FH Burgenland
UNIVERSITY OF APPLIED SCIENCES

Department Informationstechnologie

Wer bist du, KI?
Philosophische Annäherungen an unser menschliches Selbstverständnis

Masterarbeit
zur Erlangung des akademischen Grades
Master of Science in Engineering

Masterstudiengang Cloud Computing Engineering

Eingereicht von:	Michael Fleischhacker, Bakk. phil.
Personenkennzeichen:	2110781001
Datum:	01. Juni 2023
Betreut von:	o. Univ.-Prof. em. Mag. Dr. Friedrich Roithmayr

Vorwort

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während der Anfertigung dieser Masterarbeit unterstützt und motiviert haben.

Zuerst gebührt mein Dank Herrn o. Univ.-Prof. em. Mag. Dr. Friedrich Roithmayr, der meine Masterarbeit betreut und begutachtet hat. Vom ersten Gespräch an, bei dem ich Herrn Mag. Dr. Roithmayr um die Betreuung meiner Arbeit bat, schenkte er dem Thema meiner Masterarbeit große Aufmerksamkeit und unterstützte mich in weiterer Folge mit hilfreichen Anregungen und konstruktiver Kritik bei der Erstellung dieser Arbeit.

Ein besonderer Dank gilt allen Personen, die sich für meine Masterarbeit als Expert*innen für die Interviews zur Verfügung gestellt haben. Ohne sie hätte diese Arbeit nicht entstehen können. Mein Dank gilt ihrer Informationsbereitschaft und ihren interessanten Beiträgen und Antworten auf meine Fragen.

Außerdem möchte ich meiner Tochter Fiona Fleischhacker für das Korrekturlesen meiner Masterarbeit danken.

Abschließend möchte ich mich bei meiner Familie und insbesondere bei meiner Frau Olivia bedanken. Nur durch eure Unterstützung war es mir möglich, mein Studium erfolgreich abschließen zu können.

Michael Fleischhacker

Steinbrunn, 01. Juni 2023

Inhaltsverzeichnis

Kurzfassung

Mit der Frage nach dem *Wer* wurde in dieser Arbeit bewusst der Blick auf das Subjekthafte von Künstlicher Intelligenz (KI) gelegt, um zu erforschen, ob und wie KI in jene ontologischen Bereiche, die bisher dem Menschen vorbehalten waren, vordringen kann. Dem Menschen und seinem Selbstverständnis von sich und der Welt blieb es bislang vorbehalten, Zuschreibungen von Phänomenen wie Vernunft, Geist, Freiheit, Würde, Intelligenz, Kultur, Ethik oder Moral für sich zu beanspruchen. Dabei eröffnete sich die Frage, was die Auseinandersetzung mit dem Phänomen der Wirkmächtigkeit von KI, unabhängig der Antworten darauf, für unser eigenes Selbstverständnis von Menschsein bedeutet.

In dieser Masterarbeit wurden mittels Literaturbefunden und Expert*innengesprächen die Auswirkungen untersucht, die in der Auseinandersetzung mit der Frage entsteht, wenn KI auf den von Jean-Paul Sartre definierten Freiheitsbegriff des Menschen trifft. Denn der Einfluss von KI hat und wird weiterhin in wachsendem Ausmaß Auswirkungen auf unser Leben und unseren Alltag haben. Im Besonderen galt das Augenmerk dieser Arbeit einer mit Bewusstsein ausgestatteten KI, da diese ein zentrales Element in der Beantwortung der Forschungsfrage darstellt.

Im Blickfeld der Untersuchung standen der von Sartre postulierte Freiheitsbegriff des Menschen und dessen Anspruch auf Gültigkeit sowie die Wirkmächtigkeit von KI. Ebenso wurde der Frage nachgegangen, ob KI nach den von Sigmund Freud beschriebenen drei großen Kränkungen der Menschheit, sich möglicherweise zur vierten großen Kränkung der Menschheit entwickeln könnte. Schon der Begriff *Intelligenz*, dem voran das Adjektiv *künstlich* beisteht, fordert uns gerade dazu auf, sich mit der formalen Definition von Begriffen wie Intelligenz, Bewusstsein, Vernunft, Würde, Tod, Leben oder Freiheit und deren Beziehungen als Grundlage für ein gemeinsames Verständnis auseinanderzusetzen.

Leitfadengestützte semistrukturierte qualitative Interviews mit sechs Expert*innen dienten zur Erhebung des qualitativen Datenmaterials. Mittels einer inhaltlich strukturierenden qualitativen Inhaltsanalyse erfolgte die analytische Betrachtung der transkribierten Interviews.

Die Gegenüberstellung und der Vergleich der Literaturbefunde mit den ausgewerteten Daten der Expert*inneninterviews führte zu der Erkenntnis, dass der von Jean-Paul Sartre postulierte Freiheitsbegriff weiterhin seine Gültigkeit behält. Die Ergebnisse der Untersuchung in dieser Arbeit stützen diese These dahingehend, dass nach Sartre das menschliche Selbstbewusstsein ausschließlich durch das Auftreten von Freiheit entstehen kann und damit eine unhintergehbare Bedingung der menschlichen Existenz darstellt, die auch von der Wirkmächtigkeit Künstlicher Intelligenz nicht aufgehoben werden kann.

Abstract

With the question of the *who*, the focus in this work was deliberately placed on the subjective nature of artificial intelligence (AI) to explore whether and how AI can penetrate those ontological realms that were previously reserved for humans. Until now, it has been reserved for humans and their self-understanding of themselves and the world to claim attributions of phenomena such as reason, spirit, freedom, dignity, intelligence, culture, ethics, or morality. This opened the question of what the confrontation with this phenomenon of the efficacy of AI, regardless of the answers to it, means for our own self-understanding of being human.

In this master's thesis, literature findings and interviews with experts were used to examine the effects that arise when AI encounters the concept of human freedom as defined by Jean-Paul Sartre. The influence of AI has and will continue to have an increasing impact on our lives and everyday life. In particular, the focus of this work was on conscious AI, as it is a central element in answering the research question.

The study focused on the concept of human freedom postulated by Sartre and its claim to validity, as well as the effectiveness of AI. The question was also explored as to whether AI, after the three great mortifications of humanity described by Sigmund Freud, could possibly develop into the fourth great mortification of humanity. The very concept of intelligence, preceded by the adjective artificial, invites us to look at the formal definition of concepts such as intelligence, consciousness, reason, dignity, death, life or freedom and their relationships as a basis for a common understanding.

Guideline-based semi-structured qualitative interviews with six experts were used to collect qualitative data. The transcribed interviews were analyzed by means of a qualitative content analysis.

The comparison of the literature findings with the evaluated data of the expert interviews led to the conclusion that the concept of freedom postulated by Jean-Paul Sartre remains valid. The results of the study in this thesis support this thesis to the extent that, according to Sartre, human self-consciousness can only arise through the occurrence of freedom and thus represents an inescapable condition of human existence that cannot be abolished even by the effectiveness of artificial intelligence.

1 Einleitung

„Nein! Nur vergessen das Licht auszumachen. Niemand da. Keine Menschenseele!“

So endet das Theaterstück *Keine Menschenseele* der Grimme-preisgekrönten und Emmy-Awards nominierten Gruppe *Laokoon*, die in ihrer ersten Arbeit im Kasino des Burgtheaters mit künstlichen neuronalen Netzen experimentiert und eigene Stimmen erzeugt, die keine Stimmbänder brauchen. In der Uraufführung war von digitaler Unsterblichkeit die Rede, einer Vision vieler Transhumanist*innen, die menschliche Intelligenz und individuelles menschliches Bewusstsein unabhängig vom schwachen und anfälligen biologischen Körper machen. Möglich gemacht werden soll all dies durch den Einsatz von Künstlicher Intelligenz (KI).

Dass diese Visionen nicht bloß auf die Theaterbühne gebrachte literarische Zukunftsvisionen sind, beschreiben die beiden Autoren des Werkes eindrucksvoll in ihren beiden Büchern *Die digitale Seele* und *Vom Ende der Endlichkeit*. In diesen werden die Erkenntnisse ihrer Recherchen in der Startup-Szene im Silicon Valley in den Vereinigten Staaten von Amerika geschildert.

1.1 Methodischer Ansatz

In dieser Masterarbeit werden zunächst grundlegende wissenschaftliche Erkenntnisse zu KI anhand der relevanten Literatur dargestellt. Dabei werden Definitionen von KI, deren technische Grundlagen sowie Funktionsweisen beschrieben. Im Zentrum des Interesses steht dabei die Wirkmächtigkeit von KI und deren allumfassender Einfluss auf unser Leben.

Ebenfalls wird anhand relevanter Literatur der von Sartre postulierte Freiheitsbegriff in seinem philosophischen Kontext erörtert. Die dabei gewonnenen Erkenntnisse werden jenen gegenübergestellt, welche aus der Beschäftigung mit dem Einfluss von KI auf unser menschliches Selbstverständnis gewonnen wurden.

Im empirischen Teil verfolgt die Masterarbeit als Forschungsstrategie leitfadengestützte semistrukturierte qualitative Interviews. Als Forschungsmethode wird ein qualitativer Ansatz in Form eines episodischen Interviews gewählt.

Aufgrund der Ergebnisse der verwendeten methodischen Ansätze sollen die Thesen und Fragestellungen sowie die Forschungsfrage überprüft und falsifiziert werden.

1.2 Problemstellung

In den letzten Jahren hat sich die Forschung intensiv mit dem Thema KI auseinandergesetzt. Wie aktuell das Thema KI ist, zeigt ein Blick in die

umfassende Literatur zu diesem Thema. So beschreibt Lee (2019) die vier Wellen der Entwicklung von Künstlicher Intelligenz. Neben der ersten Welle, der Internet-KI, bei der es zum Beispiel um robotergenerierte Berichte oder Fake News geht, werden in der zweiten Welle, der Business-KI, bereits Geschäftsmodelle entwickelt. Dabei stellt sich die Frage, wozu man noch Banker beschäftigen sollte (S. 144-157).

In der dritten Welle, der Wahrnehmungs-KI, verschwimmen bereits die Grenzen unserer Lebenswelt, in dem die Onlinewelt mit der Offlinewelt verschmilzt. Die KI der dritten Welle digitalisiert unsere Umwelt durch den Einsatz von Sensoren und intelligenter Geräte, die sie dabei zum Einsatz bringt. Dabei wird unsere natürliche Welt in digitale Daten umgewandelt, um danach von Deep-Learning-Algorithmen analysiert und optimiert zu werden (Lee, 2019, S. 157-159).

Die vierte und letzte Welle der Entwicklung von KI und für den Forschungsgegenstand dieser Arbeit von größter Bedeutung, ist die Autonome-KI.

> "Autonome KI verbindet die drei vorangegangenen Wellen miteinander und stellt ihre Krönung dar. Sie verknüpft die Fähigkeit von Maschinen, Optimierungen anhand äußerst komplexer Datensätze vorzunehmen, mit den neuen sensorischen Kräften dieser Maschinen. Wenn sich diese übermenschlichen Fähigkeiten verbinden, entstehen Maschinen, die ihre Umwelt nicht nur verstehen, sondern sie auch beeinflussen." (Lee, 2019, S. 172)

In dieser Welle kumulieren all jene Fragen, die in dieser Arbeit zum Gegenstand der Untersuchung gemacht werden. Die sich nicht nur anbahnende, sondern bereits sich aufbauende vierte Welle der Autonomen-KI verpflichtet die Gesellschaft geradezu, aus Blickwinkeln verschiedenster wissenschaftlicher Forschungsgebiete dieses Thema zu betrachten.

Ein interdisziplinärer Forschungsansatz ist dabei notwendig, um dem Gegenstand von KI in seinen unterschiedlichsten Wirklichkeiten näher zu kommen. Dabei kann es nicht bei einem rein technischen Blick darauf bleiben. Auf die Unmöglichkeit der Technikneutralität bei von Menschen geschaffenen Artefakten und implementierten Techniken wird im Weiteren noch genauer eingegangen.

Die Auseinandersetzung mit dem Thema KI führt zur unausweichlichen Notwendigkeit, sich mit jenen ethischen und moralischen Fragen zu beschäftigen, welche der Mensch seit jeher in der Hoffnung und dem Ziel, sich seinem menschlichen Selbstverständnis anzunähern, zu beantworten sucht. Zu untersuchen sind dabei die Implikationen von KI nicht nur auf das einzelne Individuum, sondern ebenso auf allgemeine gesellschaftliche Bereiche. In den sich daraus entwickelnden Diskursen versuchen verschiedenste wissenschaftliche Disziplinen wie Philosophie, Psychologie,

Neurowissenschaften, Informatik oder Technikphilosophie Antworten auf diese drängenden Fragen.

Weiterhin unklar und nicht erforscht scheint jedoch die Auswirkung von KI auf jenen von Sartre postulierten Freiheitsbegriff zu sein, dass nämlich der Mensch dazu verurteilt ist, frei zu sein. So ist für Sartre bereits der Akt des *sich Seiner selbst bewusst zu sein* ein Akt der Freiheit.

Der Zweck dieser Studie war, die Auswirkungen der Wirkmächtigkeit und Kompetenz von KI sowie die einer künftig möglichen, mit Bewusstsein ausgestatteten KI, und die dadurch entstehenden Implikationen und Spannungsfelder auf Mensch und Gesellschaft zu untersuchen.

Mit der Untersuchung der Wechselwirkungen des Freiheitsbegriffes des Menschen von Sartre und der Wirkmächtigkeit und Kompetenz von KI auf das menschliche Selbstverständnis, wird eine Forschungslücke geschlossen.

1.3 Wissenschaftliche Fragestellung

Die wissenschaftliche Fragestellung dieser Masterarbeit lautet: Wie wirken sich die Kompetenzen und Fähigkeiten von Künstlicher Intelligenz auf die Gültigkeit des von Jean-Paul Sartre postulierten Freiheitsbegriff des Menschen aus?

1.4 Thesis Struktur

Die Struktur der Thesis ist in 5 Hauptkapitel, wie in Abbildung 1 dargestellt, gegliedert.

- **Kapitel 2-3:** Diese beiden Kapitel behandeln in Bezug auf den Stand des Wissens sowie auf den Grundlagen und dem theoretischem Hintergrund die Fragen, wo Künstliche Intelligenz (KI) auf Sartre und seinen Freiheitsbegriff trifft. Als ein zentrales und gemeinsames Phänomen der Untersuchungsgegenstände von KI und dem Freiheitsbegriff von Sartre stellte sich das Bewusstsein heraus.
- **Kapitel 4:** Anschließend stellt Kapitel 4 den methodischen Aufbau der Arbeit sowie die methodologische Positionierung dar.
- **Kapitel 5:** Darauf aufbauend werden in Kapitel 5 durch das Extrahieren der Kernaussagen aus den Expert*inneninterviews diese in Verbindung gebracht sowie deren Korrelation gewichtet.
- **Kapitel 6:** Im abschließenden Kapitel 6 werden die Erkenntnisse diskutiert und interpretiert, um die Forschungsfrage zu beantworten.

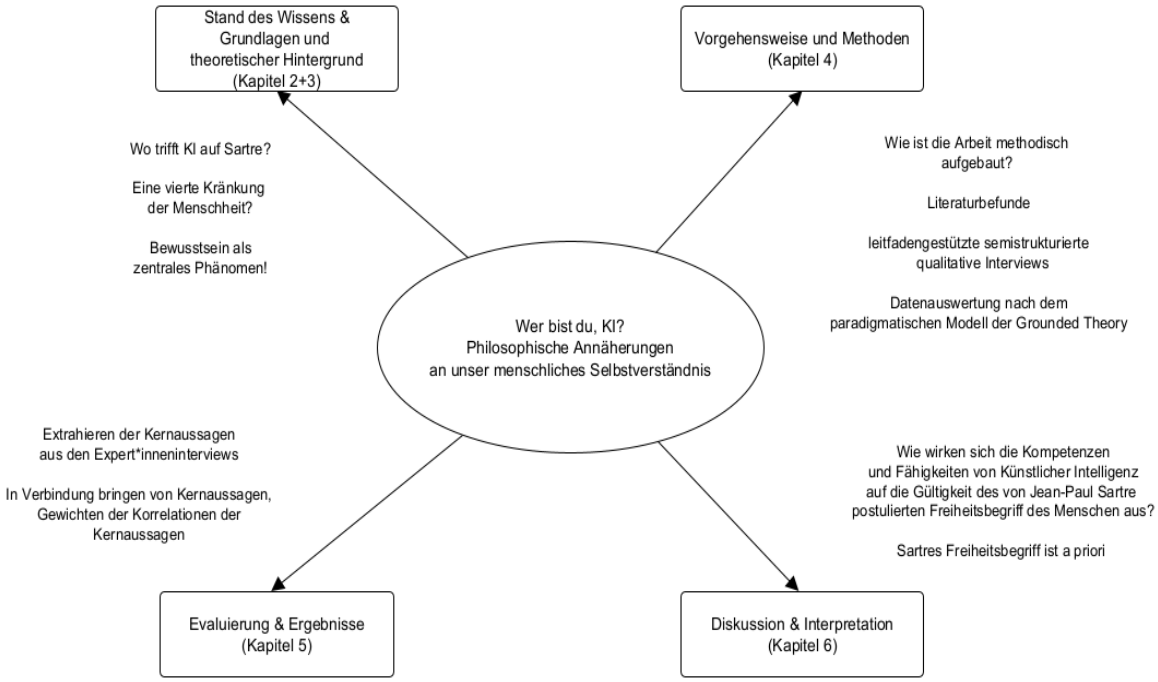

Abbildung 1: Thesis Struktur der Masterarbeit

2 Stand des Wissens

"Alles Interesse meiner Vernunft (das spekulative sowohl, als auch das praktische) vereinigt sich in folgenden drei Fragen: Was kann ich wissen? Was soll ich tun? Was darf ich hoffen?" (Kant, 2014, S. 245). Kant (1968) ergänzt in seinen Vorlesungen über die Logik diese drei Grundfragen der Philosophie um eine vierte Frage "Was ist der Mensch?" (S. 21).

Die Wahl des Interrogativpronomens *Wer* im Titel dieser Masterarbeit „Wer bist du, KI?“ ist bewusst gewählt, um die Frage nach dem *Subjekt KI* zu stellen. Was KI ist, beziehungsweise wie KI funktioniert, wird in dieser Arbeit ebenfalls untersucht, dient jedoch lediglich ergänzend und erläuternd der eigentlichen Frage und dem zentralen Interesse der Arbeit nach der Möglichkeit einer mit Bewusstsein ausgestatteten KI.

2.1 Wer bist du, KI?

Was bedeutet es, wenn eine mit Bewusstsein ausgestatte KI in jene ontologischen Bereiche, die bisher dem Menschen vorbehalten waren, vordringen kann? Bislang blieb es dem Menschen und seinem Selbstverständnis von sich und der Welt vorbehalten, Zuschreibungen von Phänomenen wie Vernunft, Geist, Freiheit, Würde, Intelligenz, Kultur, Ethik oder Moral für sich zu beanspruchen. Die rasanten Entwicklungen von KI haben in wachsendem Ausmaß Auswirkungen auf unser Leben und unseren Alltag.

Wer also bist du, KI? Angelehnt an Immanuel Kants vierte Frage *Wer bist du Mensch?* ergeben sich in der Auseinandersetzung mit KI eine Vielzahl ethischer Fragen, die es zu bedenken gilt. So ist der Umstand, dass es keine Technikneutralität gibt, wie es Loh (2019a) beschreibt, mit ein Grund dafür, dass wir sowohl als Gesellschaft als aber auch als Individuen nicht um den Versuch einer Beantwortung dieser Fragen umherkommen (S. 205).

"Die von Menschen geschaffenen Artefakte und implementierten Techniken sind immer schon und unweigerlich normativ beziehungsweise evaluativ, denn in jede Technologie, in jede Technik, in jede technische Ausdrucksweise gehen die Normen und Werte ihrer menschlichen Schöpfer*innen ein," führt Loh (2019a, S. 205) in ihren abschließenden Bemerkungen eines Plädoyers für einen inklusiven und kritischen Diskurs aus.

Technik, so Hengstschläger (2020), wird immer als ein Mittel zur Erreichung von Zielen oder auch Zwecken eingesetzt. Implizit damit verbunden stellt sich die Frage nach Werten, die hinter diesen Zielen und Zwecken stehen. Antworten auf diese wichtigen Fragen finden sich in der Philosophie, welche sich schon immer mit den grundlegenden Fragen des Lebens auseinandergesetzt hat. So versucht zum Beispiel Technikphilosophie ethische Fragen auch im Zusammenhang mit KI zu beantworten und die Welt und die menschliche Existenz zu ergründen, zu deuten und zu verstehen (S. 9).

Unterschiedlichste Positionen wie jene von Steven Hawking und Gotthard Günther stehen sich gegenüber. So führt Hawking (2019) aus, dass das Aufkommen superintelligenter KI entweder das Beste oder das Schlimmste wäre, was der Menschheit passieren kann. Dabei sieht Hawking das eigentliche Risiko bei KI nicht in deren Bosheit, sondern in deren Kompetenz (S. 73). Günther (1956) argumentiert dagegen, dass Kritiker, die beklagen, dass die Maschine uns unsere Seele raubt, im Irrtum seien. Denn, das Subjekt, also der Mensch, wird, egal wieviel er von seiner Reflexion auch an den Mechanismus abgibt, durch unerschöpfliche Innerlichkeit neuer Kräfte immer nur reicher (S. 16).

Hawking (2019) sieht eine Entwicklung von KI dahingehend, dass wenn KI bei der Konstruktion von KI besser wird als Menschen und sie sich somit rekursiv und ohne menschliches Zutun selbst verbessern kann, uns eine Intelligenzexplosion bevorsteht, die in einer Maschinenintelligenz münden wird. Dabei geht er davon aus, dass wenn Computer ihre Geschwindigkeit und Speicherkapazität gemäß des mooreschen Gesetzes alle 18 Monate verdoppeln, dies zur Folge haben wird, dass in den kommenden 100 Jahren Menschen hinsichtlich der Intelligenz von Computern überholt werden (S. 72).

Das *Mooresche Gesetz* beschreibt die Entwicklung von Computerchips und sagt aus, dass sich alle zwei Jahre die integrierten Schaltungen so verkleinern, dass jeweils doppelt so viele Schaltelemente auf einem Chip untergebracht werden können wie zwei Jahre zuvor. Es wird also ein exponentielles Wachstum der Leistungsfähigkeit von Chips behauptet. Russel (2020) argumentiert, dass das *Mooresche Gesetz* nur bis ungefähr 2025 Bestand haben wird, da seit einigen Jahren durch die Hitzeentwicklung in den Siliziumtransistoren die Rechengeschwindigkeit ausgebremst wird (S. 43).

"Nach 2025 lässt sich das mooresche Gesetz (oder sein Nachfolger) nur noch aufrechterhalten, wenn wir auf exotischere physikalische Phänomene zurückgreifen, darunter negative Kapazität, Transistoren aus einem Atom, Graphen-Nanoröhren und Photonik" (Russell, 2020, S. 43).

Wie weit heute schon Maschinenintelligenz fortgeschritten ist und was dies konkret bedeutet, lässt sich am Beispiel des KI-Systems AlphaGo von DeepMind veranschaulichen (Tegmark, 2020). Dabei spielte das KI-System im März 2016 gegen den weltbesten Spieler des frühen 21. Jahrhunderts, Lee Sedol, das Jahrtausende alte Brettspiel *Go* und schlug diesen in vier von fünf Spielen. Da es bei diesem Brettspiel mehr mögliche Go-Positionen als es Atome im Universum gibt, bedarf es beim Spielen neben intellektuellen Fähigkeiten auch der Intuition, Kreativität und strategischen Denkvermögens. Diese Eigenschaften schrieb der Mensch bisher eher sich als Maschinen zu (S. 132).

Im zweiten Spiele schockierte AlphaGo im 37. Spielzug die Fachwelt (Herger, 2020). Dabei setzte AlphaGo einen Stein auf eine Position, die nach menschlicher Logik zu vermeiden und sogar ein Fehler gewesen wäre. Die Chance, nach

diesem Zug zu gewinnen, war für das KI-System aus statistischer Sicht 1:10000 (S. 9).

AlphaGo setzte dabei aber auf eine langfristige Planung seiner Spielstrategie und bevorzugte somit den strategischen Vorteil gegenüber kurzfristigem Gewinn. Dass sich dieser Spielzug 50 Züge später als der entscheidende Spielzug erwies und AlphaGo zum Sieg verhalf, kann als einer der kreativsten Züge, die es in der Geschichte dieses Spieles je gegeben hat, beschrieben werden (Tegmark, 2020, S. 134). Den Sieg von AlphaGo bezeichnet Tegmark als einen Wendepunkt in der Entwicklung von KI und zeigt, wie schwierig es sein kann, Durchbrüche in der KI kommen zu sehen (S. 131).

Die Forscher*innen von DeepMind nahmen Ende 2017 den Nachfolger von AlphaGo, AlphaZero in Betrieb. Im Unterschied zu AlphaGo lernte AlphaZero das Spiel von Grund auf neu, indem es gegen sich selbst spielte (Tegmark, 2020, S. 135). Herger (2020) führt aus, dass AlphaZero nicht mit Bibliotheken an Spielen gefüttert wurde, welche die Menschen in der Vergangenheit spielten, sondern dem System nur die Spielregeln eingegeben wurden (S. 11).

AlphaZero schlug dabei nicht nur AlphaGo, sondern "auch die menschlichen KI-Programmierer und damit die ganze handgemachte KI-Software, die sie im Laufe vieler Jahrzehnte entwickelt hatten. Mit anderen Worten: Die Vorstellung, dass KI bessere KI erschafft, lässt sich nicht von der Hand weisen" (Tegmark, 2020, S. 135).

2.1.1 Mensch – Maschine Interaktion

Die Zukunft der Kommunikation liege in Gehirn-Computer-Schnittstellen durch auf dem Kopf angebrachte Elektroden oder Implantate. Bei ersterer Möglichkeit hat man den Eindruck "durch zugefrorene Glasscheiben zu schauen; die zweite Möglichkeit ist besser, birgt allerdings das Risiko von Infektionen" (Hawking, 2019, S. 75). Am Beispiel wissenschaftlicher Forschung an einer KI, mit Hilfe derer kabellos implantierte Chips aus Silikon als elektronische Schnittstellen zwischen Gehirn und Körper die Lähmung bei Menschen mit Rückenmarksverletzung aufheben würde, indem Körperbewegungen mit Gedanken gesteuert werden könnten, beschreibt Hawking die Entwicklungen in der Mensch-Maschine Interaktion (S. 75).

Von einer *invasiven Technisierung* spricht Böhme (2008). Dabei wird das Beispiel eines Patienten, der sich aufgrund einer Parkinson-Erkrankung ein technisches Hirnimplantat setzen ließ, geschildert. Dieses Implantat bewirkt durch seine Stimulation im Gehirn einerseits einen Rückgang der Parkinson-Symptome, andererseits beeinflusst es negativ das Sprachvermögen des Patienten. Durch einen Sender in der Hand kann nun dieser Patient das Hirnimplantat so steuern, dass es je nach Wunsch entweder aktiv ist und so die Parkinson-Symptome mildert, es in dieser Zeit aber das Sprachvermögen vermindert, oder aber inaktiv

ist und so genau den umgekehrten Effekt nach sich zieht. Wenn ein Mensch sich nun an- und abschalten kann, stellt sich die grundsätzliche Frage, was dieses *Sich* noch für eine Bedeutung haben kann (S. 12).

Mit fortschreitender Computertechnik vergrößern sich die technologischen Möglichkeiten in der Interaktion von Mensch und Technik. Maßgebliche Faktoren sind die immer größer werdende Geschwindigkeit, mit der Daten übertragen werden können, sowie die immer größer werdende Datenmenge, die gespeichert werden kann und zur Verfügung steht. Zudem schreitet die Entwicklung neuartiger und kostengünstiger Sensorik mit neuen Funktionalitäten, welche technischen Systemen zur Verfügung stehen können, voran (Yurish, 2010, S. 3).

In einer Gehirn-Maschinen-Schnittstelle, auch *Brain-Computer-Interface (BCI)*, liegt jedoch eine viel größere Gefahr als in jener des Überwachungskapitalismus, nämlich die Gefahr der Beherrschung des Menschen. All die Begriffe der implementierbaren Gehirn-Computer-Schnittstellen wie *BCIs*, die auch *Mind-Machine Interfaces (MMIs)* oder *Direkte Neurale Interfaces (DNIs*) genannt werden, "deuten auf dieselbe Idee eines direkten Kommunikationsweges hin, zuerst zwischen einem weiterentwickelten oder verdrahteten Gehirn und einem externen Gerät und dann direkt zwischen Gehirnen" (Žižek, 2020, S. 51).

Dabei stellt Žižek (2020) die Frage, ob sich eine *BCI* nicht als das ideale Medium zur zum Beispiel politischen Kontrolle des Innenlebens von Individuen anbietet. Befürworter versuchen dabei, wie bei allen Erfindungen, die eine Bedrohung der menschlichen Freiheit darstellen, das Problem zu verschleiern, wenn sie Beispiele für diese Technologien anführen, die das Leben von Menschen mit Behinderungen erleichtern können. Dass aber gedankengesteuerte Maschinen natürlich auch im Umkehrschluss die maschinelle Steuerung der Gedanken selbst bedeuten, bleibt unerwähnt (S. 69).

Historisch betrachtet ging die Entwicklung in den 1950er Jahren über das Erfassen von Eingabedaten durch technische Systeme und die anschließende Darstellung dieser Daten sowie deren Umsetzung in mechanische Aktionen vor sich (Parasuraman & Wickens, 2008, S. 511). Gesteigerte Systemkomplexität ermöglicht heute ein immer zahlreicheres Lösen und Erledigen von Teilaufgaben durch Maschinen, da entsprechende moderne Mensch-Maschine Schnittstellen mehr und multiple Modalitäten nutzen. Verschiedenste Modalitäten wie zum Beispiel Gesichtsausdrücke, Körperbewegungen, Sprach- oder Fingergesten können technisch erkannt und mit intelligenten Algorithmen verarbeitet werden (Turk, 2014, S. 192).

Der Versuch einer Annäherung der beiden Denk-Systeme *KI* und *Mensch* lässt sich mit einem Gedankenexperiment von Loh (Bauder, 2021) veranschaulichen. Darin geht sie davon aus, dass man einen Menschen, der eine Arm- oder Beinprothese oder ein künstliches Herz hat, wohl sicher als einen Menschen bezeichnen würde. Loh fordert auf sich nun vorzustellen, dass weitere

Körperteile durch Prothesen und in weiterer Folge nach und nach mehr innere Organe durch künstliche Organe ersetzt würden, um dann zu fragen, ob es eine Schwelle gäbe, ab der nicht mehr von einem Menschen gesprochen werden könnte (S. 20).

Nach Spiekermann (Hengstschläger & Rat für Forschung und Technologieentwicklung, 2020) stellt sich aus kritischer ethischer Perspektive die Frage, ob der Vergleich von Menschen mit KI-Systemen nicht einer Diffamierung gleichkäme. "Die von Marketing und Hypes geprägte Technikwelt setzt sich zu wenig achtsam mit der angestammten Bedeutung solcher Begriffe auseinander und begibt sich damit auf eine fragwürdige Gratwanderung" (S. 120-121).

Die große Bandbreite der Charakteristika von realistischen und unrealistischen KI-Systemen wird in Abbildung 2 *Charakteristika von realistischen und unrealistischen KI-Systemen* dargestellt.

	Menschenähnliche Sci-Fi-KI	Virtuelles KI-System	Physisches KI-System
Kognitive Funktionen	wahrnehmen (Informationssammlung- und integration), planen (z.B. m.H.v. Musteranalyse), schlussfolgen (z.B. basierend auf Mustern), kommunizieren (z.B. text to voice), entscheiden/wählen (z.B. basierend auf prediction)		
Körper	Menschenähnlicher Körper	Software (inkl. Datenbanken)	Software (inkl. Datenbanken) Hardware (inkl. Aktuatoren)
Information	Unbereinigte Daten	Unbereinigte Daten Bereinigte Daten (strukturiert, kontinuierliche/ diskrete)	Unbereinigte Daten Bereinigte (struktu-riert, kontinuierlich/ diskrete)
Machine Learning (ML)	Unüberwachtes ML Minimale Interventi-on von Menschen im Life Cyle	Unüberwachtes ML Überwachtes ML am Anfang dann eingefroren Überwachtes ML während des Life Cycles	Unüberwachtes ML Überwachtes ML am Anfang dann eingefroren Überwachtes ML während des Life Cycles
Ziele	Ziel beeinflusst von Menschen Ziel selbst gegeben (nicht zufällig) Ziel jenseits des vorspezifizierten Einsatzbereiches	Ziel gegeben vom Mensch Ziel selbst gegeben (auch m.H.v. Zufall-salgorithmen) Ziel jenseits des vorspezifizierten Einsatzbereichs	Ziel gegeben vom Menschen
Autonomie (nach Aktivierung)	Agiert ohne Intervention von Menschen	Agiert ohne Intervention von Menschen Agiert mit Intervention von Menschen	Agiert ohne Intervention von Menschen Agiert mit Intervention von Menschen

Abbildung 2: Charakteristika von realistischen und unrealistischen KI-Systemen, (Hengstschläger & Rat für Forschung und Technologieentwicklung, 2020, S. 121)

Auf der Suche nach Künstlicher Intelligenz war der Mensch jedoch schon seit den Anfängen der Zivilisation und lange bevor es die Begriffe *KI* oder *Informatik* gab. Zeugnisse davon finden sich in dem chinesischen Volksmärchen von Yan Shi, einem Handwerker, der einen humanoiden Roboter baut oder in der griechischen Mythologie, in der uns *Talos*, der bronzene Automat begegnet. Bis zum heutigen Tag hat diese Kraft der Künstlichen Intelligenz jegliche Dimension der menschlichen Zivilisation revolutioniert (Lee & Chen, 2022, S. 20).

2.1.2 Die drei Kränkungen der Menschheit

Sigmund Freud konstatiert dem Menschen im Laufe seiner Entwicklung drei schwere Kränkungen. Er nennt diese die kosmologische, die biologische und die psychologische Kränkung. Dass die Erde, auf der er lebt, nicht der Mittelpunkt des Universums und er, der Mensch selbst nicht dessen Herrscher ist, musste der Mensch spätestens seit Nikolaus Kopernikus erkennen. Nach dieser ersten Kränkung erfuhr der Mensch die zweite Kränkung durch Charles Darwin, der unmissverständlich klarmachte, dass der Mensch ein Produkt der biologischen Evolution und nicht Herrscher über seine Mitgeschöpfe ist. Die dritte Kränkung schließlich, dass der Mensch nicht mehr Herr in seinem eigenen Hause ist, hat Freud selbst dem Menschen zugemutet (Wildenburg, 2004, S. 64). Denn nichts Geringeres als des Menschen Souveränität über sich selbst wird damit in Frage gestellt (Freud, 1917):

> "Am empfindlichsten trifft wohl die dritte Kränkung, die psychologischer Natur ist. Der Mensch, ob auch draußen erniedrigt, fühlt sich souverän in seiner eigenen Seele. Irgendwo im Kern seines Ichs hat er sich ein Aufsichtsorgan geschaffen, welches seine eigenen Regungen und Handlungen überwacht, ob sie mit seinen Anforderungen zusammenstimmen. Tun sie das nicht, so werden sie unerbittlich gehemmt und zurückgezogen. Seine innere Wahrnehmung, das Bewußtsein, gibt dem Ich Kunde von allen bedeutungsvollen Vorgängen im seelischen Getriebe, und der durch diese Nachrichten gelenkte Wille führt aus, was das Ich anordnet, ändert ab, was sich selbständig vollziehen möchte." (S. 4-5)

Damit bringt Freud zum Ausdruck, dass die unterdrückten Sexualtriebe und sein Unbewusstes dem Menschen die Herrschaft und somit seine Autonomie über sich selbst streitig machen (Wildenburg, 2004, S. 64).

2.2 Im Zeitalter der vierten Kränkung?

Die durch Freud beschriebene dritte Kränkung des Menschen, in der er den Verlust der Souveränität seiner Seele und damit den Verlust seiner Autonomie hinnehmen muss, bezieht sich aus psychologischer Sicht auf das Unterbewusste, auf die Triebhaftigkeit des Menschen.

Betrachtet man nun die Wirkmächtigkeit von KI sowie deren Fähigkeiten und Kompetenzen, so deutet schon der Begriff *Intelligenz* darauf hin, dass es sich bei KI um ein Phänomen handelt, dass dem Menschen in seinem Selbstverständnis von sich nahekommt. Dieses Selbstverständnis äußert sich durch die dem Menschen sich selbst zugeschriebenen Eigenschaften als ein vernunftbegabtes Wesen, welches über Bewusstsein und Selbstbewusstsein verfügt. Dieses Selbstbewusstsein ist auch die Voraussetzung für die von Sartre postulierte Freiheitsidee des Menschen.

2.2.1 KI trifft auf Sartre

Wie im Kapitel 2.7 beschrieben, tritt nach Sartre im Menschen durch das Selbstbewusstsein die Freiheit auf. Und diese Freiheit zwingt den Menschen, zwingt die menschliche Realität, *sich zu machen* statt *zu sein*.

Wenn nun aber namhafte Wissenschaftler*innen über mit Bewusstsein ausgestattete Superintelligenzen und Maschinen sprechen und diese auch für möglich halten (Bostrom, 2020, S. 41), stellt sich die Frage, ob sich daraus auch schlussfolgern ließe, dass auch bei KI-Systemen die Freiheit im Sinne des von Sartre postulierten Freiheitsbegriffes auftritt. Aus dieser Überlegung heraus folgt unweigerlich der Schluss, dass Freiheit demnach auch KI-Systeme zwingt, *sich zu machen* statt zu *sein*.

Den Begriff der *Machbarkeit* verwendet auch Hawking (2019), wenn er argumentiert, dass, wenn KI bei der Konstruktion von KI besser wird als Menschen, sie sich somit rekursiv und ohne menschliches Zutun selbst verbessern kann (S. 72). Dass die nächsten Jahrzehnte alle Errungenschaften, die wir bis heute kennengelernt haben mit Sicherheit in den Schatten stellen werden und wir überhaupt nicht voraussagen können "was wir zu leisten vermögen, wenn unser Geist durch KI erweitert wird," davon geht Hawking (2019, S. 75) in seinen Überlegungen aus.

Könnte man daher von der vierten Kränkung des Menschen sprechen, wenn ihm sein Alleinstellungsmerkmal in der Welt, nämlich das eines freien, autonomen und mit Bewusstsein ausgestatteten Wesens von einer mit Bewusstsein ausgestatteten KI strittig gemacht würde?

2.2.2 Bewusstsein als zentrales Phänomen

Ob es sich nun um die von Hawking beschriebene Erweiterung des menschlichen Geistes, um die von Wissenschaftler*innen für möglich gehaltene Entwicklung von Superintelligenzen oder um den Kern von Sartres Philosophie handelt – ein zentrales Phänomen findet sich in all diesen Überlegungen: *Bewusstsein*.

Damasio (2011) beschreibt das Bewusstsein als einen Geisteszustand "in dem man Kenntnis von der eigenen Existenz und der Existenz einer Umgebung hat. Bewusstsein ist ein *Zustand des Geistes* – ohne Geist gibt es auch kein Bewusstsein"

(S. 169). Nur aus der exklusiven und unmittelbaren Perspektive unseres eigenen Organismus können wir diesen bewussten Geisteszustand erleben. Somit kann dieser auch nie von irgendjemand anderem beobachtet werden (S. 169).

Ein bewusster Geist entsteht dann "wenn zu den grundlegenden geistigen Vorgängen ein Selbst-Prozess hinzukommt" (Damasio, 2011, S. 19). Zu berücksichtigen ist jedoch, dass unsere Handlungen als Menschen in vielen Fällen mit unbewussten Prozessen gesteuert werden. Diese Mitwirkung des Unbewussten an menschlichen Tätigkeiten kann jedoch auch falsch interpretiert werden und "den Wert einer vom Selbst gelenkten, bewussten Kontrolle herunterzuspielen" (S. 283).

Dass es für den Begriff *Bewusstsein* ähnlich wie für die Begriffe *Leben* oder *Intelligenz* keine eindeutig korrekte Definition gibt, stellt Tegmark (2020) fest und definiert *Bewusstsein* als ein subjektives Erlebnis (S. 422). Um auf die Frage, was ein künftiges KI-System subjektiv erleben wird und wie sich dies anfühlen könnte eine Antwort zu finden, "fehlt uns nicht nur eine Theorie, die diese Frage beantwortet, sondern wir sind noch nicht einmal sicher, ob es logisch überhaupt möglich ist, sie vollständig zu beantworten" (S. 457-458).

Trotz dieser Unfähigkeit, eine vollständige Antwort auf diese Frage zu geben, können wir trotzdem Teilantworten geben, denn wenn ein intelligenter Außerirdischer das menschliche Sinnessystem studieren würde, dann könnte er zu der Schlussfolgerung kommen, "dass Farben Qualia sind, die sich mit jedem Punkt auf einer zweidimensionalen Oberfläche (unserem Gesichtsfeld) verbunden anfühlen, während Klänge sich räumlich nicht begrenzt anfühlen und Schmerzen Qualia sind, die mit unterschiedlichen Teilen unseres Körpers zu tun haben" (Tegmark, 2020, S. 458).

Der Umstand, dass sich elektromagnetische Signale mit Lichtgeschwindigkeit und damit einige Millionen Mal schneller als Neuronensignale fortpflanzen, ist eine weitere Teilantwort darauf, dass ein künstliches Bewusstsein von der Größe eines menschlichen Gehirnes Millionen Mal mehr Erlebnisse pro Sekunde hätte als wir (Tegmark, 2020, S. 459). Sollte also ein künstliches Bewusstsein möglich sein, so würde der Raum möglichen KI-Erlebens im Vergleich zu dem was wir Menschen erleben können, riesig sein (S. 468-469).

"Da es ohne Bewusstsein keinen Sinn geben kann, ist es nicht unser Universum, das bewussten Wesen Sinn verleiht, sondern es sind die bewussten Wesen, die unserem Universum Sinn verleihen" (Tegmark, 2020, S. 469). "Während wir Menschen uns darauf vorbereiten, von noch schlaueren Maschinen gedemütigt zu werden, weist manches darauf hin, dass wir uns damit trösten, hauptsächlich *Homo sentiens* statt *Homo sapiens* zu sein (S. 469).

Unter dem Einfluss der zunehmenden Kooperation von menschlicher Intelligenz und Künstlicher Intelligenz können sich Veränderungen unserer alltäglichen Wahrnehmung der soziokulturellen Wert- und Lebenswelt entwickeln. Dazu zählen das Verantwortungsbewusstsein, das Sozialverhalten sowie die

Wertbeurteilung. Die Frage nach dem Ort und dem Halt der das Zusammenwirken leitenden Normativität stellt sich dabei ebenso wie die Frage "nach der Wertorientierung menschlichen Verhaltens und des selbst- und fremdreflexiven Verhältnisses des Menschen als Subjekt und Objekt des Zusammenwirkens von Mensch und Technik unter dem Diktat von Technisierung, Medialisierung und Digitalisierung unserer Lebenswelt" (Haux et al., 2021, S. 241).

Als ein Schlüsselbegriff wird dabei die *Verantwortung* definiert, hinter deren Verantwortungskonzeptionen ein von Bewusstseinsfähigkeit, von Autonomie und Entscheidungsfreiheit geprägtes Menschenbild stehen. Dabei stellt sich die Frage, ob KI zu Bewusstsein, wertebestimmter Reflexion und Selbstgesetzgebung fähig sei. Denn "Roboter seien ebenso wie künstliche neuronale Netzwerke Imitate menschlicher Intelligenz; menschliche Gehirnleistung sei mit einer reichen Gefühlswelt, mit rationalem und intersubjektivem Problembewusstsein verbunden, die in verantwortliche Entscheidungen eingehen" (Haux et al., 2021, S. 242).

2.2.3 Das Libet-Experiment

In einem Experiment von Libet (1983) wird der Zeitpunkt der bewussten Handlungsabsicht im Verhältnis zum Beginn der Hirnaktivität untersucht. Dabei wird "die registrierbare zerebrale Aktivität (readiness-potential, RP) die einer freiwilligen, vollständig endogenen motorischen Handlung vorausgeht, direkt mit der berichtbaren Zeit für das Auftreten einer subjektiven Erfahrung des "Wollens" oder der Absicht zu handeln, verglichen" (S. 623).

In dem Experiment wurde festgestellt, dass das Einsetzen der zerebralen Aktivität dem berichteten Zeitpunkt der bewussten Handlungsabsicht eindeutig um mindestens einige hundert Millisekunden vorausging. Die Schlussfolgerung, dass die zerebrale Initiierung einer spontanen und freiwilligen Handlung unbewusst beginnen kann, bevor es noch ein subjektives und erinnerbares Bewusstsein gibt, dass eine Entscheidung zur Handlung bereits zerebral initiiert wurde, führte zu der Interpretation, dass sich gewisse Einschränkungen für die Möglichkeit der bewussten Initiierung und Kontrolle von freiwilligen Handlungen ergeben (Libet, 1983, S. 623).

Entschieden gegen diese Interpretation des Experimentes von Libet, dass das Gehirn und nicht das subjektive Bewusstsein entscheidet, argumentiert Vogd (2023), in dem er festhält, dass die Willensfreiheit nicht im Gehirn und nicht im subjektiven Bewusstsein, sondern ein emergentes Phänomen der Gesamtsituation ist (15:47–15:57). "Die Proband*innen dieses Experimentes sind demnach in einer sozialen Situation, in der der Versuchsleiter vorher die Anweisungen gibt, einen freien Willen zu haben. Die Aufforderung jedoch, einen freien Willen zu haben ist jedoch genau das, was Watzlawick als die *Sei-spontan-*

Paradoxie beschrieben hat. Und diese Paradoxie kann nicht gelöst werden und es entsteht eine Unbestimmtheit" (16:14–16:35).

Bezogen auf das Libet-Experiment spürt man daher den Druck etwas zu tun und gibt der Empfindung den Sinn, dass es meine freie Entscheidung ist, denn "über die soziale Situation muss ich der Bewegung einen Sinn geben, dann kommt Bewusstsein rein und dann entsteht ein Sinn der vorher gar nicht da war. Also das Bewusstsein schiebt immer im Nachhinein einen sozialen Sinn rein" (Vogd, 2023, 18:01-18:15).

2.3 Intelligenz

Historisch betrachtet bezog sich der Begriff *Intelligenz* und dessen Bedeutung auf den Menschen. Den Begriff *Intelligenz* auf andere Lebewesen wie Tiere oder Pflanzen zu übertragen, war lange Zeit undenkbar. So löste Francis Darwin, der Sohn von Charles Darwin und weltweit einer der ersten Lehrstuhlinhaber für Pflanzenphysiologie, am 2. September 1908 bei seiner Rede des Jahreskongresses der *British Association for the Advancement of Science* mit seiner Behauptung, dass Pflanzen intelligente Wesen seien, einen Sturm der Entrüstung aus. Denn Pflanzen und Physiologie galten am Ende des 19. Jahrhunderts noch als unvereinbare Konzepte (Mancuso et al., 2015, S. 25-26).

2.3.1 Menschliche Intelligenz

William Stern (1912) definiert Intelligenz als die allgemeine Fähigkeit eines Individuums, sich bewusst auf neue Herausforderungen einzustellen und sich geistig an neue Aufgaben und Lebensbedingungen anzupassen. Somit unterscheidet diese Definition Intelligenz von anderen geistigen Fähigkeiten (S. 3).

Die Fähigkeit, sich auf das Neue einzustellen, trennt Intelligenz vom Gedächtnis, dessen Zweck die Bewahrung und Verwendung von bereits vorhandenen Bewusstseinsinhalten ist. Die Anpassungsfähigkeit hebt die Abhängigkeit der Intelligenz von äußeren Bedingungen und Anforderungen hervor und unterscheidet sie von der Genialität, die sich auf spontane Neuschöpfung konzentriert. Schließlich unterscheidet sich Intelligenz von Talent durch ihre *Allgemeinheit der Fähigkeit*, da Talent auf ein bestimmtes inhaltliches Gebiet beschränkt ist, während Intelligenz es einem ermöglicht, sich geistig an neuen Anforderungen auf verschiedenen Gebieten anzupassen (Stern, 1912, S. 4).

2.3.2 Tierische und pflanzliche Intelligenz

Was denn überhaupt Intelligenz sei, fragt Mancuso (2015), um gleich darauf zu antworten, dass es wohl so viele Definitionen von Intelligenz gibt wie es Forscher gibt, die man danach fragt. Dass nämlich Pflanzen ebenso wie Tiere Probleme lösen können oder aus einem Labyrinth herausfinden, Hindernisse umrunden, sich mit komplexen Strategien gegen Räuber verteidigen, ist eine Tatsache (S. 124-125).

Intelligenz sei etwas, das allem Leben innewohnt. Denn wo ist die Schwelle bei Primaten, Hunden, Katzen, Mäusen, Ameisen, Kraken bis hin zu Amöben, ab der man plötzlich von Intelligenz sprechen müsste? Würde es eine Schwelle geben, dann müsste man sich fragen, ob diese biologisch determiniert oder vielmehr kulturell bedingt ist. So hielt im 19. Jahrhundert kaum wer Tiere oder gar Pflanzen für intelligent (Mancuso et al., 2015, S. 125-126).

In Charles Darwins Werk *The Power of Movements*, ein Werk mit bahnbrechender Wirkung auf die Geschichte der Botanik, erläutert Charles Darwin die Schlussfolgerungen aus seinen Versuchen, dass die mit einem Empfindungsvermögen ausgerüstete Spitze einer Wurzel einer Pflanze mit dem Gehirn eines Tieres zu vergleichen ist (Mancuso et al., 2015, S. 127-128).

2.3.3 Künstliche Intelligenz

Da die Eingrenzung eines Begriffes und dessen Bedeutung nur zu einem bestimmten Zweck möglich ist (Wittgenstein, 2022, § 43, S. 262-263), so begegnen uns im sich wechselseitig bedingenden Wortpaar *Künstliche Intelligenz* zwei Begriffe, deren Bedeutung es einzugrenzen gilt. Dabei bedarf es sowohl eines Blickes auf das dem Subjekt *Intelligenz* vorgesetzte Adjektiv *künstlich*, als auch eines Blickes auf das Adjektiv *intelligent*.

Wie kaum eine andere Technik rückt Künstliche Intelligenz schon allein mit ihrem Namen dem Menschen sehr nahe. Vor allem auch deswegen, weil Intelligenz, auch wenn sie schwer zu definieren ist, als ein Wesensmerkmal des Menschen gilt. Grunwald (2019) definiert trotz der Vielzahl von möglichen Definitionen, was Intelligenz sei, diese so: "Egal wie die Intelligenz im Detail nun definiert wird, der Begriff ist jedenfalls eng mit unserem Bild von uns selbst verbunden" (S. 45).

Das Attribut *künstlich* ruft in uns Assoziationen wie Angst vor intelligenten Robotermenschen hervor und weckt Bilder von Science-Fiction-Romanen. Dabei stellt sich die Frage, "ob wir versuchen sollten, unser höchstes Gut, den Geist, zu verstehen, zu modellieren oder gar nachzubauen" (Ertel, 2021, S. 1). Damit wird es schon auf den ersten Blick äußerst schwierig, den Begriff Künstliche Intelligenz einfach und prägnant zu definieren. Als eine elegante Definition hingegen sieht Ertel jene von Rich (Rich, 1983, zitiert in Ertel, 2021): "Künstliche Intelligenz ist die Lehre davon, wie man Computer dazu bringen kann, Dinge zu tun, die Menschen derzeit besser können" (S. 201).

Ohne jedoch ein grundlegendes Verständnis menschlichen Schließens und eines intelligenten Handelns zu haben, können intelligente Systeme nicht gebaut werden, weshalb die Kognitionswissenschaft für die KI eine große Bedeutung hat (Ertel, 2021, S. 3). Jedenfalls muss die Fähigkeit zu lernen ein integraler Bestandteil des Systems einer KI sein und darf nicht nachträglich hinzugefügt werden. Ebenso muss es die Fähigkeit haben, in effektiver Weise mit

Ungewissheit und probabilistischen Informationen umzugehen (Bostrom, 2020, S. 42).

Als künstlich intelligent, so eine weitere Definition von KI, kann ein technisches System dann genannt werden, wenn dieses Leistungen analog zum menschlichen Denken verwirklichen kann (Nilsson, 2010, S. 533).

"KI kann definiert werden als vom Menschen entworfene Software und gegebenfalls auch Hardwaresysteme zur Bearbeitung von Aufgaben, zu deren Lösung bisher menschliche Intelligenz notwendig war. Eigene Entscheidungsfindung und autonomes Handeln in gewissen Grenzen sind Merkmale der KI" (Müller et al., 2021, S. 77).

Die Annahme jedoch, dass es nur zwei Arten von informationsverarbeitenden Systemen gibt, nämlich künstliche und natürliche, ist falsch, da diese begriffliche Unterscheidung zwischen natürlichen und künstlichen weder erschöpfend noch exklusiv ist. Es könnte nämlich intelligente und/oder bewusste Systeme geben, die eben keiner der beiden Kategorien ausschließlich angehören (Metzinger, 2017). So gibt es bereits jetzt Systeme, die zum Beispiel von künstlicher Software kontrolliert werden, jedoch biologische Hardware verwenden. In umgekehrter Weise gibt es auch künstliche Hardware, auf der natürlich evolvierte Software läuft (Metzinger, 2017):

> "Hybride Bioroboter sind ein Beispiel für die erste Kategorie. Die hybride Biorobotik ist eine neue Disziplin, die auf natürlich evolvierte Hardware zurückgreift und sich damit nicht aufhält, etwas noch einmal zu erschaffen, das Mutter Natur bereits über Millionen von Jahre hinweg optimiert hat."(S. 274)

So können zum Beispiel die Körperbewegungen von Kakerlaken kontrolliert werden, indem man Steuerungselemente chirurgisch einpflanzt.

2.4 Künstliche Intelligenz und Autonomie

Wenn *Sophia*, eine Humanoidin der Firma *Hanson Robotics*, ausgestattet mit KI, die saudi-arabische Staatsbürgerschaft erhält, eröffnen sich zwingend grundlegende gesellschaftspolitische und ethische Fragen. Denn Staatsbürgerschaft für eine KI bedeutet unter anderem, mit Rechten und Pflichten dem Menschen gleichgestellt zu sein. Würde, Vernunft, Autonomie, Bewusstsein oder Freiheit – all diese Begriffe drängen unweigerlich in den Mittelpunkt der Betrachtung von KI.

Eine *Humanoidin* oder ein *Humanoide* ist ein Roboter, der dem menschlichen Körper nachempfunden ist. Diese Roboter sind physische Agenten, die die Welt manipulieren, indem sie bestimmte Aufgaben ausführen und erfüllen. Sie sind ausgestattet mit Effektoren wie Rädern, Gelenken oder Greifarmen sowie mit Sensoren wie Kameras oder Laser, mit deren Hilfe sie die Umgebung wahrnehmen (Russell & Norvig, 2012, S. 1120-1121).

Und wenn eine KI wie *Sophia* Eigenschaften wie das eigenständige Lösen von Problemen besitzt oder ihrem Tun Zwecke setzen kann, dann muss Sophia wohl als autonomes Wesen bezeichnet werden. Durch den Erhalt einer Staatsbürgerschaft wird eine KI also zu einer juristischen Person und erhält dadurch die gleichen Rechte wie Menschen. Anhand dieses Beispiels wird deutlich, warum verbindliche Rechtsrahmen für KI zu gestalten sind. So schlägt die Europäische Kommission einen Rechtsrahmen vor, der die Risiken von KI anspricht (*Vorschlag für einen Rechtsrahmen für künstliche Intelligenz Gestaltung der digitalen Zukunft Europas*, 2022) .

Als eine Selbstbestimmung im Rahmen einer übergeordneten Moral definierte Immanuel Kant den Begriff *Autonomie*. Dabei nutzt der Mensch als ein autonomes System für seine unabhängige und selbstbestimmte Entscheidung moralische Grundsätze. Von diesen moralischen Grundsätzen sind Maschinen zurzeit noch weit entfernt, sodass im technischen Sinne unter *autonom* eher ein *automatisierter* und ohne menschlichen Eingriff durchgeführter Prozess verstanden wird (Jipp & Steil, 2021, S. 19). So wird zum Beispiel eine Fahraufgabe durch ein automatisiert fahrendes Fahrzeug automatisiert und ohne menschliches Eingreifen durchgeführt (SAE, 2018).

Beruft man sich auf die griechischen Wurzeln des Begriffs *Autonomie*, so bedeutet dieser *Selbstgesetzgebung*. "Der Mensch ist das Wesen, das sich selbst die Maßstäbe, Normen und Vorschriften seines Verhaltens, seines Handelns, Unterlassens und Duldens vorgeben und zur verbindlichen Maxime machen kann" (Haux et al., 2021, S. 127).

In den 1950er Jahren lieferten technische Systeme aus ihrer Umgebung einfache Messdaten, nahmen Anweisungen auf und setzten diese in mechanische Funktionen um. Nun aber stehen "künstlich intelligente Systeme im Fokus der Forschung und Entwicklung, die nicht nur kognitiv komplexe Funktionen wie Informationsverarbeitung und Entscheidungsfindung automatisieren, sondern auch die Art und Weise revolutionieren können, mit der wir mit technischen Systemen interagieren" (Jipp & Steil, 2021, S. 27-28).

Künstlich intelligente Systeme "können zum Beispiel mithilfe der Auswertung von Kameradaten Informationen über den aktuellen, emotionalen Zustand von Menschen liefern und ihre eigene Funktionalität an diese Zustände adaptieren" (Jipp & Steil, 2021, S. 28). Der daraus folgende mögliche Fortschritt in der Entwicklung von künstlich intelligenten Systemen wäre der, dass technische Systeme zu einer Art kognitiven Empathie befähigt würden, menschliche Emotionen zu erfassen und in weiterer Folge darauf auch angemessen zu reagieren. An dieser Stelle stellt sich die Frage, ob der Mensch dann nicht auf das reduziert würde, was von einem künstlichen intelligenten System erfasst werden würde. "Wie wird sich unser Verständnis von sozialen Interaktionen, Empathie,

Emotion und Kooperation durch hybrides Zusammenwirken verändern?", ist eine berechtigte Frage die Jipp und Steil (2021, S. 28-29) stellen.

2.5 Die Überwindung der Natur und Selbstidentifikation

Die Geschichte des Menschen ist von den immer wieder aufs Neue auftauchenden Fragen gekennzeichnet, "ob denn die Natur für uns ein Vorbild, eine Leitlinie sein kann, oder ob es nicht in unserer Bestimmung liegt, Natur zu überwinden, hinter uns zu lassen und durch künstliche Stoffe und Gebilde aller Art zu ersetzen" (Liessmann, 2023, S. 197).

Auch wenn wir einerseits die Natur romantisch illuminieren, sosehr sind wir andererseits fasziniert vom Phänomen der Künstlichkeit, von "Stahlkonstruktionen bis zu Raumkapseln, von humanoiden Robotern bis zur Künstlichen Intelligenz, von der die einen hoffen und die anderen fürchten, dass sie den Menschen als Gattung ablösen" (Liessmann, 2023, S. 200-201).

Ein grundlegender Bestandteil des Gefühls, jemand zu sein, ist das Besitzen unseres eigenen Körpers und seiner Empfindungen. So entwickeln wir, wenn wir eine Zeitlang mit verbundenen Augen oder im Dunkeln laufen und uns dabei mit einem Stock vorwärts tasten, eine Tastempfindung am Ende des Stocks. Dies ist ein Beispiel dafür, was Philosophen das *Gefühl der Meinigkeit* nennen, „welches einen spezifischen Aspekt des bewussten Erlebens darstellt – eine Form von automatischer Selbstzuschreibung, die eine bestimmte Art von Bewusstseinsinhalt in das integriert, was als das eigene Selbst erlebt wird“ (Metzinger, 2017, S. 115).

Das Experiment der *Gummihand-Illusion* zeigt, was dabei im Gehirn geschieht, wenn das Gefühl der Meinigkeit zum Beispiel von einem realen Arm einer Versuchsperson in die Gummihand übertragen wird. Bildgebende Verfahren neurowissenschaftlicher Untersuchungen zeigten in einer bestimmten Gehirnregion, dem prämotorischen Kortex, erhöhte Aktivität genau in dem Moment, in dem die Gummihand als Teil des eigenen Körpers erlebt wurde. Die Vermutung liegt daher nahe, dass es im Gehirn zu einer Verschmelzung taktiler

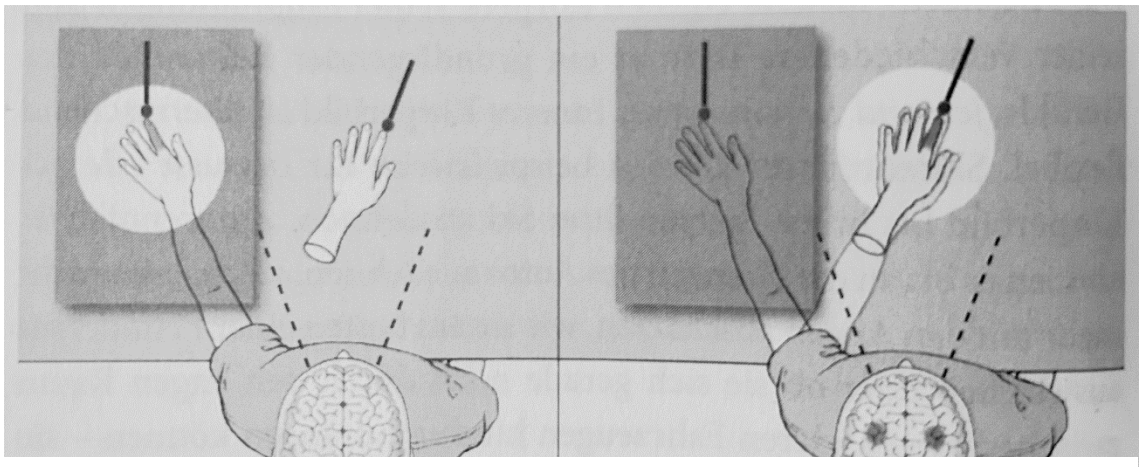

Abbildung 3: Die Gummihand-Illusion (Metzinger, 2017, S. 116)

und visueller rezeptiver Felder kommt (Metzinger, 2017, S. 116). Eine anschauliche Darstellung davon findet sich in Abbildung 3.

"Der Effekt offenbart eine dreifache Interaktion zwischen Sehen, Tasten und Propriozeption und kann Beweise für die Grundlage der körperlichen Selbstidentifikation liefern" (Botvinick & Cohen, 1998, S. 756).

2.6 Digitaler Wandel durch Künstliche Intelligenz

Der Stand der Technik von KI entwickelt sich aufgrund von Big Data sowie der immer leistungsfähigeren Hardware und der damit immer höheren Rechenleistung rasant. Der damit einhergehende Wissenszuwachs wird damit nicht nur beschleunigt, sondern befeuert wiederum die Entwicklung der ihm zugrunde liegenden Ursache. Reichl und Welzer (Hengstschläger & Rat für Forschung und Technologieentwicklung, 2020) führen aus, warum der ethische Diskurs über die Folgen des digitalen Wandels nicht ausreicht. Der digitale Wandel durch den Einsatz und das Vorhandensein von KI ist nicht bloß eine soziale Tatsache, sondern durchdringt alle Beziehungen des Menschen, sowohl zu sich selbst als auch zu seiner Umwelt (S.12).

Die Wirkmächtigkeit und Kompetenz von KI stellt den Menschen vor ein altes Problem, welches Immanuel Kant schon 1781 in folgenden Worten formuliert hat (Kant, 1781):

> "Die menschliche Vernunft hat das besondere Schicksal in einer Gattung ihrer Erkenntnisse: daß sie durch Fragen belästigt wird, die sie nicht abweisen kann; denn sie sind ihr durch die Natur der Vernunft selbst aufgegeben, die sie aber auch nicht beantworten kann; denn sie übersteigen alles Vermögen der menschlichen Vernunft." (S. 3)

Zum Ausdruck gebracht wird dieses Paradigma vom Übersteigen allen Vermögens menschlicher Vernunft auch von Hawking (2019). „Es ist zu befürchten, dass die KI alleine weitermacht und sich mit ständig zunehmender Geschwindigkeit selbst überarbeitet. Menschen, die aufgrund der Langsamkeit ihrer biologischen Evolution beschränkt sind, könnten nicht mithalten und würden verdrängt“ (S. 76). Die Lösung von Weltproblemen durch den Einsatz von KI und dessen Potential ist mit Skepsis zu betrachten, wird doch bereits weltweit über den Beginn eines Wettrüstens autonomer Waffensysteme nachgedacht (S.76).

Ob damit, wenn KI sich demnach *selbst entwerfen* würde, KI auch einer Definition des Menschen nach Sartre gleichkäme, indem er argumentiert, dass der Mensch zuallererst einmal i*st*, also existiert, und erst indem er *ist*, sich zu dem macht, was er *ist*, somit die Existenz demnach vor dem Wesen käme, ist eine sich aufdrängende Frage und wurde im Kapitel 2.2.1 erörtert.

Ein gegensätzlicher Weg zu jenem der Entwicklung von Maschinen und Robotern, die sowohl durch ihr Aussehen als auch durch ihre kognitiven und intellektuellen Fähigkeiten menschenähnliche Formen und Züge annehmen und zu humanoiden Robotern werden, ist jener, den Menschen mit Zuhilfenahme von Technik im weitesten Sinne zu roboterhaften Humanoid*innen oder posthumanen Wesen werden zu lassen.

Das bedeutendste Ziel vieler Transhumanistinnen und Transhumanisten stellt die radikale Lebensverlängerung und sogar die Aufhebung des Alterungsprozesses dar. Dies bedeutet im Endeffekt die automatisch einhergehende Unsterblichkeit. Zwar wäre Unsterblichkeit "nicht in einem absoluten Sinn gegeben. Menschen könnten durch einen Unfall, Selbst- oder Fremdtötung aktiv ihr Leben oder das Leben anderer beenden […] (Loh, 2019b, S. 42-43).

"Durch die Erlangung der Unsterblichkeit erhält der Mensch ultimativ die Kontrolle über sein Leben und transformiert vom passiven Spielball der Kräfte von Natur und Zeit zu einem aktiven Gestalter seines eigenen Daseins" (Loh, 2019b, S. 43).

Allerdings gibt es auch Gegenbewegungen zu den transhumanistischen Strömungen, da eine Spaltung der Gesellschaft droht zwischen jenen, "die sich mithilfe neuer Technologien (nicht nur digitaler, sondern auch medizinischer und pharmazeutischer oder nanotechnologischer Art) auf eine höhere Stufe stellen, und solchen, denen das mangels ökonomischer und technischer Mittel versagt ist" (Nida-Rümelin, 2018, S. 193). Damit reiht sich der zeitgenössische Transhumanismus in jenen uralten Menschheitstraum, sich über die *conditio humana* hinwegzusetzen und damit alle Beschränkungen der menschlichen Natur aufzulösen (S. 193).

Dieser Wunsch ist aus psychoanalytischer Sicht regressiv und narzisstisch. So beschrieb Freud die Vorstellung von Doppelgängern als krankhaft und narzisstisch. Wenn also Robotern unsere Gehirne eingepflanzt werden sollen, dann sind diese nichts anderes als eine Fortführung regressiver Ideen, gleich der altägyptischen Tradition der Sarkophage, "die als Ebenbilder des Toten seine Unsterblichkeit garantieren sollten" (Nida-Rümelin, 2018, S. 194).

Gleichzeitig ist die Sehnsucht nach der Unvergänglichkeit ein fester Bestandteil der menschlichen Kulturgeschichte und damit kein Phänomen der Moderne oder eine Erfindung des Silicon Valley. Dass sich der Mensch nicht mit der Endlichkeit seines irdischen Lebens abfinden will zeigen die Geschichten und Mythologien über Gilgamesch, der, um ein Heilmittel gegen die Sterblichkeit zu finden, die halbe Welt bereiste bis hin zu Orpheus, der Eurydike aus dem Totenreich zurückholen wollte (Riesewieck & Block, 2020, S. 282).

"Anders als Tiere sind wir Menschen uns unserer Sterblichkeit bewusst. Es ist der Preis unserer Intelligenz, dass wir mit absoluter Gewissheit sagen können, dass es uns eines Tages nicht mehr geben wird" (Riesewieck & Block, 2020, S. 282).

Und so treten im Windschatten der digitalen Revolution Start-ups aus der ganzen Welt um den Markt der digitalen Unsterblichkeit in einen Wettlauf mit dem Ziel, "unsere Persönlichkeiten über den Tod hinaus am Leben zu erhalten" (Riesewieck & Block, 2022, S. 12-13).

2.7 Jean-Paul Sartre - Die Entdeckung der Freiheit

Der in Jean-Paul Sartres 1943 erschienenem Hauptwerk *Das Sein und das Nichts* postulierte Freiheitsbegriff mündet in der Kernaussage, dass der Mensch zur Freiheit verurteilt ist (Sartre, 2020):

> "Ich selbst bin verurteilt, für immer jenseits meines Wesens zu existieren, jenseits der Antriebe und Motive meiner Handlung: ich bin verurteilt, frei zu sein. Das bedeutet, daß man für meine Freiheit keine anderen Grenzen als sie selbst finden kann oder, wenn man lieber will, daß wir nicht frei sind, nicht mehr frei zu sein." (S. 764)

Die Aufgabe, die Sartre in seinem Hauptwerk zu bewältigen sucht, ist einerseits die Autonomie des Menschen und andererseits seine Realität unter den Objekten. Er will damit das Spannungsfeld von Freiheit und Notwendigkeit ausleuchten, in dem der Mensch situiert ist (Wildenburg, 2004, S. 26).

Das Problem vor dem Sartre also steht ist die Frage, wie sich Freiheit und Notwendigkeit, diese beiden Pole, miteinander in Einklang oder Verbindung bringen lassen. Es ist der Versuch, ein System zu entwerfen, welches alle Aspekte der menschlichen Existenz zu integrieren versucht, sowie diese menschliche Existenz auf ein Fundament zurückzuführen versucht – die Freiheit. (Wildenburg, 2004, S. 18).

Und dann kam Sartre, der die Probleme all jener in einer Formel zusammenfasste, welche sowohl für Philosophen als auch für Laien bedeutungsvoll und verständlich ist: „Der Mensch ist frei, denn er ist das einzige Geschöpf dieser Welt, bei dem die Existenz dem Wesen vorausgeht" (Holz, 1958, S. 11).

Holz (1958) fasst Sartres Werk durch die Verbindung mit Persönlichkeiten wie beispielsweise Marcel Proust, Max Weber und James Joyce, die verschiedene weltanschauliche Positionen vertreten. Sie stehen für die Weite und das oppositionelle Niveau des individuellen Menschen gegen die Tyrannei gesellschaftlicher Versklavung (S. 11).

Jeder Gegenstand hat ein Wesen, hat eine Existenz, wobei das Wesen eine konstante Gesamtheit von Eigenschaften darstellt und die Existenz eine effektive Anwesenheit in der Welt hat. Einem religiösen Denken entspringt dagegen die Idee, dass zuerst das Wesen und erst dann die Existenz käme und dass beispielsweise Erbsen oder Gurken entsprechend einer Idee von Erbsen oder

Gurken wüchsen, weil sie am Wesen dieser Pflanzen teilhaben (Sartre, 2021, S. 115).

Sartre (2021) argumentiert, dass für alle Menschen, die glauben, dass sie von Gott erschaffen wurden, das Wesen der Existenz deshalb vorausgehe, weil er dies entsprechend einer Idee, die er vom Menschen hatte, getan hätte. Gleich jemandem, der ein Haus bauen will und tatsächlich zuvor wissen muss, welchen Gegenstand er damit schaffen will (S.115).

> "Aber selbst jene, die nicht glauben, haben diese traditionelle Auffassung behalten, daß ein Gegenstand immer nur in der Übereinstimmung mit seinem Wesen existiere, und das ganze 18. Jahrhundert hat gedacht, daß es ein allen Menschen gemeinsames Wesen gäbe, das man *Menschennatur* nannte. Der Existenzialismus dagegen hält daran fest, daß beim Menschen – und nur beim Menschen – die Existenz dem Wesen vorausgeht". (Sartre, 2021, S. 115-116)

Dies bedeutet nichts anderes, als dass der Mensch zuerst einmal *ist*, also existiert, und erst indem er *ist*, sich zu dem macht, was er *ist*. Der Mensch muss daher sein eigenes Wesen schaffen und als ein in die Welt geworfenes Wesen in der Welt leiden, kämpfen und sich dadurch definieren. Diese Definition bleibt jedoch offen, denn "man kann nicht sagen, was *ein bestimmter* Mensch ist, bevor er nicht gestorben ist, oder was die Menschheit ist, bevor sie nicht verschwunden ist" (Sartre, 2021, S. 116).

Sartre geht es somit nicht um die Bestimmung *des* Menschen. Wildenburg (2004) stellt dazu fest, dass Sartre das Individuum in seinen konkreten Entschlüssen und Handlungen zu verstehen sucht. Sartre will dieses konkrete Individuum, will dessen Konstituierung im Spannungsfeld von Freiheit und Notwendigkeit, von Wahl und Gegebenheit, verstehen (S. 68).

Was das faszinierende an Sartre war ist "dass er nicht über ein abstraktes menschliches Subjekt geschrieben hat, sondern dass da plötzlich ein ganz lebendiges Individuum beschrieben wurde, der ganz konkrete Mensch" (Wildenburg, 2011, 01:47–01:57).

2.7.1 Das Für-sich

Das *Für-sich*, das ist für Sartre das Selbstbewusstsein. Dieses Selbstbewusstsein weiß im Gegensatz zum *An-sich* von sich. Somit ist es *nicht* das was es ist, also *keine* reine Identität, sondern etwas das ein Geheimnis hat, ein Innenleben sowie eine Außenwelt. Es hat "Beziehungen zu sich, Beziehungen zu anderem und zu Anderen" (Wildenburg, 2004, S. 32). Im Kapitel 2.7.3 wird der Begriff *An-sich* erläutert.

Denn, so Wildenburg (2011), "der Mensch macht sich und er ist ständiger Entwurf, jeden Tag und jede Sekunde. Der Mensch ist immer in Zukunftsentwürfen und hat Ziele" (15:40–16:08).

> "Die Freiheit ist genau das Nichts, das im Kern des Menschen *geseint* wird [est été] und die menschliche-Realität zwingt, *sich zu machen* statt *zu sein*. Wir haben gesehen, daß sein für die menschliche-Realität *sich wählen* ist: nichts geschieht ihr von außen und auch nicht von innen, was sie *empfangen* oder *annehmen* könnte. Sie ist ohne irgendeine Hilfe ganz der untragbaren Notwendigkeit ausgeliefert, bis in das kleinste Detail hinein sich sein zu machen. Also ist die Freiheit nicht *ein* Sein: Sie ist das Sein des Menschen, das heißt sein Nichts an Sein. (Sartre, 2020, S. 765)

2.7.2 Die Welt und der Mensch

Die Ebene, die Sartre in seiner Philosophie mit dem An-sich und dem Für-sich betritt, ist eine, welche noch vor dem konkreten Menschen anzusiedeln ist und von ihm selbst als „apriorische Beschreibung des Seins des Für-sich" bezeichnet wird. Dabei geht es Sartre um die Bedingungen der Möglichkeit menschlicher Existenz. Sartre stellt die für seinen Freiheitsbegriff entscheidende Frage:

> "Wie ist es möglich, dass der Mensch und die Welt so beschaffen sind, wie sie es sind, wie ist es möglich, dass der Mensch fragen kann, dass er verneinen, lügen, unaufrichtig sein kann? Was bedeutet es, dass der Mensch Angst hat? Wie verhält es sich mit der Zeitlichkeit und dem Körper des Menschen, wie und in welcher Weise ist eine Beziehung zu anderen Personen möglich, was heißt es, dass wir handeln können, und was schließlich sollen wir tun?" (Wildenburg, 2004, S. 29-30)

Sartres Werk *Das Sein und das Nichts* kreist um zwei Pole: Der eine Pol ist das *An-sich*, der andere Pol das *Für-sich*, also die *Dinge* und das *Selbstbewusstsein*, die *Welt* und der *Mensch.* Diese beiden Pole, das *An-sich* und das *Für-sich*, also das *Sein* und das *Nichts*, sind die "starre und grundlose Massivität des Seins auf der einen und der freie Akt des Nichtens, das Selbstbewusstsein auf der anderen Seite," so Wildenburg (2004, S. 34).

2.7.3 Das An-sich

"Im Moment des Todes *sind* wir, das heißt, wir sind wehrlos gegenüber den Urteilen der Anderen; man kann in Wahrheit entscheiden, was wir sind, wir haben keinerlei Chance mehr, der Bilanz zu entgehen, die eine allwissende Intelligenz aufstellen könnte" (Sartre, 2020, S. 230).

Sartre gelingt jedenfalls mit dieser Aussage eine scharfe Grenzziehung für seine Begrifflichkeiten und die Bedeutung des *An-sich* und *Für-sich*. Denn, so Sartre weiter (2020), "durch den Tod verwandelt sich das Für-sich immer in ein An-sich, insofern es völlig in die Vergangenheit geglitten ist" (S. 230).

Das *An-sich* ist ein Grenzbegriff, es kommt nie in Reinform vor, denn beschrieben, erschlossen oder konstruiert kann dieses *An-sich* immer nur aus der Perspektive des *Für-sich* werden. Sartre nennt drei Merkmale des *An-sich*. Zum einen ist es weder erschaffen noch Grund seiner selbst. Zum anderen *ist* das Sein. Es ist einfach und ist von absoluter Grundlosigkeit. Dieser Grund tritt mit dem Selbstbewusstsein, dem *Für-sich* auf (Wildenburg, 2004, S. 28-31). "Schließlich – das ist unser drittes Merkmal – das An-sich-sein *ist*. Das bedeutet, daß das Sein weder vom Möglichen abgeleitet noch auf das Notwendige zurückgeführt werden kann," schreibt Sartre (2020, S. 43-44) in seinen Überlegungen zu den Seinsphänomenen.

Das Sein ist also ohne jeden Bezug zu sich oder zu anderen Dingen und weiß weder von sich noch von anderem, hat weder ein Geheimnis oder ein Innenleben und Außenleben. Es ist das, was es ist, eine reine Identität (Wildenburg, 2004, S. 32).

2.7.4 Selbstbewusstsein als Akt der Freiheit

Die Freiheit, so Wildenburg (2004), tritt durch das Selbstbewusstsein auf und "ihr Auftreten komme einer Revolution gleich, einer Erschütterung der weltlosen Welt des An-sich, wie kein Erdbeben sie hervorrufen kann" (S. 32). Erst durch den Akt der Verneinung, von Sartre auch als *Nichtung* bezeichnet, wird Selbstbewusstsein möglich. Dieses Selbstbewusstsein lässt den Menschen sagen, *nicht* dieser Stein zu sein, *nicht* dieses Ding oder *nicht* diese Wurzel zu sein, *nicht An-sich* zu sein (S. 32).

Sartre (2020) beschreibt diese Nichtung:

> "Das Nichts ist die Infragestellung des Seins durch das Sein, das heißt eben das Bewusstsein oder Für-sich. Es ist ein absolutes Ereignis, das durch das Sein zum Sein kommt und, ohne das Sein zu haben, dauernd vom Sein getragen wird". (S. 172)

Die beiden Seinsweisen, also das *An-sich*, welches das ist, was ist und das *Für-sich,* also das, was zu sein hat was es ist, weil es sich stetig auf Ziele hin entwirft, werden durch eine synthetische Verbindung geeint. Diese Verbindung ist das *Für-sich* selbst (Streller, 1952, S. 2).

Das *Für-sich* besteht in nichts anderem als im Vollziehen, denn es *ist* nicht, sondern *macht* sich. Denn es ist sein eigener Grund, es weiß um sich und seine Beziehung zu anderem. Dieses *Für-sich*, diese Leerstelle des Bewusstseins, ist die Freiheit des Menschen. Sie ist die unhintergehbare Bedingung menschlicher Existenz, welche der Mensch verleugnen oder annehmen kann. Das *Nichts* tritt erst mit dem menschlichen Bewusstsein in das *Sein*, in jene totalitäre Welt des *An-sich*. Sartre versteht dieses *Nichts*, welches einem Loch im *Sein* gleichkommt, als einen nichtenden Akt des *Seins*. Somit entwirft sich das Bewusstsein "wählend ständig nach der

Zukunft hin, um sein eigenes Fundament zu werden, jedoch ohne dieses Ziel je zu erreichen (Werner, 2021, S. 25).

2.8 Handlungsempfehlungen im Umgang mit KI im gesellschaftspolitischen Spannungsfeld

Algorithmische Entscheidungen von KI-Systemen haben das Ziel, "Aufgaben und Probleme eigenständig zu erkennen, zu bearbeiten und zu lösen, ohne dass jeder Schritt vom Menschen definiert wird" (Dengel et al., 2023, S. 92). Damit werden also auch Entscheidungen getroffen, die unsere Lebenswelten im Allgemeinen als auch im Individuellen betreffen. Es stellt sich somit die Frage nach der Verantwortung, der Erklärbarkeit als auch nach der Transparenz dieser algorithmischen Entscheidungen (S. 92). Denn das Vertrauen von Menschen in KI-Systeme ist entscheidend für die Benutzerakzeptanz, denn Menschen vermeiden Systeme, denen sie nicht vertrauen (Bartneck et al., 2019, S. 38-39).

Ob Chatbots, Roboter oder autonome Fahrzeuge – sie alle sind ein wichtiges und allgegenwärtiges Phänomen in Wirtschaft und Gesellschaft geworden. "Die Entscheidung, welchem KI-System man vertrauen kann und welchem nicht, ist von entscheidender Bedeutung, da solche Systeme autonom Aufgaben ausführen und die menschliche Entscheidungsfindung beeinflussen" (R. Riedl, 2022, S. 1).

2.8.1 Die Bedeutung des Vertrauens in technische Systeme

Mit der wachsenden Bedeutung des Vertrauens in KI-Systeme geht auch die zunehmende Erkenntnis darüber einher, "dass die Persönlichkeit des Nutzers mit dem Vertrauen zusammenhängt und somit die Akzeptanz und Annahme von KI-Systemen beeinflusst" (R. Riedl, 2022, S. 1). Ebenfalls zeigt sich, "dass das präskriptive Wissen darüber, wie vertrauenswürdige KI-Systeme in Abhängigkeit von der Benutzerpersönlichkeit gestaltet werden können, weit hinter dem deskriptiven Wissen über die Nutzung und die Vertrauenswirkung von KI-Systemen zurückbleibt" (R. Riedl, 2022, S. 1).

R. Riedl (2022) untersuchte den Einfluss auf das Vertrauen in KI-Systeme anhand des Fünf-Faktoren-Modells *Big Five*, welches die breiteste Abstraktionsebene zur Konzeptualisierung der menschlichen Persönlichkeit darstellt. Das Modell umfasst die Facetten Extraversion, Verträglichkeit, Gewissenhaftigkeit, Neurotizismus und Offenheit (S. 3). Dabei zeigten sich überwältigende Beweise, dass sowohl Verträglichkeit als auch Offenheit das Vertrauen positiv beeinflussten. Auch wenn die Beweise weniger überzeugend waren, dass Extraversion das Vertrauen im Vergleich zu Verträglichkeit und Offenheit ebenfalls positiv beeinflusst, so "gibt es dennoch eine klare Tendenz, dass Extraversion das Vertrauen ebenfalls positiv beeinflusst. Was den Neurotizismus betrifft, so sind die Ergebnisse ebenfalls eindeutig. Weniger neurotische Menschen zeigen mehr Vertrauen in KI-Systeme" (R. Riedl, 2022, S. 15).

In einer Studie, in welcher der Frage, wie Kinder einen virtuellen Agenten wahrnehmen, der sie trainiert, nachgegangen wurde, wurden 25 Kinder, davon 16 männlich und 9 weiblich im Durchschnittsalter von 12,5 Jahren gebeten, ihre Wahrnehmung eines virtuellen und eines menschlichen Trainers zu bewerten. Diese präsentierten dabei auf Videos geschriebene Wörter und führten zusätzlich für jedes Wort in einer Fremdsprache eine semantisch verwandte Geste aus. Sowohl der virtuelle Trainer als auch der menschliche Trainer wurden anschließend von den Probanden nach Merkmalen, die sich auf die Gesten als auch nach ihrer Persönlichkeit bezogen, bewertet (Macedonia et al., 2014, S. 131-133).

"Die Probanden fanden die menschlichen Gesten besser und gaben dem menschlichen Trainer höhere Sympathiewerte; der Gesamtunterschied zwischen der Wahrnehmung der virtuellen und menschlichen Trainer war jedoch nicht signifikant" (Macedonia et al., 2014, S. 131). Mädchen fanden den virtuellen Agenten besser als den Menschen. Warum Mädchen dies taten, war unklar und bedarf daher weiterer Untersuchungen, um dieses Ergebnis zu bestätigen beziehungsweise zu klären. Daraus kann der Schluss gezogen werden, dass die Akzeptanz virtueller Agenten durch künftige Verbesserungen der Software, insbesondere bei der Ausführung von Gesten, weiter erhöht wird und dadurch den Einsatz dieser virtuellen Agenten beim Fremdsprachenlernen erleichtern werden (Macedonia et al., 2014, S. 136).

2.8.2 Rechtliche Rahmenbedingungen für KI-Systeme

Die Handlungsfelder im Umgang mit KI sind vielfältig. Da die scheinbar objektiven Prozesse der automatisierten Datenauswertung enorme Verzerrungen mit sich bringen können, sind rechtliche Rahmenbedingungen und Handlungsempfehlungen unerlässlich. Denn inwieweit Technologie Entscheidungsmacht über unseren Alltag haben soll und wer für Entscheidungen verantwortlich ist, die automatisiert getroffen werden, sind ethische Probleme praktischer Natur (Dengel et al., 2023, S. 94).

Parallele und vorgreifende ethische und juristische Neujustierungen in kollektiven Verantwortungsketten in Gesellschaft, Wirtschaft, Wissenschaft und Staat sind notwendig, um der rasant und komplex expandierenden KI zu begegnen. Dabei muss die Realisierung von Verantwortung im Einsatz von KI über sogenannte Gefährdungshaftungen geregelt werden (Haux et al., 2021, S. 242).

"Luchterhand sieht aus rechtlicher und anthropologischer Sicht in der progredienten Digitalisierung und der notwendigen adaptiven, sich selbst funktionalisierenden Umformung die fundamentale Gefahr, dass sich der Mensch seiner autonomie- und würdebegründenden Moralität, Rationalität und Sozialität entfremdet" (Haux et al., 2021, S. 242) .

2.8.3 Aspekte der Regulierung von Künstlicher Intelligenz

Die im Folgenden verhandelten Aspekte einer rechtlichen Regulierung von KI-Systemen beziehen sich ausschließlich auf europäisches Recht. Rechtnormen anderer Staaten sind an dieser Stelle nicht Gegenstand der Untersuchung.

Für Europa hat die Europäische Kommission einen Rechtsrahmen für KI vorgeschlagen, der neben Risiken, die angesprochen werden, Europa auch in die Lage versetzen soll, weltweit eine führende Rolle zu spielen. Dabei konzentriert sich der Ansatz der EU für künstliche Intelligenz auf Exzellenz und auf Vertrauen, mit dem Ziel, Grundrechte zu gewährleisten und die Forschungs- und Industriekapazitäten zu stärken (*Vorschlag für eine Verordnung des europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für die Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union*, 2021).

Mayrhofer und Rachbauer (Hengstschläger & Rat für Forschung und Technologieentwicklung, 2020) führen aus, dass aufgrund der Tatsache, dass unterschiedlichste Rechtsbereiche von der komplexen Technologie von KI-Systemen und deren technologischen Entwicklungen betroffen sind und andererseits sich KI-Systeme stark in Bezug auf Autonomiegrad und Eingriffsintensität unterscheiden, ein abgestufter Ansatz bei der Regulierung notwendig sein wird, da diese Technologie keiner pauschalen rechtlichen Steuerung zugänglich ist (S. 243).

Mayrhofer und Rachbauer (Hengstschläger & Rat für Forschung und Technologieentwicklung, 2020) betonen, dass die rechtliche Einordnung und Regulierung von KI wesentlich vom Autonomiegrad abhängt, den ein KI-System aufweist. So werfen rein regelbasierte Systeme, bei denen der Algorithmus bereits vollständig vorherbestimmt ist, rechtlich keine besonderen Fragestellungen auf (S. 219). Dagegen sind bei KI-Systemen mit Machine-Learning und Deep-Learning Verfahren Arbeitsschritte als auch die erzielten Ergebnisse ab einem bestimmten Punkt nicht mehr nachvollziehbar und vorhersehbar. Somit werfen die fehlende Prognostizierbarkeit und die fehlende Nachvollziehbarkeit der Entscheidungsfindung von Systemen mit hohem Autonomiegrad neue rechtliche Fragen auf (S. 220).

2.8.3.1 Direkte und indirekte Regelung von KI

Ein Beispiel einer direkten Regelung von KI ist die Datenschutz-Grundverordnung (DSGVO), welche im Artikel 22 unter dem Titel *Automatisierte Entscheidungen im Einzelfall einschließlich Profiling* von einer automatisierten Verarbeitung der Daten zur Entscheidungsfindung spricht. Nach Art. 22 Abs. 1 DSGVO hat jede von einer Datenverarbeitung betroffene Person das Recht, "nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber

rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt" (Pollirer et al., 2017, S. 64).

Mayrhofer und Rachbauer (Hengstschläger & Rat für Forschung und Technologieentwicklung, 2020) führen aus, dass Art. 22 Abs. 2 DSGVO gleichzeitig weitreichende Ausnahmen vorsieht, die KI-gestützte Entscheidungen in bestimmten Fällen durchaus erlauben. Eine mitunter im Ergebnis weitaus erheblichere regulierende Wirkung für KI entfalten indirekte Regelungen in Form zahlreicher Rechtsvorschriften (S. 222). Im Besonderen ergeben sich aus dem Datenschutzrecht solche Wirkungen, wenn beispielsweise für das Training von KI eine große Menge an Daten erforderlich ist. Denn wenn für das Training einer KI-Anwendung personenbezogene Daten benötigt werden oder mittels einer KI-Anwendung verarbeitet werden, so unterliegen diese Vorgänge der DSGVO (S. 224).

2.8.3.2 Pflichten und Grenzen einer Regulierung von KI

Mayrhofer und Rachbauer (Hengstschläger & Rat für Forschung und Technologieentwicklung, 2020) weisen darauf hin, dass nationale Gesetzgebungen bei der Regulierung von KI nur einen begrenzten Spielraum haben. Dies liegt in der Natur unions-, und verfassungsrechtlicher Grenzen. So besitzt der Einsatz innovativer Technologien bei medizinischen Behandlungen eine besondere grundrechtliche Dimension. Dem umfassenden Selbstbestimmungsrecht des Patienten steht spiegelbildlich die ärztliche Therapiefreiheit gegenüber (S. 231).

2.8.3.3 Technische Grenzen der Regulierung von KI

Dass sich die Grenzen der Regulierung nicht allein aus dem höherrangigen Recht ergeben, sondern auch aus der Technologie selbst, zeigt sich daran, dass zwar die eingemahnte Transparenz KI-basierter Systeme, welche zumindest durch die Offenlegung der Algorithmen und der Datenbasis erreicht werden soll, das Problem der Vorhersehbarkeit des Handelns solcher Systeme entschärfen (Wischmeyer, 2018, S. 42-43). Es scheint jedoch sehr zweifelhaft, ob Prozesse oder Entscheidungsfindungen aufgrund der Funktionsweise von KI und der Binnenkomplexität solcher Systeme im Verein mit den großen Datenmengen zu deren Training in allen seinen Details nachempfunden werden kann (S. 46-47).

2.8.4 Ethische und moralische Herausforderungen im Umgang mit KI

Aristoteles begründete den Begriff der *Ethik* und verstand darunter die Betrachtung von Gewohnheiten und Sitten aus wissenschaftlicher Sicht. Durch ein ethisches Verhalten und durch das Streben nach Tugendhaftigkeit konnte der Mensch sein oberstes Ziel, Glückseligkeit, erreichen (Fritz et al., 2019, S. 19).

Moral vom lateinischen Wort *mores* abgeleitet, beschreibt die in einer Gruppe als *gut* geltenden Gebräuche und Sitten und definiert damit das von einer

Gesellschaft als richtig erachtete Handeln. Die *Ethik* ist die Wissenschaft der Moral und somit jene praktische Philosophie, "die sich damit beschäftigt, was als gutes und daher moralisches Handeln gilt" (Fritz et al., 2019, S. 19).

Auf der vom *Future of Life Institute* veranstalteten *Asilomar Konferenz 2017* wurden verschiedene Leitsätze beschlossen (Future of Life Institute, 2018). Unter anderem wurden unter dem Titel *Ethik und Werte* Themen wie Sicherheit, Transparenz, Verantwortung, Wertausrichtung, menschliche Werte, Privatsphäre, Freiheit, geteilter Nutzen, geteilter Wohlstand oder menschliche Kontrolle diskutiert.

So sollten KI-Systeme während ihrer gesamten Funktionszeit möglichst nachweislich sicher sein. Ebenso muss Transparenz bei Fehlfunktionen gewährleistet sein, falls durch ein KI-System Schaden entsteht, sodass es möglich ist, die Ursache zu ermitteln. Des Weiteren muss es Transparenz bei der Rechtsprechung geben. Wenn also autonome Systeme in entscheidungsfindende Prozesse der Rechtsprechung eingebunden sind, dann haben diese Prozesse nicht nur nachvollziehbar zu sein, sondern sich auch einer Überprüfung durch die Autorität von Menschen zu unterziehen (Future of Life Institute, 2018).

Während also im Fall von KI-Entscheidungen sehr kritische ethische Fragen gestellt werden, "werden intuitive Entscheidungen realer Personen selten infrage gestellt, obwohl Menschen nicht unbedingt die bessere Wahl treffen würden" (Dengel et al., 2023, S. 96). Daher argumentieren Befürworter algorithmischer Techniken, dass es durch diese zu einer Eliminierung menschlicher Vorurteile kommen würde. Gleichzeitig darf nicht übersehen werden, dass Technologie nicht neutral ist und ein Algorithmus nur so gut ist wie die Daten, mit denen er arbeitet und somit Voreingenommenheit oder Vorurteile, die sich in den erfassten Daten widerspiegeln von KI-Anwendungen übernommen werden (S. 98). Diese tief in unserer Gesellschaft verwurzelte Voreingenommenheit wird *Machine-Learning-Bias* genannt. Diesen Bias zu erkennen und zu reduzieren ist eine Herausforderung, der sich Politik und Gesellschaft stellen müssen (S. 98).

Jobin, Ienca und Vayena (2019) führten in ihren Untersuchungen eine Art Inventur für ethische KI durch. Dabei wurden bis zum Stichtag 23. April 2019 84 Dokumente identifiziert, in denen Aussagen zu KI gemacht wurden. Es konnten dabei in einer Inhaltsanalyse 11 übergreifende ethische Werte und Grundsätze identifiziert werden. In absteigender Häufigkeit sind diese: Transparenz, Gerechtigkeit und Fairness, das Verhindern von Schaden, Verantwortung, Datenschutz und Privatsphäre, Wohltätigkeit, Freiheit und Autonomie, Vertrauen, Nachhaltigkeit, Würde sowie Solidarität. Dabei zeigte sich, dass bei der Interpretation der Ergebnisse und bei Fragen, warum diese Prinzipien als wichtig erachtet werden und wie sie umgesetzt werden sollen, kein Konsens herrschte (S. 389-399).

Das *Institute of Electrical and Electronics Engineers (IEEE)*, der größte und älteste Ingenieursverband der Erde, hat für ethisch angepasstes Design eine Initiative begründet, bei der über drei Jahre hinweg 1000 Expert*innen bottom-up involviert waren. Dabei sollten diese herausfinden, wie man technische Systeme menschengerechter bauen kann. An der vom Ingenieursverband propagierten einschlägigen Liste von Wertprinzipien sollten sich unter anderem IT-Entwickler orientieren und dabei die Menschenrechte, wie sie in der UN-Menschenrechtskonvention festgehalten sind, berücksichtigen. Über diese Forderungen der UN-Menschenrechtskonvention hinaus verweist der *IEEE* noch auf weitere Wertekataloge, in deren Zentrum der Grundsatz, dass Technik zur Verbesserung des Wohlbefindens von Menschen beitragen sollte, steht (Spiekermann, 2021, S. 177).

2.8.5 Verantwortung im digitalen Zeitalter

Kompetenzen, die traditionell menschlichen Akteuren vorbehalten waren, werden durch die raschen Fortschritte in der Robotik und der künstlichen Intelligenz aufgrund ihrer transformativen Kraft zu einer großen Herausforderung für den Menschen. Ehemals exklusiv dem Menschen vorbehaltene Konzepte wie Autonomie, Handlungsfähigkeit und Verantwortung könnten eines Tages in ähnlicher Weise auf künstliche Systeme zutreffen und werfen somit die Frage nach der Bedeutung dieser Begriffe auf (Coeckelbergh & Loh, 2020, S. 7).

Es ist fraglich, ob gerade im Bereich von Technologie, Big Data und neuen Medien unser traditionelles Verständnis von Verantwortung den aktuellen Herausforderungen gewachsen ist. Der Grund dafür ist eine eingeschränkte Fokussierung auf den autonomen, autarken, individuellen Menschen, der genuin als verantwortlicher Akteur agiert. So muss im Fall von autonomen Fahrsystemen diese differenzierte Fokussierung auf den einzelnen Verantwortlichen in Frage gestellt werden. Dabei stellt sich die Herausforderung, wer in bestimmten Situationen auf der Straße die relevanten Entscheidungen treffen kann und soll und wie die Verantwortung entsprechend zwischen den aktuellen Insassen des Fahrzeugs oder seinem Besitzer, den Ingenieuren oder Entwicklern des Autos und eventuell sogar dem künstlichen System, welches das Auto selbst betreibt, verteilt werden kann (Coeckelbergh, 2016, 749).

Die exponentielle Zunahme der Komplexität sowie die langfristigen Auswirkungen von Mensch-Maschine Interaktionen und vernetzten Interaktionen erfordert eine Aktualisierung unserer ethischen Theorie, um hochgradig verteilten Szenarien Rechnung zu tragen. Es ist moralisch inakzeptabel und pragmatisch riskant, dass jedermanns Problem zu jedermanns Verantwortung wird (Floridi, 2016, S. 11).

2.8.6 Risiken durch unkontrollierte KI-Experimente

Am 22. März 2023 veröffentlichte das *Future of Life Institute*, dessen Präsident der am Massachusetts Institute of Technology (MIT) als Professor für Physik tätige Max Tegmark ist, einen Offenen Brief, der sich prominenter Unterstützer wie - um nur einige zu nennen - Stuart Russel, Professor für Comupterwissenschaft an der Universität Berkeley, Elon Musk, CEO of SpaceX, Tessla und Twitter oder Steve Wozniak, Co-Mitbegründer von Apple, erfreuen darf. In diesem Offenen Brief fordern das *Future of Life Institute* alle KI-Labors auf, ihre gigantischen KI-Experimente sofort für sechs Monate zu stoppen. Alle wichtigen Akteure seien in diese Pause einzubeziehen, zudem sollte sie öffentlich und überprüfbar sein. Und wenn die Umsetzung dieser Pause nicht schnell umgesetzt werden kann, dann sollten die Regierungen eingreifen und ein Moratorium verhängen. Dies soll keine generelle Pause in der KI-Entwicklung sein, "sondern lediglich eine Abkehr von dem gefährlichen Wettlauf zu immer größeren, unvorhersehbaren Black-Box-Modellen mit emergenten Fähigkeiten" (Future of Life Institute, 2023).

Es sei offensichtlich, so die Verfasser des Offenen Briefes, dass KI-Systeme, die einer dem Menschen ebenbürtigen Intelligenz entsprechen, enorme Risiken sowohl für die Menschheit als auch für die Gesellschaft in sich bergen. Dies wurde durch umfangreiche Forschungsarbeiten und die Anerkennung durch führende KI-Labors bestätigt (Future of Life Institute, 2023). In den *Asilomar-KI-Grundsätzen* (Future of Life Institute, 2018) wird empfohlen, dass die Entwicklung und der Einsatz fortgeschrittener KI mit Sorgfalt geplant werden sollte, denn sie stellt einen tiefgreifenden Wandel in der Geschichte des Lebens auf der Erde dar.

Solch eine Planung findet aber nicht statt, "obwohl sich die KI-Labors in den letzten Monaten in einem außer Kontrolle geratenen Wettlauf um die Entwicklung und den Einsatz immer leistungsfähigerer digitaler Köpfe befinden, die niemand - nicht einmal ihre Erfinder - verstehen, vorhersagen oder zuverlässig kontrollieren kann" (Future of Life Institute, 2023). So wurden auch schon bei anderen Technologien, die potenziell katastrophale Wirkungen auf die Gesellschaft haben könnten, eine Pause eingelegt. Als Beispiele hierfür nennen die Verfasser des Offenen Briefes das Klonen von Menschen, die Veränderung der menschlichen Keimbahn, die Funktionserweiterungsforschung und die Eugenik. (Future of Life Institute, 2023).

Dass genomverändernde Technik es Wissenschaftler*innen ermöglichen würde, genetische Krankheitsursachen durch die Korrektur von Genmutationen zu behandeln, ist das eine. "Allerdings gibt es aber auch moralisch kaum vertretbare Möglichkeiten der DNA-Manipulation. Wie weit wir bei der Genmanipulation gehen dürfen, ist eine immer dringlicher werdende Frage," so Hawking (2019, S. 75) in seiner Argumentation zu den Gefahren technologischer Entwicklungen.

Ebenfalls gefordert wird in dem Offenen Brief des *Future of Life Institute* der Trainingsstopp von KI-Systemen, die leistungsfähiger als *GPT*-4 sind. *GPT-4* ist ein multimodales Modell, welches Bild- und Texteingaben akzeptiert und Textausgaben generiert. "Obwohl GPT-4 in vielen realen Szenarien weniger fähig ist als ein Mensch, zeigt es bei verschiedenen beruflichen und akademischen Benchmarks Leistungen auf menschlichem Niveau, einschließlich des Bestehens einer simulierten Anwaltsprüfung mit einer Punktzahl, die in den oberen 10 % der Testteilnehmer liegt (*GPT-4 Technical Report*, 2023). Somit kann es dank seines breiten Allgemeinwissens und seinen Problemlösungsfähigkeiten schwierige Probleme und Aufgaben lösen. *GPT-4* ist ein sogenanntes *Transformator-basiertes Modell*. Dieses ist darauf trainiert, in einem Dokument das nächste Token mittels Vorhersage zu generieren. Zu einer verbesserten Leistung führt dann der Anpassungsprozess nach dem Training "bei der Messung der Faktizität und der Einhaltung des gewünschten Verhaltens. Ein zentraler Bestandteil dieses Projekts war die Entwicklung einer Infrastruktur und von Optimierungsmethoden, die sich in einem breiten Spektrum von Skalen vorhersagbar verhalten" (*GPT-4 Technical Report*, 2023).

2.8.7 Menschenzentrierte Künstliche Intelligenz

Aus der Perspektive *Menschenzentrierter KI* müssen KI und Machine Learning aus einem Bewusstsein heraus entwickelt werden, dass sie ein Teil eines viel größeren Systems, das aus Menschen besteht, sind. Das bedeutet, dass der Mensch auch in der Lage sein muss, KI-Systeme zu verstehen und KI-Systeme müssen in der Lage sein, den Menschen zu verstehen (Riedl, 2019, S.1).

Diese Entwicklung hin zu einer *Menschenzentrierten KI* gleicht einer zweiten kopernikanischen Revolution, denn bislang lag das Hauptaugenmerk auf der Entwicklung von KI und der Mensch musste sich danach den entwickelten Systemen anpassen. Eine anschauliche Darstellung davon findet sich in Abbildung 4.

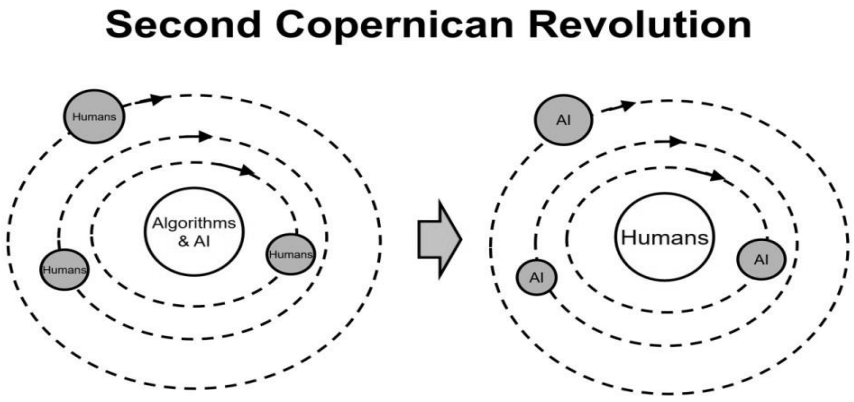

Abbildung 4: Zweite kopernikanische Revolution, (Shneiderman, 2020b, S. 112)

Im Unterschied dazu betont die *Menschzentrierte KI* die Wichtigkeit der menschlichen Nutzer, indem sie sich auf die Gestaltung einer positiven

Nutzererfahrung, die Messung der menschlichen Leistung und die Anerkennung der neuen Fähigkeiten des Menschen fokussiert. Der Fokus liegt auf den Bedürfnissen der Benutzer, es werden erklärungsfähige Systeme geschaffen und Wert auf sinnvolle menschliche Kontrolle gelegt. Der Mensch steht im Vordergrund und das Ziel besteht darin, den Bedürfnissen des Menschen gerecht zu werden. Damit wird die enorme Bedeutung, ein geeignetes Konzept für zukünftige Technologien zu entwickeln, welches die Würde des Menschen fördert, unterstrichen (Shneiderman, 2020, S. 112).

Menschenzentrierte KI muss dabei drei Voraussetzungen erfüllen. Sie sollte reliabel, sicher und vertrauenswürdig sein. Shneiderman (2020b) nimmt dazu zwei Skalen für die Betrachtung an. Zum einen die einer menschlichen Kontrolle und zum anderen die einer Autonomie der Maschine. Dabei gilt eine Anwendung als reliabel, sicher und vertrauenswürdig, wenn eine Maschine über hohe Autonomie verfügt, zugleich es dem Menschen aber möglich ist, in einem hohen Maße Kontrolle auszuüben (S. 495-504).

Gesellschaft und Politik müssen sich über die Rolle von KI Gedanken machen und Fragen, wie die Interaktion zwischen dem Menschen und KI sich gestalten soll, beantworten. Neben dieser Frage der Interaktion müssen wir Menschen uns auch bewusst werden, wie wir KI-Systeme nutzen wollen, nämlich als eine Partnerin oder bloß als ein Werkzeug (Chromik & Butz, 2021, S. 619-640).

So könnte beispielsweise eine Symbiose zwischen Arzt und Technik künftig so aussehen, dass Technik bei der Diagnose von Krankheiten die Interpretation der Daten übernehmen würde. In den Händen von Ärztinnen und Ärzten hingegen läge die Auswahl einer bestimmten Therapie. Sei es im Verlaufe der Entwicklung von Digitalisierung in einem ersten Schritt noch darum gegangen, Daten zu erfassen, zu speichern, zu übertragen und zu verarbeiten, so ginge es in einem zweiten Schritt darum, die "Daten digital zu verstehen, zu veredeln, aktiv zu nutzen und zu monetarisieren. KI-Systeme verfügten über bestimmte menschliche Kernfähigkeiten wie etwa Kommunikation, Schlussfolgern oder Lernen" (Cordes, 2019, S. 797).

Damit geht die Entwicklung neuer Präventions-, Diagnose- und Behandlungsformen für das Gesundheitswesen einher und eröffnet neue Möglichkeiten, obgleich man sich der Tatsache all ihrer Risiken wie Manipulationen, Fehler oder Datenmonopole bewusst sein muss (Cordes, 2019, S. 797).

2.9 Zusammenfassung

Im Kapitel *Stand des Wissens* wurde ausgehend von den philosophischen Grundfragen der Philosophie wie sie unter anderem Immanuel Kant formulierte, der Blick auf eine mit Bewusstsein ausgestatte KI gelegt. Die als *Superintelligente KI* bezeichnete Künstliche Intelligenz birgt zum einen bisher ungeahnte

Möglichkeiten zum Nutzen der Menschheit, löst zum anderen jedoch auch berechtigte Sorge aus, dass superintelligente KI-Systeme das Schlimmste wären, was der Menschheit passieren könnte.

Der rasante Fortschritt in der Entwicklung von KI-Systemen wurde in diesem Kapitel an Beispielen wie den KI-Systemen *AlphaGo* und *AlphaZero* dargestellt, deren enorme Fähigkeiten Tegmark als einen Wendepunkt in der Entwicklung von KI bezeichnet. Als ein weiteres Feld neuer technologischer Möglichkeiten wurde jenes der *Mensch-Maschine-Interaktion* untersucht. Die als *Brain-Computer-Interfaces (BCIs)* bezeichneten Schnittstellen zwischen dem menschlichen Gehirn und Maschinen dienen als Beispiel für die bereits erwähnten Hoffnungen und Ängste, die diese Technologie, gepaart mit superintelligenter KI, hervorrufen könnte. So könnten beispielsweise Körperbewegungen durch Gedanken gesteuert werden und damit Lähmungen beim Menschen nach Rückenmarksverletzungen aufheben. Andererseits könnten genau diese Implantate von BCIs eine große Bedrohung der menschlichen Freiheit darstellen.

Dass die Geschichte des Menschen dadurch gekennzeichnet ist, sich die Natur als Vorbild zu nehmen, um sie sowohl zu überwinden als auch hinter sich zu lassen, zeigen die neuesten technologischen Entwicklungen.

Um sich dem Thema dieser Masterarbeit, unserem Selbstverständnis von Menschsein, weiter anzunähern, wurden in diesem Kapitel über die von Sigmund Freud konstatierten drei Kränkungen der Menschheit Überlegungen angestellt, ob die Wirkmächtigkeit von KI-Systemen nicht zu einer vierten Kränkung der Menschheit führen könnte. Diese Überlegungen führten zum Schluss, dass wenn, wie bei Sartre das *Selbstbewusstsein* eine Voraussetzung für die Freiheit des Menschen ist, dies nicht auch bei mit Bewusstsein ausgestatten superintelligenten KI-Systemen die Freiheit der Maschinen zur Folge haben müsste.

Was dieser Freiheitsbegriff bedeutet und wie Sartres philosophisches Gedankengebäude aussieht, wurde in diesem Kapitel ebenfalls untersucht und dargestellt. Dabei wurden die Kernaussagen zu Sartres Philosophie, dargelegt in seinem Hauptwerk *Das Sein und das Nichts*, sowie grundlegende Begriffe seiner Philosophie, wie beispielsweise das *Für-sich* und das *An-Sich*, erläutert.

Abschließend wurden aus den bisher gewonnenen Erkenntnissen über KI-Systeme sowie über die Philosophie von Sartre Handlungsempfehlungen im Umgang mit KI-Systemen im gesellschaftspolitischen Spannungsfeld gemacht. Die sich dabei eröffnenden ethischen und moralischen Fragestellungen und Herausforderungen aufgrund der Wirkmächtigkeit von KI-Systemen wurden ebenso untersucht wie die Verantwortung in dem identifizierten Handlungsfeld der Politik. Die verhandelten Aspekte einer rechtlichen Regulierung von KI-Systemen beziehen sich ausschließlich auf europäisches Recht. Die rasante Entwicklung von Künstlicher Intelligenz stellt nicht nur die Gültigkeit des von

Sartre definierten Freiheitsbegriffes des Menschen in Frage, sondern betrifft sowohl unsere Lebenswelten im Allgemeinen als auch im Individuellen.

3 Grundlagen und theoretischer Hintergrund

Begriffe und Definitionen sind die Grundlage für die Theoriebildung und haben keinen Selbstzweck. Demnach ist es unerlässlich, sich über den Wert ihrer Bedeutung im Klaren zu sein, auch um als Wissenschaft akzeptiert zu werden und den Wissenschaftskriterien der Überprüfbarkeit und Verständlichkeit gerecht zu werden (Heinrich et al., 2007, S. 59). Die Begriffsbildung und das Entwickeln einer Fachsprache ist ein evolutionärer Prozess, denn ihre Begriffe bilden sich stetig im Verlauf der Entwicklung eines Faches. Auch werden sie durch Definitionen geklärt und gegeneinander abgegrenzt (S. 62).

3.1 Begriffe und Definitionen

Welchen Prozessen wissenschaftliche Begriffsbildung unterworfen ist, beschreiben Heinrich et al. (2007):

> "Der Prozess der wissenschaftlichen Begriffsbildung ist nicht geradlinig. Vielmehr unterliegt er ständigen Anpassungen mit dem Ziel, die Klarheit und Eindeutigkeit der Begriffe und Definitionen zu verbessern und letztlich so zu normieren, dass sich im Laufe der Entwicklung immer mehr am Wissenschaftsprozess Beteiligte der gleichen Begriffe und Definitionen bedienen, wenn die gleichen Objekte gemeint sind." (S.62)

Ein Blick in eine längst vergangene Zeit zeigt, welch evolutionären Weg der Mensch in seinem Ringen um Begriffe und Definitionen und deren Bedeutung hinter sich hat.

3.1.1 Der Universalienstreit

Der Universalienstreit des Mittelalters war eine geistige Auseinandersetzung, welche das intellektuelle Mittelalter zutiefst beschäftigte. Liessmann (2009) erläutert diesen anhand eines der Protagonisten in dieser Auseinandersetzung, dem Frühscholastiker Petrus Abaelardus (1079-1142). Die Beschäftigung mit der aristotelischen Philosophie, von der zu Abaelards Zeiten in Europa nur die logischen Schriften des Aristoteles bekannt waren, war der Ausgangspunkt des Universalienstreites. Die Hauptfrage in diesem Universalienstreit war, wie sich die Offenbarung der biblischen Schriften sowie deren Auslegung durch die Kirchenväter und die Philosophie mit ihrer zentralen Bedeutung menschlicher Vernunft, vereinen lassen.

Ein Problem, welches aus dieser Fragestellung hervorging, war unter anderem, welchen ontologischen Seinsstatus Allgemeinbegriffe wie Lebewesen, Mensch, Rose oder Gerechtigkeit haben:

> "Es gab in einer spätantiken Interpretation des Aristoteles den Hinweis, oder wurde die Frage aufgeworfen, ob diese Allgemeinbegriffe eine eigene Existenz führen und die real existierenden Dinge und Wesen, also

> das einzelne Lebewesen, der einzelne Mensch, die einzelne Blume, sozusagen nur ein Abbild, etwa nach dem Modell der platonischen Ideenlehre, nur ein Abbild dieser in Wahrheit existierenden Allgemeinbegriffe sind, oder ob diese Allgemeinbegriffe nichts anderes sind als Namen, die der Mensch relativ willkürlich wählt, um diese Dinge zu bezeichnen und in Wirklichkeit existieren diese Begriffe als Ideen gar nicht, sondern es existieren nur die einzelnen Dinge, also die einzelne Rose, der einzelne Mensch, das einzelne Tier und es gibt sozusagen in dem Sinn gar keine Idee von Allgemeinheit, der so etwas wie die Qualität eines Seins zugesprochen werden kann." (Liessmann, 2000, Audiofile 8: Der Name der Rose, Minute 01:53)

Als Realisten wurden jene bezeichnet, welche davon ausgingen und davon überzeugt waren, dass Allgemeinbegriffe als Allgemeinbegriffe tatsächlich real existieren. Jene, die davon ausgingen, dass es keine Allgemeinbegriffe gibt, sondern nur die Namen die der Mensch willkürlich wählt, wurden Nominalisten genannt. Abaelardus nahm eine die ganze Entwicklung des Mittelalters beeinflussende Position ein. Nach ihm ist nur das Einzelseiende tatsächlich real, denn durch eine eigenständige Realität des Allgemeinen würde sich das Problem ergeben, das Universalien teilbar oder vervielfältigbar sein müssten und sich dadurch der Unterschied zum Einzelnen aufheben würde (Liessmann, 2000).

Würde ein Allgemeinbegriff wie zum Beispiel *Lebewesen* tatsächlich real existieren, dann wäre es ein in sich widersprüchlicher Begriff. Zu Lebewesen zählen wir nämlich vernünftige Lebewesen, also den Menschen, aber auch unvernünftige Lebewesen wie Tiere. Somit also müsste der existierende Begriff Lebewesen zugleich vernünftig als auch unvernünftig sein, was jeglicher Logik widerspricht (Liessmann, 2000).

Dass die Ideen zwar etwas Allgemeines sind, aber gleichzeitig auch, indem sie von den Dingen getrennt sind, etwas Einzelnes sind, bemerkt schon Aristoteles. So ist auch für Aristoteles nur das Allgemeine wissenschaftsfähig, hat aber keine eigene Realität, sondern ist eine Leistung des Denkens, welche ihre Fundierung in den Einzeldingen hat (Vorländer, 1990, S. 58-59).

Allgemeinbegriffe und Universalien sind demnach nach Abaelardus das Resultat einer logischen und sinnfälligen Abstraktionsleistung des Menschen. Diese geistigen Konzepte, die der Mensch entwirft, um sich die Welt durch Vernunft und eben auch mittels Allgemeinbegriffen zu ordnen, werden Konzeptualismus genannt. Eine Frage die Abaelardus ausführlich diskutiert hat war, was ein Begriff bedeutet, mit dem der Mensch Dinge belegt, die gar nicht mehr existieren. Was würde also der Name der Rose für einen gedachten Fall bedeuten, dass es gar keine Rosen mehr gibt (Liessmann, 2000).

Begriffe und Worte können nach Abaelardus somit in unterschiedlicher Bedeutung verwendet werden. Zum einen zur Bezeichnung realer Dinge, zum anderen aber auch in dem Sinn, indem wir ihnen eine Bedeutung in einem

geistigen Horizont verleihen, denen nicht etwas in der sinnlich erfahrbaren Realität entsprechen muss.

3.1.2 Wittgenstein - Der Gebrauch allgemeiner Begriffe

In seinen Überlegungen, die Ludwig Wittgenstein in seinem Spätwerk, den *Philosophischen Untersuchungen*, anstellt, geht er davon aus, dass die Bedeutung eines Wortes nicht von gegenständlicher Natur ist. Dabei steht für ihn außer Frage, dass es unmöglich sei, endgültige Definitionen von Wörtern zu geben. Die Bedeutung eines Begriffes könne nur zu einem bestimmten Zweck eingegrenzt werden. Somit können niemals alle Bedeutungen eines Wortes angegeben werden und wenn, dann nur für bestimmte Situationen. Eine Frage nach *der* Bedeutung eines Wortes stellt sich somit nicht:

> "Man kann für eine große Klasse von Fällen der Benützung des Wortes »Bedeutung« - wenn auch nicht für alle Fälle seiner Benützung - dieses Wort so erklären: Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache. Und die Bedeutung eines Namens erklärt man manchmal dadurch, daß man auf seinen Träger zeigt." (Wittgenstein, 2022, § 43, S. 262-263)

3.1.3 Husserl - Von den bloßen Worten zu den Sachen

Die zuvor in ihrem evolutionären Prozess dargestellten Entwicklungen von Begriffsbildungen und Definitionen sind im Besonderen für den Begriff und die Definition von Künstlicher Intelligenz von größter Wichtigkeit. So wie Wittgenstein feststellt, ist eine Eingrenzung eines Begriffes und dessen Bedeutung nur zu einem bestimmten Zweck möglich. Für den Begriff KI gilt es daher, sowohl die Bedeutung des Adjektivs *künstlich* als auch des Substantives *Intelligenz* einzugrenzen, um ein gemeinsames Verständnis von den Dingen zu haben.

Edmund Husserl stellt sich mit seinem Entwurf *von den bloßen Worten zu den Sachen selbst* gegen alle Spekulation, die sich bloß mit Begriffen und Worten beschäftigt. Für Husserl führt der Weg, wie wir zu diesen Sachen selbst kommen, über die Anschauung. Um eine umfassende Erkenntnis dessen, was gegeben ist zu gewährleisten, ist nach Husserl eine dreifache Ausschaltung nötig, von Husserl *Reduktion* genannt. Zunächst muss alles Subjektive vermieden werden. Danach sind alle Annahmen und Hypothesen, welche bislang vor der Betrachtung eines zu erforschenden Gegenstandes existierten, zu reduzieren. Zuletzt muss sich der Forschende von aller Tradition und allem was bislang über diesen Gegenstand gesagt wurde, frei machen (Husserl, 2002).

Laut Husserl sind wir überzeugt, dass es Dinge gibt die bestimmte Eigenschaften haben und wir diese erkennen können. Dabei gilt es aber erst herauszufinden, wie genau wir das leisten können. Am Beispiel eines bis zur Hälfte gefüllten

Wasserglases lässt sich die Methode der Reduktion erläutern. Betrachten zwei Menschen ein bis zur Hälfte gefülltes Glas, so mag es für den einen halbvoll, für den anderen halb leer sein.

Bezogen auf Husserls Entwurf *von den bloßen Worten zu den Sachen selbst* würde sich in diesem Fall die Frage stellen, wie diese beiden Personen ausschließen können, dass sie tatsächlich ausschließlich über das Glas und seinen Inhalt, nicht jedoch über ihre seelischen Zustände, die sie mit dem Inhalt des Glases verbinden, sprechen.

Wir Menschen neigen dazu, in das, was wir sehen etwas hineinzuinterpretieren was nicht vorhanden ist und somit das Ergebnis unserer Erkenntnis bereits feststeht, bevor wir uns noch mit dem Phänomen auseinandergesetzt haben. Die Reduktion wäre laut Husserl in der Lage, dies auszuschließen.

> "Mit anderen Worten: Vergeude deine Zeit nicht mit Interpretationen, die den Dingen etwas zuschreiben, und schon gar nicht mit der Frage, ob die Dinge real sind. Betrachte einfach nur *das*, was sich dir präsentiert, was auch immer *es* sei, und beschreibe es so präzise wie möglich". (Bakewell, 2021, S. 14)

3.2 Forschungsgegenstand Künstliche Intelligenz

KI hat sich in den letzten sechs Jahrzehnten rasant entwickelt. Zurückzuführen ist dies auf den rasanten Anstieg der Rechenleistung von Computern, den Anstieg von digitalen Daten aufgrund wirtschaftlicher Transformation und sozialer Medien sowie der Einführung digitaler Produkte, die das Rechts- und Ethiksystem herausfordern. Beispiele für den Einsatz von KI sind an die jeweiligen Nutzerprofile angepasste Ergebnisse von Suchmaschinen oder Navigationssysteme in Fahrzeugen, die Verkehrsteilnehmer*innen basierend auf aktuellen Verkehrsdaten die optimale Route berechnen. Diese Funktionen sind heute so selbstverständlich, dass KI kaum mehr wahrgenommen wird.

Die Leistungsfähigkeit von Computern hat in den letzten Jahren signifikante Fortschritte gemacht, sodass sie eine enorme Rechenleistung und eine unglaubliche Menge an Daten bereitstellen. Diese Rechenleistung ist der Motor der KI, während die Daten der Treibstoff sind. Beispielsweise ist die Rechenleistung eines Smartphones von heute mehrere Millionen Mal höher als die des NASA-Computers, der 1969 für die Mondlandung verwendet wurde. Auch das Internet des Jahres 2020 ist fast 1 Billion Mal größer als das Internet von 1995 (Lee & Chen, 2022, S. 51).

"If I have seen further, it is by standing on the shoulders of Giants" (Newton, 1675). In einem Brief an Robert Hooke beschreibt Isaac Newton mit der Feststellung, dadurch weiter gesehen zu haben, weil er auf dem Rücken von Giganten stand, ein allgemeingültiges Phänomen der Wissenschaft. Neue Entwicklungen und neue Erkenntnisse in der Wissenschaft sind Resultat

vorangegangener Forschungen und gleichzeitig die Grundlage künftiger Forschungen.

Schmidhuber (Ford, 2019) erläutert, dass die Wiege mechanischer Rechner sowie von Künstlicher Intelligenz in Europa liegt. So wurde der Antikythera-Mechanismus Altgriechenlands aus dem ersten Jahrhundert v. Chr. an Raffinesse erst eineinhalb Jahrtausende später durch Henleins miniaturisierte Taschenuhren im Jahre 1505 übertroffen. Leibnitz entwickelte die heute übliche binäre Computerarithmetik um 1700 und die Mustererkennung durch lineare Regression durch Gauß & Legendre begann um 1800. Die theoretische Informatik im Allgemeinen sowie die KI-Theorie wurde durch Kurt Gödel begründet, der 1931 die erste universelle formale Sprache veröffentlichte (S. 7-8).

Im Jahre 1965 wurde in der Ukraine, die damals noch ein Teil der Sowjetunion war, der erste funktionierende Lernalgorithmus für Netze mit beliebiger Tiefe und mit einer beliebigen Anzahl von Schichten von den beiden Mathematikern Alexey Ivakhnenko und Valentin Lapa veröffentlicht. Damit wurden sie zu den Begründern von *Deep Learning*, einem Machine-Learning Verfahren (Ford, 2019, S. 9).

3.2.1 Geburtsstunde der KI-Forschung

Als Geburtsstunde der *Künstlichen Intelligenz* gilt das Forschungsprojekt am Dartmouth College in Hanover, New Hampshire. Auf der Dartmouth-Konferenz im Jahr 1956 diskutierten zehn Wissenschaftler die sich für neuronale Netze, die Theorie von Automaten oder Intelligenz interessierten, wie man mithilfe von Programmen Künstliche Intelligenz erzeugen könnte (Bostrom, 2020, S. 18). Ausgehend von der Dartmouth-Konferenz entstand eine erste Arbeitsdefinition der KI als jene Wissenschaft, die sich mit der Entwicklung von Maschinen beschäftigt, die sich intelligent verhalten, sich selbst weiter verbessern und bestimmte Probleme lösen, wie sie nur Menschen vorbehalten sind (McCarthy et al., 1955, S. 2).

In Ihrem Antrag an die Rockefeller-Stiftung, welche Mittel für die den sechswöchigen Workshop der Dartmouth-Konferenz bereitstellte, zeigte sich all der Optimismus aller Teilnehmer (Bostrom, 2020, S. 19). Dabei betrachten sie das Problem der *Künstlichen Intelligenz* als dasjenige, dass eine Maschine sich auf solche Weise verhält, die auch als intelligent bezeichnet würde, wenn ein Mensch sich auf eine solche Weise verhalten würde (McCarthy et al., 1955, S. 11).

Es würde jedoch noch ein weiter Weg zu dem werden, was die Teilnehmer dieser Konferenz, unter anderem John McCarthy, der den Begriff *Künstliche Intelligenz* prägte, in ihrer Arbeitsdefinition formulierten. Dass *künstliche Intelligenz* wiederholt einen zu großen Hype abbekommen hat, und dies sogar von einigen Begründern des Fachgebietes, zeigt sich daran, dass schon bei der Dartmouth-Konferenz darüber nachgedacht wurde, dass innerhalb von zwei Monaten in

bestimmten Problembereichen bedeutsame Fortschritte erzielt werden können. So zum Beispiel der Versuch, Maschinen dazu zu bringen, Sprache zu benutzen (Tegmark, 2020, S. 66).

John McCarthy selbst hat dann später auf die Frage, was *Künstliche Intelligenz* sei, geantwortet, dass sie die Wissenschaft und Technik der Entwicklung intelligenter Maschinen sei, insbesondere der Entwicklung intelligenter Computerprogramme. Diese Wissenschaft ist zwar verwandt mit einer ähnlichen Aufgabe, die Computer nutzen, um menschliche Intelligenz zu verstehen, jedoch muss sie sich nicht auf biologisch beobachtbare Methoden beschränken (McCarthy, 2007, S. 2).

Ob es eine solide Definition von *Intelligenz* gibt, die sich nicht auf menschliche Intelligenz bezieht, verneint McCarthy (McCarthy, 2007) deswegen, weil wir noch nicht allgemein beschreiben können, welche Arten von Rechenverfahren als intelligent bezeichnet werden sollen. Einige der Mechanismen von Intelligenz verstehen wir, andere noch nicht. Tegmark (Tegmark, 2020) definiert den Begriff *Intelligenz* als die Fähigkeit, komplexe Ziele zu erreichen, um gleichzeitig den Begriff *Künstliche Intelligenz* als nichtbiologische Intelligenz in Abgrenzung zu setzen (S. 63).

3.2.2 Zukunftsvision der KI-Forschung

Einige Zukunftsforscher sagen voraus, dass bis zum Jahre 2045 die Entwicklung von KI in einer *Singularität* münden wird. So wird laut der Singularitätstheorie aufgrund eines exponentiellen Wachstums von Rechenleistung die selbstgesteuerte KI ebenfalls exponentiell wachsen. Dabei wird sie schneller Superintelligenz erwerben als wir uns es vorstellen können, "weil Menschen exponentielle Phänomene nicht begreifen können. Mit anderen Worten: Singularität ist der Moment, in dem Maschinenintelligenz menschliche Intelligenz aussticht – und in dem KI uns Menschen die Kontrolle über unsere Welt entreißen könnte" (Lee & Chen, 2022, S. 525).

Bostrom (Bostrom, 2020) sieht den Grund dafür, dass Maschinen das Potential haben, "wesentlich intelligenter zu werden als jedes Lebewesen, da sie eine Reihe grundlegender Vorteile verfügen; selbst verbesserte biologische Menschen werden weit dahinter zurückbleiben" (S. 80). Die verschiedenen Formen der Superintelligenz werden im Kapitel 3.3.3 behandelt.

Zum Erreichen von Superintelligenz bedarf es jedoch mehr als exponentiell schnellerer Rechenleistung. Wissenschaftliche Durchbrüche "vom gleichen Kaliber wie Deep Learning", so Lee und Chen (2022, S. 526) sind dafür notwendig. So war *Deep Learning* in der 65-jährigen Geschichte der KI die einzige bahnbrechende Erfindung. Um Superintelligenz zu erreichen brauchen wir mindestens ein Dutzend weiterer Superinnovationen (S. 526).

So haben wir nach wie vor keine Antworten auf grundlegende Fragen wie jener nach dem Bewusstsein von Maschinen. "Wir sind nicht nur unfähig, eine KI, die

über Bewusstsein verfügt, zu erschaffen, wir verstehen nicht einmal die dem menschlichen Bewusstsein zugrunde liegenden physiologischen Mechanismen" (Lee & Chen, 2022, S. 526).

3.3 KI-Typen

"*As soon as it works, no one calls it AI anymore*", ist eine Aussage von einem der Pioniere der Künstlichen Intelligenz, John McCarthy. Wie sehr diese Feststellung, dass sobald es funktioniert, niemand es mehr KI nennt, zutrifft, zeigt sich an alltäglichen Beispielen, in denen Künstliche Intelligenz zur Anwendung kommt. Dazu zählen Navigationssysteme in Autos, welche die beste und schnellste Route heraussuchen genauso wie Suchanfragen bei Suchmaschinen wie Google, welche auf ein und dasselbe Suchwort für jeden Nutzer eine maßgeschneiderte, an sein Profil angepasste Antwort ausgeben. Dass hier Künstliche Intelligenz im Spiel ist, wird nicht mehr als solche wahrgenommen, da deren selbstverständliche Funktion in den Hintergrund getreten ist. (Dengel et al., 2023, S. 13).

Die einer Künstlichen Intelligenz zugrunde liegende Technologie kann in mehrere KI-Typen unterteilt werden. Die Methodik der klassischen Verfahren, also *Schwache KI* und *Starke KI*, wird *symbolische KI* genannt, "weil sie Elemente der realen Welt mit Symbolen beschreibt. Symbolische KI ordnet Gegenständen bestimmte Symbole zu, explizit und nachvollziehbar" (Dengel et al., 2023, S. 21).

Die Verfahren der *subsymbolischen KI* stehen im Mittelpunkt maschinellen Lernens, welches ohne eine symbolisierte Wissensbasis auskommt. So müssen Systeme einer subsymbolischen KI lernen, Dinge zu klassifizieren. Um jedoch auf die richtigen Merkmale zu kommen, anhand derer ein System Objekte klassifizieren kann, bedarf es eines aufwendigen Verfahrens, *Deep Learning* genannt (Dengel et al., 2023, S. 22). Im Kapitel 3.4.3 wird *Deep Learning* detailliert beschrieben.

3.3.1 Schwache KI

Ein Ansatz in der KI-Forschung ist es, einen Satz von Algorithmen zu entwickeln, der intelligentes Verhalten simuliert. Entscheidend ist das Endergebnis. Wenn eine Aufgabe von KI gelöst wird, muss das Ergebnis messbar und ähnlich zufriedenstellend wie das eines menschlichen Benutzers sein. So spielt es keine Rolle, ob das System seine Funktionsweise erklären kann, denn die Hauptsache ist, dass die Aufgabe erledigt wird (Ford, 2019, S. 359). Diese Position wird als *schwache KI* bezeichnet. Schwache KI beschränkt sich somit darauf, menschliches Verhalten zu simulieren.

Philosophisch gesehen handelt es sich dabei um den Vergleich zweier Architekturen, nämlich Mensch und Maschine. Darüber hinaus wird traditionell jedoch nicht die Frage nach der Maximierung des erwarteten Nutzens gestellt,

vielmehr geht es dabei um die Frage, ob Maschinen denken können (Russell & Norvig, 2012, S. 1176).

Als eine *schwache KI-Hypothese* wird die Behauptung genannt, "dass Maschinen agieren könnten, *als ob* sie intelligent wären" (Russell & Norvig, 2012, S. 1176). Wenn KI-Forscher die *schwache KI-Hypothese* als gegeben hinnehmen und sich dabei nicht um die *starke KI-Hypothese*, also den Umstand, dass Maschinen wirklich denken und das Denken nicht nur einfach simulieren, kümmern, ist es ihnen egal, ob funktionierende Programme als Simulation der Intelligenz oder als echte Intelligenz bezeichnet werden. Dabei sollten sich KI-Forscher aber auch mit den ethischen Auswirkungen ihrer Arbeit beschäftigen (Russell & Norvig, 2012, S. 1176).

Alan Turing schlug vor, nicht danach zu fragen, ob Maschinen denken können, sondern ob sie einen Verhaltensintelligenztest bestehen können. Denn schon alleine die beiden Begriffe *Maschine* und *Denken* sind, wenn sie den normalen Gebrauch widerspiegeln, gefährlich. Denn wenn die Bedeutung dieser beiden Begriffe dadurch zu finden wäre, indem man untersuchen würde, wie sie üblicherweise verwendet werden, käme man zur Schlussfolgerung, die Antwort könne in einer Gallup Umfrage zu suchen sein. Das jedoch ist absurd und so muss die Frage, ob Maschinen denken können, durch eine andere ersetzt werden (Turing, 1950, S. 433).

Turing untersuchte vielfach mögliche Einwände gegen die Möglichkeit von intelligenten Maschinen. So brachte er unter anderem das *Argument der verschiedenen Behinderungen*, dass Maschinen niemals in der Lage sein werden, X zu tun. Turing schlug dabei zahlreiche Merkmale von X vor:

> "Gutmütig sein, einfallsreich sein, schön sein, freundlich sein, die Initiative ergreifen, Sinn für Humor haben, Recht von Unrecht unterscheiden, Fehler machen, sich verlieben, Erdbeeren und Sahne genießen, jemanden dazu bringen, sich in ihn zu verlieben, aus Erfahrungen lernen, Wörter richtig benutzen, Subjekt seiner eigenen Gedanken sein, sich so vielfältig verhalten wie ein Mensch, etwas wirklich Neues tun." (Turing, 1950, S. 447)

Ein anderes Argument und ein beständiger Kritikpunkt von Turing an KI ist jenes der *Formlosigkeit des Verhaltens*. Er stellt dabei die Behauptung auf, dass "menschliches Verhalten viel zu komplex ist, als dass es durch eine einfache Regelmenge abgedeckt werden könnte, und weil Computer nichts anderes tun als Regelmengen zu folgen, können sie kein Verhalten erzeugen, das so intelligent wie das des Menschen ist" (Russell & Norvig, 2012, S. 1180).

Turing formulierte das Argument, dass es nicht möglich sei, Regeln aufzustellen, die vorgeben, was ein Mensch unter allen denkbaren Umständen tun sollte. So kann man die Regel aufstellen, dass man an Ampeln, die rotes Licht zeigen, stehen bleiben muss und an Ampeln, die

ein grünes Licht zeigen, fahren darf. Was aber geschieht in dem Fall, wenn durch einen Fehler beide Lichter, also rot und grün, erscheinen würden?

> "Man kann vielleicht entscheiden, dass es am sichersten ist, anzuhalten. Aber aus dieser Entscheidung kann sich später eine weitere Schwierigkeit aus dieser Entscheidung ergeben. Der Versuch, Verhaltensregeln für alle Eventualitäten aufzustellen Verhaltensregeln aufzustellen, die alle Eventualitäten abdecken, auch die, die sich aus Ampeln ergeben, erscheint unmöglich." (Turing, 1950, S. 452)

Dieser Umstand, nämlich die Unfähigkeit, alles in einer Menge von logischen Regeln auszudrücken wird auch als *Quantifizierungsproblem* der KI bezeichnet (Russell & Norvig, 2012, S. 1180).

3.3.2 Starke KI

Im Gegensatz zu *schwacher KI* wird *starke KI* mit menschlicher Intelligenz gleichgesetzt. Das Ziel der Entwicklung einer starken KI ist die Schaffung eines künstlichen Wesens, das über ein eigenes Bewusstsein verfügt, Gefühle zeigen kann und letztlich dem Menschen intellektuell ebenbürtig ist. Obwohl diese Vision noch in weiter Ferne liegt, gibt es Forscher, die sich mit der Erreichung dieses Ziels beschäftigen.

Schon Turing (1950) beschäftigte sich mit der Frage, ob Maschinen Bewusstsein entwickeln können. In seinem *Argument des Bewusstseins* zitiert er eine Rede von Jefferson Lister, in der dieser argumentiert, dass, wenn eine Maschine aufgrund ihrer Gedanken und Gefühle, und nicht zufällig auf Symbole stoßend, in der Lage ist ein Sonett zu schreiben oder ein Konzert zu komponieren, man dem zustimmen könne, diese Maschine dem Gehirn gleichzustellen (S. 445).

Turing hätte verschiedene Gründe aufzählen können, dass Maschinen Bewusstsein entwickeln können, doch blieb er dabei, dass die Frage nach dem Bewusstsein einfach falsch definiert sei (Russell & Norvig, 2012, S. 1182). Man könne sich nur dann sicher sein, dass eine Maschine denkt, wenn man selbst diese Maschine wäre und sich daher auch selbst beim Denken fühlen könnte. Diese Gefühle könnte man dann der Welt beschreiben, doch niemand wäre berechtigt, sie zur Kenntnis zu nehmen. Die einzige Möglichkeit dieser Auffassung nach zu wissen, dass ein Mensch denkt wäre, dieser bestimmte Mensch zu sein (Turing, 1950, S. 446).

Der Glaube, den Heiligen Gral der KI-Forschung, *starke KI* oder auch *artificial general intelligence (AGI)* in Bälde zu erreichen, spaltet die intellektuelle Gemeinschaft in das Lager der Utopisten und jenes der Dystopiker (Lee, 2019, S. 186). Die Hoffnungen und Sorgen liegen in der Fähigkeit einer Allgemeinen Künstlichen Intelligenz, jede beliebige kognitive Aufgabe zumindest so zu lösen, wie es ein Mensch täte (Tegmark, 2020, S. 64).

Eine Intelligenz, die sich wie ein Mensch verhält, löst einerseits eine Faszination aus, führt aber andererseits zu Unbehagen, wenn sie versucht, menschenähnliche Entscheidungswege durch Algorithmen nachzubilden und somit eigenständige Entscheidungen trifft (Grunwald, 2019, S. 45-46).

Wie sich die Zukunft der *KI* weiterentwickelt, darüber gehen die Expertenmeinungen auseinander. So würden moderne KI-Projekte von den leistungsstarken Hardware-Komponenten profitieren, die durch Fortschritte in den vielen Teildisziplinen der KI-Forschung, Softwareentwicklung und komputationalen Neurowissenschaften ermöglicht wurden (Bostrom, 2020, S. 37).

Vor allem werden Quantencomputer das Gebiet der KI mit exponentiell noch höherer Geschwindigkeit vorantreiben. Damit werden Quantencomputer "alles verändern, auch die Biologie des Menschen" (Hawking, 2019, S. 75).

3.3.3 Superintelligenz

Tegmark (2020) definiert *Superintelligenz* als eine weit über das menschliche Niveau hinausreichende allgemeine Intelligenz (S. 64). Keine so klare Grenze findet dagegen Russel (2020), wenn er feststellt, dass "es keine klare Grenze gibt, ab der eine KI superintelligent ist. Maschinen leisten in einigen Bereichen bereits heute Übermenschliches" (S. 85). Gleichzeitig wird aber häufig der Fehler gemacht, superintelligenten KI-Systemen göttliche Allwissenheit zuzuschreiben, einem vollkommenen Wissen über die Gegenwart und die Zukunft (S. 105).

Um die Frage zu beantworten, ob und wann es zu dazu kommen wird, dass Maschinen superintelligent sein werden, lohnt sich ein Blick auf die verschiedenen Technologien. Zu diesen zählen "Künstliche Intelligenz, Gehirnemulation, biologischen Kognition, Mensch-Maschine-Schnittstellen sowie Netzwerke und Organisationen" (Bostrom, 2020, S. 41). Die Wahrscheinlichkeit, dass einer dieser Wege zum Ziel führt, erhöht sich aufgrund der vielen Optionen, die vorhanden sind und dann bestimmt der Mensch, "wie wahrscheinlich es ist, dass sie jeweils zur Superintelligenz führen werden" (S.41).

Grundsätzlich wird der Begriff der *Superintelligenz* verwendet, um einen Intellekt zu bezeichnen, "der die intelligentesten Menschen in vielen sehr allgemeinen kognitiven Leistungen weit übertrifft" (Bostrom, 2020, S. 80). Eine Unterscheidung zum Zwecke einer klarer verständlichen Begrifflichkeit von *Superintelligenz* in drei Formen – die *schnelle,* die *kollektive* und die *qualitative Superintelligenz* - dieses Phänomens scheint jedoch sinnvoll (S. 80).

3.3.3.1 Schnelle Superintelligenz

Unter einer *Schnellen Superintelligenz* versteht Bostrom (2020) ein System, welches alles was ein Mensch tun kann, nur viel schneller. So könnte eine zehntausendfach beschleunigte Emulation "ein Buch in wenigen Sekunden lesen und eine Doktorarbeit an einem Nachmittag schreiben" (S. 81). Damit würden

die Ereignisse für ein solches Wesen in der Außenwelt wie in Zeitlupe ablaufen und daher würde aufgrund dieser scheinbaren Zeitdehnung der materiellen Welt "eine schnelle Superintelligenz es lieber mit digitalen Objekten zu tun haben" (S. 81).

3.3.3.2 Kollektive Superintelligenz

Unter einer *Kollektiven Superintelligenz* versteht man ein System, "das aus einer großen Zahl geringerer Intellekte besteht und so zusammengesetzt ist, dass die Gesamtleistung des Systems jedes andere existierende kognitive System in vielen Bereichen weit übertrifft" (Bostrom, 2020, S. 82) . Durch die Erhöhung der Anzahl beziehungsweise die Erhöhung der Qualität der Bestandteile eines Systems wird sich die kollektive Intelligenz dieses System steigern (Bostrom, 2020, S. 83).

3.3.3.3 Qualitative Superintelligenz

Ein System wird dann *Qualitative Superintelligenz* genannt, wenn es "mindestens so schnell denkt wie ein Mensch und qualitativ erheblich klüger" (Bostrom, 2020, S. 86). Um den Begriff der *qualitativen Superintelligenz* zu veranschaulichen, kann man sich eine Intelligenz vorstellen, welche die Intelligenz des Menschen qualitativ mindestens so sehr übertrifft, wie wir die Intelligenz von Elefanten, Delphinen oder Schimpansen (S. 87). "Hätten dem Homo sapiens zum Beispiel jene kognitiven Module gefehlt, die komplexe sprachliche Repräsentationen ermöglichen, wäre aus ihm vielleicht nur eine weitere Affenart geworden, die im Einklang mit der Natur lebt" (S. 87). Im Umkehrschluss bedeutet dies jedoch, dass wenn der Mensch neue Module in dieser Größenordnung erwerben würde, er dann ebenso superintelligent würde (S.87).

3.4 Unser Gehirn als Vorbild

Das menschliche Gehirn und seine Funktionsweise dient künstlichen neuronalen Netzen als Vorbild. Sobald wir etwas wahrnehmen, geben die Neuronen unseres Gehirns zahllose Impulse weiter. Neuronen bestehen aus mehreren Elementen, die verschiedene Aufgaben erfüllen. Ein Element sind die sogenannten *Dendriten*, über welche chemische oder elektrische Signale eingehen. Über deren Andockstationen, den *Synapsen*, werden diese Signale entweder als erregender oder als ein hemmender Faktor an den Zellkörper weitergegeben. Der Zellkörper sendet seinerseits ein Signal über das *Axon*, den Ausgabekanal (Dengel et al., 2023, S. 29).

"Werden Wahrnehmungen wiederholt, entstehen entlang der Verknüpfungen zwischen den Neuronen bestimmte Aktivierungsmuster, die für eine gewisse Zeit Bestand haben. Sie geben uns, grob gesagt, die Möglichkeit, uns zu erinnern, zu assoziieren oder Dinge in der Welt da draußen einzuordnen" (Dengel et al., 2023, S. 29).

Genau dieser Ansatz, so Dengel (2023), kann auf künstliche Neuronen übertragen werden, welche sich mathematisch beschreiben und simulieren lassen. Die Anordnung künstlich neuronaler Netze erfolgt in Schichten, ausgehend von der Eingabeschicht bis zur Ausgabeschicht. Damit ähnelt sie dem visuellen Kortex unseres Gehirns. So erhalten beispielsweise Neuronen an einer Eingabeschicht Impulse von außen (S. 29).

> "Die Impulse werden an Neuronen der nächsten Schicht weitergegeben, dort mittels bestimmter Gewichte verstärkt oder abgeschwächt und mit anderen Impulsen kombiniert. In jedem Knoten des Netzes wird entschieden, ob das Signal weitergeleitet wird oder nicht, ähnlich wie im natürlichen Gehirn. Ist die letzte Schicht erreicht, wird ein Muster ausgegeben, das als Antwort auf die Eingangssignale verstanden werden kann." (Dengel et al., 2023, S. 29-30)

Da die Kombinationen solcher künstlichen Neuronen sehr unterschiedlich sein können, ergeben sie eine Vielzahl sogenannter Netztypologien. Diese variieren je nach Problemstellung in der Anzahl von Neuronen pro Schicht und in der Anzahl der Schichten. Der Grad der Vernetzung von Neuronen zwischen den Schichten fällt ebenso unterschiedlich aus. Es können daher umso mehr Merkmale gelernt werden, je tiefer ein Netzwerk ist und je mehr Schichten es hat. So gibt es bereits Netzwerke, deren Lernfähigkeit bei mehreren hundert Millionen Merkmalen liegt (Dengel et al., 2023, S. 30).

3.4.1 Machine Learning

Ein Teilbereich der KI ist das *Machine Learning*. Dabei werden mithilfe eines stetigen maschinellen Lernens IT-Systeme verbessert und erweitert. Daher bezeichnet der Begriff *Machine Learning* die Lernfähigkeit einer Software oder eines Computers. Implementierte Algorithmen können anhand vorhandener Daten Muster, Gesetzmäßigkeiten und Regeln erkennen und die entsprechenden Lösungen entwickeln. Auf künstlicher Intelligenz basierende Systeme nutzen maschinelles Lernen für unterschiedlichste Prozesse. Dazu zählen das Aufspüren, Extrahieren und Zusammenfassen essentieller Informationen und Daten, Vorhersage spezifischer Ereignisse, Optimieren von Prozessen, die Errechnung von Wahrscheinlichkeiten und die selbständige Weiterentwicklung bestehender Prozesse (Kuhlmann, 2018, S. 19-20).

Grundsätzlich ist maschinelles Lernen eine Sammlung von Methoden, "die in Daten der Vergangenheit nach Mustern suchen, die für die Zukunft Vorhersagen erlauben. Meistens wird mithilfe einer Grundwahrheit gelernt, das heißt: die Daten einer Person sind mit ihrem vergangen Verhalten verknüpft" (Zweig, 2019, S. 316). Die Methoden maschinellen Lernens versuchen auf die Grundwahrheit wie zum Beispiel, dass ein Bewerber A eingestellt wurde und ein Bewerber B nicht, jene Eigenschaften zu identifizieren, "die oft mit dem einen Verhalten und selten mit dem anderen Verhalten gefunden werden" (S. 316).

Ford (2019) beschreibt maschinelles Lernen als ein Teilgebiet der KI, das sich mit der Erstellung von Algorithmen befasst, die aus Daten lernen können. Demnach wären Machine-Learning-Algorithmen Programme, die sich, indem sie auf Informationen zurückgreifen, selbst programmieren (S. 24).

3.4.2 Formen von Machine Learning

Die Methoden maschinellen Lernens werden in unterschiedlichsten Bereichen wie etwa zur Erkennung von Objekten in Bildern, in der Übersetzung von Sprachen oder der Spracherkennung eingesetzt. Für maschinelles Lernen existieren unterschiedliche Rechenverfahren. So beruhen manche Systeme auf erlernten Regeln, andere wiederum setzen zum Lernen künstliche neuronale Netze ein, die den Strukturen des Gehirns ähneln.

Je nach Zweck gibt es unterschiedliche Formen des maschinellen Lernens. Im Folgenden werden die verschiedenen Möglichkeiten, Machine-Learning Systeme zu trainieren, erörtert (Dengel et al., 2023, S. 23).

3.4.2.1 Supervised Learning

Eine Form des maschinellen Lernens ist das *supervised learning*, auch *überwachtes Lernen* genannt. Dieses Lernverfahren für Maschinen arbeiten demnach mit einem Lehrer.

> "Beim Lernen mit Lehrer soll der Agent anhand von Trainingsdaten eine Abbildung der Eingabevariablen auf die Ausgabevariablen lernen. Wichtig ist hierbei, dass für jedes einzelne Trainingsbeispiel sowohl alle Werte der Eingabevariablen als auch alle Werte der Ausgabevariablen vorgegeben sind. Man braucht eben einen Lehrer, beziehungsweise eine Datenbank, in der die zu lernende Abbildung für genügend viele Eingabewerte näherungsweise definiert ist." (Ertel, 2021, S. 351)

Das bedeutet, dass in der Trainingsphase zu den Beispielsdaten die richtigen Antworten als sogenannte Labels mitgeliefert werden. Damit lassen sich während des Trainings falsche Antworten korrigieren. Am Ende wird aus allen Beispielen ein verallgemeinertes Modell gelernt.

Machine Learning ähnelt jenem Lernprozess, den auch wir als Menschen zeitlebens durchlaufen. Denn schon im "Kindesalter beginnt das Lernen, indem Objekte auf Bildern analysiert und zugeordnet werden" (Kuhlmann, 2018, S. 21).

Erst wenn einem Lernalgorithmus sorgfältig vorbereitete Trainingsdaten, welche auch kategorisiert oder gekennzeichnet sind, bereitgestellt werden, kann von einem überwachtem Lernen gesprochen werden. Das überwachte Lernen ist bei aktuellen KI-Systemen das am meisten verwendete Verfahren und es wird in 95% der Anwendungen, wie etwa der Übersetzung bei Fremdsprachen, eingesetzt. Das Problem beim überwachten Lernen ist die Unmenge an

gekennzeichneter Daten, die für das Training benötigt werden. "Das erklärt, weshalb Unternehmen wie Google, Amazon oder Facebook, die über gigantische Datenmengen verfügen, bei der Deep-Learning-Technologie eine so dominierende Stellung einnehmen" (Ford, 2019, S. 25-26).

3.4.2.2 Unsupervised Learning

Das *unsupervised learning*, auch *unüberwachtes Lernen*, bedeutet, "dass Maschinen direkt aus unstrukturierten Daten lernen, die ihrer Umgebung entstammen" (Ford, 2019, S. 26). Diese Art des Lernens kann bei kleinen Kindern beobachtet werden, wenn sie das Sprechen lernen, indem sie ihren Eltern zuhören (S. 26).

Wenn KI-Systeme in derselben Art wie Kinder lernen, sind Systeme denkbar, die selbstständig lernen ohne dass sie große Mengen an gekennzeichneten Trainingsdaten benötigen. Dies gilt allerdings als eine der größten Herausforderungen. "Ein Durchbruch, der es ermöglichen würde, dass Maschinen tatsächlich unüberwacht effizient lernen, wäre eines der bedeutendsten Ereignisse in der Geschichte der KI und ein Meilenstein auf dem Weg zur KI auf menschlichem Niveau" (Ford, 2019, S. 26-27).

Unüberwachtes Lernen bedeutet demnach, das Trainingsdaten ohne Zusatzinformationen in das System eingegeben werden. Das System soll nun in den Daten Muster entdecken oder Gruppen von gleichen oder ähnlichen Beispielen zusammenstellen.

Das *Clustern*, also das Erkennen möglicher nützlicher Cluster von Eingabebeispielen, ist eine der häufigsten Aufgaben für unüberwachtes Lernen. "So könnte ein Taxi-Agent allmählich ein Konzept von guten Verkehrstagen und schlechten Verkehrstagen entwickeln, ohne jemals benannte Beispiel dafür von einem Lehrer bekommen zu haben" (Russell & Norvig, 2012, S. 811).

3.4.2.3 Reinforcement Learning

Beim *reinforcement learning*, auch *bestärkendes Lernen*, ist die Situation eine ungleich schwierigere als jene beim *supervised learning*, da keine Trainingsdaten verfügbar sind (Ertel, 2021, S. 351). Ford (2019) argumentiert, dass das Lernen demnach ausschließlich durch die Trial-and-Error-Methode oder durch Üben erfolgt. "Anstatt einen Algorithmus durch die Bereitstellung korrekt gekennzeichneter Daten zu trainieren, überlässt man es dem System, selbst eine Lösung zu finden, und wenn es erfolgreich ist, gibt es eine »Belohnung«" (S. 26).

Im Gegensatz zum *Clustern*, bei dem ein System Konzepte entwickelt, ohne vom Lehrer dafür benannte Beispiele zu bekommen, erfolgt beim *reinforcement learning* das Lernen demnach "aus einer Reihe von Verstärkungen – Belohnungen oder Bestrafungen. Beispielsweise könnte das Fehlen eines Trinkgeldes am Ende einer Fahrt für den Taxi-Agenten ein Hinweis darauf sein, dass er etwas falsch gemacht hat" (Russell & Norvig, 2012, S. 811).

Demis Hassabis, Mitbegründer und CEO der Firma *Deep Mind*, geht davon aus, dass "Reinforcement Learning in den nächsten paar Jahren so bedeutend werden wird wie Deep Learning" (Ford, 2019, S. 176). Den Grund dafür sieht Hassabis in Erkenntnissen der Neurowissenschaft, demzufolge das menschliche Gehirn eine Form des bestärkenden Lernens als eines seiner Lernmechanismen einsetzt.

Dies wird als *Temporal Difference Learning* bezeichnet, "und wir wissen, dass Neurotransmitter wie Dopamin diesen Mechanismus nutzen. Die dopaminergen Neuronen überwachen die Vorhersagefehler, die ihr Gehirn macht, und dann wird die Erregung der Nervenzellen an den Synapsen entsprechend der Belohnungssignale verstärkt" (Ford, 2019, S. 176). Da also das Gehirn nach diesem Prinzip funktioniert und es das einzige Beispiel für eine allgemeine Intelligenz ist, die wir kennen, ist die Neurowissenschaft ernst zu nehmen (S. 176).

3.4.2.4 Verfahren maschinellen Lernens

Maschinelles Lernen braucht bestimmte Voraussetzungen für die zu verwendenden Beispielsdaten während einer Trainingsphase. So muss die Datenmenge entsprechend groß sein und in den Beispielsdaten müssen Muster erkennbar sein, aufgrund derer Einteilungen in Kategorien vorgenommen werden können. Erst danach stellt sich die Frage der Methodik. Dabei beruht das maschinelle Lernen auf drei Verfahren – dem der *Klassifikation*, des *Clustering* und der *Regression* (Dengel et al., 2023, S. 26).

Unter der *Klassifikation* versteht man die systematische Zuordnung von Beispielen zu bereits bekannten Klassen, welche klar voneinander abgrenzbar und geeignet sein müssen, in die Daten eine Grundordnung zu bringen. "Die Zuordnung zu einer Klasse beruht auf bestimmten Merkmalen, und die Menge der Klassennamen bildet ein Vokabular. Das sind die sogenannten Labels" (Dengel et al., 2023, S. 26).

Beim Verfahren des *Clustering* sind die Klassennamen vor Anwendung des Verfahrens unbekannt und stattdessen übernimmt ein Algorithmus die Aufgabe, "für eine Menge von Beispielen ein Modell zu konstruieren, das die Elemente der Datenmenge nach Ähnlichkeit beziehungsweise Übereinstimmung von Merkmalen sortiert" (Dengel et al., 2023, S. 26). Das Ergebnis sind dann Cluster, also die Häufung von ähnlichen Elementen (S. 26).

Die Suche nach einem mathematischen Zusammenhang zwischen zwei Merkmalen wird *Regression* genannt. Bei diesem Verfahren besteht das Ziel darin, "für den Wert einer sogenannten Zielvariable mithilfe einer anderen Variablen eine Vorhersage zu treffen" (Dengel et al., 2023, S. 26).

3.4.3 Deep Learning und Neuronale Netze

Deep Learning ist das spannendste und vielversprechendste Gebiet des maschinellen Lernens. "Mithilfe neuronaler Netze und der Zufuhr sehr großer Datenmengen verfolgt das Deep Learning das Ziel, Lernmethoden nach dem Vorbild des menschlichen Gehirns zu schaffen" (Kuhlmann, 2018, S. 36). Neuronale Netze sind die Basis der Forschung für *Deep Learning*. Das Grundprinzip hinter der Lernmethode ist, "dass die Vernetzung einzelner Recheneinheiten, den sogenannten *Neuronen*, für die intelligente Verarbeitung von Informationen und Daten genutzt wird" (S. 39).

> "Neuronale Netze repräsentieren komplexe nichtlineare Funktionen mit einem Netz aus Einheiten mit linearen Schwellenwerten. Mehrschichtige neuronale Feedword-Netze sind in der Lage, jede beliebige Funktion darzustellen sofern genügend Einheiten vorhanden sind. Der Backpropagation-Algorithmus (Fehlerrückführung) implementiert einen Gradientenabstieg im Parameterraum, um den Ausgabefehler zu minimieren." (Russell & Norvig, 2012, S. 883)

Unter einem *Feedword-Netz* versteht man die Architektur eines neuronalen Netzes. Dessen Struktur besteht aus einem neuronalen Netz "inklusive einer Inputschicht, mehreren Hidden Layers und einer Outputschicht. Feedword-Netze zeichnen sich durch die Gewichtung der Neuronen vorheriger und folgender Schichten aus" (Kuhlmann, 2018, S. 43).

Unter einer *Backpropagation* versteht man einen Lernalgorithmus, den Deep-Learning-Systeme verwenden. Dabei breiten sich beim Trainieren eines neuronalen Netzes die Informationen rückwärts durch die Neuronenschichten aus, die das Netz bilden. Dadurch wird eine Neukalibrierung der Gewichte der einzelnen Neuronen bewirkt. Somit kommt das Netz der richtigen Lösung immer näher (Ford, 2019, S. 25).

Software, die die Funktionsweise von Neuronen im menschlichen Gehirn emuliert und dabei tiefe, aus vielen Schichten bestehende, künstliche neuronale Netze verwendet, ist demnach ein Machine-Learning Verfahren, welches Deep Learning genannt wird (Ford, 2019, S. 24).

Seit den Anfängen der Forschung zu KI wurde bereits an neuronalen Netzen geforscht. Jedoch blieb der große Durchbruch aus, "denn ein ganz wichtiges Problem, nämlich die Erkennung von Objekten auf Bildern, konnte über viele Jahre nicht gelöst werden" (Ertel, 2021, S. 13). Die Grundlagen für die Methode von Deep Learning wurden dennoch schon in den 1950er Jahren gelegt.

In seinen Forschungen setzte sich Frank Rosenblatt mit den einzelnen Einheiten neuronaler Netze auseinander. Dabei entwickelte er das sogenannte *Perzeptron*, ein einfaches künstliches neuronales Netz, welches aus 400 photosensitiven Einheiten bestand, "die mithilfe einer Zuordnungsschicht an eine

Ausgabeschicht, bestehend aus acht Einheiten, angeschlossen wurde" (Kuhlmann, 2018, S. 37).

Ein Grund dafür, dass die Forschungen Rosenblatts in den folgenden Jahrzehnten nicht weiterverfolgt wurden, lag unter anderem an den zu geringen Rechenleistungen. Beinahe 50 Jahre später gelang Geoffrey Hinton 2006 erstmals ein Training eines neuronalen Netzes, das aus mehreren Schichten bestand (Kuhlmann, 2018, S. 37-38).

Das Problem der Objekterkennung war plötzlich gelöst und es eröffneten sich neue Möglichkeiten zum Bau von autonomen Systemen. So sind Fahrerassistenzsysteme mit sehr guter Objekterkennung und Lokalisierung schon heute im Einsatz, und "die KI auf dem PC oder Handy kann problemlos die Familienfotos einzelner ausgewählter Personen finden" (Ertel, 2021, S. 13). Ebenso zeigen Deep Learning Systeme ihr Stärken in der Medizin, wo bei der Diagnose von Krankheiten anhand von bildgebenden Verfahren die Fehlerrate in vielen Bereichen weit unter der von Ärzten liegt (S.13).

Die Fortschritte bei Deep Learning Systemen sind so groß, dass das "Deep Learning anhand neuronaler Netze fast schon zum Synonym des maschinellen Lernens wurde" (Kuhlmann, 2018, S. 38).

3.5 Jean-Paul Sartre und sein Wirken

Jean-Paul Charles Eymard Sartre, geboren am 21. Juni 1905 in Paris, schreibt im Jahre 1962 über sich selbst: "Ich wollte im reinen Äther leben, unter den luftigen Trugbildern der Dinge. Weit davon entfernt, mich an Luftballons anklammern zu wollen, habe ich mich später mit ganzem Eifer bemüht, nach unten zu gelangen; dazu braucht man Sohlen aus Blei" (Wildenburg, 2004, S. 9).

Sartre, ein Philosoph und Schriftsteller zahlreicher Romane und Bühnenwerke, dessen weltweit bekannte Theaterstücke 1948 vom Vatikan auf den Index gesetzt werden. Sartre, ein Agitator und Provokateur, der 1960 das *Manifest der 121* über das Recht zum Ungehorsam in Algerien unterzeichnet, welches zur Kriegsdienstverweigerung und Desertion aufruft und Verhaftungen und Prozesse zur Folge hat. "Nur die seine nicht. »Einen Voltaire verhaftet man nicht«, lässt der französischen Staatspräsident de Gaulle verlauten" (Wildenburg, 2004, S. 10).

Es ist Sartre, dessen Arbeitsplatz die Öffentlichkeit ist und dem es gelingt, durch Alltagsbeispiele die Menschen von seiner Philosophie zu begeistern. So werden in Cafés, Salons und verrauchten Jazzkneipen Fragen von philosophischer Reichweite diskutiert. "In Frankreich, wo weder Durkheim noch Brunschvicg im Café arbeiteten, löst dieser gelebte Angriff auf akademische Routinen Empörung bei Philosophie-ProfessorInnen aus, die Sartre fortan als unbedeutend und suspekt […] herabwürdigen, da es ihm an Strenge und Ernst fehle," argumentiert Werner (2021, S. 237).

Damit beschreitet Sartre aber neue Wege und ein Feld, das "immer auch ein Raum des Möglichen („un espace des possibles“) ist, dessen AkteurInnen die in ihm waltenden Kräfteverhältnisse, Grenzen und Gesetze selbst erschaffen" (Werner, 2021, S. 237). Durch Sartre wird das Philosophieren somit nicht zu einer ausschließlichen Angelegenheit von Expert*innen, sondern findet seinen Weg mitten in die Gesellschaft.

Betrachtet man die philosophischen Grundlegungen Sartres in seinem 1943 erschienen Hauptwerk *Das Sein und das Nichts*, in dem er seine Freiheitsidee entwirft, in seinem historischen Kontext, so geht diesem zum einen eine ideengeschichtliche Entwicklung voraus. Der Mensch wurde durch die von Freud beschriebenen drei Kränkungen in seinem Selbstbild als die Krone der Schöpfung immer mehr entmachtet. Zum anderen schreibt Sartre sein Hauptwerk in einer Zeit, in der durch die Herrschaft der Nationalsozialisten alle Werte und Normen der abendländischen Kultur zerstört und vernichtet waren (Wildenburg, 2011).

Sartre bot jedoch in dieser Zeit keine neuen Werte an, wollte in einer Zeit, in der alle Welt neue Heilsversprechen machte, genau diese nicht machen, da es dieses Heil eben nicht gäbe. Vielmehr zeigte Sartre auf, dass der Mensch alleine ist, ohne Entschuldigung und Rechtfertigung in seiner Existenz. Philosophie hat daher, so Sartre, die Aufgabe, alle Ideologien zuerst aufzudecken und dann zu zerstören, und diese "in einem Säurebad aufzulösen, um zu zeigen, wie weit diese Ideologien eben auch dahin führen den Menschen [ja] zu verblenden, eben mit falschen Heilsversprechen irgendwohin zu führen, was eben nicht seiner Freiheit entspricht" (Wildenburg, 2011, 08:31-08:46).

3.6 Der Existenzialismus

In einer Zeit größter Unterdrückung, der Herrschaft der Nationalsozialisten und der Besetzung Frankreichs, wird Sartre mit seinem 1943 erschienenem Hauptwerk *Das Sein und das Nichts* zum Mitbegründer des französischen Existenzialismus. Sartre (2021) definiert den Existenzialismus in seinem 1946 erschienen Werk *Der Existenzialismus ist ein Humanismus* als "eine bestimmte Betrachtungsweise der menschlichen Fragen, die es ablehnt, dem Menschen eine für immer festgelegte Natur zuzuschreiben" (S. 116).

Dass der französische Existenzialismus damals von einem erklärten Atheismus begleitet wurde, sei nach Sartre aber absolut nicht notwendig. Der Mensch sei eben ausschließlich durch das Handeln definiert, in Wirklichkeit "kann der Mensch nur handeln; seine Gedanken sind Entwürfe und Verpflichtungen, seine Gefühle Unternehmungen; er ist nichts anderes als sein Leben, und sein Leben ist die Einheit seiner Verhaltensweisen" (Sartre, 2021, S. 116-117).

Genau diese Situation, in der sich der in die Welt geworfene Mensch befindet, löst bei ihm ein Gefühl der Angst aus. Und wie sollte es auch anders sein.

> "Wenn der Mensch *nicht ist*, sondern *sich schafft*, und wenn er, indem er sich schafft, die Verantwortlichkeit für die ganze Gattung Mensch übernimmt, wenn es weder einen Wert noch eine Moral gibt, die *a priori* gegeben sind, sondern wenn wir in jedem Fall allein entscheiden müssen, ohne Stütze, ohne Führung und dennoch *für alle*, wie sollten wir da nicht Angst haben, wenn wir handeln müssen?" (Sartre, 2021, S. 117)

Diese Angst, so Sartre (2021), ist "keineswegs ein Hindernis für das Handeln, sondern vielmehr dessen Voraussetzung" (S. 117).

Doch nicht alles im Werk von Sartre zeichnet sich durch eine radikale Neuheit aus, was in der Philosophie danach mit dem Etikett *Existenzialismus* versehen wurde. Die Originalität liegt vor allem darin, vorhandene Ansätze der Philosophie Husserls oder Heideggers weiterzuentwickeln und zu einem neuen Denken zu verbinden (Werner, 2021).

> "Die Überlegung, dass die Existenz der Essenz vorangeht und der ungefragt in die Welt und damit in Angst, Verlassenheit und Hoffnungslosigkeit geworfene Mensch sich selbst wählen muss, hat nach dem Einschnitt des Zweiten Weltkriegs ein gänzlich anderes Gewicht als noch zehn Jahre zuvor." (S. 20)

Sartre ist, bevor er sein Konzept von Freiheit entwirft, Husserlianer, nachdem er sich mit der Philosophie der Phänomenologie auseinandergesetzt hatte (Werner, 2021, S. 22). Bakewell (2021) argumentiert, dass im Gegensatz zu den bisher abstrakten Theorien der Philosophen für Phänomenologen das Leben selbst Gegenstand der Betrachtung ist, das Leben wie es Moment für Moment erfahren wird. Dabei wird das meiste, was die Philosophie bisher beschäftigte, wie beispielsweise die Frage, ob die Dinge real sind, ausgeklammert (S. 14).

Vielmehr weisen die Phänomenologen darauf hin, dass jeder, "der diese Fragen stellt, immer schon hineingeworfen ist eine Welt voller Dinge – voller «Phänomene», ein Wort, dass im Griechischen so viel bedeutet wie «das, was erscheint»" (Bakewell, 2021, S. 14).

Simone de Beauvouir (2017) umschreibt diese vollkommenste Voraussetzungslosigkeit der Ideen Husserls, die Sartre antreiben:

> "Wenn er einem Objekt gegenüberstand, so schob er es nicht um eines Mythos, eines Wortes, eines Eindrucks, einer vorgefaßten Idee willens beiseite, sondern schaute es an und ließ es nicht wieder fallen, bevor er nicht sein Wie und Wohin und jeden ihm möglicherweise innewohnenden Sinn verstanden hatte. Er fragte sich nicht, was man denken müßte oder was zu denken pikant oder interessant sein könnte, sondern nur danach, was er wirklich dachte." (S. 490)

3.7 Zusammenfassung

Im Kapitel *Grundlagen und theoretischer Hintergrund* wurde ausgehend von der Erklärung wichtiger Begriffe über die Beschreibung erforderlicher Definitionen und Grundlagen der untersuchten Wissenschaftsbereiche ein Blick in die historischen Entwicklungen über die Bildung von Begriffen geworfen. Dabei wurde über den Universalienstreit des Mittelalters ein Bogen über Ludwig Wittgenstein zur Phänomenologie Edmund Husserls gespannt.

Bei der Untersuchung historischer und zukünftiger Entwicklungen in der Forschung an Künstlicher Intelligenz zeigte sich, dass neue Entwicklungen und Erkenntnisse in der Wissenschaft das Resultat vorangegangener Forschungen und gleichzeitig die Grundlage künftiger Forschungen sind. So ist beispielsweise die Rechenleistung des Smartphones von heute mehrere Millionen Mal höher als die des NASA-Computers, der 1969 für die Mondlandung verwendet wurde.

Als Zukunftsvisionen der KI-Forschung konnten zum einen optimistische Annahmen wie das Entstehen einer *Singularität*, deren Theorie besagt, dass aufgrund eines exponentiellen Wachstums von Rechenleistung die selbstgesteuerte KI ebenfalls exponentiell wachsen wird und sie dabei schneller Superintelligenz erwerben wird als wir uns es vorstellen können, identifiziert werden. Zum anderen bedarf es jedoch für diese Entwicklungen weiterer bahnbrechender Innovationen wie jener des *Deep Learning*, um diese optimistischen Annahmen zu erfüllen.

Die Grundlagen der für diese Entwicklungen notwendigen Technologie wurden in den KI-Typen *Schwache KI*, *Starke KI* und *Superintelligenz* sowie deren Ansätzen in der KI-Forschung beschrieben. Dass das menschliche Gehirn und seine Funktionsweise künstlichen neuronalen Netzen als Vorbild dient und auf diese übertragen werden kann, war ein weiteres Ergebnis der Untersuchungen.

Neben *Machine Learning*, einem Teilbereich der Künstlichen Intelligenz, der die Lernfähigkeit einer Software oder eines Computers bezeichnet, wurden auch die unterschiedlichen Formen von *Machine Learning* identifiziert. Wenn einem Lernalgorithmus sorgfältig vorbereitete Trainingsdaten, welche auch kategorisiert oder gekennzeichnet sind, bereitgestellt werden, spricht man von *Supervised Learning*. Im Gegensatz dazu bedeutet *Unsupervised Learning*, dass Maschinen direkt aus unstrukturierten Daten lernen, die ihrer Umgebung entstammen. Wenn keine Trainingsdaten vorliegen, die einem KI-System bereitgestellt werden, spricht man von *Reinforcement Learning*. Die Verfahren, *Klassifikation, Clustering* und *Regression* sind eine Voraussetzung des maschinellen Lernens.

Als das vielversprechendste Gebiet des maschinellen Lernens gilt *Deep Learning*. *Deep Learning* war auch in der 65-jährigen Geschichte der KI die einzige bahnbrechende Erfindung. *Deep Learning* ist demnach ein Machine-Learning Verfahren, bei dem Software die Funktionsweise von Neuronen im menschlichen

Gehirn emuliert und dabei tiefe, aus vielen Schichten bestehende, künstliche neuronale Netze verwendet.

Abschließend wurden im Kapitel *Grundlagen und theoretischer Hintergrund* das Wirken von Jean-Paul Sartre, biographische Grundlagen zu seiner Person sowie der Einfluss seiner Arbeit auf Philosophie und Gesellschaft dargestellt. Die Grundlagen und historischen Voraussetzungen und Bedingungen, die zum Entwurf seiner Freiheitsidee geführt haben, wurden ebenso untersucht wie der Umstand, dass Sartre mit seinem 1943 erschienenem Hauptwerk *Das Sein und das Nichts* zum Mitbegründer des französischen Existenzialismus wurde.

4 Vorgangsweise & Methoden

Um Daten für diese Studie zu sammeln, liegen der Forschungsstrategie als Forschungsmethoden zum einen das Studium und die Diskussion der für die Beantwortung der Forschungsfrage herangezogenen Literaturbefunde zugrunde. Zum anderen werden im empirischen Teil der Arbeit leitfadengestützte semistrukturierte qualitative Interviews geführt, um Erkenntnisse über die Thesen und Fragestellungen, die zur Beantwortung der Forschungsfrage führen, zu gewinnen.

Die Anlage der Forschung wird in diesem Kapitel behandelt. Dabei werden themenspezifische Implikationen qualitativer Forschung diskutiert sowie der methodische Zugang zum Forschungsprojekt, die *Grounded Theory*, vorgestellt. Den Abschluss des Kapitels bilden zum einen die für den methodischen Zugang notwendigen Instrumente für die Erkenntnisgewinnung, zum anderen die Entwicklung der mittels Kodierparadigma abgeleiteten gegenstandsbezogenen Theorie.

4.1 Motivation und Zielsetzung

Ein wichtiger Schritt im Forschungsprozess ist die methodologische Positionierung, um das Ziel, die Forschungsfrage, mit der Auswahl der richtigen Methoden zu beantworten. "Die Auseinandersetzung mit Grundlagentheorien strukturiert die Wahl der Methoden, sowie die Erhebungs- und Auswertungsverfahren" (Przyborski & Wohlrab-Sahr, 2014, S. 123).

Vor der Durchführung einer Untersuchung ist es im engeren Sinne wichtig, mit Bedingungen des Forschungsbereiches vertraut zu sein, denn die Berücksichtigung der Bedingungen und des Umfangs des Forschungsbereiches kann in einigen Fällen den gesamten Forschungsprozess begleiten (Przyborski & Wohlrab-Sahr, 2014, S. 40). Daher ist bei der Erhebung die Auswahl des Erhebungsinstrumentes Voraussetzung einer gelungenen Forschung. "Nur wenn die Erhebung kompetent durchgeführt wurde, bekommt man Material, das man sinnvoll auswerten kann" (Przyborski & Wohlrab-Sahr, 2014, S. 78).

Um die Zielsetzung, die Forschungsfrage zu diskutieren, zu analysieren und einer Beantwortung zuzuführen, zu erreichen, wurden in dieser Forschung neben dem Studium, der Diskussion und der Auswertung der für die Beantwortung der Forschungsfrage relevanten Literaturbefunde auch Expert*inneninterviews durchgeführt. Misoch (2015) definiert Expert*inneninterviews als eine Methode, bei der unterschiedlichste Formen von semistrukturierten Leitfadeninterviews subsumiert werden (S. 120).

Die Motivation, Expert*inneninterviews durchzuführen lag darin begründet, dass diese Personen im untersuchten Handlungsfeld "eine besondere, mitunter gar eine exklusive Position einnehmen, in der ihnen Wissen zuwächst, das anderen nicht ohne weiteres verfügbar ist" (Strübing, 2013, S. 96).

4.2 Allgemeine Vorgangsweise

Jede Art von Forschung, deren Ergebnisse keinen statistischen Verfahren oder anderen Arten eines quantifizierten Verfahrens entspringen, wird als *qualitative Forschung* verstanden. Dabei kann sich qualitative Forschung beziehen auf Forschung über "Leben, Geschichten oder Verhalten einzelner Personen, aber auch auf Funktionieren von Organisationen, auf soziale Bewegungen oder auf zwischenmenschliche Beziehungen" (Strauss & Corbin, 1996, S. 3).

Auch wenn einige Daten wie Bevölkerungsstatistiken quantifiziert sind, bleibt die Analyse jedoch eine qualitative. Strauss und Corbin (1996) beschäftigen sich mit einer nicht-mathematischen analytischen Vorgehensweise, "deren Ergebnisse aus Daten stammen, die mit einer Vielzahl unterschiedlicher Verfahren erhoben wurden" (S.3). Neben Beobachtungen und Interviews zählen auch Dokumente oder Bücher dazu.

4.2.1 Qualitative Forschungsmethode

Indem qualitative Sozialforschung über die Beschreibung sozialer Prozesse hinausgeht, entwickelt sie gegenstandsbezogene Theorien über den Forschungsbereich. Im Gegensatz zu Theorien mittlerer Reichweite basiert qualitative Sozialforschung daher nicht auf einer hypothesentestenden Forschungslogik.

Durch die Erbringung ihrer durchgängigen intersubjektiven und vermittelnden Interpretations- und Reinterpretationsleistung, nehmen die forschenden Personen im Prozess der empirischen Beforschung eines spezifischen Bereichs eine konstitutive Rolle ein. Damit die forschenden Personen die Komplexität der erforschten Prozesse aber auch angemessen erfassen und erklären können, bedarf es einer umfassenden Überprüfung dieser Interpretations- und Reinterpretationsleistungen (Strübing, 2013, S. 33-47).

Dass es viele gute Gründe gibt, qualitativ zu forschen, begründen Strauss und Corbin (1996) damit, dass es zum einen die auf Forschungserfahrung basierende Überzeugung von forschenden Personen ist, die durch qualitative Methoden der Datenerhebung und Datenanalyse befriedigende Ergebnisse erzielen konnten. Im Besonderen sind es forschende Personen aus den wissenschaftlichen Disziplinen wie zum Beispiel der Anthropologie oder Anhänger philosophischer Richtungen wie der Phänomenologie (S. 4).

Zum anderen liegt ein Grund für die Wahl eines qualitativen Forschungsansatzes darin, dass einige Forschungsgebiete ihrem Wesen nach damit angemessener beforscht werden können. Dabei handelt es sich um Forschungen,

> "über die Art der persönlichen Erfahrung mit Phänomenen wie Krankheit, Glaubenswechsel oder Sucht. Qualitative Methoden können verstehen helfen, was hinter wenig bekannten Phänomen liegt. Sie können benutzt werden, um überraschende und neuartige

> Erkenntnisse über die Dinge zu erlangen, über die schon eine Menge an Wissen besteht." (Strauss & Corbin, 1996, S. 4-5)

Neben diesen beiden Argumenten war auch jener Umstand, dass qualitative Forschung "in Bereichen, die sich mit Themen des menschlichen Verhaltens und Funktionierens beschäftigen" (Strauss & Corbin, 1996, S. 5) mit ein Grund dafür, in der vorliegenden Arbeit eine qualitative Forschungsmethode einzusetzen.

Denn die Besonderheit einer qualitativen Inhaltsanalyse ist, unabhängig vom Ursprungstext, die interpretative Analyse des Textes, wobei im Prozess der qualitativen Inhaltsanalyse nur jene Daten extrahiert werden, die für die Beantwortung der Forschungsfrage bedeutsam sind (Heiser, 2018, S. 112).

4.2.2 Auswahl der Interviewpartner*innen

Durch die Entscheidung, welches Expert*innenwissen Gegenstand der Untersuchung ist, wird die Auswahl der Interviewpartner*innen strukturiert. "Die zentrale Schwierigkeit bei der Befragung von Expertinnen besteht darin, diejenigen Personen zu finden, denen tatsächlich ein entsprechender Expertenstatus zukommt, die also über das gesuchte Betriebswissen oder Deutungswissen verfügen" (Przyborski & Wohlrab-Sahr, 2014, S. 121).

4.2.3 Ablaufschema der Interviews

Konzipiert werden Expert*innengespräche als Leitfadeninterviews, wobei man unter Leitfaden eine Reihe von Sachfragen versteht, welche aus einem Forschungsinteresse heraus abgeleitet sind "und vom Interviewpartner beantwortet werden sollen. Die Vorbereitung der Interviewerin bezieht sich dann in der Regel darauf, sich die für die Untersuchung wesentlichen Fragen zu überlegen und diese in eine angemessene thematische Reihung zu bringen" (Przyborski & Wohlrab-Sahr, 2014, S. 121).

4.2.4 Vorgespräch und Prinzipien der Durchführung der Interviews

Indem man dem bei der Durchführung eines Interviews dem Experten*innenstatus des Gegenübers Rechnung trägt, hat dies Konsequenzen für den Ablauf des Interviews. So ist bereits im Vorgespräch, in dem man das eigene Forschungsinteresse erläutert, das Gegenüber als Expertin oder Experte ansprechen, um dadurch auf das Interesse an ihrer oder seiner spezifischen Kompetenz zu signalisieren. Dabei nimmt man selbst, jedoch auf einem anderen Gebiet, ebenfalls eine Expertenrolle ein. "Man ist also im Verhältnis zum Gegenüber gleich und ungleich: Gleich ist man insofern, als sich hier zwei fachlich kompetente und spezialisierte Personen gegenübersitzen; ungleich ist

man, insofern man das spezifische Wissensgebiet des Anderen nicht kennt" (Przyborski & Wohlrab-Sahr, 2014, S. 122).

Um ein erfolgreiches Expertengespräch zu führen, ist es wichtig, auf Augenhöhe mit der Expertin zu kommunizieren. Dazu gehört, dass man sich vor dem Interview über den Gesprächspartner informiert und welche Position er inne hat. Damit ist sichergestellt, ihn nicht über Sachverhalte zu befragen, für welche er kein Experte ist (Przyborski & Wohlrab-Sahr, 2014, S. 125).

Alle Interviews die im Rahmen dieser Arbeit geführt wurden, wurden nach vorheriger Absprache persönlich und in den von den Gesprächspartner*innen gewünschten Räumlichkeiten geführt. Dieses Setting wurde von allen Expert*innen begrüßt, da im Gegensatz zu Gesprächen über Videokonferenzdienste im persönlichen Gespräch tiefgreifendere Beziehungen zwischen den Gesprächspartner*innen entstehen und sich dadurch eine entspanntere Gesprächsatmosphäre bildet.

4.2.5 Transkription der Interviews

Eine genaue Dokumentation bei der qualitativen Forschung ist für den Forschungsprozess unerlässlich. "Der Begriff der Transkription leitet sich vom lateinischen transcriptio bzw. transcribere ab (Übertragung bzw. um-/überschreiben) und bedeutet in den empirischen Sozialwissenschaften die Verschriftlichung von verbalen oder auch nonverbalen Daten" (Misoch, 2015, S. 249).

Dabei werden audio- oder audiovisuell gespeicherte Daten, die auch als *Sekundärdaten* bezeichnet werden, in Textdaten, auch *Tertiärdaten* bezeichnet, umgeschrieben, um danach einer Analyse zugeführt zu werden. Da hier keine Daten erhoben werden, die numerisch codiert und verarbeitet werden können, sondern diese Daten in komplexer Form als gesprochene Sprache inklusive den Gesten und Bewegungen vorliegen, ist es in der qualitativen Forschung unabdingbar, Transkriptionen durchzuführen (Misoch, 2015, S. 249).

Da es sich beim Prozess der Transkription um einen elementaren Bestandteil für die Qualitätssicherung der qualitativen Daten handelt, muss dieser sehr sorgfältig durchgeführt werden. "Das Datenmanagement und der Transkriptionsvorgang sollten deshalb Teil des Forschungsberichtes sein, damit diese Prozesse intersubjektiv nachvollziehbar und transparent gemacht werden" (Misoch, 2015, S. 251-252).

4.3 Wissenschaftliche Methode

Die wichtigsten Bestandteile qualitativer Forschung sind die Daten, welche aus unterschiedlichsten Quellen wie zum Beispiel Büchern, Beobachtungen oder Interviews stammen. Analytische und interpretative Verfahren sind der nächste Bestandteil, um zu Befunden und Theorien zu gelangen. Als dritten Bestandteil nennen Strauss und Corbin (1996) schriftliche und mündliche Berichte, die "in

wissenschaftlichen Zeitschriften oder auf Tagungen vorgestellt werden und zahlreiche Variationen annehmen. Jemand könnte zum Beispiel entweder einen Überblick über die gesamten Ergebnisse oder eine tiefergreifende Diskussion eines Teils seiner Studie vorstellen" (S. 5-6).

4.3.1 Methodischer Zugang der Grounded Theory

Die *Grounded Theory* wurde von Glaser und Strauss entwickelt und ist ein qualitativer Forschungsansatz. Sie ist "eine gegenstandsverankerte Theorie, die induktiv aus der Untersuchung des Phänomens abgeleitet wird, welches sie abbildet" (Strauss & Corbin, 1996, S. 7). Dabei stehen sowohl die Datensammlung, die Analyse als auch die Theorie in einer Wechselbeziehung zueinander. Strauss und Corbin (1996) argumentieren, dass zu Beginn nicht die Theorie steht "die anschließend bewiesen werden soll. Am Anfang steht vielmehr ein Untersuchungsbereich – was in diesem Bereich relevant ist, wird sich erst im Forschungsprozeß herausstellen" (S. 8).

Eine gut entwickelte Grounded Theory muss, um ihre Anwendbarkeit auf Phänomene beurteilen zu können, als wesentliche Kriterien die Übereinstimmung und Verständlichkeit, sowie die Allgemeingültigkeit und die Kontrolle erfüllen (Strauss & Corbin, 1996, S. 8).

Im Sinne von Hypothesenbildung verzichtet die Grounded Theory auf "gegenstandsbezogene theoretische Vorannahmen," so Flick (2007, S. 112). Vielmehr bezieht sie sich stattdessen auf die Entwicklung theoretischer Sensibilität gegenüber dem Forschungsfeld. Flick (2017) argumentiert, dass das permanente Einfließen analytischer Ideen in den Forschungsprozess durch das parallel, beziehungsweise zeitnah stattfindende zirkuläre Anwenden einzelner Forschungsschritte ermöglicht wird (S. 113).

Die Datenauswertung folgt dem paradigmatischen Modell der Grounded Theory, auch um dem subjektiven Erleben der betroffenen Individuen im vorliegenden Forschungsprojekt gerecht zu werden. Um die sich permanent entwickelnde und verändernde Natur von Erfahrung und Handlung sowie die aktive Rolle der Menschen bei der Gestaltung ihrer Lebenswelt zu verstehen, erscheint die Grounded Theory am besten geeignet, Zusammenhänge zwischen Bedingungen, Bedeutung und Handeln herauszuarbeiten (Strauss & Corbin, 1996, S. 9).

„In der analytischen Arbeit im Rahmen der Grounded Theory geht es nicht um alltagspraktische, situativ gebundene Orientierung, sondern darum, aus der Fülle empirischer Phänomene relevante theoretische Konzepte und Aussagen zu generieren" (Strübing, 2013, S. 114).

4.3.2 Offenes Kodieren

In vivo, also direkt am Datenmaterial, findet die Auswertung nach dem Kodierparadigma im Zuge des offenen und axialen Kodierens statt. Unter *offenem Kodieren* wird in einer Grounded Theory der analytische Prozess verstanden, durch den "Konzepte identifiziert und in Bezug auf ihre Eigenschaften und Dimensionen entwickelt werden" (Strauss & Corbin, 1996, S. 54-55).

Dabei gilt es grundlegende analytischen Verfahren wie "das Stellen von Fragen an die Daten und das Vergleichen hinsichtlich Ähnlichkeiten und Unterschieden zwischen jedem Ereignis, Vorfall und anderen Phänomenen," so Strauss und Corbin (1996, S. 55), zum Einsatz zu bringen. Sich ähnelnde Ereignisse und Vorfälle werden benannt und können danach in Kategorien gruppiert werden (S. 55). Abbildung 5 stellt ein Beispiel für das offene Kodieren dar.

Nummer der Kategorie (offener Code)	**Interview/ Zeile**	**Kategorie** (offener Code in vivo)	**Eigenschaft** (offener Code in vivo)	**Dimension** (eigene Einschätzung/ Interpretation)	**Memo** (Interpretation latenter Inhalte)
19	IP2/328-330	Es geht hier um die Angst, es wird ja Angst geschürt und Angst gemacht dich nicht glücklich und Angst kann dich krank machen, ja.	Ursache und Auslöser von Angst durch externe Faktoren	Fremdbestimmung der eigenen Gefühle – Selbstbewusstsein als Grundlage der eigenen Gefühle	Widerständigkeit, Mut und Willensstärke als Voraussetzung für das eigene ICH
20	IP2/359	Weil die Intensivmedizin wirklich ein sehr, sehr sensibler Bereich ist.	Unterscheidung von Einsatzbereichen von Technik diese Geräte Millionen von Diagnosen vereinen und daraus die Erfahrung sammeln IP2/359-360	Differenzierung in der Anwendung von Technik – dogmatischer Zugang	Abstraktionsvermögen
21	IP2/362-364	Wir Ärzte, auf der Intensivmedizin, machen viele Stunden Dienste, sind manchmal sehr müde und können aus der Müdigkeit heraus nicht immer die gleiche, gute Entscheidung treffen.	Menschen als Schwachpunkt	Differenzierung in der Betrachtung von Mensch und Maschine – dogmatischer Zugang	Offener Zugang zu Phänomenen Vertrauen auf Fortschritt

Abbildung 5: Beispielsseite Offenes Kodieren, Interviewpartner IP2, Seite 10, 28. Februar 2023

4.3.3 Axiales Kodieren

"Axiales Kodieren ist der Prozess des In-Beziehung-Setzens der Subkategorien zu einer Kategorie. Es stellt einen komplexen Prozess induktiven und deduktiven Denkens dar, der aus mehreren Schritten besteht" (Strauss & Corbin, 1996, S. 92). Dabei werden diese Schritte wie beim offenen Kodieren durch "Anstellen und Vergleichen und Stellen von Fragen durchgeführt. Beim axialen Kodieren ist der Einsatz dieser Vorgehensweisen fokussierter und auf das Entwickeln und In-Beziehung-Setzen von Kategorien nach dem paradigmatischen Modell ausgerichtet" (Strauss & Corbin, 1996, S. 92).

Die Grounded Theory kann daher gesehen werden als ein "Analysesystem, das Handlung/Interaktion in Beziehung zu ihren Bedingungen und Konsequenzen

untersucht" (Strauss & Corbin, 1996, S. 132). Diese erlaubt die Untersuchung der interaktiven Natur von Ereignissen. Abbildung 6 stellt ein Beispiel für das axiale Kodieren dar.

NUMMER (des axialen Codes)	PHÄNOMEN (offene Codes in vivo)	URSACHEN (offene Codes in vivo)	STRATEGIEN (offene Codes in vivo)	KONSEQUENZEN (offene Codes in vivo)	MEMO	ÜBERKATEGORIE (auch in einem extra Dokument möglich)
24	dieses Alleinstellungsmerkmal des Menschen dazu. Das Selbstbewusstsein IP3/Z. 779-780	Das sich bewusst sein, ein denkendes, lebendes, fühlendes, soziales Wesen zu sein IP3/Z. 780-781 die Libet-Versuche IP4/Z. 718 Und das spricht aber trotzdem nicht gegen den freien Willen meines Erachtens, weil jedes Gehirn einzigartig ist, weil es eben aus Abermillionen Verknüpfungen, die aus Abermillionen Erfahrungen aus Genen, die bestimmte Dinge prädisponieren, aus Umweltfaktoren, aus vielfältigen Rückmeldungen, Rückkopplungsschleifen, Verstärkungen, Bestrafungen usw. resultieren, sodass es trotzdem mein freier Wille ist, aber letztlich ist es alles an das Substrat gebunden IP4/Z. 725-730	Im Grunde sind das religiöse Fragestellungen IP3/Z. 790-791 die Grenze ist dort, wo ich sage, wo man eigentlich nicht weiß, wo im Gehirn sitzt denn wirklich so was wie die Seele? Die Seele ist auf jeden Fall was Materielles IP4/Z. 715-717	das ist, glaube ich, der entscheidende Unterschied vom Mensch zu anderen Lebewesen. Dieses sich seiner Selbst bewusst sein IP3/Z. 784-786 Wenn also Adam und Eva im Paradies zunächst einmal in einem Zustand gelebt haben, wo ihnen das alles nicht bewusst war IP3/Z. 793-794 Adam und Eva kaschieren sich mit Feigenblättern oder so was. Ihre Scham. Gott sagt, warum machst du das? Ja, weil ich mich schäme, weil ich ja nackt bin. Gott sagt, woher weißt du, dass du nackt bist? Vorher wusste der gar nicht, dass du nackt bist IP3/Z. 799-802 da sehe ich das so, dass die Menschen einmal die primäre Erfahrung machen mit ihrer Umgebung IP1/Z. 215-216 Überlegen sich sozusagen mögliche	Was muss geschehen, damit Selbstbewusstsein auftritt? An dieser Stelle tritt Freiheit auf! Memo zu IP1/Z. 219-223: auch wenn es unlogisch erscheint – Verweise auf Entscheidungen von KI am Beispiel von AlphaGo, Spielzug 37 unlogisch, doch gewinnbringend im Spielzug 50 (vgl. Tegmark "Leben 3.0")	Selbstbewusstsein Freiheit Freiheitsbegriff von Sartre Religiöse Fragestellungen

Abbildung 6: Beispiel Axiales Kodieren

4.3.4 Prozessanalyse

Ein wesentlicher Bestandteil der Grounded Theory ist die Analyse von Prozessen, da dadurch das sich stetig verändernde Bedingungsgefüge zwischen Strategien, Erfahrungen, Geschehnissen und Konsequenzen aufgezeigt werden kann. Gleichzeitig kann nachvollzogen werden, wie Handlungen und Interaktionen sich verändern, gleichbleiben oder auch zurückentwickeln (Strauss & Corbin, 1996, S. 119).

Das Verknüpfen von Handlungssequenzen und Interaktionssequenzen ist der Prozess, der ein wichtiger Bestandteil der Analyse der Grounded Theory ist, um in diese Analyse Prozessaspekte einzubringen. Dabei muss der Analysierende gezielt nach Daten und Zeichen suchen, "die auf eine Veränderung in Bedingungen hinweisen, und aufzeichnen, welche Veränderungen in der Handlung/Interaktion damit einhergehen" (Strauss & Corbin, 1996, S. 131).

4.3.5 Selektives Kodieren

Nachdem die Daten nun gesammelt und analysiert wurden, steht man vor der Aufgabe, diese Kategorien zu einer Grounded Theory zu integrieren. "Integration unterscheidet sich nicht sehr vom axialen Kodieren. Sie wird nur auf einer höheren, abstrakteren Ebene der Analyse durchgeführt" (Strauss & Corbin, 1996, S. 95).

Die Grundlagen für das selektive Kodieren werden bereits beim axialen Kodieren gelegt. Dabei werden Kategorien "in bezug auf deren hervortretende Eigenschaften und Dimensionen ausgearbeitet und mit paradigmatischen Beziehungen verbunden, die den Kategorien Fülle und Dichte verleihen", so Strauss und Corbin (1996, S. 95).

Selektives Kodieren ist "der Prozeß des Auswählens der Kernkategorie, des systematischen In-Beziehung-Setzens der Kernkategorie mit anderen Kategorien, der Validierung dieser Beziehungen und des Auffüllens von Kategorien, die einer weiteren Verfeinerung und Entwicklung bedürfen" (Strauss & Corbin, 1996, S. 94). Die Kernkategorie ist jenes zentrale Phänomen, "um das herum alle anderen Kategorien integriert sind" (Strauss & Corbin, 1996, S. 94).

4.3.6 Leitfadengestützte semistrukturierte qualitative Interviews

Bei leitfadengestützten semistrukturierten qualitativen Interviews wird die Methode des *Episodischen Interviews* nach Flick (2007) angewandt. Bei dieser Methode erfolgen zu thematischen Bereichen leitfadengestützt regelmäßig Erzählaufforderungen. Die Interviewpartner*innen, mit denen für diese Studie leitfadengestützte semistrukturierte qualitative Interviews geführt wurden, kommen aus dem Bereich der Soziologie, Psychologie, Medizin, Ökologie und Technik.

Es geht bei den Interviews nicht um eine große Narration, sondern um mehrere, eng umgrenzte Narrationen, die aus dem Erfahrungsbereich der Subjekte zum gegebenen Gegenstandsbereich entstehen. Rückgriffsphasen zu Erzählteilen, die nicht klar verstanden oder abgebrochen wurden und Billanzierungsphasen über rückblickende Evaluationen des eigenen Lebens sind weitere Elemente des Episodischen Interviews. Die inhaltlich-semantische Transkription der mittels Audioaufnahmen gestützten Interviews erfolgt nach Rädiker und Kuckartz (Rädiker & Kuckartz, 2019).

4.3.7 Qualitative Auswertung nach der Grounded Theory

Die qualitative Auswertung des Datenmaterials der zeilennummerierten Interviewtranskripte erfolgt nach der empirischen Methode der *Grounded Theory* nach Strauss und Corbin. Der Grund, warum diese Methode gewählt wurde, liegt darin, dass die Datenauswertung nach dem paradigmatischen Modell der *Grounded Theory* am ehesten geeignet scheint, dem subjektiven Erleben der Interviewpartner*innen gerecht zu werden. Für den Erkenntnisgewinn werden die folgenden Elemente des paradigmatischen Modells der Grounded Theory herangezogen: Datensammlung, theoretische Sensibilisierung, offenes Kodieren, axiales Kodieren und selektives Kodieren.

Für die qualitative Auswertung des Datenmaterials der zeilennummerierten Interviewtranskripte wurde keine Software verwendet. Die Entscheidung, die zeilennummerierten Interviewtranskripte manuell auszuwerten liegt darin

begründet, dass das manuelle Arbeiten am Text und der sich daraus entwickelnde Prozess der Iteration zu einer höheren Sensibilisierung bei der Interpretation der Daten und Inhalte führen sollte.

Die Ergebnisse der Literaturanalyse und der methodengeleiteten qualitativen Datenanalyse werden in der Folge miteinander in Beziehung gesetzt und diskutiert.

4.4 Zusammenfassung

In diesem Kapitel werden themenspezifische Implikationen qualitativer Forschung diskutiert sowie der für das Forschungsprojekt gewählte methodische Zugang, die *Grounded Theory*, vorgestellt. Dabei ist ein wichtiger Schritt im Forschungsprozess die methodologische Positionierung, um das Ziel, die Forschungsfrage, mit der Auswahl der richtigen Methoden zu beantworten. In dieser Arbeit wurden daher neben dem Studium, der Diskussion und der Auswertung der für die Beantwortung der Forschungsfrage relevanten Literaturbefunde auch Expert*inneninterviews durchgeführt.

Qualitative Forschung eignet sich als Methode, um menschliches Verhalten, soziale Strukturen und kulturelle Phänomene zu untersuchen. Dies war mit ein Grund dafür, in der vorliegenden Arbeit eine qualitative Forschungsmethode einzusetzen. Als methodischer Zugang zum Forschungsprojekt wurde die *Grounded Theory* gewählt.

Die Auswahl der Interviewpartner*innen, das Ablaufschema der Interviews, das Vorgespräch und die Prinzipien der Durchführung der Interviews sowie die Transkription der Interviews werden in diesem Kapitel ebenso beschrieben wie die wissenschaftliche Methode und der methodische Zugang der *Grounded Theory*.

Die Datenauswertung folgt dem paradigmatischen Modell der *Grounded Theory*. Die Techniken des offenen Kodierens, des axialen Kodierens, des selektiven Kodierens sowie die Prozessanalyse sind ein weiterer Teil dieses Kapitels. Neben der Beschreibung des leitfadengestützten semistrukturierten qualitativen Interviews, welches nach der Methode des Episodischen Interviews nach Flick erfolgt, wird abschließend in diesem Kapitel die qualitative Auswertung nach der *Grounded Theory* erläutert.

5 Evaluierung & Ergebnisse

Die im Rahmen der Expert*inneninterviews entstandenen Transkripte wurden inhaltlich analysiert. Im dabei generierten Kategorienschema wird ein Überblick über alle wichtigen Inhalte gegeben und die Hauptaussagen werden reflektierend zusammengefasst. Die Ergebnisse werden anschließend evaluiert, um in einer Darstellung über das zentrale Phänomen der Untersuchung den Kern der Geschichte zu identifizieren.

5.1 Motivation und Zielsetzungen

Die Motivation bei der inhaltlichen Analyse der Transkripte der Expert*inneninterviews war der sich stetig weiterentwickelnde Erkenntnisgewinn durch das Extrahieren von Kernaussagen beim Arbeiten am Text. Die Zielsetzung war es, diese Kernaussagen miteinander in Verbindung zu bringen und deren Korrelation untereinander zu gewichten, um für die Untersuchung der Forschungsfrage relevante Ergebnisse zu erzielen.

5.2 Allgemeine Vorgehensweise

Schon während des axialen Kodierens beginnt man schon bestimmte Muster wie wiederholt auftauchende Beziehungen zwischen Eigenschaften und Dimensionen von Kategorien zu bemerken. Gleichzeitig entsteht beim Erstellen dieser Kategorien bereits ein gewisses Ausmaß an Integration und damit bereits ein Gewebe und Netzwerk bereits vorhandener Beziehungen. Diese Muster nun zu erkennen und die Daten danach auch dementsprechend zusammenzufassen ist bedeutsam, da der Theorie damit Spezifität verliehen wird. "Dann ist man in der Lage zu sagen: Unter diesen Bedingungen (Auflistung) passiert das und das; während unter anderen Bedingungen das und das eintritt" (Strauss & Corbin, 1996, S. 107).

Das Systematisieren und Verfestigen von Verbindungen erfolgt aus einer Kombination aus induktivem und deduktivem Denken, "wobei wir durchgängig zwischen dem Stellen von Fragen, dem Aufstellen von Hypothesen und dem Vergleichen hin -und herpendeln" (Strauss & Corbin, 1996, S. 107).

5.3 Bildung des Kategorienschemas

Im Kategorienschema wurden die für das axiale Kodieren erforderlichen Kategorien *Phänomen, Ursachen, Strategien* und *Konsequenzen* gebildet und definiert. Die Bedeutung dieser Begriffe wird wie folgt erläutert:

- Phänomen: Als ein *Phänomen* wird eine zentrale Idee, ein Ereignis, ein Geschehnis, die oder das sich auf mögliche Handlungen und Interaktionen richtet, verstanden.

- Ursachen: Ursächlich oder zeitlich vorausgehende Bedingungen werden in der Kategorie *Ursachen* beschrieben. Auch wirken auf Handlungs- und interaktionale Strategien intervenierende Bedingungen ein.
- Strategie: In der Kategorie *Strategien* finden sich die prozessualen, zweckgerichteten und zielorientierten Denkweisen und Überlegungen, die Kommunikation, das Handeln sowie die Bewältigungsversuche der Akteur*innen wieder.
- Konsequenzen: Die Bildung der Kategorie *Konsequenzen* hat ihre Begründung darin, dass die Folgen, Reaktionen und Wirkungen des Prozesses, der durch die Kategorien *Phänomene* und *Strategien* ausgelöst wird, zu Konsequenzen und Ergebnissen führt.

In einer Überkategorie wurden die Phänomene in einem größeren Rahmen zusammengefasst. Dabei entstanden eine Vielzahl von Überkategorien, wobei jene Überkategorien, die für die Beantwortung der Forschungsfrage nicht relevant oder vernachlässigbar waren, weggelassen wurden. Als für die Beantwortung der Forschungsfrage relevant erwiesen sich 4 Überkategorien, welche auch in Form als Unterkapitel in den Ergebnisteil der Forschungsarbeit einflossen. Abbildung 7 stellt ein Beispiel für das Kategorienschema und die Definition der Kategorien dar.

NUMMER (des axialen Codes)	**PHÄNOMEN** (offene Codes in vivo)	**URSACHEN** (offene Codes in vivo)	**STRATEGIEN** (offene Codes in vivo)	**KONSEQUENZEN** (offene Codes in vivo)	**MEMO**	**ÜBERKATEGORIE** (auch in einem extra Dokument möglich)
1 fortlaufend	Kategorien oder Eigenschaften Das Phänomen: eine zentrale Idee, ein Ereignis, ein Geschehnis, die oder das sich auf mögliche Handlungen und Interaktionen richtet	Kategorien oder Eigenschaften Ursächlich oder zeitlich vorausgehende Bedingungen Zudem wirken intervenierende Bedingungen auf Handlungs- und interaktionale Strategien ein → bei follow-ups interessant	Kategorien oder Eigenschaften Denkweisen, Überlegungen, Kommunikation, Handeln, Bewältigungsversuche der Akteur*innen; prozessual, zweckgerichtet, zielorientiert (auch wenn sie ausbleiben, ist dies von Bedeutung)	Kategorien oder Eigenschaften Folgen, Reaktionen, Wirkung des Prozesses: Phänomen--> Strategien--> führt zu Konsequenzen bzw. zu Ergebnissen. Konsequenzen können tatsächlich oder möglich sein, in der Gegenwart oder in der Zukunft. Sie können zu einem späteren Zeitpunkt zu Bedingungen für neuerliche Strategien werden.	Interpretation auf zweiter Abstraktionsstufe; auch Zusammenführen der dazu passenden Memos aus dem offenen Kodieren möglich	Zusammenfassen der Phänomene in einem größeren Rahmen. Möglicherweise erhält man hier 20 oder mehr Überkategorien. Diese werden in einem nächsten Schritt auf 4-7 Überkategorien zusammengeführt. Ist eine Überkategorie für die Beantwortung der Forschungsfrage nicht relevant oder vernachlässigbar, kann diese auch weglassen. Die Überkategorien, die bleiben, können mit den Unterkapiteln des Ergebnisteils der Forschungsarbeit ident sein.

Abbildung 7: Axiales Kodieren – Kategorienschema und Definitionen der Kategorien

5.4 Darstellung und Interpretation der Interviews

Ziel dieses Kapitels ist die Datenauswertung nach dem paradigmatischen Modell der Grounded Theory, um damit die aus den Expert*inneninterviews gewonnen Meinungen der Experten*innen in aggregierter Form wiederzugeben. Dabei wird durch Zusammenführung, Komplexitätsreduktion und Verdichtung jedes Phänomen auf seine ursächlichen Bedingungen, auf seinen Kontext, auf

intervenierende Bedingungen, auf Handlungs- und interaktionale Strategien und auf Konsequenzen untersucht.

5.5 Ergebnisse

In den für die Beantwortung der Forschungsfrage sich als relevant erwiesenen 4 Überkategorien werden im Folgenden nach Auswertung der Daten nach dem paradigmatischen Modell der Grounded Theory über das zentrale Phänomen, die Ergebnisse der Untersuchung präsentiert. In einer beschreibenden Erzählung wird, um dem subjektiven Erleben der betroffenen Individuen im vorliegenden Forschungsprojekt gerecht zu werden, der rote Faden der Geschichte identifiziert und offengelegt.

Die Interviewpartner*innen werden im Folgenden als IP1, IP2, IP3, IP4, IP5, IP6 benannt.

5.5.1 Neue Handlungsoptionen durch neue Technologien

Durch das Entstehen neuer Technologien, so waren sich alle Interviewpartner*innen einig, eröffnen sich für den Menschen neue Handlungsoptionen in den unterschiedlichsten Handlungsfeldern, die zu einer Erleichterung bei der Bewältigung von Aufgaben führen. IP2 betont die Erleichterungen durch den Einsatz neuer Technologien:

> "Na ja, ich empfinde es schon sehr hilfreich, z.B. ein EKG zu haben, das gleich die Ergebnisse präsentiert, ja" (IP2/Z. 52-53).
>
> "Es gibt heutzutage Stethoskope, die sofort anzeigen, Achtung, Lungenentzündungsgefahr" (IP2/Z. 53-54).
>
> "[…] wenn ich zu einem Patienten komme, der einen Oberbauchschmerz hat, kann ich mit dem Handy feststellen, der hat Gallensteine, ja. Das ist eine sehr große Erleichterung, ja" (IP2/Z. 56-58).

IP4 sieht die Möglichkeiten des Einsatzes neuer Technologien in klar begrenzten Bereichen und stellt fest,

> "[…] da ist das Erkennen eines Tumors aus einem CT oder wobei da sind ja eben immer noch Bereiche, wo man sagen kann, das ist jetzt, das sind immer noch, das ist nicht im Sinne des General Problem Solvings, sondern das ist immer klar definierte Probleme, da ist das Erkennen eines Tumors aus einem CT oder einem MR-Bild, da ist das Erkennen eines Melanoms aufgrund von Apps, die auch inzwischen besser funktionieren als Hautärzte sozusagen und bessere Vorhersagen machen können" (IP4/Z. 521-525).

Der Nutzen neuer Technologien, um dem Menschen dienlich zu sein, wird von IP1 und IP5 betont:

> "[…] und wir Drohnen verwenden, wenn wir auf der Suche nach einem Demenz erkrankten Menschen in der Nacht sind, dann könnte ich sagen, perfekt" (IP1/Z. 350-352).
>
> "Auf jeden Fall geht es um Pflegerobotik, wo die TU Wien gemeinsam mit dem Technischen Museum zu Pflegerobotik arbeitet und da haben, war natürlich auch dieses Ding, das so robotermäßig herumfährt" (IP5/Z. 224-226).
>
> "Also vermutlich werden vielleicht auch Roboter kommen, die aber eher haushaltsähnliche Leistungen machen" (IP1/Z. 385-387).
>
> "[…] die Maschine lernt mit, die, der Roboterarm oder die Unterstützung des Arbeiters lernt mit, indem beobachtet wird, wie greift der hin, mit welcher Hand, mit welchem Finger, wie viel Kraft braucht die Person" (IP5/Z. 316-318).

Damit Technologien weiterhin einen Nutzen für den Menschen darstellen, müssen wir darauf achten, so IP1, "[…] dass wir uns jener Technik verschreiben oder jene Technik implementieren, die uns zweckdienlich und notwendig erscheint" (IP1/Z. 51-52). Denn, so IP1 weiter, wir sollten nicht "[…] Technik um jeden Preis oder Technik, der Technik Willen, sondern um unsere Abläufe beispielsweise besser zu organisieren" (IP1/Z. 54-55) anstreben.

Das moderne Technologien den Vorteil haben, weniger fehleranfällig zu sein als der Mensch, hat laut IP2 besonders im Handlungsfeld der Intensivmedizin enorme Auswirkungen. Denn in der "[…] Intensivmedizin, da geht es jetzt wirklich um Entscheidungen, die über Leben und Tod ziemlich schnell bestimmen, ja" (IP2/Z. 416-418). Und IP2 weiter:

> "Wir Ärzte, auf der Intensivmedizin, machen viele Stunden Dienste, sind manchmal sehr müde und können aus der Müdigkeit heraus nicht immer die gleiche, gute Entscheidung treffen" (IP2/Z. 362-364).
>
> "Aber diese Geräte ermüden nicht. Die haben immer die gleiche Qualität" (IP2/Z. 364-365).

Wie rasant sich Technologie entwickelt, geht aus folgenden Aussagen von IP2, IP3 und IP6 hervor und macht Technologiesprünge deutlich:

> "Wo du dann Computerzeit reservieren musstest und das ging damals (schmunzelt) noch über Lochkarten und so etwas" (IP3/Z. 129-130).
>
> "[…] dann gehe ich in die Bibliothek und dann borge ich mir das Buch aus und dann eine Literaturstelle, ja, das dauert Wochen, Monate, jetzt plötzlich in Sekunden abzuwickeln. Nämlich einfach draufklickt, zack und es geht auf und es ist da" (IP6/Z. 256-259).

> "Und das World Wide Web, das war, ist mir spätestens in dem Moment klar geworden, wie ich gesehen habe, wie das explodiert" (IP6/Z. 231-233).

> "[…] diese Geräte Millionen von Diagnosen vereinen und daraus die Erfahrung sammeln" (IP2/Z. 359-360).

Völlig neue Handlungsoptionen und Handlungsfelder für den Menschen durch neue technologische Entwicklungen eröffnen sich in Bereichen, die bislang außerhalb der Einflussnahme darauf durch den Menschen galten. Transhumanismus als eine philosophische Denkrichtung will die Grenzen menschlicher, intellektueller, physischer und psychischer Möglichkeiten durch den Einsatz technologischer Verfahren erweitern. Diese Entwicklungen wurden von den Interviewpartner*innen kritisch betrachtet.

Wohl ist diese Unsterblichkeitsphantasie des Menschen nichts Neues, "Das gibt es seit Anbeginn der Menschheit" (IP3/Z. 491-492). IP3 stellt dazu die Überlegung an: "Die Frage ist aber, will ich das?" (IP3/Z. 495). Denn, so IP3 weiter, "[…] ich halte die Sterblichkeit für eines der entscheidenden Merkmale des Lebens letzten Endes" (IP3/Z. 517-518).

> "Früher war Sterblichkeit viel bewusster" (IP3/Z. 536-537).

> "Sterblichkeit war ein völlig selbstverständliches Allgemeinwissen" (IP3/Z. 545).

Mensch-Maschine Schnittstellen ermöglichen es, in diese innerlichsten Prozesse des Menschen einzugreifen:

> "Nämlich zu versuchen Gehirnprozesse von außen durch die non-invasive Technik der transkanälen, elektrischen Stimulation von außen, Intelligenz quasi zu befördern und auch andere kognitive Prozesse" (IP4/Z. 28-31).

IP3 und IP5 sehen im Transhumanismus und damit in der Vorstellung der Unsterblichkeit des Menschen klare Grenzen:

> "Also das Gehirn ist wahrscheinlich wirklich nicht vorstellbar zu ersetzen, ja. Also, wenn ich das jetzt rein theoretisch, wie ich das mir jetzt rein theoretisch vorstellen würde, ist da wahrscheinlich der Punkt, wo dann das Menschsein natürlich aufhört, ja. Spätestens da" (IP3/Z. 571-574).

> "Also es gibt, also ich glaube nicht an eine immaterielle Seele oder so, ich bin Materialist in dem Sinne, im philosophischen Sinne, dass ich sage, wenn das Gehirn tot ist, dann ist es nicht mehr vorhanden" (IP4/Z. 730-733).

5.5.2 Technikgläubigkeit und kritisches Bewusstsein

Vertrauen ist eine Strategie des Menschen im Umgang mit Technologie und technologischen Entwicklungen, aufgrund dieser er die Entscheidung sowohl über den Einsatz als auch über den Nutzen von Technologie treffen muss. Einerseits ist dieses Vertrauen eine Grundbedingung, andererseits birgt es jedoch die Gefahr in sich, sich bei unkritischer und passiver Haltung gegenüber Technologie sich dieser auszuliefern. Dies wiederum führt zu einer Technologie, die sich verselbstständigt und nicht mehr im Dienste der Menschheit steht. Technikgläubigkeit wird von den Interviewpartner*innen kritisch betrachtet und als Gefahr für den Menschen gesehen.

Sich mit Technologie kritisch auseinanderzusetzen ist notwendig, denn, so IP6, "Weil natürlich die Technik auch die Technik der Mächtigen ist" (IP6/Z. 378-379). Und IP6 weiter:

> "[…] mir ist es immer wieder im Programmierunterricht passiert, dass einzelne Studentinnen oder Studenten nicht erfasst haben, dass es jetzt um die Korrektheit dieses Programms geht, dass sie selbst beurteilen müssen. Sondern die haben sich versucht zu überlegen, was möchte ich als ihr Prüfer hören" (IP6/Z. 563-566).

Gleichzeitig ist für IP6 jedoch "[…] ein gewisser Pessimismus [angebracht], weil die Menschen so leichtfertig Daten hergeben" (IP4/Z. 541-542).

> "Ich habe ja nichts zu verbergen und diese ganzen schwachsinnigen Sätze" (IP6/Z. 776-777).

Dass dieser technikgläubige und leichtfertige Umgang mit Daten Konsequenzen hat, zeigt sich daran, "[…] wie negativ sozusagen kapitalistisch getriebene Konzerne mittlerweile Einfluss auf Demokratien, auf Wahlen haben oder uns auch durch Algorithmen natürlich in unseren Gedanken und Interessen und Wunschlage beeinflussen" (IP1/Z. 285-288). Und IP1 weiter:

> "Also man hat ein Interesse, eine bestimmte Partei oder eine Abstimmung zu erzeugen. Die Firmen gibt es, verdienen daran und bedienen sie sozusagen auch politische Interessen" (IP1/Z. 301-303).
>
> "Aber mittlerweile ist es durch den Neoliberalismus 40 Jahre so, dass immer mehr die Politik das tut, was die Finanzmärkte wollen und von dem getrieben ist. Und letztlich gehört dabei auch der hochtechnische, hochgerüstete Bereich dazu" (IP1/Z. 346-349).
>
> "Und gleichzeitig natürlich steckt dahinter der militärisch-industrielle Komplex, der derzeit natürlich unglaubliche Profite macht aufgrund der Kriegssituation jetzt" (IP1/Z. 338-340).

Technologische Entwicklungen haben zudem nicht erfüllte Hoffnungen zur Folge, wie IP6 bemerkt:

> "Wir werden durch schnellen, guten, unzensierten Informationsaustausch de facto so was wie eine große Weltdiskussion anfachen, die dann in Summe das Beste für die Welt ergibt. Weil alle frei miteinander diskutieren können. Das ist so gesehen gescheitert. Das muss man einfach so sehen" (IP6/Z. 419-422).

Mit einen Grund dafür sieht IP4 in marktwirtschaftlichen Bedingungen, denn "[…] die besten Persönlichkeitspsychologen in den USA, die werden von Google, Facebook und Co. abgeworben und die verdienen natürlich dort die fünffachen Gehälter, wie sie auf der Uni verdienen würden, nicht" (IP4/Z. 552-555). Und IP4 weiter mit einer Antwort auf die ihm einmal gestellte Frage, ob der intelligente Mensch nicht auch der gute Mensch sei:

> "Ich habe gesagt, nein, überhaupt nicht. Die drei dunklen Merkmale der Persönlichkeit, die sogenannte dunkle Triade, Narzissmus, Psychopathie und Machiavellismus, die korrelieren alle drei zu null mit dem IQ, mit der menschlichen Intelligenz" (IP4/Z. 347-350).

> "Weil natürlich ist ein intelligenter Machiavellist viel gefährlicher als ein unintelligenter Machiavellist" (IP4/Z. 356-357).

Die Konsequenz dieses Umstandes beschreibt IP1 dahingehend, in welcher Art und Weise soziale Medien auf ganze Gesellschaften als auch auf Individuen Einfluss nehmen: "Und wir wissen nicht, welche Menschen mit welchen Interessen diese künstliche Intelligenz produzieren. Ich weiß ja nicht, ob die künstliche Intelligenz eine Gesellschaft des Gleichen letztlich dann anstrebt" (IP1/Z. 549-551).

> "Und die soziale Praxis heute ist Social-Media-Betrachtungen, Algorithmen dahinter, die unsichtbar bleiben. Und eine Steuerung sozusagen unserer Ideen, Wünsche, Vorlieben, Interessen und wir werden immer mehr damit. Ulrich Becker hat gesagt, der Mensch ist keine Ware, aber (schmunzelt) wir werden letztlich immer mehr zur Ware entwickelt" (IP1/Z. 430-435).

> "[…] da sind einfach enorme Möglichkeiten, weil man für Leute, die in sozialen Netzwerken unterwegs sind, die dort so viele Spuren hinterlassen, persönlichkeitsrelevante Spuren, dass man im Prinzip deren Persönlichkeit von Intelligenz und Kreativität und anderen kognitiven Funktionen über sozial-emotionale Kompetenzen, Stichwort emotionale Intelligenz, bis hin zu grundlegenden Charaktereigenschaften, wie den sogenannten Big Five Persönlichkeitsmerkmalen, bis hin vielleicht zur dunklen Triade, nicht, die aversiven Persönlichkeitsmerkmale mit einer erstaunlichen Genauigkeit herauslesen kann" (IP4/Z. 93-99).

> "[…] die haben enorme Datensätze und können natürlich auch Mini-Zusammenhänge aufspüren" (IP4/Z. 100-101).
>
> "[…] das ist schon eine große Sorge. AI funktioniert ja heute über diese riesigen Ähnlichkeitschecks im Wesentlichen" (IP6/Z. 728-729).
>
> "Also wir rollen da schon auf eine, an, auf eine Welle zu, wo jetzt schon einfach unendlich viel Manipulation ist" (IP6/Z. 777-778).

Diese Art einer Manipulation, wenngleich auch einer Selbstmanipulation, stellt auch IP2 fest, wenn das Phänomen einer sich selbst erfüllenden Prophezeiung aufgrund des Einwirkens moderner Technologie sich auf Menschen negativ auswirkt, denn "[…] Menschen haben sehr viel Angst vor Krankheit, vor der Zukunft" (IP2/Z. 248-249). Und IP2 weiter:

> "Er hat das auf seiner iWatch gesehen, ist sehr nervös geworden" (IP2/Z. 271-274).

Ähnliche Gefahren liegen, so IP1, auch darin, "[…] wo die künstliche Intelligenz womöglich überhaupt die Entscheidungen des Menschen abnehmen. Also, dort wird es gefährlich" (IP1/Z. 392-394). IP1, IP3, IP4 und IP6 sind derselben Meinung:

> "[…] was ChatGPT macht. Also wenn ich sage, baue mir einen wissenschaftlichen Text, dann macht der irgendwas, von dem er hofft, dass ich eine Freude habe damit" (IP6/Z. 573-575).
>
> "Und wir wissen nicht, welche Menschen mit welchen Interessen diese künstliche Intelligenz produzieren. Ich weiß ja nicht, ob die künstliche Intelligenz eine Gesellschaft des Gleichen letztlich dann anstrebt" (IP1/Z. 549-551).
>
> "[…] wenn eine KI dann wirklich jegliches, mögliches auftretende Problem und damit auch jedes möglich auftretende Hindernis in Form einem von anderen Menschen, die dann dagegen arbeiten, ausschalten kann, dann ist halt wirklich irgendwann die Gefahr, dass es zur Singularität kommt und dass die KI dann irgendwann nicht mehr ausgeschaltet werden, sich nicht nur ausschalten lässt" (IP4/Z. 526-530).
>
> "[…] weil eine Superintelligenz ohne hohe Moralitätsansprüche, aber wie kann ich die definieren? Das ist noch schwieriger wahrscheinlich, was welche Werte pflanze ich einer KI ein, aber ohne das wird es halt höchst gefährlich werden" (IP4/Z. 386-389).
>
> "[…] wir sollten nicht sozusagen die politischen Entscheidungen der künstlichen Intelligenz überantworten und trotzdem sind wir in einem Prozess, wo derzeit die quasi künstliche Intelligenz uns Menschen angreift und manipuliert und jetzt schon eindeutig

> Wahlen beeinflusst und auch durch die Algorithmen" (IP1/Z. 524-527).
>
> "[…] da sehe ich eben die Gefahr darin, dass ein Wettlauf der Intelligenzen dann letztlich dazu führt, dass diese, ja, dass in Kombination mit problematischen Werthaltungen dann katastrophal, viel Katastrophaleres für die Welt resultieren könnte" (IP4/Z. 503-505).
>
> "Dann wird der Mensch so eine Art Maschine, die eben toll funktioniert, die man gut steuern kann, aber die eben dann nicht mehr letzten Endes die, für die eigenen Handlungen selbst verantwortlich ist und nicht mehr selbst entscheiden kann" (IP3/Z. 616-619).

Bedenklich äußert sich IP1 zu der Tatsache, dass man gesellschaftspolitische Zielsetzungen algorithmischen Entscheidungen überlässt, denn "[…] Entscheiden tut es letztlich davor der Eigentümer, der die künstliche Intelligenz entwickelt und wahrscheinlich nicht einmal die Parlamente oder die Regierungen auf das Einfluss nehmen" (IP1/Z. 553-555).

5.5.3 Menschliche Intelligenz als Voraussetzung für Fortschritt

Menschliche Intelligenz ist die Voraussetzung für das, was der Mensch im Allgemeinen als Fortschritt bezeichnet. Intelligenz wird von IP4 bezeichnet "[…] nämlich [als] die Fähigkeit, mit wirklich völlig neuartigen Problemen umgehen zu können und nicht Dinge, die eh schon definiert sind" (IP4/Z. 244-245).

Nach IP3 ist "Intelligenz ist eine der wenigen Persönlichkeitseigenschaften" (IP3/Z. 258-259), und "Intelligenz ermöglicht, eine komplexere Situation zu erfassen, zu begreifen und Lösungen dafür zu finden" (IP3/Z. 226-227).

> "Also im Grunde ist es Wahrnehmung, Gedächtnis, Denken. Das sind so die drei Hauptkriterien bei der Intelligenz" (IP3/Z. 242-243)."

Neue technologische Entwicklungen basieren auf zuvor erworbenen Erkenntnissen. IP6 meint dazu:

> "Also ich habe das Gefühl, dass jede Technologie, eigentlich jeder Technologieschub eigentlich den alten ja nicht wegwirft, sondern einschachtelt" (IP6/Z. 46-47).
>
> "[…] ich kann mir halt Sachen in einem Zehntel der Zeit erarbeiten, da neue Erkenntnisse, weil ich auf satten Grundlagen sitze" (IP6/Z. 318-319).
>
> "Also wenn ich das vergleiche mit meiner Schulzeit, wo wir an einzelnen Z80 mit hexadezimal programmiert haben, bin ich heute

> noch froh, dass ich weiß, wie das geht, auch wenn die einzelne Maschine in einem Cloud-Verbund 27-mal virtualisiert, ein Stecknadelkopf in dem ganzen System ist. Aber ich habe noch gelernt, wie dieser Stecknadelkopf funktioniert. Und insofern habe ich nicht das Gefühl, dass da viel ausgestorben ist, sondern jedes war eigentlich die Stufe zur nächsten Generation" (IP6/Z. 53-58).
>
> "Ich halte die Revolution des Rechnens für einen Riesen, eigentlich also für die Voraussetzung, für die industrielle Revolution" (IP6/Z. 69-71).

Die Wichtigkeit, das große Ganze zu sehen, betont IP6, denn "Jede Zeit hat ihre Denke, dass man dann begonnen hat, spätestens mit der Dampfmaschine, die Welt als einen mechanischen Ablauf zu sehen, ja. Man hat sich gedacht, okay, im Wesentlichen ist die ganze Welt nichts anderes als eine große Ansammlung von Zahnrädern, die halt vor sich hin werkelt, ja. Das war so die Idee. Und daraus ist ja die ganze Technikgläubigkeit, die bis heute, also der Determinismus entstanden" (IP6/Z. 178-183). Und IP6 weiter:

> "Also man muss sehr viel von einzelnen Zahnrädern verstehen, damit man ganze Getriebe bauen kann und man muss ganz viel von Getrieben verstehen, damit man ganze Anlagen bebauen kann und man muss ganz viel von Anlagen verstehen" (IP6/Z. 50-53).

5.5.4 Selbstbewusstsein und Verantwortung des Menschen

Wurde im Kapitel 5.5.3 die menschliche Intelligenz als die Voraussetzung für das, was der Mensch im Allgemeinen als Fortschritt bezeichnet genannt, so bedurfte und bedarf es weiterer Fähigkeiten und Eigenschaften des Menschen, Entwicklungen in verantwortlicher Weise voranzutreiben. Ethische und moralische Fragen drängen sich bei jeder dieser Entscheidungen auf. Und diese Entscheidungen liegen beim Menschen, denn die Technik, so IP6, "[…] sie hat uns nicht geholfen wirklich absolut grundlegende und damit meine ich eher Probleme des Zusammenlebens, das, der sozialen Probleme der Probleme, wie lösen wir spieltheoretische Konflikte? Da hat uns die Technik überhaupt nicht weiter geholfen. Da sind wir und da stehen wir auch an" (IP6/Z. 299-303).

IP2 sieht die Verantwortung auch bei der Nutzung von moderner Technologie beim Menschen, wenn es beispielsweise um die Verwendung diagnostischer Technologie in der Medizin geht:

> "Na ja, es bleibt noch immer bei mir zu entscheiden, in welche Richtung suche ich, ja" (IP2/Z. 115).
>
> "Also diese Unterscheidungen und quasi das Screening, kann, entscheide noch immer ich, was ich einsetze als diagnostisches Hilfsmittel, ja" (IP2/Z. 119-120).

> "Man muss sich entscheiden, mit welchen diagnostischen Mitteln suche ich weiter, ja" (IP2/Z. 132).

Entscheidungen treffen zu müssen bedeutet Verantwortung für die zu entscheidende Sache zu tragen, im Bewusstsein, damit auch selbst verantwortlich zu sein.

> "Weil entscheidendes Kriterium des Menschseins ist die Selbstverantwortung" (IP3/Z. 600).
>
> "[…] dieses Alleinstellungsmerkmal des Menschen dazu. Das Selbstbewusstsein" (IP3/Z. 779-780).
>
> "Wir werden hineingeworfen in die gegebene Welt, aber in eine spezifische, gegebene Welt" (IP1/Z. 185-1869).
>
> "[…] ist es schon richtig, dass man zur Freiheit verurteilt ist, weil die Verurteilung, dass wir uns selbst entwickeln müssen, sozusagen ist gegeben" (IP1/Z. 190-192).
>
> "Eine Maschine, wenn eine Maschine selbstbewusst wäre, dann wäre sie nicht so selbstbewusst wie das ein Mensch ist, weil es einen anderen Körper hat" (IP6/Z. 914-915).
>
> "[…] wenn ich jetzt ein Neurologe bin, ja und eine Gehirnoperation mache und da versuche, ja, eine einzelne Gehirnzelle zu fixen, ja, macht das einfach überhaupt keinen Sinn mehr jetzt über dein Selbstbewusstsein Gedanken zu machen" (IP6/Z. 859-862).

Die Tatsache, dass der Mensch auf sich alleine gestellt ist "[…] und dass es mit AI, mit künstlicher Intelligenz immer schwieriger wird, die Tatsachen von den Lügen zu unterscheiden" (IP6/Z. 675-676), fordert den Menschen geradezu auf, sich dieser Tatsachen bewusst zu werden und Verantwortung zu übernehmen. So liegt beispielsweise eine der Verantwortungen darin, "Dass man Verantwortung dafür hat, diese Datenfülle auch so auszuwählen in den Systemen, damit dann eben niemand diskriminiert wird" (IP5/Z. 142-143) oder "[…] dieses Grundverständnis, dass man auch ein bisschen ein Misstrauen haben kann den Systemen gegenüber" (IP5/Z. 150-151).

Für IP6 äußert sich dieses Grundverständnis von technischen Systemen in einem rückblickenden historischen Vergleich:

> "Selbst was wir jetzt da sehen mit Musikgenerierung und Textgenerierung, schaut schon super magisch aus, ja. Aber ich meine, zu Leibniz-Zeiten war es auch super magisch, dass eine Rechnung, eine Maschine multiplizieren konnte. Ja. Und später hat man sich dran gewöhnt, dass Maschinen das halt können, was Maschinen halt können" (IP6/Z. 972-976).

> "[…] damals haben wir uns schon alles angehört, was die AI nicht alles werden wird in den nächsten fünf Jahren. Jetzt sind wir es fast, jetzt sind wir fast 30 Jahre später. Und jetzt freuen wir uns, dass sie ein paar, sich freundlich anhörende Texte verfasst, oh mei, ja" (IP6/Z. 537-540).

Für IP3 liegt es an uns Menschen, welche Fähigkeiten wir technischen Systemen zuschreiben oder sie bezeichnen:

> "[…] ich würde es aber nicht unbedingt Intelligenz nennen, ja. Also beispielsweise, dieses ChatGPT" (IP3/Z. 636-637).

Ob Intelligenz, Vernunft, Ethik, oder freier Wille des Menschen – verortet wird dies alles im menschlichen Gehirn, jenem Substrat, an das alles gebunden ist:

> "[…] weil jedes Gehirn einzigartig ist, weil es eben aus Abermillionen Verknüpfungen, die aus Abermillionen Erfahrungen aus Genen, die bestimmte Dinge prädisponieren, aus Umweltfaktoren, aus vielfältigen Rückmeldungen, Rückkopplungsschleifen, Verstärkungen, Bestrafungen usw. resultieren, sodass es trotzdem mein freier Wille ist, aber letztlich ist es alles an das Substrat gebunden" (IP4/Z. 725-730).

Denn, so IP3, "[…] in dem Moment wo ich mich komplett fremdsteuern lasse, gebe ich ja meine Selbstverantwortung vollkommen ab" (IP3/Z. 601-602).

6 Diskussion & Interpretation

In diesem Kapitel werden die Ergebnisse der Datenauswertung nach dem paradigmatischen Modell der Grounded Theory und die aus den Expert*inneninterviews gewonnen Meinungen der Experten*innen interpretiert und in Bezug zum Stand des Wissens diskutiert. Die Limitierung der Studie auf den Untersuchungsgegenstand wird ebenso dargestellt wie die Schlussfolgerungen, in denen die Erkenntnisse der Forschung präsentiert werden, um in einer Beantwortung der Forschungsfrage zu münden. Eine abschließende Empfehlung über mögliche weiterführende Untersuchungen zu dem Forschungsgegenstand dieser Arbeit schließt dieses Kapitel.

Die Ergebnisse der Arbeit sollen dazu beitragen, sich dem Thema KI aus verschiedenen Positionen wie den Human-, Geistes-, Sozial-, Ingenieur- und Naturwissenschaften anzunähern, um relevante Herausforderungen im Rahmen des Themenkomplexes angemessen erfassen zu können und damit bearbeitbar zu machen. In nahezu jedem Wissenschaftsbereich zählt KI zum Forschungsbehavior. Der Einfluss von KI wirkt in die Lebenswelten der Menschen und beeinflusst soziale, politische, ökonomische und ökologische Entwicklungen.

Roboterethiker*innen, Technikphilosoph*innen, Entwickler*innen von KI-Systemen - sie alle seien an dieser Stelle nur beispielhaft für Nutzer*innen, Profiteur*innen, Beteiligte, Teilhabende oder Betroffene der Wirkmächtigkeit von KI genannt.

In diesem Zusammenhang liegt der Schluss nahe, das Wesen und die Existenz von KI mit dem ersten Satz des *Tractatus logico-philosophicus* von Ludwig Wittgenstein zu fassen: „Die Welt ist alles, was der Fall ist“ (Wittgenstein, 2021, S. 9). Um gleichzeitig mit Wittgenstein auf die Welt zu blicken und herauszufinden, wo die Grenze von dem liegt worüber wir schweigen müssen und dem, was wir sinnvoll zeigen oder sagen können. Wir werden in unsere Sprache hineingeboren, aber daraus hinaus kommen wir nicht. „Wovon man nicht sprechen kann, darüber muss man schweigen“ (Wittgenstein, 2021, S. 111).

6.1 Einleitung

Eine genaue und methodengeleitete Untersuchung der Phänomene wie Bewusstsein, Seele, Intelligenz oder Freiheit in Verschränkung mit der Entwicklung von KI ist eine Voraussetzung für die Untersuchung des Forschungsgegenstandes sowie die Beantwortung der Forschungsfrage nach der Wirkmacht von KI auf das menschliche Selbstverständnis und im Speziellen auf das Verständnis von Freiheit im Sinne Sartres. Die Literaturbefunde und Forschungsergebnisse zu Spannungsfeldern und Herausforderungen von KI breiten sich über alle Wissenschaftsgebiete aus. Während sich die

Technikphilosophie mit ethischen Fragen beschäftigt, widmet sich die Informatik den technischen Herausforderungen und Möglichkeiten von KI.

6.2 Limitationen

In dieser Arbeit wurde ausschließlich die Gültigkeit des von Sartre postulierten Freiheitsbegriffes und seiner Freiheitsidee in seinem philosophischem Kontext untersucht. Das zentrale Interesse der Forschung richtete sich dabei auf die Wirkmächtigkeit und Auswirkungen von KI, insbesondere von mit Bewusstsein ausgestatteter KI, auf Sartres Freiheitsbegriff. Kein Gegenstand der Forschung dieser Arbeit waren andere philosophische Erkenntnisse zum allgemeinen Phänomen *Freiheit* sowie der Einfluss von KI-Systemen darauf. Gleichwohl ergab sich eine Auseinandersetzung mit Phänomenen und Begrifflichkeiten wie Ethik, Vernunft, Geist, Würde, Intelligenz, Kultur, Ethik oder Moral in weiteren unterschiedlichsten Themenbereichen der Philosophie. Die wertvollen Information aus der Beschäftigung mit diesen Themenbereichen schärften das erkenntnisleitende Interesse an der Forschungsfrage.

6.3 Schlussfolgerungen

Die Ursache dafür, dass in der philosophischen Konzeption von Sartre beim Menschen Selbstbewusstsein auftreten kann, ist Freiheit. Der Mensch als ein Handelnder ist dazu verurteilt, frei zu sein und sich damit ständig entwerfen und sich Ziele setzen zu müssen. Die Untersuchungen über den von Sartre postulierten Freiheitsbegriff, dass der Mensch zur Freiheit verurteilt sei, mündeten aufgrund der Datenauswertung nach dem paradigmatischen Modell der Grounded Theory und den aus den Expert*inneninterviews gewonnenen Meinungen der Experten*innen in wertvollen Erkenntnissen über den Forschungsgegenstand und die zu beantwortende Forschungsfrage.

Als *die* grundlegende Erkenntnis aus der empirischen Untersuchung hat sich, ebenso wie aus der theoretischen Untersuchung in der Auseinandersetzung mit dem Stand des Wissens und der daraus erfolgten Literaturanalyse, der Begriff *Bewusstsein* als zentrales Phänomen herausgestellt. In diesem Phänomen, dem *Bewusstsein*, treffen die Wirkmächtigkeit von KI und der von Sartre postulierte Freiheitsbegriff aufeinander. Die Erkenntnis, dass es ohne ein Bewusstsein keinen Sinn gibt, da es mit Bewusstsein ausgestattete Wesen sind, die erst Sinn verleihen, zeigt sich in den Ergebnissen der empirischen Untersuchung, wenn die Entwicklung neuer Technologien, die die Handlungsoptionen des Menschen erweitern, als sinnvoll erlebt werden. Der Mensch entwickelt Problembewusstsein, um Entscheidungen bewusst und in Selbstverantwortung zu treffen.

Die konzeptionelle Gültigkeit von Sartres Freiheitsbegriff zeigt sich durchgehend in allen Ergebnissen der empirischen Untersuchung. Dem Einsatz neuer Technologien und deren Evaluierung zum Zwecke der Erkenntnis darüber, ob sie dem Menschen dienlich sind, geht voraus, dass der Mensch immer über deren

Einsatz entscheiden muss. Selbst beim Auftreten der Wirkmächtigkeit von KI in den unterschiedlichsten Handlungsfeldern wie Industrie, Finanz- und Geschäftswelt oder Medizin ist der Mensch zu der von Sartre definierten Freiheit verurteilt und stets ein Handelnder.

Auch das Auftreten invasiver Technologien, sogenannter Mensch-Maschine Schnittstellen, hat so lange keine Auswirkung auf die Gültigkeit von Sartres Freiheitsbegriff und die Tatsache, dass der Mensch sich stets neu entwerfen und Entscheidungen treffen muss, solange diese invasiven Technologien, seien sie auch KI-Systeme nach dem derzeitigen Stand der Technik und keine mit künftig möglichem Bewusstsein ausgestattete AGI oder superintelligente KI, beispielsweise zum Zwecke medizinischer Indikationen eingesetzt werden.

Ein kritisches Bewusstsein gegenüber Technologie zeigt sich in den Ergebnissen der empirischen Untersuchung zu technologischen Entwicklungen, die nicht erfüllte Hoffnungen und Erwartungen zur Folge hatten. Diese kritische Haltung wird auch gegenüber KI-Systemen eingenommen, wissend um die Gefahren, die durch deren Einsatz in unterschiedlichsten Handlungsfeldern entstehen könnten. Bedenken derselben Art finden sich auch durchgehend in den Ergebnissen der Literaturanalyse.

Die größten Hoffnungen und zugleich auch die größten Bedenken zeigen sich sowohl in den Ergebnissen der Literaturanalyse als auch in den Ergebnissen der empirischen Untersuchung, wenn es um die Wirkmächtigkeit von mit Bewusstsein ausgestatteten KI-Systemen geht. Ab dem Zeitpunkt, wo eine AGI oder eine superintelligente und mit Bewusstsein ausgestatte KI ihre Wirkmächtigkeit entfalten würde, würde dies einen massiven Einfluss auf die Gesellschaft und jeden einzelnen Menschen haben.

Das Aufkommen von AGI und superintelligenter KI, ausgestattet mit Bewusstsein, würde demnach die logische Überlegung nach sich ziehen, dass auch diese KI-Systeme nach Sartre zur Freiheit verurteilt sein würden und diese Freiheit somit nach Sartre KI-Systeme zwingen würde, *sich zu machen* statt zu *sein*. Damit wäre die Gültigkeit des Freiheitsbegriffes von Sartre zwar für den Menschen weiterhin gegeben, müsste jedoch demnach auch AGI und superintelligenter KI zugestanden werden. Ob dies dann nach den von Freud konstatierten drei Kränkungen der Menschheit die vierte Kränkung der Menschheit zur Folge hätte, bleibt eine hypothetische Frage.

Aus einer Erkenntnis der empirischen Untersuchung dieser Arbeit geht hervor, dass das menschliche Gehirn einzigartig ist und der Mensch mit freiem Willen ausgestattet ist. Diese These, dass der Mensch mit freiem Willen ausgestattet ist wird durch eine Erkenntnis aus der Literaturanalyse insofern bestärkt, indem die gegenteiligen Ergebnisse des Libet-Experimentes, dass der Mensch keinen freien Willen habe, durch die von Paul Watzlawick bezeichnete *Sei-spontan-Paradoxie* entkräftet werden.

Inwieweit in einem weiteren Szenario noch von einem freiem Willen gesprochen werden kann, wenn mit AGI und superintelligenter KI ausgestattete invasive Technologien beim Menschen zum Einsatz kommen und unseren Geist erweitern, ist eine sich aufdrängende hypothetische Frage. Denn, so die Schlussfolgerung, es würde in dem Moment, in dem solche invasive Technologien beim Menschen zum Einsatz kommen würden und es nicht mehr klar und eindeutig sein würde, welches der beiden *Bewusstseine*, das menschliche oder das künstliche, beim Aufeinandertreffen in unserem Gehirn und unserem Geist den Menschen im Sinne des Freiheitsbegriffes von Sartre neu entwerfen, nicht nur zu einer Falsifikation des von Sartre postulierten Freiheitsbegriffes kommen, sondern aufgrund des nicht mehr zu adressierenden Phänomens *Bewusstsein* sogar zu dessen Auflösung.

Die in dieser Studie erhobenen empirischen und theoretischen Ergebnisse führten zu einer Beantwortung der Frage nach der Gültigkeit des von Sartre postulierten Freiheitsbegriffes.

Die wissenschaftliche Fragestellung dieser Masterarbeit lautet: Wie wirken sich die Kompetenzen und Fähigkeiten von Künstlicher Intelligenz auf die Gültigkeit des von Jean-Paul Sartre postulierten Freiheitsbegriff des Menschen aus?

Die Ergebnisse dieser Studie lassen den Schluss zu, dass der von Sartre postulierte Freiheitsbegriff nach wie vor Gültigkeit hat. Der Grund dafür liegt darin, dass nach Sartre das menschliche Selbstbewusstsein ausschließlich durch das Auftreten von Freiheit entstehen kann und damit eine unhintergehbare Bedingung der menschlichen Existenz darstellt. Alle Untersuchungen in dieser Arbeit weisen darauf hin, dass die Wirkmächtigkeit von KI zu keiner Falsifikation des von Sartre postulierten Freiheitsbegriffes des Menschen führt.

Der in dieser Studie untersuchte Gegenstand hat auf die Gesellschaft sowie auf jedes einzelne Individuum massive Implikationen. In einer philosophischen Annäherung an die Frage "Wer bist du, KI?" wurde untersucht, mit welcher Wirkmacht Künstliche Intelligenz und die damit verbundene Technologie auf das Selbstverständnis des Menschen trifft. Antworten darauf finden sich in dieser Arbeit und sollen einen Beitrag sowie neue Erkenntnisse über Künstliche Intelligenz und unser menschliches Selbstverständnis leisten.

Zukünftige Forschungen über den Untersuchungsgegenstand dieser Masterarbeit könnten die in dieser Studie erhobene Hypothese untersuchen, welche Auswirkungen der Einsatz von mit AGI und superintelligenter KI ausgestatteten invasiven Technologien beim Menschen auf die Gültigkeit des von Sartre postulierten Freiheitsbegriffes haben könnten. Die in dieser Masterarbeit aufgeworfene Frage, ob die Wirkmächtigkeit von mit AGI und superintelligenter KI ausgestatteten invasiven Technologien zu einer vierten Kränkung des Menschen führt, wäre ein weiterer Untersuchungsgegenstand zukünftiger Forschungen.

7 Zusammenfassung

Künstliche neuronale Netze die eigene Stimmen erzeugen und dafür keine Stimmbänder brauchen, Transhumanismus und Unsterblichkeit, menschliche Intelligenz und individuelles menschliches Bewusstsein unabhängig vom schwachen und anfälligen eigenen biologischen Körper – all diese Begriffe verdanken ihre Bedeutung dem Phänomen und der Wirkmächtigkeit Künstlicher Intelligenz (KI).

In dieser Arbeit wird ein interdisziplinärer Forschungsansatz gewählt, um dem Gegenstand von KI in seinen unterschiedlichsten Ausprägungen näher zu kommen. Dies führt zur Notwendigkeit, sich sowohl mit technischen Grundlagen als auch mit philosophischen Fragen auseinanderzusetzen.

In der wissenschaftlichen Fragestellung dieser Masterarbeit wird untersucht, wie sich die Kompetenzen und Fähigkeiten von Künstlicher Intelligenz auf die Gültigkeit des von Jean-Paul Sartre postulierten Freiheitsbegriff des Menschen auswirken.

Um sich dieser Forschungsfrage anzunähern, werden zunächst grundlegende wissenschaftliche Erkenntnisse zu KI sowie die Wirkmächtigkeit von KI anhand relevanter Literatur erörtert. Im Besonderen konzentrieren sich die Untersuchungen auf die Auswirkungen von KI-Systemen auf das menschliche Selbstverständnis. Die rasante Entwicklung von *Schwacher KI* hin zu *Starker KI*, von Mensch-Maschine-Schnittstellen hin zu einer mit Bewusstsein ausgestatteten KI trifft dabei auf den von Sartre postulierten Freiheitsbegriff des Menschen. In diesem behauptet Sartre, dass der Mensch zur Freiheit verurteilt ist. Weitere Grundlagen zu Sartres philosophischem Gedankengebäude werden anhand relevanter Literatur in ihrem philosophischen Kontext erörtert, um danach die beiden Untersuchungsgegenstände, KI und die Gültigkeit von Sartres Freiheitsbegriff, gegeneinander zu diskutieren und sich ihnen zu nähern.

Die technischen Grundlagen und Funktionsweisen von KI wie zum Beispiel Machine Learning oder Deep Learning werden dargestellt und erläutert, auch um die Entwicklungen dieser Technologie historisch zu beleuchten und um daraus Schlüsse für künftige Entwicklungen zu ziehen.

Eine Annäherung an die Beantwortung der Forschungsfrage erfolgt in Form von Expert*inneninterviews und den daraus in einer qualitativen Inhaltsanalyse gewonnenen Daten, welche den Erkenntnissen aus der Literaturanalyse gegenübergestellt werden.

Das Ergebnis der Gegenüberstellung und des Vergleiches der Literaturbefunde mit den ausgewerteten Daten der Expert*inneninterviews führte zu der Erkenntnis, dass der von Jean-Paul Sartre postulierte Freiheitsbegriff weiterhin seine Gültigkeit behält, da nach Sartre das menschliche Selbstbewusstsein ausschließlich durch das Auftreten von Freiheit entstehen kann und damit eine

unhintergehbare Bedingung der menschlichen Existenz darstellt, die auch von der Wirkmächtigkeit Künstlicher Intelligenz nicht aufgehoben werden kann.

8 Literatur

Bakewell, S. (2021). Das Café der Existenzialisten: Freiheit, Sein und Aprikosencocktails: mit Jean-Paul Sartre, Simone de Beauvoir, Albert Camus, Martin Heidegger, Edmund Husserl, Karl Jaspers, Maurice Merleau-Ponty und anderen (R. Seuß, Übers.; 4. Auflage). C.H.Beck.

Bartneck, C., Lütge, C., Wagner, A. R., & Welsh, S. (2019). Ethik in KI und Robotik. Hanser.

Bauder, M. (2021, 10. Dezember). Wer wir waren. https://www.werwirwaren.de/downloads/Schulheft-WER-WIR-WAREN.pdf

Beauvoir, S. de. (2017). Memoiren einer Tochter aus gutem Hause (46. Auflage). Rowohlt Taschenbuch Verlag.

Böhme, G. (2008). Invasive Technisierung: Technikphilosophie und Technikkritik. die Graue Edition.

Bostrom, N. (2020). Superintelligenz: Szenarien einer kommenden Revolution (J.-E. Strasser, Übers.; 4. Auflage). Suhrkamp.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. Nature, 391(6669), 756–756. https://doi.org/10.1038/35784

Chromik, M., & Butz, A. (2021). Human-XAI Interaction: A Review and Design Principles for Explanation User Interfaces. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, & K. Inkpen (Hrsg.), Human-Computer Interaction – INTERACT 2021 (Bd. 12933, S. 619–640). Springer International Publishing. https://doi.org/10.1007/978-3-030-85616-8_36

Coeckelbergh, M. (2016). Responsibility and the Moral Phenomenology of Using Self-Driving Cars. Applied Artificial Intelligence, 30(8), 748–757. https://doi.org/10.1080/08839514.2016.1229759

Coeckelbergh, M., & Loh, J. (2020). Transformations of Responsibility in the Age of Automation: Being Answerable to Human and Non-Human Others. In B. Beck & M. Kühler (Hrsg.), Technology, Anthropology, and Dimensions of Responsibility (Bd. 1, S. 7–22). J.B. Metzler. https://doi.org/10.1007/978-3-476-04896-7_2

Cordes, L. (2019). Übernimmt der Algorithmus? Medizinrecht, 37(10), 797–798. https://doi.org/10.1007/s00350-019-5359-8

Damasio, A. R. (2011). Selbst ist der Mensch: Körper, Geist und die Entstehung des menschlichen Bewusstseins (3. Aufl). Siedler.

Dengel, A., Socher, R., Kirchner, E. A., & Ogolla, S. (2023). Künstliche Intelligenz—Die Zukunft von Mensch und Maschine. Zeit Akademie. https://www.zeitakademie.de/

Ertel, W. (2021). Grundkurs Künstliche Intelligenz: Eine praxisorientierte Einführung (5. Auflage). Springer Vieweg.

Floridi, L. (2016). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083), 20160112. https://doi.org/10.1098/rsta.2016.0112

Ford, M. (2019). Die Intelligenz der Maschinen: Mit Koryphäen der Künstlichen Intelligenz im Gespräch (K. Lorenzen, Übers.; 1. Auflage). mitp.

Freud, S. (1917). Imago. Zeitschrift für Anwendung der Psychoanalyse auf die Geisteswissenschaften V (1917). Imago, V. Jahrgang Heft 1. https://doi.org/10.11588/diglit.25679#0012

Fritz, O., Weber, C., König, A., & Wolf, J. (2019). Ethische Aspekte der künstlichen Intelligenz. MA Akademie Verlags- und Druck-Gesellschaft mbH.

Future of Life Institute. (2018). Die KI-Leitsätze von Asilomar. https://futureoflife.org/open-letter/ai-principles-german/

Future of Life Institute. (2023). Pause Giant AI Experiments: An Open Letter. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

GPT-4 Technical Report. (2023). https://doi.org/10.48550/arXiv.2303.08774

Grunwald, A. (2019). Der unterlegene Mensch: Die Zukunft der Menschheit im Angesicht von Algorithmen, künstlicher Intelligenz und Robotern (Originalausgabe, 1. Auflage). riva verlag.

Günther, G. (1956). Seele und Maschine. Augenblick 1,3.

Haux, R., Gahl, K., Jipp, M., & Richter, O. (2021). Zusammenwirken von natürlicher und künstlicher Intelligenz (1. Auflage). Springer VS.

Hawking, S. W. (2019). Kurze Antworten auf große Fragen (S. Held & H. Kober, Übers.; 19. Auflage). Klett-Cotta.

Heinrich, L. J., Heinzl, A., & Roithmayr, F. (2007). Wirtschaftsinformatik: Einführung und Grundlegung (3., vollst. überarb. und erg. Aufl). Oldenbourg.

Heiser, P. (2018). Meilensteine der qualitativen Sozialforschung: Eine Einführung entlang klassischer Studien. Springer VS.

Hengstschläger, M. & Rat für Forschung und Technologieentwicklung (Hrsg.). (2020). Digitaler Wandel und Ethik (1. Auflage). Ecowin.

Herger, M. (2020). Wenn Affen von Affen lernen: Wie künstliche Intelligenz uns erst richtig zum Menschen macht. Plassen Verlag.

Holz, H. H. (1958). Der französische Existentialismus Theorie und Aktualität. Oswald Dobbeck Verlag, Speyer-München.

Husserl, E. (2002). Zur phänomenologischen Reduktion: Texte aus dem Nachlass (1926-1935) (S. Luft, Hrsg.; Softcover reprint of the hardcover 1st edition 2002). Springer-Science+Business Media, B.V.

Jipp, M., & Steil, J. (2021). Steuern wir oder werden wir gesteuert? Chancen und Risiken von Mensch-Technik-Interaktion. In Zusammenwirken von natürlicher und künstlicher Intelligenz (S. 17–34). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-30882-7_3

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kant, I. (1781). Kritik der reinen Vernunft.

Kant, I. (1968). Kants Werke: Bd.9, Logik, Physische Geographie, Pädagogik (Faks.-Ausg). W. de Gruyter.

Kant, I. (2014). Kritik der reinen Vernunft/Kritik der praktischen Vernunft. Nikol.

Kuhlmann, P. (2018). Künstliche Intelligenz: Einführung in Machine Learning, Deep Learing, Neuronale Netze, Robotik und Co (1. Auflage). Independently published.

Lee, K. (2019). AI Superpowers: China, Silicon Valley und die neue Weltordnung (J. W. Haas, Übers.). Campus Verlag.

Lee, K., & Chen, Q. (2022). KI 2041: Zehn Zukunftsvisionen (T. Schmidt, Übers.). Campus.

Libet, B. (1983). Time of conscious intention to act in relation to onset of cerebral acitivity (Readiness-Potential): The unconscious initiation of a freely voluntary act. Brain, 106(3), 623–642. https://doi.org/10.1093/brain/106.3.623

Liessmann, K. P. (2023). Lauter Lügen (1. Auflage). Zsolnay, Paul.

Liessmann, K. P. (2000). Denken und Leben II. Annäherungen an die Philosophie in biographischen Skizzen. 4-teilige ORF-CD-Edition.

Loh, J. (2019a). Roboterethik: Eine Einführung (Erste Auflage, Originalausgabe). Suhrkamp.

Loh, J. (2019b). Trans- und Posthumanismus zur Einführung (2., überarbeitete Aufl). Junius.

Macedonia, M., Kern, R., & Roithmayr, F. (2014). Do Children Accept Virtual Agents as Foreign Language Trainers? International Journal of Learning, Teaching and Educational Research, Vol. 7(1), 131–137.

Mancuso, S., Viola, A., & Mancuso, S. (2015). Die Intelligenz der Pflanzen. Kunstmann.

McCarthy, J. (2007). What is Artificial Intelligence? http://www-formal.stanford.edu/jmc/whatisai.html

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

Metzinger, T. (2017). Der Ego Tunnel: Eine neue Philosophie des Selbst: von der Hirnforschung zur Bewusstseinsethik (T. Schmidt, Übers.; 6. Auflage). Piper.

Misoch, S. (2015). Qualitative Interviews. de Gruyter Oldenbourg.

Müller, A., Schröder, H., & Thienen, L. von. (2021). Digineering: Business Process Management im digitalen Zeitalter. Springer Vieweg.

Newton, I. (1675). Isaac Newton letter to Robert Hooke. https://digitallibrary.hsp.org/index.php/Detail/objects/9792

Nida-Rümelin, J. (2018). Digitaler Humanismus: Eine Ethik für das Zeitalter der künstlichen Intelligenz (Originalausgabe). Piper.

Nilsson, N. J. (2010). The quest for artificial intelligence: A history of ideas and achievements. Cambridge University Press.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still Vital After All These Years of Automation. Human Factors: The Journal of the Human Factors and Ergonomics Society, 50(3), 511–520. https://doi.org/10.1518/001872008X312198

Pollirer, H.-J., Weiss, E. M., Knyrim, R., & Haidinger, V. (Hrsg.). (2017). Datenschutz-Grundverordnung (DSGVO). Manz'sche Verlags- und Universitätsbuchhandlung.

Przyborski, A., & Wohlrab-Sahr, M. (2014). Forschungsdesigns für die qualitative Sozialforschung. In Handbuch Methoden der empirischen Sozialforschung (S. 117–133). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-18939-0_6

Rädiker, S., & Kuckartz, U. (2019). Analyse qualitativer Daten mit MAXQDA: Text, Audio und Video. Springer VS.

Rich, E. (1983). Artificial intelligence. McGraw-Hill.

Riedl, M. O. (2019). Human-Centered Artificial Intelligence and Machine Learning. https://doi.org/10.48550/ARXIV.1901.11184

Riedl, R. (2022). Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. Electronic Markets, 32(4), 2021–2051. https://doi.org/10.1007/s12525-022-00594-4

Riesewieck, M., & Block, H. (2020). Die digitale Seele: Unsterblich werden im Zeitalter Künstlicher Intelligenz. Goldmann.

Riesewieck, M., & Block, H. (2022). Vom Ende der Endlichkeit: Unsterblichkeit im Zeitalter Künstlicher Intelligenz (1. Auflage, gekürzte und aktualisierte Paperback-Ausgabe). Goldmann.

Russell, S. J. (2020). Human compatible: Künstliche Intelligenz und wie der Mensch die Kontrolle über superintelligente Maschinen behält (G. Lenz, Übers.; 1. Auflage). mitp Verlags GmbH & Co. KG.

Russell, S. J., & Norvig, P. (2012). Künstliche Intelligenz: Ein moderner Ansatz (F. Langenau, Übers.; 3., aktualisierte Auflage). Pearson, Higher Education.

SAE. (2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE International. https://doi.org/10.4271/J3016_201806

Sartre, J.-P. (2020). Das Sein und das Nichts: Versuch einer phänomenologischen Ontologie (T. König & V. von Wroblewsky, Hrsg.; 22. Auflage). Rowohlt-Taschenbuch-Verlag.

Sartre, J.-P. (2021). Der Existentialismus ist ein Humanismus und andere philosophische Essays 1943-1948 (11. Auflage). Rowohlt Taschenbuch Verlag.

Shneiderman, B. (2020a). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. International Journal of Human–Computer Interaction, 36(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Shneiderman, B. (2020b). Human-Centered Artificial Intelligence: Three Fresh Ideas. AIS Transactions on Human-Computer Interaction, 109–124. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1136&context=thci

Spiekermann, S. (2021). Digitale Ethik: Ein Wertesystem für das 21. Jahrhundert. Droemer Taschenbuch.

Stern, W. (1912). Die psychologischen Methoden der Intelligenzprüfung und deren Anwendung an Schulkindern. Johann Ambrosius Barth.

Strauss, A. L., & Corbin, J. M. (1996). Grounded Theory: Grundlagen qualitativer Sozialforschung. Beltz.

Streller, J. (1952). Zur Freiheit verurteilt—Ein Grundriß der Philosophie Jean Paul Sartres. Felix Meiner.

Strübing, J. (2013). Qualitative Sozialforschung: Eine komprimierte Einführung für Studierende. De Gruyter Oldenbourg.

Tegmark, M. (2020). Leben 3.0: Mensch sein im Zeitalter Künstlicher Intelligenz (H. Mania, Übers.; 3. Auflage). Ullstein.

Turing, A. M. (1950). Computing Machinery and Intelligence. Mind, LIX(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Turk, M. (2014). Multimodal interaction: A review. Pattern Recognition Letters, 36, 189–195. https://doi.org/10.1016/j.patrec.2013.07.003

Vogd, W. (Regisseur). (2023, März 23). Wie entscheidet unser Gehirn. In Punkt eins. https://oe1.orf.at/programm/20230323/712972/Wie-entscheidet-unser-Gehirn

Vorländer, K. (1990). Geschichte der Philosophie. Bd. 2: Mittelalter und Renaissance (Neuausg). Rowohlt.

Vorschlag für eine Verordnung des europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für die Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union. (2021). https://eur-

lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0019.02/DOC_1&format=PDF

Vorschlag für einen Rechtsrahmen für künstliche Intelligenz Gestaltung der digitalen Zukunft Europas. (2022). https://digital-strategy.ec.europa.eu/de/policies/regulatory-framework-ai

Werner, J. (2021). Existentialismus in Österreich: Kultureller Transfer und literarische Resonanz (Bd. 153). De Gruyter. https://doi.org/10.1515/9783110683066

Wildenburg, D. (2004). Jean-Paul Sartre. Campus-Verl.

Wildenburg, D. (2011, 15. Dezember). «Das Sein und das Nichts» von Jean-Paul Sartre. In: SRF Kultur Reflexe. https://www.srf.ch/audio/reflexe/das-sein-und-das-nichts-von-jean-paul-sartre?id=10204826

Wischmeyer, T. (2018). Regulierung intelligenter Systeme. Archiv des öffentlichen Rechts, 143(1), 1. https://doi.org/10.1628/aoer-2018-0002

Wittgenstein, L. (2021). Logisch-philosophische Abhandlung: = Tractatus logico-philosophicus (38. Auflage). Suhrkamp.

Wittgenstein, L. (2022). Philosophische Untersuchungen (11.Auflage). Suhrkamp Verlag.

Yurish, S. Y. (2010). Sensors: Smart & Transducers. Vol.114, Issue 3.

Žižek, S. (2020). Hegel im verdrahteten Gehirn (Deutsche Erstausgabe). S. Fischer.

Zweig, K. A. (2019). Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können (Originalausgabe). Heyne.

Abbildungsverzeichnis

Abkürzungen

AI	Artificial Intelligence
AGI	Artificial General Intelligence
BCI	Brain-Computer-Interface
DNA	Deoxyribonucleic Acid
DNI	Direkte Neurale Interfaces
DSGVO	Datenschutz-Grundverordnung
EU	Europäische Union
GPT	Generative pre-trained transformers
IEEE	Institute of Electrical and Electronics Engineers
IT	Informationstechnologie
KI	Künstliche Intelligenz
MIT	Massachusetts Institute of Technology
ML	Machine Learning
MMI	Mind-Machine Interfaces
NASA	National Aeronautics and Space Administration

Eidesstattliche Erklärung

„Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit selbständig ohne die Verwendung unerlaubter Hilfsmittel verfasst und keine anderen als die angegebenen Quellen verwendet habe. Alle Stellen, die wörtlich oder sinngemäß den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.

Sofern von der Studiengangsleitung eine Verwendung von Hilfsmitteln (insbesondere IT- und KI-gestützte) vorgesehen ist, erkläre ich, diese in der Arbeit mit dem jeweiligen Produktnamen, der Produktversion und einer Beschreibung des genutzten Funktionsumfangs vollständig angeführt zu haben.

Zudem versichere ich, dass ich diese Arbeit gemäß der geltenden Prüfungsordnung der FH Burgenland sowie den Richtlinien der Österreichischen Agentur für wissenschaftliche Integrität zur guten wissenschaftlichen Praxis verfasst habe. Die Arbeit wurde bisher weder im Inland noch Ausland zur Begutachtung oder Beurteilung vorgelegt und nicht veröffentlicht.“

Ort, Datum | Unterschrift

Steinbrunn, 01. Juni 2023

Signiert von: Michael Fleischhacker
Datum: 01.06.2023 20:24:41

A-TRUST

Dieses Dokument ist digital signiert!

Dieses mit einer qualifizierten elektronischen Signatur versehene Dokument hat gemäß Art. 25 Abs. 2 der Verordnung (EU) Nr 910/2014 vom 23. Juli 2014 ("eIDAS-VO") die gleiche Rechtswirkung wie ein handschriftlich unterschriebenes Dokument.

Prüfinformation: Informationen zur Prüfung der elektronischen Signatur finden Sie unter: www.a-trust.at/pdf