

Projektbescrieb «Gefahren der AI-Entwicklung»

Beschreibung:

Die Artificial Intelligence (AI) 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 die AI schon in der heutigen Welt nicht mehr wegdenken lässt. Zukünftige Generationen der AI-Technologien werden wirtschaftlich effizienter und mächtiger werden.

Viele Personen aus der Politik, Industrie und Akademie haben sich dazu 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» einen «Open-Letter»; das Papier «Research Priorities for Robust and Beneficial AI», welches auf die verschiedenen Gefahren und Chancen in der AI-Entwicklung hinweist. Dieses 10-seitige Dokument benutzen wir als Einstiegshilfe und mithilfe von Artikeln und Papern aus der ausführlichen Bibliografie, auf die sich dieser "Open Letter" stützt und als Textgrundlage dient, können wir die verschiedenen Argumente rekonstruieren. Wir untersuchen dabei die Gefahren der AI-Entwicklung.

Learning outcomes:

Wir zielen auf folgende Erkenntnisse ab:

- Wir wollen sehen, wie weit die AI schon heute verwendet wird und wie weit sie in Gebieten wie Militär, Börse und Wirtschaft eine Rolle spielen. 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 besteht die Gefahr einer Singularität (Ein Zeitpunkt, ab dem Maschinen sich selbst verbessern können, den technischen Fortschritt massiv beschleunigen und so unkontrollierbare und irreversible Folgen mit sich bringt). Stellt diese Entwicklung ein existenzielles Risiko für die Menschheit dar?

Teilnehmer:

- Sebastian Flick – 7. Semester BA Major Philosophie, Minor Linguistik
- Claude Aebersold – 4. Semester BA Major Philosophie, Minor Geschichte

Textgrundlage:

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.

Alle Texte sind auf Github verfügbar:

<https://github.com/flicksolutions/musk/tree/master/quellen>

Zotero: Auf Zotero haben wir eine Bibliothek angelegt, um die Quellenverweise dynamisch verwalten zu können:

https://www.zotero.org/groups/2463181/musk_argumentationsanalyse/collections/2VWWS9ZF

Argdown-Karte:

Die Argdown-Karte ist auf Github verfügbar:

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

Einzelne Argdown-Dateien:

<https://github.com/flicksolutions/musk/blob/master/argdown/ research-priorities.argdown>

https://github.com/flicksolutions/musk/blob/master/argdown/ argumente_claude_200510.ad

https://github.com/flicksolutions/musk/blob/master/argdown/research-priorities_hirarchy.argdown

Ein PDF der Argumentationskarte ist auf Github verfügbar:

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

Projektstatus 14.05.2020:

Wir haben durch die Auswahl der Texte, welche wir bearbeitet und Argumente daraus rekonstruiert haben, einen guten Überblick der Gefahren der AI-Entwicklung gewonnen. Die Kernaussagen der einzelnen Texte sind analysiert und als Argumente soweit rekonstruiert worden, dass die meisten Argumente an der Sitzung vom 19.05.2020 präsentiert werden können. Durch die breite Natur der Thematik lassen sich eine Vielzahl an weiteren Artikeln und Papern finden, doch mussten wir die Recherche dem Zeitaufwand im Rahmen dieses Projektes anzupassen. So erachten wir das Spektrum der Texte als nahezu komplett. Die Argumente im Einzelnen werden iterativ weiter verbessert und ergänzt. Die Erkenntnisse aus der Sitzung werden zu weiteren Anpassungen führen.

Als momentane Schwierigkeiten im Projekt haben sich drei Punkte herauskristallisiert:

- Übersetzung aus dem Englischen: Wir haben uns entschieden, alle Argumente zu übersetzen. Durch die teilweise komplexe Formulierung mit vielen spezifischen Fachausdrücken in den Texten hat sich die Übersetzung als teilweise schwierig herausgestellt. Gewisse (Fach-) Ausdrücke lassen sich entweder gar nicht oder nur schwer übersetzen, ohne den Inhalt geringfügig zu verändern.
- Konjunktiv und Stärke der Argumentation: Teilweise werden Argumente aufgestellt, welche auf Auswirkungen in der Zukunft abzielen. Dies führt dazu, dass viele Texte auf «es könnte sein», «es kann Gefahren beinhalten», «die Zukunft kann so aussehen», etc. aufgebaut sind.
- Verknüpfung der einzelnen Argumente und Darstellung als Argdown-Karte. Es gibt verschiedene Möglichkeiten, die Argumente zu gruppieren und miteinander zu verknüpfen. Auch hier wird kontinuierlich daran gearbeitet, um die beste Balance zwischen Information und Übersichtlichkeit in der Darstellungsweise zu erlangen.
- Manipulation einzelner Nodes, z.B. des Hauptargumentes.

Schritte nach dem 19.05.2020:

- Die Formulierung weiter verbessern.
- Die Argumente finalisieren, dass alle deduktiv gültig sind.
- Fehlende oder ergänzende Argumente implementieren.
- Die Struktur der Karte festigen.
- Projektbericht erstellen, und somit
- Projekt abschliessen.