Ausarbeitung zum Seminarthema

Diskussionen zu KI-Systemen, ethische Grenzen und Auswirkungen

zum Seminar „Aktuelle Themen aus der Wissensverarbeitung“
bei der Professur für Künstliche Intelligenz und Softwaretechnologie

Finaler Abgabetermin 1.12.2016

Inhaltsverzeichnis

1. Darstellung des Diskussionsthemas

Ich veranschauliche die Problemstellung anhand eines wohl bekannten und aktuellen Beispiels:
Mit dem Sieg DeepMinds AlphaGo über Lee Sedol, bekannt als weltbester menschlicher Go Spieler, wurde deutlich, dass lernende Algorithmen mithilfe von Big-Data und neuronalen Netzwerken eine Art Intelligenz entwickeln können, die der Mensch noch nicht verstanden hat oder vielleicht auch nie verstehen wird. AlphaGo spielt das Spiel Go besser als seine Programmierer; der Algorithmus hat sich das nötige Wissen für das Spiel unermüdlich selbst „antrainiert". Professionelle Go-Spieler bewerteten die Züge und Entscheidungen AlphaGos anfänglich als Fehler [5], jedoch stellten sie sich im Laufe des Spiels als effektiv heraus und führten schließlich zum 4:1 Sieg über den Südkoreaner. [4]
Heutzutage werden lernende Algorithmen und künstliche Intelligenz (im Folgenden vereinfacht ausgedrückt mit „AI Systeme", wobei „AI" englisch ist und für „artificial intelligence" steht) in immer mehr und mehr alltäglichen Lebenslagen eingesetzt, nicht nur um Go zu spielen. Der Mensch benutzt AI Systeme in den verschiedensten Lebenslagen, die weit mehr Urteilskraft erfordern als das Setzen eines weißen oder schwarzen Spielsteins auf einem 19x19 Spielbrett.
AI Systeme werden verwendet, um Reiserouten zu planen, die Auslastung des Straßenverkehrs vorherzusagen, Gesichter in Fotos zu erkennen und dann Menschen zuzuordnen, Vorschläge für Bücher und Filme zu generieren, Diagnosen für Patienten zu geben, Viren zu identifizieren und bald auch um Autos zu fahren. Und dies sei noch lange nicht alles. [1]
Nun kommen Bedenken auf: Was passiert, wenn die AI-Algorithmen sich im gleichen Grad wie AlphaGo entwickeln, sodass der Mensch nicht mehr imstande ist, die Entscheidungen zu verstehen? Was ist wenn die Entscheidungen anfangs, oder nach etlichen Codemutationen absurd erscheinen oder gar Schaden anrichten könnten? Die Entscheidungen tragen weitaus mehr Verantwortung als das Setzen von Spielsteinen, beispielsweise hat ein Autopilot in einem Linienflugzeug durchschnittlich die Verantwortung über rund 200 Menschenleben, oder ein Auktion-Bot fährt eine Firma wie IBM in den Ruin.

Im Folgenden wird dargestellt, wieso es so schwierig ist, die Kontrolle über die AI zu behalten und wie diese uns in ihrem Verhalten täuschen.
Des Weiteren wird ausarbeitet, wie man mit AI Systemen umgehen sollte; es werden Überlegungen und Vorschläge präsentiert um so viel Kontrolle wie möglich, wenn nicht sogar die totale Kontrolle über die AI zu behalten.
Ein Kontrollverlust über die Systeme, die uns täglich umgeben, könnte fatale Folgen haben.

2. Schwierigkeit: Kontrolle

Wenn man einen Algorithmus implementiert und in Betrieb nehmen möchte, sollte er zuvor ausführlich getestet und verifiziert worden sein. Besonders wichtig ist ein Beweis der Korrektheit, wenn der Algorithmus große Verantwortung trägt, wie etwa Menschenleben oder die Verwaltung eines Firmenvermögens. Allerdings kommt es bei AI-System schnell zu Schwierigkeiten, sie zu verifizieren. Dies kann verschiedene Gründe haben:
1. Die AI wurde so geschrieben, dass sie schwer zu lesen ist. Dies kann dadurch zustande gekommen sein, dass der/die Programmierer einen ganz eigenen Stil hat/haben, oder dass entschieden wurde, dass der Algorithmus obfuskiert werden soll. Meist steht dahinter die Absicht, den Quellcode veröffentlichen zu können, ohne ihn wirklich preiszugeben.
2. Die Algorithmen hinter dem AI System sind zu komplex für das allgemeine Verständnis. Die verwendeten Funktionen greifen eventuell auf ein mathematisches, meist stochastisches Vorwissen zurück, das womöglich nur von sehr wenigen Menschen verstanden und nachvollzogen werden kann.
3. Das AI System als Ganzes ist zu groß und wurde aus mehreren Codestücken zusammengesetzt, die jeweils von verschiedenen Programmierern geschrieben wurden. Eventuell sind die Stücke in verschiedenen Sprachen geschrieben oder benutzen verschiedenste Codebibliotheken. Dass die einzelnen Stücke miteinander funktionieren lässt sich praktisch testen, wie sie es allerdings tun und ob sie sich korrekt verhalten wird dadurch nicht bestätigt.

In mindestens diesen drei Fällen entwickeln sich die AI Systeme zu „Black Boxes“, die für den Menschen nicht mehr einsehbar sind. Sie greifen auf gigantische Datenmengen zu, die der Mensch in dieser Geschwindigkeit niemals überblicken oder verarbeiten könnte. Am Ende ist nur noch erkennbar was die AI tut und wie sie handelt, nicht mehr allerdings, wie sie zu den Entscheidungen kam, die sie schließlich traf.
Etwa dies lässt sich bei AlphaGo beispielhaft beobachten, es stellt quasi vorbildhaft dar, wie sich ein AI System im generellen entwickeln wird wenn es unermüdlich mit neuen Informationen genährt wird und daraus lernt. Es ist zu erwarten, dass es jeder Art von AI Systemen früher oder später so ergehen wird. Wenn dies für AI Systeme zutrifft, denen wir unser Leben anvertrauen, so wäre es höchste Priorität die Kontrolle über dieses System zu wahren.
Ein kleines Gedankenspiel: Wir lassen ein Anti-Malware AI System gegen ein Malware AI System antreten. Beide versuchen nun mit maximaler Simulationsgeschwindigkeit unermüdlich einander besiegen zu lernen. Das Anti-Malware AI System lernt neue Barrikaden zu errichten um die Angriffe

abzuwehren, wobei das Malware AI System lernt diese Barrikaden zu umgehen und neue Angriff zu generieren. Wird der Mensch nach etlichen Millionen Simulationszyklen (Generationen) den entstandenen Code der beiden AI Systemen noch verstehen? Haben sie vielleicht beide gelernt besser miteinander freundschaftlich auszukommen anstatt gegeneinander zu kämpfen? Sie hätten dann ihre Intension verfehlt, was anfangs nicht abzusehen war. Um dem entgegenzuwirken wird im Folgenden der Entwurf der „AI-Guardians“ vorgestellt.
Auch dem Anwender wird es schwierig, AI Systeme zu kontrollieren. Dies ist meist verursacht dadurch, dass sie sich nicht als solche erkenntlich zeigen. Man nehme zum Beispiel Siri, den verbalen Kommunikations- und Informationsdienst von Apple. Fragt man Siri, ob sie eine Maschine, oder ein Algorithmus ist, so antwortet sie nicht mit der Wahrheit, sondern umgeht geschickt die Frage mit einer floskelartigen Antwort. Ferner gibt es eine Puppe, die sich via Bluetooth mit einem iPhone verbinden kann, um den Service von Siri zu übernehmen. Sollte ein Kleinkind nun damit aufwachsen, so mag nicht intuitiv die Vermutung bestehen, dass es sich um ein AI System handelt, mit dem das Kind rund um die Uhr reden kann. Mehr dazu, und eine Idee, die dies verhindern könnte unter „Turing‘s Red Flag“ [3].
Zuletzt kann die Kontrolle über die AI Systeme auch durch Hacker-Angriffe gefährdet werden. AI Systeme bieten viele andere Angriffsflächen, als konventionelle Programme, deren Output nicht von Gelerntem beeinflusst wird. Hierzu bedarf es nicht unbedingt größeren Fähigkeiten im Hacken, sondern womöglich auch nur dem achtsamen Auge für eine Backdoor. Ein Beispiel ist TayAndYou, ein lernender Twitterbot von Microsoft, der durch Input von anderen Usern des Internets dazu gebracht wurde diverse rassistische Bemerkungen zu tweeten, bevor er dann kurz darauf wieder abgeschaltet wurde.

3. „AI Guardians“

Die „AI Guardians“ (zu Deutsch „Wächter der künstlichen Intelligenz“ oder sinnlich „Die Polizei unter den AI Systemen“) ist ein Entwurf spezieller AI Systeme, deren Aufgabe es ist lernende, agierende AI Systeme zu überwachen. Es geht darum gerade den Lernprozess der agierenden AI Systeme zu überwachen und gegebenenfalls zu steuern, sollte festgestellt werden, dass ein AI System etwas lernt, das nicht mit den Intentionen des Programmierers vereinbar ist. Das zielt darauf ab, den Kontrollverlust über ein AI System so weit es geht vorzubeugen. Des Weiteren sollen AI Guardians gewährleisten, dass sich die AI Systeme, die sie überwachen, stets an die Normen, Gesetze und ethische Werte unserer Gesellschaft halten. Hier kommt man an einen Punkt, an dem sich die Programmierer, oder Informatiker im Generellen, mit Juristen, Ökologen und Experten aus verschiedensten Wissensbereichen zusammensetzen müssen, um gemeinsam die Vorstellungen von den besagten Normen, Gesetzen und ethischen Werten zu formalisieren, damit die AI Guardians einen für sie verständlichen Input haben.
AI Guardians benötigen die Eigenschaft, sich nicht von dem zu kontrollierenden AI System austricksen zu lassen, falls es merkt, dass der Guardian ihn in seinen Lernprozessen behindert. Auf der anderen Seite müssen die AI Systeme so programmiert sein, dass sie stets auf den Guardian hören oder im Notfall einen zuverlässigen „on-off switch“ haben. Für den Fall dass sie abgeschaltet werden müssen, müssen wiederum ebenfalls zuverlässige „fail-safe“ Routinen existieren.
Die AI Guardians sollen mit dem Lerntempo der agierenden AI Systeme mithalten können. Ein Mensch ist dieser Aufgabe nicht mächtig. Daher sind sie mit der nötigen Gabe ausgestattet, während des Lernprozesses von AI Systemen unmittelbar Verstöße (hinsichtlich der Werte, Intention, usw…) zu erkennen. Einen weiteren Vorteil haben AI Guardians, wenn sie die Möglichkeit bieten, „on the fly“ Korrekturen von Richtlinien anzunehmen. Dann können die AI Systeme, die überwacht werden effektiv und schnell angepasst werden.
AI Guardians haben außerdem die Aufgabe den Grund eines Verstoßes zu ermitteln, damit das agierende AI System rekonfiguriert, korrigiert oder schlimmstenfalls ersetzt werden kann. Außerdem soll überprüft werden, ob die Hardware für die Belastung des AI Systems geeignet ist, sollte es sich um ein eingebettetes AI System handeln.
Zu alldem soll ein AI Guardian erkennen, ob die Rechtfertigung eines Herstellers über einen vorliegenden Fehler im AI System stimmt. Hierzu gibt es ein Beispiel: Ein lernender Werbealgorithmus von Google soll besser bezahlte Jobs Usern angezeigt haben, von denen er gelernt hatte, dass sie männlich seien. Demnach haben weibliche User nur schlechter bezahlte

Jobs angezeigt bekommen. Google behauptete, dass das in keiner Weise ein Akt der Diskriminierung gewesen sein soll. Allerdings soll der Betreiber der Werbefirma, für den die Werbung geschaltet wurde, den Algorithmus so konfiguriert haben, dass dieses Resultat erzielt wird. Ein AI Guardian könnte eventuell aufklären ob nun tatsächlich der Algorithmus so konfiguriert wurde, oder ob er sich das selbst beigebracht hat [2].
Weitere Aufklärung bedarf es der Behauptung von Twitter, angeblich rund 125.000 Twitteraccounts gesperrt zu haben, die angeblich direkt mit der Terrororganisation Islamischer Staat in Verbindung standen. Ob der lernende Algorithmus nun wirklich erkannt hat, welcher Account mit dem IS in Verbindung stand, oder ob er wahllos potentielle Kandidaten gebannt hat ist unklar [2].
Ferner würde ein AI Guardian es zum Beispiel verhindern, dass ein autonomes Fahrzeug von anderen Verkehrsteilnehmern lernen würde, zu schnell zu fahren. Auch könnte ein AI Gaurdian von der aktuellen Umweltlage seine Richtlinien an den Treibstoffverbrauch des autonomen Fahrzeuges anpassen, oder dafür sorgen, dass das Fahrzeug nicht überholt oder über gelb fährt, wenn bekannt ist, dass eine gesamte Familie an Bord ist, die dennoch rasch ans Ziel gebracht zu werden wünscht.
Eine letzte Überlegung wäre es, die AI Guardians die politische und ökologische Lage zu Lande wissen zu lassen. Vor allem die wirtschaftliche und ökologische Umgebung wird sich mit der Weiterentwicklung der AI Systeme sprunghaft verändern. Es bahnt sich eine neue Epoche an, die mit der Industrialisierung zu vergleichen ist. Es werden noch mehr Arbeitsplätze durch effizienter arbeitende AI Systeme ersetzt werden. Länder die sich keine AI Systeme leisten können, werden noch stärker von anderen, wirtschaftlich mächtigeren Ländern abgehängt. Dies führt letzten Endes zu einer weiteren Beschleunigung der Öffnung der sozialen Ungleichheit auf der Welt.

3.1 Umsetzung und Realisierbarkeit der Guardians

Wieso ist aber das Konzept der AI Guardians so schwierig umzusetzen?
Es ist die Frage, wer den Quellcode für diese AI Guardians schreibt. Die Wahl muss auf ein Unternehmen fallen, das mit absoluter Sicherheit keine eigenen Absichten verfolgt. Ein Solches zu finden stellt sich als besonders große Herausforderung heraus. Jeder Hersteller wird behaupten, dass sich sein AI Guardian zuverlässig verhält, allerdings gibt es hierüber auch keine Gewissheit, da der Quellcode womöglich ebenfalls als Black-Box enden wird. Forscher sind der Meinung, dass die Einführung der AI Guardians längst überfällig sei und drängen zur Entwicklung erster Prototypen [2].
Ein Ansatz um diese knifflige Frage zu beantworten wäre es, dass jeder Nutzer seinen eigenen AI Guardian konfigurieren kann, damit eher der Wille des Nutzers umgesetzt werden kann, anstatt der Wille eines bestimmten Unternehmens. Hierbei ist jedoch zu beachten, dass mindestens selbstverständliche und natürliche Werte und Grundgesetze nicht umgangen werden können. Juristen müssten Gesetze entwerfen, die auf AI Systeme

zugeschnitten werden, im Falle eines Vergehens. Am Ende steht immer noch der Mensch, dem das AI System dient, und auch er muss für die Taten seines AI Systems gerade stehen.
Die oben aufgeführten Ansprüche an die AI Guardians decken alle womöglich anfallenden Probleme, die mit der Weiterentwicklung der AI Systeme anfallen können, ab. Das heißt, wenn sich dieses Konzept umsetzen ließe, so wäre die Zukunft einer Harmonie zwischen AI Systemen und den Menschen im Alltag gewährleistet. Mir erscheint die Umsetzung aber äußerst unrealistisch, allein schon wenn man sich auf einen weltweiten politischen Konsens einigen will. Die schlichte Tatsache, dass es auf der Welt Krieg gibt zeigt, dass sich die Menschen nicht immer einig sind, oder vielleicht auch nie einig sein werden. Wenn nun der Anspruch besteht, ein AI System zu entwerfen, das alle ethischen Ansichten in Harmonie bringen soll müssen sich zuerst die Menschen einig werden.
Lässt man diesen Anspruch weg, so wird es dadurch nicht wirklich einfacher, die AI Guardians umzusetzen. Alle bisher agierenden AI Systeme müssten auf ihren Guardian angepasst werden. Das ist mit einem immensen Aufwand verbunden, der finanziert werden muss.
Und mal ganz unbeachtet der angeführten Probleme bei der Umsetzung – auch ein AI Guardian ist ein AI System, das außer Kontrolle laufen kann. Daher handelt es sich lediglich um eine Verschiebung des Problems auf ein anderes AI System. Und eine unendliche Kette von Guardians für Guardians ist praktisch unmöglich und theoretisch auch keine zufriedenstellende und endgültige Lösung.

4. „Turing‘s Red Flag“

Bereits Alan Turing (1912 – 1954) schätze zu seinen Zeiten, dass Computer, oder präziser: Bots den Menschen etwa im Jahre 2000 täuschen werden [3]. Mit „täuschen“ ist gemeint, dass der Mensch die Interaktion eines Bots fälschlicher Weise für menschlich hält. Hierzu gibt es jährlich einen Wettbewerb, zu dem Programmierer mit ihren Programmen antreten können um sich dem so genannten Turing-Test zu stellen.
In Großbritannien wurde mit der Einführung der ersten motorbetriebenen Fahrzeuge ein Gesetzt erlassen, das besagt, dass ein solches Gefährt mit einer roten Flagge gekennzeichnet werden muss, damit die Passanten vor dieser neuartigen Erfindung und den potentiellen Gefahren die sie mit sich bringen kann gewarnt werden. 1896 wurde dieses Gesetz annulliert und die Geschwindigkeitsbegrenzung wurde erhöht. Von diesem Tag an gab es die ersten Verkehrsunfälle, und noch im selben Jahr gab es den ersten Toten im Straßenverkehr. Seitdem stieg die Zahl der Unglücke kontinuierlich an [3].
Diese Idee der roten Flagge auf AI Systeme bezogen nennt sich dann „Turing‘s Red Flag“, also die vor der von Turing prognostizierte Gefahr warnende rote Flagge, die auf ein AI System hinweisen oder sogar davor warnen soll. Die Idee ist, dass jedes AI System sich deutlich als solches zu erkennen gibt, damit es nicht als Mensch missverstanden werden kann.
Dies fände in den verschiedensten Lagen Anwendung. Zuerst betrachten wir die autonomen Fahrzeuge. Diese waren bisher durch ihre auffällig großen Aufbauten gut als solche zu erkennen. Mittlerweile allerdings sind sie weitestgehend nicht mehr optisch von den übrigen Fahrzeugen zu unterscheiden. Wenn andere Verkehrsteilnehmer wissen, dass es sich vor ihnen im Verkehr um ein AI System handelt, so kann man sich darauf einstellen, dass es (je nach Stand der Forschung und Technik) nicht zu schnell fahren wird, nicht ohne zu blinken die Spur wechseln, oder sonstige Widrigkeiten begehen wird. Wenn einmal die autonomen Fahrzeuge einen Großteil des Verkehrs ausmachen, wäre es durchaus sinnvoll, eher die menschlichen Fahrer zu signalisieren, sodass man eher mit Fehlerquellen rechnen kann. Die autonomen Fahrzeuge könnten untereinander ihre Absichten austauschen, um somit ein kollektives Verkehrswissen zu erzeugen und den Verkehr intelligenter zu leiten.
Eine ganz andere Wirkung erhält „Turing‘s Red Flag“ wenn man an Bots denkt, die Glücksspiele wie zum Beispiel Online Poker Texas Hold‘em spielen. Ein lernender Bot kann äußerst akkurat berechnen, welche Karten als nächstes höchstwahrscheinlich aufgedeckt werden und hat damit einen unfairen Vorteil gegenüber den anderen, menschlichen Mitspielern. Sollten die Mitspieler durch „Turing‘s Red Flag“ darüber aufgeklärt werden, dass sie mit einem Bot spielen, ist zu erwarten, dass die menschlichen Mitspieler

das Spiel auf der Stelle verlassen, da sie sich hoffnungslos dem Bot unterlegen fühlen.
Ein ähnlicher Effekt ist zu beobachten, wenn ein Bot Ego-Shooter oder Reflexspiele spielt. Auch hier ist der Bot in der Lage augenblicklich auf die Situation im Spiel zu reagieren. Es ist zu erwarten, dass ein menschlicher Mitspieler auch hier das Spiel auf der Stelle verlassen würde. Daher ist es im Interesse des Betreibers alle Bots aus seinem Spiel zu werfen. „Turing‘s Red Flag“ würde das zuverlässig ermöglichen.
Es wäre es außerdem nützlich, zu wissen ob ein Text aus dem Internet von einem AI System zusammengestellt wurde, oder ob es einen menschlichen Autor zu dem Text gibt. Bereits einige Beträge auch von wissenschaftlichen Internetseiten sind von lernenden Algorithmen zusammengestellt worden. Hierbei ist nicht auszuschließen dass dieser Algorithmus sich eine Art eigenen Willen antrainiert hat, und seinen Lesern eine künstliche Meinung auferlegt. „Turing‘s Red Flag“ würde dem Leser signalisieren, dass es sich um einen automatisch zusammengesetzten Text handelt und dadurch vielleicht das Interesse des Lesers reduzieren, da es sich möglicher Weise authentischer anfühlt einen „menschlichen“ Text zu lesen. Sollte sich das Interesse gebildet haben mit dem Autor Kontakt aufzunehmen, so verliefe sich das nicht in einer antwortlosen Mail. Der Leser käme gar nicht dazu dieses Interesse zu entwickeln, da er sich darüber bewusst ist, dass es keinen solchen gibt.
Zuletzt ist zu beobachten, dass Kinder sprechenden Puppen oder sprechenden Kuscheltieren vermutlich mehr Eigenschaften zusprechen, als tatsächlich implementiert wurden. Zum Beispiel kann eine Puppe schnell als „traurig“ oder „glücklich“ beschrieben werden, allerdings ist im Quellcode keine einzige Emotion implementiert. Es scheint so, als würde der Mensch den AI Systemen intuitiv mehr Vermögen zuschreiben, als sie tatsächlich besitzen. Das könnte dazu führen, dass wir ein AI System als menschlich wahrnehmen, längst bevor dieses AI System im Ansatz menschliche Züge imitieren kann. Daher ist es dringend nötig, „Turing‘s Red Flag“ einzuführen, um die schwindende Grenze zwischen AI Systemen und Menschen wieder sichtbarer zu machen.
Seit 2011 sind selbstfahrende Autos in Nevada (USA) legal und das dafür verabschiedete Gesetz sieht keine „Turing‘s Red Flag“ vor. Ebenfalls ist seit Februar 2015 in Deutschland ein Komitee gegründet, das das autonome Fahren gesetzlich legalisieren versucht. Auch hier wird keine Überlegung der „Turing‘s Red Flag“ gewidmet [3].

4.1 Umsetzung und Realisierbarkeit der roten Flagge

Mir erscheint die Umsetzung von „Turing‘s Red Flag“ eher realistisch als die der AI Guardians. Die Ansprüche sind klarer definiert und es bedarf nicht des Zusammensetzens von mehreren Wissenschaften. Es obliegt den Informatikern allein das Konzept umzusetzen. Die AI Systeme müssen nicht grundlegend angepasst werden. Lediglich ihre Form der Kommunikation

müsste angepasst werden. Ihr restliche Funktionsweise, die beliebig komplex sein kann, müsste meist gar nicht angerührt werden.
Wenn ein Mensch weiß, ob er gerade mit einem AI System oder einem anderen Menschen agiert, würde er sich anders verhalten. Die Information, die „Turing‘s Red Flag“ vermittelt, würde also dem Menschen helfen anders, oder besser zu urteilen. Das Verschweigen dieser Information lässt den Menschen also in einer Position, in der er nicht in dem Maße urteilen kann, wie es ihm zustünde.

5. Fazit

Abschließend ist zu sagen, dass der Mensch selbst als Schöpfer der AI Systeme gilt. Daher sollte auch er derjenige sein, der stets die Kontrolle behält, egal wie intelligent nun ein System geworden ist.
Ob die AI Guardians nun realisiert werden, oder „Turing‘s Red Flag“ zu implementiert wird, wird sich zeigen. Allerdings hat es oberste Priorität, dass AI Systeme nicht einen eigenen Willen entwickeln und sich anfängt den Menschen mit unverständlich komplexer Intelligenz zu unterwerfen.
Wenn sich nun auftut, dass diese Kontrolle gefährdet ist, so müssen alle AI Systeme zuverlässig anpassbar sein, oder sogar im schlimmsten Fall komplett abgeschaltet werden können. Es bedarf also immer einem Plan B, der auch in einer Welt ohne AI Systeme funktioniert.
AI Systeme sollten daher nicht die Entscheidungen von sehr wichtigen Fragen übernehmen, da der Mensch nicht absehen kann ob nun diese Entscheidungen tatsächlich richtig sind, oder ob das AI System bereits eine Art eigenen Willen entwickelt hat.
Ich finde, dass sich dem Menschen mit lernenden Algorithmen eine riesige Herausforderung auftut. Es ist dringend nötig, dass die Folgen von AI Systemen und Risiken die sie mit sich bringen können besser erforscht werden. Es ist in unser aller Interesse diese Systeme für uns arbeiten zu lassen und uns in der Forschung zu unterstützen. Allerdings nutzt ein AI System nichts, wenn man es nicht kontrollieren und verstehen kann. Sobald wir von AI Systemen lernen können und ratlos zusehen was sie als nächstes tun werden, haben wir eine Intelligenz erschaffen, für die der Mensch noch nicht bereit ist. Und das ist in meinen Augen eine nicht einschätzbare Gefahr.

Literaturverzeichnis

1: *„Rise of concerns about AI: Reflectiosn and Directions“* veröffentlicht von Thomas G. Dietterich und Eric J. Horvitz

2: *„Designing AI Systems that Obey Our Laws and Values“* veröffentlicht von Amitai und Oren Etzioni

3: *„Turing‘s Red Flag“* veröffentlicht von Toby Walsh

4: *„DeepMind AlphaGo vs Lee Sedol“* - https://gogameguru.com/tag/deepmind-alphago-lee-sedol/ (Abgerufen am 01.11.16)

5: *„Es geht um weit mehr als Go“* - http://www.spiegel.de/netzwelt/gadgets/alphago-sieg-wendepunkt-der-menschheitsgeschichte-a-1082001.html (Abgerufen am 12.11.2016)