich. 1963. *Das Land der unbegrenzten Zumutbarkeiten: deutsche Reflexionen*. Reinbek bei Hamburg: Rowohlt.
- Sonnemann, Ulrich. 1965. „Spontaneität und Apparat: Über die Freiheit als Eigenschaft und als Einrichtung: Ein Dialog“. In *Land der unbegrenzten Zumutbarkeiten*, Schriften in 10 Bänden, Band 4, Paul Fiebig (Hrsg.). Springe: zu Klampen, 2014.
- Sonnemann, Ulrich. 1969. *Negative Anthropologie: Vorstudie zur Sabotage des Schicksals*. Reinbek bei Hamburg: Rowohlt.
- Sonnemann, Ulrich. 1987. „Zeit ist Anhörungsform: Über Wesen und Wirken einer kantischen Verkenntung des Ohrs“. In *Die sterbende Zeit: 20 Diagnosen*, hrsg. von Dietmar Kamper und Christoph Wulf, 202–220. Darmstadt: Luchterhand.
- Visser, Edward J. de, Richard Pak, und Tyler H. Shaw. 2018.: „From ‚automation‘ to ‚autonomy‘: the importance of trust repair in human-machine interaction.“ *Ergonomics*, Vol. 61, No. 10, 1409–1427.

- Wiedebach, Hartwig. 2014. *Pathische Urteilskraft*. Freiburg i. Br.: Karl Alber.
- Zhang, Chiyuan, Samy Bengio, Moritz Hardt, Benjamin Recht und Oriol Vinjals. 2021. „Understanding deep learning (still) requires rethinking generalization.“ *Communications of the ACM* 64, 107–115.
- Zhu, Meixin, Xuesong Wang, Yinhai Wang. 2018. „Human-like autonomous car-following model with deep reinforcement learning.“ *Transportation Research Part C: Emerging Technologies*, 348–368.

