

Sommersemester 2021

LVA: „Technik für Menschen 2040“

Literaturarbeit

Chan Hyuk Kong Matr. 01427388

Natasa Nikic Matr. 01325974

Florian Nikolaus Schager Matr. 11819578

Bianca Träger Matr. 01326815

2. Mai 2021

Inhaltsverzeichnis

1	Literaturarbeit	3
1.1	Gruppe	3
1.2	Nick Bostrom, Superintelligence	3
1.2.1	Synopsis	3
1.2.2	Erkenntnisse	4
1.2.3	Kritik	5
1.3	Nick Bostrom, Superintelligence	5
1.3.1	Synopsis	5
1.3.2	Erkenntnisse	6
1.3.3	Kritik	7
1.4	Julien Offray de La Mettrie, Der Mensch eine Maschine	7
1.4.1	Synopsis	7
1.4.2	Erkenntnisse	8
1.4.3	Kritik	8
1.5	Richard David Precht, Künstliche Intelligenz und der Sinn des Lebens	9
1.5.1	Synopsis	9
1.5.2	Erkenntnisse	9
1.5.3	Kritik	10
2	Gegenüberstellung	11
3	Fragen und Diskussion in der VU	12
4	Literaturverzeichnis	13

1 Literaturarbeit

1.1 Gruppe

Die Aufteilung der Bücher innerhalb der Gruppe war wie folgt:

- 01427388, Chan Hyuk Kong:
Nick Bostrom, Superintelligence [1]
- 01325974, Natasa Nikic:
Julien Offray de la Mettrie, Der Mensch eine Maschine [2]
- 11819578, Florian Nikolaus Schager:
Nick Bostrom, Superintelligence [1]
- 01326815, Bianca Träger:
Richard David Precht, Künstliche Intelligenz und der Sinn des Lebens [3]

1.2 Nick Bostrom, Superintelligence

Im Rahmen der Literaturkritik wurde das Buch Superintelligence von Nick Bostrom aus dem Jahr 2014 gelesen.

1.2.1 Synopsis

Das Buch versucht mit einer Lesung der Geschichte des globalen BIP in Verbindung mit der Entwicklung der Geschwindigkeit in der Produktivität einen weiteren plausiblen Baustein für die Zukunft zu legen. Dabei fokussiert er sich auf die mögliche Entwicklungswege der künstlichen generellen Intelligenz, die möglicherweise die des durchschnittlichen Menschen um ein vielfaches übertreffen könnte und zeigt, wie diese Entwicklung eine existentielle Krise für den Menschen bedeuten, gleichzeitig aber auch wie man dieses fatale Ende vermeiden kann. Bostrom denkt dazu viele mögliche Formen der Superintelligenz (Definition: any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest) und deckt auf, welche Formen größere Gefahren tragen könnten oder welche weniger.

Man kann das Buch in 3 Teile zerlegen. Der erste Teil behandelt die mögliche technische Entwicklung einer Superintelligenz. Der zweite Teil beschreibt den „takeoff“, die Zeit die benötigt wird um anfangen von der künstlichen generellen Intelligenz eines durchschnittlichen Menschen die Superintelligenz erreicht und wie man diese Geschwindigkeit kontrollieren kann, da eine schnelle Geschwindigkeit gewisse Gefahren beinhalten. Als letzten Punkt und auch der wichtigste Punkt um ein fatales Ergebnis zu vermeiden, behandelt Bostrom Methoden um die Superintelligenz zu kontrollieren, als auch die Schwierigkeit friedliche Werte zu programmieren.

Als erstes die technische Wege eine Superintelligenz zu kreieren:

- A.I
- Emulation des gesamten Gehirns (das Gehirn als Software simulieren und optimieren im Computer)
- Biologische Kognition (bspw. Optimierung der Wahrnehmung mit Hilfe von Drogen; genetische Selektion...)
- Gehirn - Computer Interface (Implantation der Computerelemente in den Menschen; Bildung eines Mensch-Maschine Hybrids)
- Netzwerk und Organisationen (Optimierung der quantitativen Vernetzung der Intelligenzen, wie den Menschen mit verschiedenen Artefakten und Bots um eine „kollektive Superintelligenz“ zu erreichen)

Auf diesen möglichen Wegen können verschiedene Formen von Superintelligenz erreicht werden, wie bspw. „Speed superintelligence“, ein System, das all das was der Mensch kann, schneller kann oder „Quality superintelligence“, ein System, das all das was der Mensch kann, mindestens gleich schnell kann, nur qualitativ „intelligenter“ und so weiter. Eine Sache betont Bostrom stark, nämlich, dass die Hardware der digitale Intelligenz ist um einiges vorteilhafter, als die des menschlichen Gehirns ist, weil sie vor allem flexibel gestaltet werden kann.

Im zweiten Teil vertieft sich Bostrom auf die Gefahren, die entstehen können, während dem „takeoff“. Hierbei unterscheidet er den Takeoff in langsam, schnell und moderate Geschwindigkeiten und zählt auf welche Vor- und Nachteile jede Geschwindigkeit hat und wie man sich zu vorbereiten hat, wie bspw. bei einem schnell Takeoff, beschreibt der Autor, wie in Minuten, Stunden oder Tagen die Superintelligenz entwickelt werden kann, nachdem man das Niveau der menschlichen Intelligenz erreicht hat. Da die Reaktionszeit zu kurz ist um ein fatales Ende zu vermeiden, muss man beim schnellen Takeoff eine gründliche Vorbereitung einplanen.

Im letzten Teil kommen alle erwähnten Kriterien einer Entwicklung der Superintelligenz zusammen. Wie gewisse Wege zu welchen Gefahren führen könnten, wie man diese Gefahren verhindern kann, wie man der Intelligenz beibringen kann, harmlos für den Menschen zu sein - wie Lernen von Moral - oder wie man, falls es zu einer Katastrophe oder Anzeichen auf eine katastrophale Übernahme der Menschheit durch die Maschine kommen sollte, die Superintelligenz ausgeschaltet/vernichtet werden kann.

1.2.2 Erkenntnisse

Das Thema der übermenschlichen Intelligenz oder künstlichen Intelligenz generell ist ein beliebter Zugang für viele Romane oder Filme. In diesem Hinblick war es für mich interessant zu überlegen, wie man den Film eventuell weiter zurückspulen könnte, zu den verschiedenen Szenarien die Bostrom bspw. aufgezählt hat. Wie genau hätte die Übernahme der Matrix sein können, wie wurde „Samantha“ im Film „Her“ entwickelt? Usw.

In der Zeit, wo Normen in der Architektur überhand gewinnen und der funktionelle, effektive Entwurf fast schon algorithmisch erarbeitet werden kann, in der Zeit, wo man für

die Generierung der Form unbedingt eine Maschine braucht, was für eine Rolle spielt der Architekt, der eben diese Überlegungen nicht besser könnte, als die künstliche Intelligenz?

1.2.3 Kritik

Wie Bostrom in der Einleitung erwähnt, dass viele Punkte, die er verwendet für seine Argumente, wahrscheinlich falsch sein werden, finde ich das Buch, nicht auf der faktischen Ebene betrachtet, sondern auf dem Narrativ, nicht besonders bereichernd. Alle Hypothesen, die auf Hypothesen aufgebaut sind, sind wackelig und die Fundamente, die der Autor aus Statistiken oder vergangene Experimente gebaut hat, ist viel zu eingeschränkt. Außerdem halte ich das Thema der brutalen Übernahme der Menschheit über die Maschinen nicht als das vordergründige Existenzproblem unsere Zeit, auch nicht die der Zukunft. Mehr im Vordergrund sollte das Lösen von sei es umwelttechnische Probleme, als auch die politische Stabilität, ein soziales Miteinander und die kapitalistische Konsumideologie stehen. Denn wir haben es heute auch mit einer existentiellen Not zu tun, die Teils unserer Arbeitskultur zu danken ist oder unserem hyperrealen Alltag.

1.3 Nick Bostrom, Superintelligence

Im Rahmen der Literaturkritik wurde das Buch Superintelligence von Nick Bostrom aus dem Jahr 2014 gelesen.

1.3.1 Synopsis

Der Autor baut strukturiert sein Argument dafür auf, dass die potentielle zukünftige Entwicklung von „Superintelligence“, also übermenschlicher künstlicher Intelligenz, eine existentielle Gefahr für die Menschheit darstellt. Daraus leitet er den Handlungsauftrag ab, bereits jetzt Forschung zur Sicherheit dieser Technologien priorisiert voranzutreiben.

Im ersten Teil des Buches argumentiert der Autor anhand der geschichtlichen Entwicklungen, sowie möglichen zukünftigen Innovationen, dass die Entwicklung von human-level artificial general intelligence (AGI) zwar wahrscheinlich nicht unmittelbar bevorsteht, aber eine reale Chance besteht, dass innerhalb des 21. Jahrhunderts human-level AGI erreicht werden könnte. Er beschreibt dabei in dem Zusammenhang auch die große Uneinigkeit unter Experten, was Zukunftseinschätzungen zum Thema AGI angeht. Des weiteren beschreibt er unterschiedliche mögliche Wege zu einer potentiellen Superintelligenz und welche möglichen „Takeoff“-Szenarien damit einhergehen könnten. Als „Takeoff“ bezeichnet er dabei den Prozess von human-level AGI zu vollständiger Superintelligenz, einer KI, die die menschlichen intellektuellen Fähigkeiten um Größenordnungen übersteigt, und somit das Potential hätte, die Welt sich ihrem Willen zu unterwerfen. In einem schnellen Szenario, welches er für durchaus wahrscheinlich hält, könnte dieser Prozess innerhalb weniger Stunden passieren, während ein langsamer „Takeoff“ sich über Monate bis Jahre hinziehen könnte.

Darauf aufbauend untermauert er seine These, dass eine Maschine, welche der Menschheit in genereller Intelligenz deutlich überlegen ist, definitiv in der Lage wäre, die Weltherrschaft an sich zu reißen. Dabei stellt er auch das Konzept von „instrumental convergence“ vor: Wir können zwar in der Theorie eine KI mit beliebigen Werten und Zielen („final goals“) versehen, die wenig miteinander gemein haben. Dennoch können wir heute schon „instrumen-

tal goals“ identifizieren, welche instrumentell hilfreich für eine breite Anzahl an möglichen „final goals“ sind. Ein wichtiges Beispiel dafür ist das Ziel der „goal-integrity“. Ein hinreichend intelligenter Agent wird mit allen Mitteln verhindern wollen, dass seine finalen Ziele geändert werden, da dies mit sehr hoher Wahrscheinlichkeit dazu führen würde, dass seine momentanen finalen Ziele nicht erreicht werden würden. Weitere Beispiele für allgemeine „instrumental goals“ sind der Erwerb von Ressourcen und der Drang zur Selbstverbesserung. Daher schließt er, dass die einzige Chance eine sichere Superintelligenz zu entwickeln, daraus besteht, bereits im Vorhinein adäquate Sicherheitsmaßnahmen zu treffen.

Im zweiten Teil beschäftigt sich Bostrom mit dem Problem, wie wir eine kontrollierte Intelligenzexplosion starten können, ohne die Menschheit dabei in die Sklaverei unter einer algorithmischen Wirtschaft zu verkaufen. Dabei geht es ihm besonders darum, wie wir eine KI mit menschlich verträglichen Werten und Zielen versehen können. Einerseits geht er hier auf mögliche technische Pitfalls bei der Implementierung ein, andererseits geht er aber auch auf das Problem ein, welche Ziele wir uns überhaupt für eine KI wünschen würden, wenn wir in der Lage wären, diese uneingeschränkt und ohne Informationsverlust der KI anzueignen. Insbesondere argumentiert er anhand von Beispielen, dass eine direkte Spezifikation unserer Ziele höchstwahrscheinlich katastrophale Folgen haben könnte. Zum Beispiel könnten wir unserer KI das Ziel geben das menschliche Glück zu maximieren. Die Superintelligenz könnte daraufhin Elektroden in unser Gehirn implantieren, welche permanent unser Lustzentrum stimulieren würden.

Er stellt in diesem Zusammenhang mehrere interessante Konzepte vor. Sie alle bauen darauf auf, dass wir der KI nicht selbst direkt unsere fehlerhaften Ziele mitgeben, sondern stattdessen der KI das Ziel geben unsere implizit definierten Werte zu lernen. Im Grunde genommen wollen wir der KI das Ziel „Do what I mean“ anstatt „Do what I say“ mitgeben. Eine Beispiel für dieses Konzept stellt er mit „Coherent extrapolated volition“, kurz CEV vor. Die Idee dahinter ist, dass die KI unsere Wünsche, wenn wir mehr wüssten, schneller denken und bessere Entwicklungsmöglichkeiten hätten, lernt. Dabei soll die KI nur dann handeln, wenn es eine breite Mehrheit unter allen menschlichen extrapolierten Wünschen gibt. Dieses System der Zielsetzung soll im Idealfall sicherstellen, dass die Menschheit trotz der Existenz einer super-intelligenten KI, welche unter anderen Zielsetzungen das Potential hätte, die Weltherrschaft an sich zu reißen, immer noch die Kontrolle über ihr eigenes Schicksal behält. Eine Möglichkeit in diesem Szenario wäre dann, dass die Menschheit über ihren extrapolierten Willen, dass die KI auf unserer Welt keine Rolle spielen sollte, indirekt dafür sorgt, dass sich die KI friedlich selbst herunterfährt.

1.3.2 Erkenntnisse

Ich fand das Thema künstliche Intelligenz bereits bevor ich das Buch gelesen hatte, spannend. Der strukturierte, aufbauende Stil des Buches hat mir einen tieferen Einblick in die Möglichkeiten, Gefahren und Wege zu dieser Technologie geliefert. Das Buch zeigt auf, warum Sicherheit die erste und wichtigste Frage darstellen sollte, wenn es um die Entwicklung super-intelligenter KI geht und warum wir die Dinge, wie in der Vergangenheit so häufig, nicht einfach auf uns zukommen lassen können, selbst wenn wir diese Entwicklung zum momentanen Zeitpunkt für sehr unwahrscheinlich halten.

Für mich wirft das Buch interessante Fragen, vor allem im Bereich der Zielsetzung von KIs auf, unter anderem:

- Müssen wir mit der Entwicklung super-intelligenter Maschinen rechnen oder bleibt diese Technologie auch noch langfristig außerhalb unserer Reichweite?
- Was können wir heute tun, um die Sicherheit einer potentiellen Intelligenzexplosion in der Zukunft zu erhöhen?
- Welche Ziele wollen wir einer potentiellen super-intelligenten KI mitgeben?
- Sollten wir versuchen, die Entwicklung von AGI aktiv zu verhindern?
- Wie können wir eine globale Zusammenarbeit zu diesem Thema fördern, um ein gefährliches Rennen um die erste KI zu vermeiden?

1.3.3 Kritik

Bostrom geht das Thema Superintelligenz aus einer relativ technischen Perspektive an. Oftmals wird mit physikalischen Limitationen argumentiert, er referenziert zahlreiche Quellen und scheut sich nicht auf fachliche Details einzugehen, um seine Argumente zu untermauern. Während es ihm weitestgehend gut gelingt, den Inhalt seiner Thesen verständlich zu vermitteln, erschwert die sprachliche Struktur und der verschachtelte Satzbau oftmals die Lesbarkeit. Des weiteren verwendet er fast schon exzessiv Fußnoten, welche gesammelt am Ende des Buches über 40 Seiten ausfassen und wenn man sie nicht komplett ignorieren möchte, den Lesefluss weiter stören.

Besonders gestört hat mich das bei seinen teilweise langatmig gestalteten Szenarien, welche wohl seine Thesen illustrieren sollten, sich allerdings meistens eher wie ein sehr langweiliges Science-Fiction Buch lesen. Er gibt ja selbst zu, dass die von ihm konstruierten Szenarien eine reine Spekulation darstellen und keinesfalls Prognosecharakter haben. Daher erschließt sich mir der Sinn eher weniger, dass er sich beispielsweise mit Möglichkeiten in der Organisationsstruktur von Unternehmen bestehend aus simulierten Gehirnen näher auseinandersetzt.

Von der stilistischen Kritik abgesehen kann ich dem Buch dennoch eine klare und konzise Aufarbeitung dieser komplexen Themenstellung attestieren. Er beleuchtet das Thema aus zahlreichen Blickwinkeln und bringt interdisziplinäre Aspekte ein, welche Biologie, Physik, Psychologie, Mathematik und Informatik verbinden. Gleichzeitig gelingt es ihm seine Argumente und Szenarien verständlich zu machen, ohne tiefes Wissen in den jeweiligen Spezialdisziplinen voraus zusetzen, unter anderem indem er das inhaltlich dicht gepackte Buch an entscheidenden Stellen mit kurzen Zusammenfassungen auflockert, welche dazu beisteuern, nicht den Blick auf das Wesentliche zu verlieren.

1.4 Julien Offray de La Mettrie, Der Mensch eine Maschine

1.4.1 Synopsis

La Mettrie hat bereits im 18 Jh. viele Thesen in „L’homme machine“ zu Deutsch „Der Mensch eine Maschine“ aufgestellt, die in der damaligen Zeit für viel Unmut sorgten. Der Text hat keine Unterteilung in Kapitel und ist sehr provokativ geschrieben. Argumentiert wird mit Beispielen aus zum Teil eigens gemachten Erfahrungen und Beobachtungen des Autors.

Gleich zu Anfang geht La Mettrie auf den Menschen als Maschine ein. Trotz des damals herrschenden Ansatzes, dass es sowohl Körper als auch Seele/Geist gibt (Dercates), vertritt La Mettrie die Ansicht, dass dies nicht stimmt. Schließlich hat die Seele ihren Sitz in den Organen und kann auf vier zusammengesetzte Körpersäfte zurückgeführt werden. Diese Mischung der Säfte ist individuell für jeden Menschen und bestimmt, ob dieser ein sanguinisches, cholerisches, melancholisches oder phlegmatisches Temperament besitzt. Außerdem behauptet er, dass Liebe oder Neid, die zu Schlafmangel führen, durch körperliche Vorgänge, wie beispielsweise gesteigerter Blutdruck, und dieser Wechselwirkung zwischen Körper und Seele entstehen.

„Wenn der Blutumlauf mit zu grosser Schnelligkeit von Statten geht, kann die Seele nicht schlafen. Wenn die Seele zu aufgeregt ist, kann das Blut sich nicht beruhigen, es jagt durch die Adern mit hörbarem Geräusch.“

Er zieht viele Parallelen zwischen Menschen und Tieren, fast so als wenn sie aus demselben Stoff gemacht wären und kaum einen Unterschied nachweisen, bis auf die Sprache und die Größe des Gehirns. Immerhin können Tiere auch unterscheiden zwischen „Gut“ und „Böse“. Als Beispiel wird hier ein Hund genommen, der seinen Herren beißt und anschließend Reue zeigt.

Vieles wirkt wie eine Satire, wie beispielsweise sein Relativieren Gottes, jedoch wenn man genauer nachdenkt, versucht er die Leute des 18. Jh. dazu zu bewegen, weniger Annahmen zum Ursprung des Lebens, der Seele oder des Jenseits zu machen und mehr das Leben zu genießen. Schuldgefühle und Reue sind seiner Meinung nach nutzlos und sollten vom Streben nach Glück ersetzt werden.

1.4.2 Erkenntnisse

La Mettrie hat in meinen Augen einen sehr mutigen Schritt gewagt, durch die Veröffentlichung dieses Werkes. Er wurde von vielen geachtet und kritisiert aufgrund seiner Sichtweisen, hat diese aber trotzdem vertreten und kein Blatt vor den Mund genommen.

„Es genügt nicht, daß ein Weiser die Natur und die Wahrheit erforscht, er muß auch den Mut haben, sie zu sagen.“

1.4.3 Kritik

Das Buch war anstrengend zu lesen, zugegeben, es war eine Übersetzung - vielleicht wäre das beim Original nicht der Fall. Mir würde eine Unterteilung in Kapitel helfen, die Themen als in sich geschlossen zu sehen, hier hatte ich teilweise das Gefühl, dass der Autor „wild drauf los geschrieben hat“ und nicht wirklich auf die Struktur geachtet hat. Dies könnte aber auch an der Zeit liegen, zu der es geschrieben wurde.

Ansonsten finde ich, dass La Mettrie sehr revolutionäre Ansichten vertreten hatte, die das Bild des Menschen geprägt haben.

1.5 Richard David Precht, Künstliche Intelligenz und der Sinn des Lebens

1.5.1 Synopsis

Dem Titel des Buches entsprechend beschäftigt sich Richard David Precht in diesem Werk mit Künstlicher Intelligenz (KI), wie aber schon im Vorwort verdeutlicht wird, geht es hier keinesfalls um eine fundiert technische Darstellung des Themas, viel mehr beschäftigt er sich mit der Frage wie Künstliche Intelligenz unser Selbstbild, unser Gesellschaftsbild oder auch unseren „Sinn des Lebens“ wandeln könnte. Die Struktur des Buches wird zwar in Kapitel gegliedert, welche das aktuelle Problem beschreiben aber die Argumentation innerhalb des Kapitels hat keine Struktur und besteht vor allem aus Gegenargumenten zu Thesen anderer Philosophen.

Zu allererst hält er dem Leser vor Augen, dass in unserer Welt gerade zwei Vorgänge geschehen, die zunächst völlig unabhängig erscheinen aber bei genauerer Betrachtung total abhängig voneinander sind. Während auf der Welt vor allem im Silicon Valley (eines der großen Feindbilder Prechts in diesem Buch) von der Erschaffung einer Superintelligenz geträumt wird, wird die ökologische Krise in der sich unser Planet zurzeit befindet kaum in diesem Kontext erwähnt, gar ignoriert. An dieser Stelle vergleicht Precht, die Menschheit mit einem Verrückten, der weiß, dass sein Keller brennt und statt innezuhalten und dieses Feuer zu löschen, seinen Dachstuhl ausbaut um dem Himmel näher zu kommen. Er zeichnet eine schwarze Zukunft in dem der Mensch den Planeten völlig ausbeutet und damit unbewohnbar macht. Precht stellt hierzu die These auf, dass wir KI gar nicht wirklich brauchen, denn anders wie von großen technologischen Konzernen behauptet, dient die KI-Forschung nicht dem Wohle der Menschheit, sondern ist ein Produkt des Kapitalismus, welches möglichst viel Gewinn in die Taschen der Großkonzerne spülen soll. Außerdem kann das Ziel des technischen Fortschrittes allein dem Menschen, keinen Lebenssinn geben, da der Sinn des Lebens etwas Subjektives und Individuelles ist, das jeder Mensch für sich selber finden muss.

Er zeigt außerdem die großen Unterschiede des maschinellen und menschlichen Denkens auf und warum man deshalb einer Maschine niemals eine Moral geben kann, denn diese entsteht aus Gefühlen, Empfindungen sowie Intuition, Eigenschaften, die man keiner Maschine „einpflanzen“ kann. Treffende Worte findet auch Donald Knuth: „Die KI kann mittlerweile so ziemlich alles, was 'Denken' erfordert, aber kaum etwas, was Tier und Menschen 'gedankenlos' tun“.

Weiters zeigt er die 2 Hauptströmungen auf, die im Kontext Superintelligenz zu nennen sind, zum einen den Transhumanismus (Verschmelzung von Menschen und Maschine also ein optimierter Mensch 2.0) zum anderen den Posthumanismus (Ein Intelligenz wird erschaffen, die den Menschen irgendwann übertrumpft und die Weltherrschaft an sich reißt)

Am Schluss findet er doch noch einige positive Aspekte am Thema KI wie etwa, dass die KI den Menschen irgendwann vom Arbeitszwang befreien könnte.

1.5.2 Erkenntnisse

Ich habe mich im Zuge meines Studiums schon mit Künstlicher Intelligenz beschäftigt und war sehr gespannt diese aus der Sicht eines Philosophen zu betrachten, wie zu erwarten

gab es einige ethische und moralische Denkanstöße. Das Buch hat mich aber nicht davon überzeugen können, meine Meinung gegenüber der KI-Forschung infrage zu stellen, die durchaus positiv ist. In meinen Augen kann KI von großem Wert für die Menschheit sein, wenn sie unter den richtigen Rahmenbedingungen geschaffen wird. Dass die Rahmenbedingungen unter dem herrschenden Kapitalismus wohl nie ganz richtig sein können, da nicht der Mensch, sondern das Geld im Fokus stehen, war wohl die wichtigste Erkenntnis des Buches.

1.5.3 Kritik

Ich fand das Buch extrem mühsam zu lesen, da Precht in weiten Teilen vor allem auf Aussagen anderer Autoren reagiert (Z.B. Bostrom), diese widerlegt oder als Unterstützung seiner Meinung einbindet, ohne diese dabei in sinnvollem Ausmaß zu behandeln. Precht greift zwar auch auf Beispiele aus dem wahren Leben auf um seine Thesen zu untermauern, aber jede seiner Formulierungen ist sprachlich enorm überladen. Die Kernaussagen der einzelnen Kapitel sind schwierig aus den vielen Argumenten herauszulesen, ein wirkliches Fazit zu den Überlegungen ist meist auch nicht vorhanden. Precht argumentiert sehr provokant, teilweise muss man sich aus der Sicht eines/r Technikers/Technikerin fast etwas ärgern, er stellt zum Beispiel die Behauptung auf, dass Menschen, die feinfühlig sind, selten Informatiker werden. Precht bringt aber auch durchaus sinnvolle Argumente und appelliert an die Menschen, welche sich am technischen Fortschritt beteiligen, sich kritisch mit den Zielen ihrer Arbeit auseinander zu setzen und Folgenabschätzung zu betreiben.

2 Gegenüberstellung

Die Grundbausteine der uns heute bekannten KI scheint durch La Mettrie's „Der Mensch eine Maschine“ und seiner Ansicht, dass die Seele im Körper vorhanden ist, gelegt worden zu sein. Auf diesem Prinzip basieren die bei Bostrom und Precht auseinandergehenden Meinungen, der Quantifizierbarkeit der Moral.

Während Bostrom der Meinung ist, dass wir mithilfe von Superintelligenz unsere momentanen moralischen Probleme und Fragestellungen algorithmisch besser lösen könnten, als wir fehlerhafte Menschen jemals dazu in der Lage wären, meint Precht, dass einer künstlichen Intelligenz fundamental keine Vorstellung von Ethik und Moral eingepflanzt werden kann.

Die Gefährlichkeit einer Superintelligenz ist für Precht kaum bis gar nicht vorhanden, da diese weder Bedürfnisse noch Machtstreben hat. Seiner Meinung nach kann die KI keinen eigenen Willen entwickeln, wobei Bostrom einige Wege aufdeckt, die das Gegenteil zeigen.

Bostrom selbst relativiert den Großteil seiner Thesen bereits mit seinem Vorwort, dass ein Großteil seines Buches in der Zukunft sich höchstwahrscheinlich als falsch oder unvollständig herausstellen wird. Dagegen stellt Precht oftmals sehr selbstbewusst seine Thesen in den Raum, auch ohne die Hintergründe weiter zu beleuchten.

3 Fragen und Diskussion in der VU

Folgende Fragen wurden für die VU vorbereitet und diskutiert:

- Ist KI eine existentielle Bedrohung für den Menschen?
- Lässt sich Moral algorithmisch berechnen bzw. sollte eine Maschine so etwas können?
- Sollte die Menschheit eine Superintelligenz überhaupt anstreben oder ist das etwas wogegen wir uns wehren sollten?

4 Literaturverzeichnis

- [1] Nick Bostrom. *Superintelligence*. Oxford University Press, first edition, 2014.
- [2] Julien Offray de la Mettrie. *Der Mensch eine Maschine*. JMV, 1747.
- [3] Richard David Precht. *Künstliche Intelligenz und der Sinn des Lebens*. Goldmann, fifth edition, 2020.