Projektbericht «Gefahren der AI-Entwicklung»

Inhalt

Einleitung

Die Künstliche Intelligenz (Im Projekt und Bericht wird im Folgenden die Abkürzung «AI» verwendet) gewinnt zunehmend an Wichtigkeit in unserer Gesellschaft. Obschon die Forschung bezüglich AI schon in den 50er-Jahren begonnen hatte, ist das Thema erst in den letzten Jahren hochaktuell geworden. Gründe dafür sind technologischer Fortschritt, aber auch Notwendigkeit in Industrie, Kommerz, Militär und Gesellschaft. Das positive Ertragspotential solcher Technologie ist immens - und so sind auch die Gefahren. Es lässt sich nur schwer prognostizieren, wie sich die Menschheit mit der Verwendung von AI entwickeln wird. Klar ist aber, dass sich AI schon in der heutigen Welt nicht mehr wegdenken lässt. Zukünftige Generationen der AI-Technologien werden wirtschaftlich noch effizienter und fähiger werden.

Viele Personen aus der Politik, Industrie, Wissenschaft und den Medien[1] haben sich zur Verwendung von AI geäussert. Neben Proponenten der AI gibt es auch sehr kritische Ansichten, die eine oft dystopische Zukunft voraussagen. Im Jahr 2015 verfassten Wissenschaftler des Institutes «Future of Life»[2] einen «Open-Letter»; den Bericht «Research Priorities for Robust and Beneficial AI», welcher auf die verschiedenen Gefahren und Chancen in der AI-Entwicklung hinweist. Dieses 10-seitige Dokument benutzten wir als Einstiegshilfe in das Projekt und mithilfe von Artikeln und Papiern aus der ausführlichen Quellenangabe, auf die sich dieser «Open Letter» stützt, konnten wir verschiedene Argumente rekonstruieren. Wir beschränkten uns dabei auf die Gefahren der AI-Entwicklung.

Folgende Ziele standen im Zentrum unserer Arbeit:

- Wir wollten sehen, wie weit die AI schon heute verwendet wird und wie weit sie in Gebieten wie Wirtschaft, Militär und Gesellschaft eine Rolle spielt. Durch die rapide Erweiterung der Einsatzgebiete der AI nimmt sie immer stärkeren Einfluss und hat Auswirkungen auf unseren Alltag. In welchen Bereichen stellt die wachsende Zunahme der Verwendung von AI eine Gefahr dar und hat Auswirkung auf die Zukunftsgestaltung?
- Durch die Entwicklung einer Superintelligenz[3] besteht die Gefahr einer Singularität[4]. Stellt diese Entwicklung ein existenzielles Risiko für die Menschheit dar?

[1] Ein paar prominente Namen sind: Mark Zuckerberg, Elon Musk, Stephen Hawking, Steve Wozniak, Bill Gates und viele weitere.
[2] Future of Life-Institute: Online unter: https://futureoflife.org/ (Zugriff: 10.08.2020).
[3] Eine «Superintelligenz» ist ein hypothetischer Agent, welcher eine Intelligenz besitzt, die die menschliche Intelligenz bei weitem überschreitet.
[4] Ein Zeitpunkt, ab dem Maschinen sich selbst verbessern können, den technischen Fortschritt massiv beschleunigen und so unkontrollierbare und irreversible Folgen mit sich bringen.

- Sind die Argumente der Kritiker von AI miteinander verknüpft und laufen sie auf gemeinsame Punkte zu?
- Des Weiteren wollten wir lernen, wie eine Argumentationsanalyse das Verständnis von komplexen Zusammenhängen verbessern kann.
- Der Umgang mit der Software «Argdown» und allgemein der Einsatz von Software in der Philosophie war für uns von grossem Interesse.

Themenfindung und Textgrundlage

Als spannendes und aktuelles Thema bietet sich das Gebiet der AI, deren Anwendungspotenzial sehr weitläufig ist und die dazugehörenden Herausforderungen fundamentale Veränderungen in der Gesellschaft hervorbringen kann, für eine Argumentationsrekonstruktion bezüglich der Gefahren der AI-Entwicklung an. Da die Thematik sehr breit ist und auch stark in die Tiefe geht, war eine Einschränkung der Fragestellungen im Rahmen dieses Projektes sehr wichtig. Das Projekt sollte eine gute Übersicht über potenzielle Gefahren aus verschiedenen Thematiken wie Wirtschaft, Militär, Gesellschaft und Ethik liefern. Dabei wurden grundlegende Fragen aus der Philosophie des Geistes bezüglich *Bewusstsein*, *Identität* und dem *Turing-Test* ausgeklammert.

Unser erster Schritt war es, eine prominente Person zu finden, welche aktiv an der Diskussion um die Gefahren der AI-Entwicklung teilnimmt. Eine solche Person, die zusätzlich eine kritische Position zu AI vertritt und uns auch auf mögliche Gefahren hinweist, ist Elon Musk. Durch seine Geschäftstätigkeiten in der autonomen Verkehrsführung mit *Tesla*, der Gründung von *OpenAI* (Entwicklung einer Open-Source-AI) und der Firma *Neuralink* (Entwicklung einer Gehirn-Maschine-Schnittstelle), sowie seiner Präsenz in den Medien, dachten wir, er biete sich hervorragend für eine Argumentationsanalyse an. Doch wurde nach eingehender Recherche festgestellt, dass Elon Musk nur sehr wenige Argumente vorträgt. Er beschränkt sich im Rahmen der Interviews und Symposien, welche wir untersucht haben, auf oberflächliche Aussagen und für Medien attraktive «Sound-bites». Ausserdem gibt es kaum schriftliche Quellen, die Musk selbst verfasst hat. Deshalb wurde bald klar, dass wir eine viel fundiertere Textgrundlage finden mussten.

Durch weitere Recherche stiessen wir auf den «Open Letter» des «Future of Life»-Institutes, welcher von vielen Grössen der Branche, auch Elon Musk, unterzeichnet wurde. Der 10-seitige Bericht mit dem Titel «Research Priorities for Robust and Beneficial AI» bietet eine hervorragende Übersicht zur Thematik und enthält sehr viele Quellen zu wissenschaftlichen Artikeln, Büchern und Papieren und fand grosse Unterstützung unter Wissenschaftlern, Politikern, den Medien und

Technologieschaffenden. Wir haben diesen Bericht genau durchgearbeitet und die Quellen gesammelt, auf welche sich der Bericht stützt. Diese Quellen wurden auf Thesen untersucht und auf Argumente geprüft, welche sich für eine Rekonstruktion eignen. Der «Open Letter» und die Quellen ergaben unsere Textgrundlage. Nach genauem Durchlesen und Strukturieren des «Open Letters» hat sich aber gezeigt, dass der Umfang dieses Berichtes den Rahmen des Projektes sprengt. Deshalb musste eine Auswahl getroffen werden, die einen guten Überblick zur Thematik lieferte, eine spannende Argumentationsanalyse ermöglichte und im Zeitplan durchführbar war.

Quellenbeschreibung

Die folgenden Quellen wurden im Rahmen dieser Arbeit untersucht und verwendet, um die Argumentationskarte zu erstellen.

«Research Priorities for Robust and Beneficial AI» des Institutes «Future of Life»[5]. Dieser «Open Letter» diente als Textgrundlage und Basis für unser Projekt. Aus den verschiedenen Themengebieten selektierten wir verschiedene relevante Aussagen, welche wir anhand der im «Open Letter» referenzierten Quellen weiter untersuchten und die daraus analysierten Argumente rekonstruierten.

«Superintelligence»[6] von Nick Bostrom. In diesem Buch wird in einem ersten Teil die Entwicklung der AI von den Anfängen in den 50er-Jahren bis heute (Erscheinungsdatum 2014) verfolgt und in einem zweiten Teil eine Prognose erstellt, welche Auswirkungen die Entwicklung zukünftiger AI haben wird. Besonders die Kapitel 5 bis 9 wurden zum Zweck des Projektes untersucht. In diesen Kapiteln geht es darum, welche Gefahren die zukünftige Entwicklung der AI mit sich bringt und wie eine Singularität unsere Gesellschaft grundlegend verändern kann.

«The Second Machine Age»[7] von Erik Brynjolfsson und Andrew McAfee. Ein grosser Teil der Argumentation im Bereich *AI's Economic Impact* stammt aus diesem Buch. Das Buch beschreibt die Veränderungen, die Wirtschaft und Gesellschaft gerade durchleben und zieht auch immer wieder Parallelen zu anderen Wirtschaftsepochen – wie der industriellen Revolution. Brynjolfsson und McAfee beschäftigen sich beide seit vielen Jahren mit dem Einfluss moderner Technik auf die Wirtschaft und sind ausgewiesene Experten auf dem Gebiet. Beide sind am MIT tätig.

[5] Future of Life-Institute, Open Letter, «Research Priorities for Robust and Beneficial AI», 2015.
[6] Bostrom, Nick, «Superintelligence: Paths, Dangers, Strategies», 2014.
[7] Brynjolfsson, Erik and McAfee, Andrew, «The Second Machine Age», 2014.

«The Case for a Federal Robotics Commission»[8] von Ryan Calo. Dieser Artikel befasst sich mit der Legislatur und dem Gebrauch von AI und Robotern in der Wirtschaft. Aus diesem Artikel wurde die Argumentation zum Gebrauch von AI an der Börse analysiert und rekonstruiert.

«Losing humanity: the case against killer robots»[9] von Bonnie Docherty. Das Buch erschien unter dem Label *Human Rights Watch,* welches eine NGO ist, die sich für die Wahrung der Menschenrechte einsetzt. Das Buch dreht sich daher auch um die Kritik am Einsatz von Maschinen in bewaffneten Konflikten. Die Argumentation in *Autonomous Weapons* stammt aus diesem Werk.

«Secular stagnation? Not in your life»[10] von Joel Mokyr. Neben Brynjolfsson und McAfee ist Joel Mokyr der wichtigste Autor für die Argumentation im ökonomischen Teil unserer Argumentationsanalyse. Das Buch, welches grundsätzlich eher eine positive Sicht auf AI in der Wirtschaft vertritt, zeigt doch einige Schwierigkeiten auf.

«Moral Machines: Teaching Robots Right from Wrong»[11] von Wendel Wallach und Colin Allen. Dieses Buch diente als Quelle für den grundlegenden Teil unserer Analyse. Es geht um die theoretischen Grundlagen, beispielsweise ob eine Maschine überhaupt (ethisch relevant) handeln kann oder nicht. Wenn die Grundlagen geklärt sind, beschäftigen sich Wallach und Allen mit den Auswirkungen des Geklärten.

Organisation und technische Hilfsmittel

Um die interne Kommunikation zu vereinfachen, das Material übersichtlich geordnet zu halten und unsere Zusammenarbeit und Versionierung effizient zu gestalten, richteten wir auf *Github*[12] ein Repository ein. So blieben wir immer auf dem neusten Stand und konnten unseren Arbeitsverlauf kontrollieren. Auf *Zotero*[13] haben wir eine Bibliothek angelegt, um die Quellenverweise dynamisch verwalten zu können.

Argdown

Die Arbeit mit Argdown gestaltete sich dank der ausführlichen Dokumentation recht gut. Die sehr grosse Flexibilität nutzten wir anfangs aus, um schlicht Karten zu zeichnen und Relationen manuell zu

[8] Calo, Ryan, «The Case for a Federal Robotics Commission», 2014.
[9] Docherty, Bonnie, «Losing humanity: the case against killer robots», 2012.
[10] Mokyr, Joel, «Secular stagnation? Not in your life», 2014.
[11] Wallach, Wendell & Allen, Colin, «Moral Machines: Teaching Robots Right from Wrong», 2008.
[12] Github repository, online unter: https://github.com/flicksolutions/musk (Zugriff: 10.08.2020).
[13] Zotero, online unter: https://www.zotero.org/groups/2463181/musk_argumentationsanalyse/collections/2VWWS9ZF (Zugriff: 10.08.2020).

erstellen und dann später saubere deduktive Argumente aufzustellen und Argdown möglichst selbst die Relationen ziehen zu lassen. Dies funktionierte erstaunlich gut und einfach. Einzig die Darstellung der Karte war zeitraubend. Da es beispielsweise nicht möglich ist, den Rang von Argumenten oder Thesen selbst zu setzen, ist man auf die Graphviz-Engine angewiesen. Nachdem von uns ein Fehler identifiziert wurde, haben wir diesen direkt auf Github dem Entwickler gemeldet und boten an, bei der Lösung des Problems zu helfen. Beim Fehler handelt es sich um eine Falsch-Setzung des Rangs eines Punktes auf der Karte, wenn die Beziehung dieses Punktes an der falschen Stelle im Code geschieht. Details zum Fehler finden sich im *Issue* auf Github. (vgl. Flick & Voigt 2020)

Bis relativ weit im Projektverlauf (Bis nach v1.0) benutzten wir für unsere Arbeit in Argdown eine Struktur von 3 Dateien. Dabei gab es eine Datei, an der Claude arbeitete, eine weitere, an welcher Sebastian arbeitete und eine dritte Datei, die die beiden Dateien zusammenfügte und einige Thesen mit der Hauptthese verband. In einem .config.json-File wurden Metadaten und verschiedene andere Optionen festgelegt. Karten wurden mittels der Kommandozeile produziert. Diese Struktur wurde später aufgegeben, da wir es sinnvoll fanden, alle Argumente in einer Datei zu haben und auch die Arbeit mit dem Visual-Studio-Code-IDE erleichtern wollten.

Git

Wir bereuen die Entscheidung nicht, ein Git-Repository für das Projekt erstellt zu haben. Dies ermöglichte uns auf jede frühere Version der Karte zurückzuschauen, es erlaubte uns, unsere eigenen Wege zu gehen und in einem anderen Branch etwas auszuprobieren, um vielleicht später Teile daraus zu übernehmen und schlussendlich liefert es auch ein objektives Bild für Aussenstehende über unsere Arbeit und es kann transparent nachvollzogen werden, wie wir gearbeitet haben. Wir freuen uns auch in Zukunft wieder Projekte und Arbeiten in Verbindung mit Git zu schreiben.

Markdown

Der etwas ambitionierte Plan, in diesem Projekt direkt jeden Text in Markdown zu schreiben, mussten wir leider aufgeben. Zu gross war die Gewohnheit eine Word-Datei zu erstellen. Dennoch konnten wir zumindest das Journal bis zu Version 1, unsere Notizen und den Projektbeschrieb in Markdown verfassen und haben auch gute Erfahrungen damit gemacht. Wir sind der Überzeugung, dass Markdown schon bald verbreitet auch in der akademischen Welt überall dort eingesetzt wird, wo Latex als Textsatz-Engine nicht unbedingt nötig ist.

Arbeitsteilung und Herangehensweise an die Textgrundlage

Im Team wurden die zu untersuchenden Artikel grob nach kurzfristigen und längerfristigen Gefahren aufgeteilt. Während Claude zuerst das Buch «Superintelligence» des Philosophen und Futurologen Nick Bostrom untersuchte und sich somit in die längerfristigen Gefahren der AI-Entwicklung einarbeitete, analysierte Sebastian das Buch «The Second Machine Age» von Erik Brynjolfsson und Andrew McAfee, welches den Fokus auf die wirtschaftlichen Auswirkungen der AI legt. Zudem haben wir verschiedene Themengebiete aus dem «Open Letter» gewählt, welche uns relevant erschienen. Es war klar, dass wir nicht auf alle Themen eingehen können, da dies den Rahmen der Arbeit gesprengt hätte. So entschieden wir uns, selektiv weitere Artikel zu den Themen Militär, Wirtschaft, und Computer Science zu analysieren.

Erste Erkenntnisse

Wir haben durch die Auswahl der Texte, welche wir bearbeitet und aus welchen wir Argumente rekonstruiert haben, einen guten Überblick der Gefahren der AI-Entwicklung gewonnen. Die Kernaussagen der einzelnen Texte wurden analysiert und als Argumente rekonstruiert.

Die Rekonstruktion hat aber einige Schwierigkeiten aufgezeigt. So ist es oft nicht einfach ein komplettes Argument aus einer spezifischen Textstelle herauszukristallisieren. Die Argumente aus dem Buch «Superintelligence» strecken sich über mehrere Kapitel hinweg. Zudem ist es nicht einfach zu erarbeiten welche Argumente in andere greifen. Weil wir möglichst exakt arbeiten und so nah wie möglich am Wortlaut bleiben wollten, ergaben sich einige Argumentreihen, die für sich stehen. Zudem haben wir beim Verknüpfen der Argumente auf eine starke Abhängigkeit geachtet. So unterstützt das Argument «Strategischer Vorteil» einer Superintelligenz nur «Zukunft verändern» und «Globale Zerstörung». Auch wenn strategische Vorteile in vielen weiteren Gebieten wichtig sind, haben wir davon abgesehen, den strategischen Vorteil, um welchen es im Argument geht, auch in weitere Argumente einfliessen zu lassen.

Unser Entscheid, die Arbeit auf Deutsch zu verfassen, führte dazu, dass einige Ausdrücke bei der Übersetzung die Gefahr mit sich brachten, nicht genau dasselbe auszudrücken. Durch die teilweise komplexe Formulierung mit vielen spezifischen Fachausdrücken in den Texten hat sich die Übersetzung als schwierig herausgestellt. Gewisse (Fach-) Ausdrücke lassen sich entweder gar nicht oder nur schwer übersetzen, ohne den Inhalt geringfügig zu verändern. Ein solches Wort war das englische «Emergence», welches je nach Themengebiet (Botanik, Technologie, Philosophie des Geistes, etc.) verschieden aufgefasst werden kann. Wir haben uns auf den Ausdruck «Entstehen»

geeinigt. Dies umfasst zwar nicht die ganze Definition des Wortes, aber ist im Kontext des Argumentes verständlicher.

Die Hauptherausforderung dieser ersten Phase war die logische Form der Argumente. Es war zum Teil nicht einfach, ein deduktiv gültiges Argument aus den Textstellen herauszuziehen. Die Schwierigkeit ergab sich dann, aus den Texten genügend implizit angenommene Prämissen zu erkennen und in einem zweiten Schritt diese zu ergänzen, ohne die Aussage des Autors zu verändern und zu stark eingreifen zu wollen. Durch die Übersetzung von Englisch nach Deutsch verstärkte sich diese Gefahr, aber gab uns auch die Möglichkeit, mit den einzelnen Prämissen besser umgehen zu können. Die Übersetzung vereinfachte die weitere Umformulierung, so dass die Argumente auch deduktiv gültig sind.

Ein weiterer Punkt, welcher uns Schwierigkeiten bereitete, war die Stärke der Argumente und der Umgang mit Konjunktiv-Formulierungen und Hypothesen. Da sich die AI rapide entwickelt und besonders die in der Zukunft liegenden Gefahren nicht genau absehbar sind, wurden in den Texten vielfach «kann eine Gefahr sein» geschrieben. Dies konstant als «ist eine Gefahr» auszulegen, wollten wir nicht. Dadurch sind verschiedene Argumente nicht so stark, wie sie sein könnten.

Die resultierende Karte zeigte auch ein weiteres Problem auf: Wir hatten nun eine sehr breite Karte mit verschiedenen Gruppen, welche alle auf eine einzelne Hauptthese (AI ist gefährlich) zuliefen. Wir konnten zwar in den Argumenten eine gute Hierarchie erstellen, aber die Karte selbst war noch sehr flach.

Restrukturierung und Fertigstellung der Karte

Das Feedback der Präsentation am 19. Mai 2020 spiegelte viele der Punkte wider, welche wir auch erkannt hatten. Die Menge der Argumente wurde als gut angesehen, doch die Struktur der Karte als mangelhaft. Prof. Dr. Gregor Betz schlug uns vor, die Hauptthese weiter auszudifferenzieren. Diesen und weitere Vorschläge setzten wir in der folgenden letzten Phase des Projekts um. Es gab verschiedene Ideen, wie der Vorschlag umgesetzt werden sollte, wobei wir einen guten Kompromiss gefunden haben und die Hauptthese stehen liessen, jedoch wichtige Thesen zweiter Ordnung einführten. Wir achteten nun auch darauf, die temporale Komponente, die Teil von vielen unserer Argumente ist, auch in die neuen Thesen einfliessen zu lassen und zwischen zukünftigen und aktuellen Gefahren zu unterscheiden. So ergab sich eine übersichtliche und nun auch viel aussagekräftigere Karte. Der wichtigste und aufwändigste Schritt war dann, unsere Argumente auf die nun ausdifferenzierten Hauptthesen zu beziehen. Neue Verbindungen ergaben sich und wir erkannten,

dass die Gebiete doch sehr verknüpft sind. Eine weitere scheinbar kleine Korrektur war die Veränderung der Art von Beziehungen, die wir bei sich widersprechenden Thesen setzten. Neu wurde >< gesetzt, welches einen kontradiktorischen Widerspruch bezeichnet, anstatt nur -, welches einen konträren Widerspruch anzeigt. In der Karte mag es nur eine Pfeilspitze in einen Diamanten verwandelt haben, wir sind aber der Meinung, dass die Beziehungen in Argdown möglichst korrekt eingegeben werden sollten, auch wenn dies visuell keinen grossen Effekt hat.

Ergebnisse

Gerne möchten wir den ungewöhnlichsten Teil unserer Karte kurz beschreiben. Die These *[Machtkonzentration Aktuell]* ist eine der Hauptthesen in unserer Karte und entspringt aus Brynjolfsson & McAfee 2014 und Mokyr 2014. Die These besagt, dass durch die Entwicklung von AI ein labiles Wirtschaftssystem entsteht, in dem einzelne Akteure extreme Macht besitzen. Brynjolfsson und McAfee erläutern in ihrem Buch aber bereits Einwände der Proponenten der AI, wonach diese Machtkonzentration keine schlechten Folgen für die Gesellschaft hätte. Es handelt sich um das *<strong bounty>-Argument*. Es versucht zu zeigen, dass das entstehende Ungleichgewicht nicht relevant ist, da auch die unteren Schichten der Gesellschaft einen grossen Nutzen aus dem Ungleichgewicht ziehen. Es gehe also allen besser, wenn die Macht konzentriert wäre, dem Teil, der die Macht an sich reisst einfach in einem grösseren Ausmass. Die wichtigste Prämisse des Arguments – *[Arme profitieren]* – wird von zwei separaten Gegenargumenten – *<Potenzgesetz ist schlecht für die Armen>* und *<Potenzgesetz ist schlecht für die Armen2>* – von Brynjolfsson und McAfee angegriffen, wobei das Erste auch direkt von den Proponenten widerlegt wird. Schlussendlich bleibt *<Potenzgesetz ist schlecht für die Armen2>* unwiderlegt stehen und wehrt somit den Angriff auf *[Machtkonzentration Aktuell]* ab. In der Karte zeigt sich die Debatte sehr schön und es wird klar, an welchen Prämissen die gesamte Argumentkette aufgebaut ist. Gegenargumente wie die besprochenen gibt es mit grosser Wahrscheinlichkeit zu jeder unserer Hauptthesen von verschiedenen anderen Akteuren. Wir sahen es als Chance, dass Brynjolfsson und McAfee Gegenargumente gegen ihre eigenen Thesen vorbringen und haben diese wahrgenommen und sie in unsere Karte einfliessen lassen, um zu zeigen, wie komplex die Debatte aufgebaut ist und wie viel übersichtlicher sie mit Hilfe der Argumentationsanalyse wird.

Eine weitere wichtige Erkenntnis ist, dass *<Vorteile verhindern Moratorium>* ein absolutes Schlüsselargument ist und die Konklusion in mehreren Bereichen, welche aussagt, dass eine AI, welche moralische Entscheidungen trifft, auch tatsächlich entwickelt wird. Wendell Wallach und Colin Allen

treffen also in ihrem Werk *Moral Machines: Teaching Robots Right from Wrong* eine Aussage, die Grundlage für die Argumentation vieler weiterer Autoren und Autorinnen ist:

> *The social issues we have raised highlight concerns that will arise in the development of AI, but it would be hard to argue that any of these concerns leads to the conclusion that humans should stop building AI systems that make decisions or display autonomy. Nor is it clear what arguments or evidence would support such a conclusion.*[14]

Fazit

Es war überraschend zu sehen, in welchen Bereichen die AI heute schon verwendet wird. Von Automatisierungsprozessen bis zu spezialisierten Agenten hat die AI schon heute kritische Einsatzgebiete in Militär und Wirtschaft. Die Verwendung von AI an der Börse wie auch in der Produktion von Gütern, als Datensammlungs- und Verarbeitungsbots in Sozialen Medien und Marketing, aber auch in der Funktion als automatisierte Maschinen im militärischen und zivilen Bereich haben grosse Auswirkungen auf unseren Alltag. Die Untersuchung der einzelnen Themengebiete zeigte auf, dass diese Prozesse zu Gefahren führen, auf die wir in vielen Bereichen nur ungenügende Antworten haben. So hinkt die Legislative der Entwicklungsgeschwindigkeit stark hinterher und es besteht auch keine Bundesämter, welche sich mit der Gesetzeslage und Sicherheit befassen. Dies muss in den nächsten Jahren stark verbessert werden, um den Anschluss an die rapide technologische Entwicklung und Verwendung nicht vollständig zu verlieren.

Das Konzept einer Superintelligenz und die mögliche Realisierung einer Singularität bleibt zu einem grossen Teil eine Hypothese. Es werden von Futurologen wie Nick Bostrom viele Annahmen getroffen, welche zwar plausibel, aber definitiv nicht unumgänglich sind. Dieses Thema ist sehr komplex und bleibt zurzeit in einem theoretischen Bereich. Doch sind die aufgeworfenen Gedankenexperimente Grund genug, den Dialog aufrechtzuerhalten und schon heute Vorbereitungen und Sicherheitsmassnahmen zu treffen. Die Diskussion um eine Superintelligenz wirft viele fundamentale philosophische Fragen auf, welche im Rahmen dieser Arbeit nur ansatzweise untersucht werden konnten. Doch sind Untersuchungen bezüglich Identität, was es bedeutet, eine Person zu sein oder was ein Leben ist, mitunter die wichtigsten Prozesse zum Umgang mit einer Superintelligenz, welche einen moralischen Status erreicht hat.

[14] Wallach, Wendell & Allen, Colin, «Moral Machines: Teaching Robots Right from Wrong», 2008. S. 52.

Es ergaben sich viele Argumente der verschiedenen Kritiker*innen, die auf gemeinsame Punkte hinauslaufen. Die Karte widerspiegelt sehr gut, wie verschiedene Argumente aus unterschiedlichen Themenbereichen in gemeinsame Thesen zusammenlaufen. Eine Hauptschwierigkeit ergab sich, diese so zu sortieren, dass die Karte im Bereich der Thesen übersichtlich blieb.

Ein Punkt, welchen wir bedauern ist, dass wir nur eine Seite der Thematik erarbeiten konnten. Sehr spannend wäre es gewesen auch die Argumentationen der AI-Befürworter*innen zu untersuchen. Dies hätte einerseits zu einer viel komplexeren Karte führen können und hätte die Einsicht in die Thematik stark erhöht. Andererseits aber hätte es den Arbeitsrahmen dieses Projektes überschritten. Jedes dieser einzelnen Themenbereiche aus dem «Open Letter» bieten sich an, sich viel intensiver damit zu beschäftigen. So betrachten wir dieses Projekt als einen Einstieg und generellen Überblick über die Gefahren der AI als eine vollständige Auseinandersetzung mit der Entwicklung und Anwendung von AI. Es bieten sich viele Bereiche in der AI-Entwicklung zu weiterer Auseinandersetzung und genaueren Analyse an und wir erachten den philosophischen Diskurs in der Thematik als weiterhin spannend und sehr wichtig.

Der theoretische Teil des Kurses und die somit erlernte Methodik und die Prozesse der Argumentationsanalyse und -rekonstruktion erachten wir als fundamental wichtig. Wir haben viele der Informationen aus dem Kurs angewendet, um zu unseren Ergebnissen zu gelangen. Da diese Arbeitsweise neu für uns war, ist das Resultat der Rekonstruktionsarbeit noch nicht dort, wo wir uns damit sicher fühlen. Insbesondere der Umgang mit Texten von anderen Autoren zeigte eine Unsicherheit, wie weit sich die Argumentationsformulierung vom Quelltext entfernen darf und soll, um aus zum Teil fragmentierten und unvollständigen Aussagen ein gültiges Argument erstellen zu können. Die Kernaussage zu finden, welche ein starkes Argument liefern kann, ist nicht immer sofort offensichtlich. Die kontinuierlichen Iterationen und Revisionen der aufgestellten Argumente, aber auch der Kontext der Debatte sind dabei extrem wichtig. Der weitere Verlauf des Studiums und die philosophische Tätigkeit allgemein wird diese Unsicherheit sicher entschärfen.

Argdown ist ein sehr hilfreiches Werkzeug, um komplexe Debatten zu analysieren und die Zusammenhänge zwischen verschiedenen Argumenten erkennen zu können. Die einfache Syntax macht die Benutzung angenehm und nicht aufdringlich, so, dass wir uns auf die Argumentationen konzentrieren konnten, ohne von technischen Aspekten gross aufgehalten zu werden. Die Layout-Funktionen und generelle Gestaltungsmöglichkeiten der Software lässt aber noch stark zu wünschen übrig. Dort zeigt sich, dass Gestaltung durch Programmieren nicht intuitiv ist. Ein Node-basierter Umgang mit den Komponenten auf der Karte würde das Erstellen erleichtern. Alinierfunktionen

würden die Lesbarkeit verbessern und visuelle Gruppier- und Anordnungsmöglichkeiten würde die Gestaltung nicht nur vereinfachen, sondern gäbe dem Autor oder der Autorin mehr Freiheit zur Darstellung des Informationsflusses. Die weitere Entwicklung der Software wird dies vermutlich weiter verbessern und wir betrachten Argdown jetzt schon als integralen Teil für jede philosophische Arbeit.

Das Projekt war für uns eine spannende Herausforderung. Unter anderem auch wegen vielen Hindernissen, die nicht direkt mit dem Projekt verbunden sind. Durch verschiedene Umstände, unter anderem die COVID-19-Pandemie, wurde unsere Arbeitsplanung immer wieder umgeworfen und die Arbeit am Projekt wurde mehrmals für längere Zeitperioden unterbrochen. Die Zusammenarbeit gestaltete sich manchmal schwierig wegen privater Verpflichtungen. Schliesslich konnten wir trotzdem sehr viel lernen. Die aktuellen Befürchtungen in der Entwicklung von AI kennen wir nun und wir können auch in Zukunft der Diskussion folgen. Wir werden auch für zukünftige Projekte die Argumentationsanalyse verwenden, um uns ein Bild zu machen über eine laufende oder abgeschlossene Debatte und wir werden Argdown dazu verwenden. Der Github-Account wird für weitere Projekte verwendet werden und wir werden weiterhin Dokumente im Markdown-Format schreiben. Somit ist also der Nutzen, den wir aus diesem Projekt ziehen, gross.

Bibliografie:

Future of Life-Institute, Open Letter, «Research Priorities for Robust and Beneficial AI», 2015.

Bostrom, Nick, «Superintelligence: Paths, Dangers, Strategies», 2014.

Brynjolfsson, Erik and McAfee, Andrew, «The Second Machine Age», 2014.

Calo, Ryan, «The Case for a Federal Robotics Commission», 2014.

Docherty, Bonnie, «Losing humanity: the case against killer robots», 2012.

Mokyr, Joel, «Secular stagnation? Not in your life», 2014.

Wallach, Wendell & Allen, Colin, «Moral Machines: Teaching Robots Right from Wrong», 2008.

Internetlinks:

Future of Life-Institute, online unter: https://futureoflife.org/ (Zugriff: 10.08.2020).

Github repository, online unter: https://github.com/flicksolutions/musk (Zugriff: 10.08.2020).

Zotero:
https://www.zotero.org/groups/2463181/musk_argumentationsanalyse/collections/2VWWS9ZF
(Zugriff: 10.08.2020).

Alle Quelltexte sind auf Github verfügbar:

https://github.com/flicksolutions/musk/tree/master/quellen (Zugriff: 10.08.2020).

Argdown-Karte:

Die Argdown-Karte ist auf Github verfügbar:

HTML: https://flicksolutions.github.io/musk/output/research-priorities.html

PDF: https://github.com/flicksolutions/musk/blob/master/output/research-priorities.pdf

Argdown: https://github.com/flicksolutions/musk/blob/master/argdown/allInOne.argdown