Martin-Luther-Universität Halle-Wittenberg
Naturwissenschaftliche Fakultät III
Institut für Informatik

Seminar

Informatik und Gesellschaft

Sommersemester 2018

geleitet durch Prof. Dr. Paul Molitor

Was kann, soll und darf Künstliche Intelligenz

Toni Holger Müller & Niklas Jordi Freymuth

Inhaltsverzeichnis

Gender-Hinweis: Die weibliche Form ist der männlichen Form in dieser Arbeit gleichgestellt; lediglich aus Gründen der Vereinfachung wurde die männliche Form gewählt.

Überarbeitung durch den Dozenten: Der vorliegende Text entspricht im Wesentlichen dem ursprünglichen durch Herrn Toni Holger Müller und Herrn Niklas Jordi Freymuth erstellten Bericht. Der Dozent hat lediglich (die sehr wenigen) Schreib- und Kommatafehler korrigiert, einige Formatierungen angepasst und Einträge in der Literaturliste vervollständigt.

1 Einleitung

Was kann, soll und darf künstliche Intelligenz? Diese zunächst einfach wirkenden Fragen erweisen sich in jüngster Zeit als ein vieldiskutiertes Thema, dessen Aspekte von der Informatik über die Rechtswissenschaft, die menschliche Psychologie, bis hin zur Ethik und Philosophie reichen.
Die Frage teilt sich in drei Teilfragen nach dem *Können*, *Sollen* und *Dürfen* künstlicher Intelligenz ein, die im Rahmen dieser Ausarbeitung in genannter Reihenfolge diskutiert werden sollen. Der Frage nach dem *Können* wird dazu zunächst eine kurze historische Zusammenfassung der Forschung und des Fortschritts künstlicher Intelligenz vorangesetzt, um eine Perspektive für den aktuellen Forschungsstand zu bekommen. Darauf folgt die Frage danach, *was* künstliche Intelligenz (KI) eigentlich ist, sowie zwei Beispiele künstlicher Intelligenzen. Anschließend geht es um die Präsenz von KI im heutigen Leben und dem generellen Können heutiger künstlicher Intelligenz.
Im nächsten Kapitel werden die Fragen nach dem *Sollen* und *Dürfen* diskutiert. Es liegt in der Natur der Sache, dass im Gegensatz zum *Können* keine definitive Antwort auf die Fragen möglich ist. Das Kapitel wird kein festes Ergebnis präsentieren, sondern verschiedene Standpunkte, Argumente und Fallbeispiele erläutern und beleuchten. Dabei geht die Thematik weit über die reine Informatik hinaus, weil Fragen wie „Ab wann fühlt eine Maschine?“ sich leider nicht in ein theoretisches Modell zwängen lassen.

2 Grundlagen

2.1 Historisches

„Wer die Vergangenheit nicht kennt, kann die Gegenwart nicht verstehen. Wer die Gegenwart nicht versteht, kann die Zukunft nicht gestalten.“ - Hans-Friedrich Bergmann

Die Geschichte künstlicher Intelligenz beginnt, je nachdem, wen man fragt, bereits in der Antike, als Homer in seiner Odyssee von mechanischen Wesen schrieb, die an der Tafel der Götter die Speisen servieren [1]. Schon hier könnte man anfangen, über die Darstellungsweise und den vermeintlichen Nutzen dieser ersten „denkenden Maschine“ zu diskutieren, im Rahmen dieser Ausarbeitung wird es jedoch hauptsächlich um real existente künstliche Intelligenz gehen. Die Geschichte dieser beginnt deutlich später, nämlich irgendwann in den späten 1940er und früher 1950er Jahren. Ein Nachweis dafür ist beispielsweise der 1945 Vannevar Bush erschienene Aufsatz „As We May Think“

[2], in dem es unter anderem darum geht, in welcher Art und Weise der Computer den Alltag des Menschen vereinfachen kann. 1950, also wenige Jahre später, erschien Alan Turings berühmtes Paper „Computing Machinery and Intelligence“ aus dem Jahre 1950, das treffend mit den Worten „I propose to consider the question, 'Can machines think'“ beginnt [3]. Diese beiden Fälle waren keine Einzelerscheinungen, und so nahm das akademische Interesse an denkenden Maschinen weiter zu, bis im Sommer 1956 am Dartmouth College in Hanover, New Hampshire, ein zweimonatiges Treffen anerkannter Wissenschaftler dieses Themengebiets zu der Prägung des Terms „künstliche Intelligenz“ führte [1, 4]. Einige Autoren geben an, dass dieses Treffen die künstliche Intelligenz zur Forschungsdisziplin erhoben hat [5].

Nach dem Treffen von 1956 folgte der erste große „KI-Hype“. Die Öffentlichkeit war von den neuen Fähigkeiten der Computer gleichermaßen überrascht und begeistert, die US-Regierung finanzierte maßgebliche Teile der Forschung und die Experten überschlugen sich geradezu mit hochtrabenden Prognosen nahender Errungenschaften [6, 7, 8]. Es schien, als wäre den Möglichkeiten künstlicher Intelligenz während dieser goldenen Jahre keine Grenze gesetzt.

Auf den ersten großen Aufschwung folgte der erste große Sturz. Die optimistischen Schätzungen konnten durch mangelnde Rechenleistung und die unerwartete Komplexität einiger Probleme nicht eingehalten werden. Einige Berichte, insbesondere ein Bericht von 1966 des „Automatic Language Processing Advisory Committee“ (kurz ALPAC) warfen ein deutlich negativeres Licht auf die Dinge [9]. In dem Bericht geht es um die Fortschritte künstlicher Intelligenz im Bereich der automatischen Übersetzung von Sprache, einem Thema, das im Kalten Krieg von hoher Relevanz war. Unter anderem hieß es dort im Abschnitt „The Present State of Machine Translation: [...] there has been no machine translation of general scientific text, and none is in immediate prospect“ [10]. Der wenige Jahre später (1973) erschienene Lighthill Report stellte künstliche Intelligenz ähnlich schlecht dar, wodurch die amerikanische Regierung letztendlich veranlasst wurde, die für KI-Forschung vorgesehenen Gelder zu streichen [9]. Die Konsequenzen waren enorm, und so kam die KI-Forschung in den 70er Jahren für mehrere Jahre mehr oder weniger zum Stehen [9, 6]. Während dieses sogenannten „KI-Winters“ passierte nicht viel. Mangelnde Rechenleistung und ineffiziente Systeme führten zu nicht zufriedenstellenden Ergebnissen, die sich in keiner Weise mit den überaus optimistisch gesetzten Zielen messen konnten. So verlor künstliche Intelligenz bis zum Beginn der 80er Jahre immer mehr an Bedeutung.

Um 1982 herum gab es den nächsten großen Aufschwung. Die meisten Autoren at-

tribuieren das neu entfachte Interesse an künstlicher Intelligenz der Entwicklung von Expertensystem, durch die der Computer in der Lage war, aus vorherigen Fehlern und Situationen zu lernen, sowie dem Einfluss der japanischen Regierung, die die USA durch ihre Bemühungen indirekt zwang, mit ihr mitzuhalten [6, 11]. Dieses Mal hielt der Hype nicht lange an, denn bereits Ende der 80er Jahre war klar, dass die erneut zu hoch gesteckten Erwartungen nicht erreicht werden würden. So begann aus verschiedenen Gründen der sogenannte „zweite KI-Winter“ [6].
Doch wieder wurde das Interesse an künstlicher Intelligenz wenig später neu geweckt. Im Jahr 1993 gab das MIT COG Project, ein Projekt mit dem Ziel, einen humanoiden Roboter zu bauen, Grund zur Hoffnung [11]. Als wenige Jahre später „DeepBlue“ als erster Schachcomputer öffentlich gegen einen Großmeister gewann, war klar, dass die Grenzen künstlicher Intelligenz noch lange nicht erforscht waren [12]. Seitdem wächst die Forschung um künstliche Intelligenz ständig weiter an und hat verschiedene bemerkenswerte Ergebnisse erzielt. Gleichzeitig sorgt die steigende Relevanz von Big Data und die Verfügbarkeit immer leistungsfähigerer Rechner für die nötigen Daten und die nötige Prozessorkraft, um bereits heute eine Vielzahl komplexer Probleme elegant durch künstliche Intelligenz lösen zu können.
Obwohl das alles sehr positiv klingt, lässt sich zum momentanen Zeitpunkt noch keine Prognose treffen, wie die KI sich in den nächsten Jahren entwickeln wird. Obwohl beispielsweise IBM derzeit Milliarden in die Entwicklung von „Watson“, einer KI, die auf in natürlicher Sprache gestellte Fragen antworten kann, investiert [13], hat die Vergangenheit gezeigt, dass der so oft versprochene nächste große Durchbruch meistens leider einige Jahre weiter entfernt ist, als es zunächst den Anschein haben mag.
Zusammengefasst lässt sich ein einfaches Bild erkennen: Künstliche Intelligenz ist vielversprechend, aber kein Hexenwerk. Obwohl verschiedene Errungenschaften im Laufe der Zeit erreicht wurden und in den kommenden Jahren erreicht werden, steht der Forschung noch ein weiter Weg bevor. Die Vergangenheit zeigt, dass optimistische Schätzungen zur Entwicklung von KI sich leider in den seltensten Fällen bewahrheiten, und so ist es über alle Maßen unwahrscheinlich, dass sich die Welt, wie man sie kennt, in den nächsten Jahre durch künstliche Intelligenz signifikant verändert. Trotzdem sind die Themen, die im späteren Verlauf dieser Ausarbeitung besprochen werden, von hoher Relevanz, da ein beständiger Fortschritt der KI im Allgemeinen nicht zu leugnen ist. Auch wenn der „Terminator“ in den nächsten Jahrzehnten eine Filmfigur bleiben wird, gibt es andere, unscheinbarer wirkende Bereiche, in denen man sich bereits jetzt mit dem Sollen und Dürfen von KI auseinandersetzen muss.

2.2 Was ist KI?

„Intelligenz ist die zusammengesetzte oder globale Fähigkeit des Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinanderzusetzen.“ - David Wechsler

Um das Konzept künstlicher Intelligenz zu verstehen, muss man zunächst wissen, was Intelligenz im Allgemeinen ist. Die oben genannte Definition ist dabei leider nur eine von vielen [14], denn wie bei vielen anderen Themen der Psychologie ist es auch hier nicht möglich, einen Begriff hart zu definieren. Trotzdem haben alle Versuche, den Begriff „Intelligenz“ in Worte zu fassen, einiges gemeinsam: Intelligenz bedeutet, Zusammenhänge zu verstehen und auf Basis dieser zu handeln. Im weiteren Sinne ist dazu ein Lernprozess beziehungsweise Erfahrung notwendig.
Ganz ähnlich verhält es sich mit künstlicher Intelligenz, die den Versuch darstellt, menschliche Intelligenz auf den Computer zu übertragen. Das Ziel künstlicher Intelligenz ist, ein System zu entwickeln, das, ganz ähnlich zu menschlicher Intelligenz, Sachverhalte verstehen und über diese nachdenken kann, und das aufgrund vorhergegangener Erfahrungen Entscheidungen trifft. Es ist schwierig bis unmöglich zu sagen, an welchem konkreten Punkt ein Algorithmus aufhört, nur ein Konstrukt aus Schleifen, Abfragen und gegebenenfalls rekursiven Aufrufen zu sein, und stattdessen zur künstlichen Intelligenz wird. Im Sinne dieser Ausarbeitung ist „künstliche Intelligenz“ deswegen als Algorithmus definiert, der in der Lage ist, komplexe Probleme selbstständig und denkend zu lösen. Auch hier ist die Definition unscharf, jedoch sollte man eine intuitive Vorstellung davon haben, wann ein Algorithmus für die Lösung eines Problems „nachdenken“ muss. Es gibt andere Ansätze zur Messung oder Abgrenzung künstlicher Intelligenz, beispielsweise den berühmten Turing Test [3]. Auf diese wird im weiteren Verlauf nicht explizit eingegangen, da ein intuitives Verständnis künstlicher Intelligenz völlig ausreicht.
Bisher ging es nur um die Historie und Definition künstlicher Intelligenz, aber nicht darum, *was* genau KI eigentlich ist. Künstliche Intelligenz wird in den Medien gerne als unaufhaltbare Tötungsmaschine im *Terminator*-Stil dargestellt [15, 16, 17], die Realität sieht aber - zumindest heute noch - etwas anders aus. Im Folgenden gibt es deswegen zwei Beispiele aus unterschiedlichen Bereichen der künstlichen Intelligenz, die die Frage danach, was KI ist, zumindest ansatzweise beantworten sollen. Anschließend folgt eine Auflistung künstlicher Intelligenzen im Alltag, um die eigentliche Frage nach dem *Können* künstlicher Intelligenz zu beantworten.

2.2.1 Beispiel: Entscheidungsbäume

Ein Entscheidungsbaum ist die wohl einfachste Unterart künstlicher Intelligenz. Er besteht aus einem Baum im mathematischen Sinne, also einem gerichteten, schwach zusammenhängenden und kreisfreien Graphen. Jeder Knoten des Baumes ist entweder ein Ergebnis oder stellt eine Frage, wobei die von ihm ausgehenden Kanten mögliche Antworten repräsentieren. Die Fragen müssen nicht binär sein, aber zu jeder Frage muss es immer genau eine Antwort, also eine ausgewählte Kante geben. In diesem Sinne müssen die Ausgangskanten eine Partitionierung des Antwortraums der gestellten Frage sein. Die Anwendung von Entscheidungsbäumen funktioniert so, dass man beim Wurzelknoten anfängt und sich Frage für Frage durch den Baum arbeitet. So landet man zwangsweise irgendwann in einem Knoten mit Ausgangsgrad 0, der per Definition ein Ergebnis darstellt [18].
Das Erstellen eines möglichst guten Entscheidungsbaumes kann mitunter sehr komplex werden: Hierzu gibt es verschiedene Algorithmen, der bekannteste ist der ID3-Algorithmus von J.R. Quinlan [19], der mehrmals modifiziert oder verbessert wurde (beispielsweise 2010 von L. Yuxun und X. Niuniu [20]). Im Gegensatz dazu kann ein fertiger Entscheidungsbaum ganz einfach über verschachtelte if-else Anweisungen, oder eben durch eine namensgebende Baumstruktur realisiert werden. Dadurch sind Entscheidungsbäume für den Menschen gut nachvollziehbar, weil man den zur Entscheidung führenden Pfad verfolgen kann, um herauszufinden, *warum* der Entscheidungsbaum ein bestimmtes Ergebnis gewählt hat. Diese Lesbarkeit ist ein riesiger Vorteil gegenüber anderen künstlichen Intelligenzen [21].
Entscheidungsbäume finden in einem weiten Spektrum Anwendung. Als Beispiel seien hier die Diagnose medizinischer Probleme [22], das Erkennen 3-dimensionaler Objekte [23] und die Verbesserung von Sprachsynthese (der Versuch, dem Computer eine möglichst menschenähnliche Aussprache zu geben) [24] genannt. Das folgende Beispiel ist dagegen vergleichsweise einfach: Wir haben uns mit Freunden zu einem Rollenspielabend verabredet und müssen uns nun entscheiden, welche Klasse von Charakter wir spielen wollen. Zur Auswahl stehen ein heilender Kleriker, ein bogenschießender Waldläufer, ein heimtückischer Assassine und ein muskelbepackter Krieger. Der in Abbildung 1 abgebildete Entscheidungsbaum kann uns helfen, eine dieser Möglichkeiten zu wählen. Wir beginnen ganz links bei der *Will ich heilen können?* und arbeiten uns durch den Baum, bis wir an einem der farbigen Knoten ankommen. Dieser steht dann für die am besten zu uns passende Klasse.

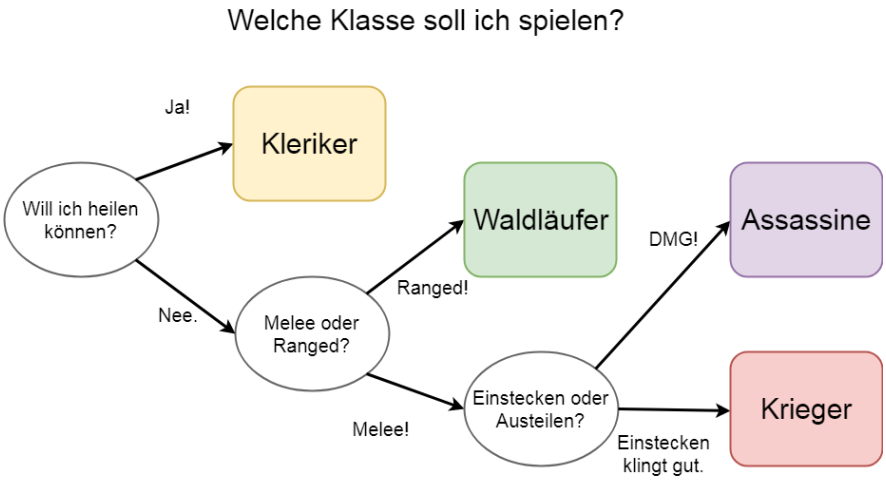

Abbildung 1: Ein Entscheidungsbaum zur Wahl der am besten geeigneten Klasse am nächsten Rollenspielabend.

Der gezeigte Entscheidungsbaum ist natürlich noch relativ klein, kann aber durch die Struktur des Baumes leicht um weitere Klassen wie dem Magier oder dem Barbaren erweitert werden. Letzterer könnte beispielsweise zusammen mit dem Assassinen von einem neu erstellten Knoten *Schleichen oder Reinstürmen?* aus erreichbar sein, zu dem jetzt die Kante *DMG!* führt.

2.2.2 Beispiel: Neuronale Netze und Go

Entscheidungsbäume sind eine recht einfache, gut nachvollziehbare und verständliche Art künstlicher Intelligenz. Neuronale Netze hingegen haben zwar auch eine einfache zugrunde liegende Mathematik, arbeiten jedoch sehr schnell auf einer für den Menschen unverständlichen Abstraktionsebene.
Neuronale Netze wurden zwar bereits in den 60er Jahren entwickelt, gelangten jedoch gemeinsam mit der restlichen Forschung um künstliche Intelligenz alsbald in Vergessenheit [25]. Doch da Neuronale Netze besonders stark von immer größer werdenden Datenmengen und höherer Rechenleistung profitieren, sind sie der momentan populärste und erfolgreichste Vertreter maschinellen Lernens [26].
Die genaue Funktions- und Arbeitsweise neuronaler Netze zu erklären würde den Rahmen dieser Ausarbeitung sprengen, deswegen seien an dieser Stelle die folgenden Quellen empfohlen:

- „Neural Networks: A Systematic Introduction“ von R. Rojas[27]

- „Neural Networks and Deep Learning“ von M. Nielsen [28]
- „Deep Learning“ von A. Gibsn und J. Patterson [29]

Sehr kurz zusammengefasst simulieren neuronale Netze die namensgebenden Strukturen im (menschlichen) Gehirn. In diesem Kontext wurden die sogenannten Neuronen erstmals in den späten 50er und frühen 60er Jahren von F. Rosenblatt erwähnt [30, 31].

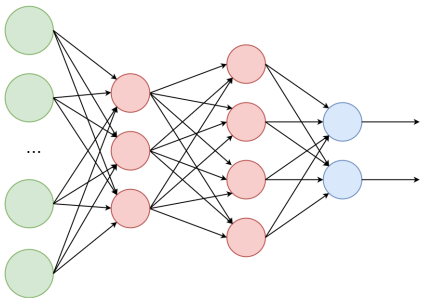

Abbildung 2: Ein neuronales Netz

Sie bestehen aus Neuronen, die auf Basis mehrerer Eingaben einfache Entscheidungen in Form von Berechnungen treffen und als Ergebnis ausgeben. Dabei kann die Ausgabe eines Neurons als Eingabe anderer Neuronen genutzt werden, wodurch die ursprünglich simplen Berechnungen zu sehr komplexen Abstraktionen und Ergebnissen führen können. Die (relative) Korrektheit der berechneten Ergebnisse wird dadurch erreicht, dass die Gewichtung der Eingaben der einzelnen Neuronen durch ein Lernverfahren Stück für Stück optimiert wird. Abbildung 2 zeigt ein beispielhaftes neuronales Netz. Die farbigen Kreise repräsentieren Neuronen, die ankommenden Pfeile ihre Eingaben, die ausgehende ihr berechnetes Ergebnis. Man sieht auf der Abbildung sehr schön, wie die Ergebnisse der grünen Neuronen von Rot weiterverarbeitet werden, um schließlich bei den Blauen zu landen.

Die Arbeitsweise eines neuronalen Netzes ist so abstrakt, dass es dem Menschen bereits bei kleineren Vertretern nicht mehr möglich ist, genau nachzuvollziehen, wie genau sie ein bestimmtes Ergebnis erzielen. Zwar kann man sehen, *was* das Netz berechnet, jedoch ist das *wieso* gerade bei schwierigeren Aufgaben wie der Bild- oder Spracherkennung quasi unmöglich. Das Verständnis der inneren Arbeitsweise der neuronalen Netzwerke ist deswegen Gegenstand aktueller Forschung [32, 33]. Ein eindrucksvolles Beispiel für das Potential moderner neuronaler Netzwerke gibt es im Zusammenhang mit dem Spiel Go, das in ausführlicher Form in einer früheren Ausarbeitung des Autors [34] zu finden ist: „Go ist ein aus China stammendes Strategiespiel für zwei Spieler. Ziel des Spiels ist es, durch geschicktes Setzen der eigenen Spielsteine einen möglichst großen Teil des (in der Regel 19×19 Felder großen) Spielbretts einzukreisen. Den umschlossenen Bereich nennt man Gebiet. Es wird abwechselnd gesetzt, am Ende gewinnt

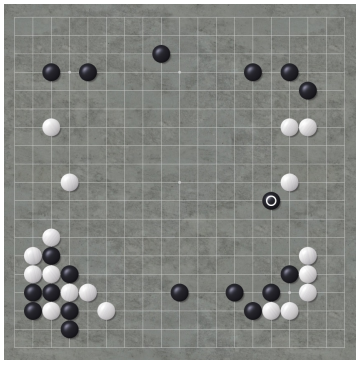

Abbildung 3: Partie 2 aus dem „Google DeepMind Challenge Match“ [35]. Schwarz hat den markierten Stein gespielt, Weiß ist am Zug. Gespielt auf einem Online-Go Server [36].

der Spieler mit mehr Gebiet“ [34]. Abbildung 3 zeigt eine laufende Partie.
Trotz weniger, überschaubarer Regeln ist Go ein kombinatorischer Albtraum; insgesamt gibt es $2,08 \cdot 10^{48}$ gültige Stellungen auf einem klassischen Go-Brett [37], was ein einfaches Durchrechnen unmöglich macht. Hier kommen die neuronalen Netzwerke und insbesondere die künstliche Intelligenz AlphaGo [38] ins Spiel: Durch das Anschauen unzähliger Profipartien lernte AlphaGo, welche Züge in welcher Situation meist gut funktionieren. Diese Züge probierte die KI in Partien gegen sich selbst aus, und behielt sie dann in ihrem Repertoire, wenn sie zu einem gewonnenen Spiel führten. Als Konsequenz spielte AlphaGo immer weniger schlechte Züge, bis es 2016 eine übermenschliche Spielstärke erreichte, und entgegen allgemeiner Erwartung einen der stärksten professionellen Go-Spieler bezwang [39]. Wichtig ist hier, dass AlphaGo zwar in der Lage ist, den besten (oder zumindest einen sehr guten) Zug zu spielen, aber selbst nicht weiß, *warum* dieser Zug der beste ist. Bis heute rätseln viele Go-Spieler an den „irgendwie merkwürdigen“ Zügen von AlphaGo und seinen in den letzten Monaten immer häufiger werdenden Konkurrenten. Zwar funktionieren gerade diese seltsamen, unerwarteten Züge oft erschreckend gut, jedoch widersprechen sie seit Jahrhunderten bewährten Grundsätzen des menschlichen Spielens.

2.2.3 Was kann KI heute?

Natürlich sind die praktischen Anwendungen einer Go-KI per Definition auf das Go-Spiel begrenzt, aber trotzdem sollte dieses Beispiel eine Vorstellung für die Möglichkeiten künstlicher Intelligenz vermitteln. Als weiteres eher akademisches Beispiel sei die ImageNet-Challenge genannt: Ziel der ImageNet-Challenge ist es, aus hunderten ver-

schiedener Objektkategorien und Millionen von Bildern möglichst oft das richtige zu erkennen [40]. 2014 haben Szegedy et al. in [41] eine künstliche Intelligenz entworfen, die bei der Bilderkennung eine Genauigkeit von über 90% hat, die richtige Kategorie als eine von 5 Vorschlägen zu nennen. Durch die hohe Ähnlichkeit der Kategorien untereinander ist das ein mit der Leistung eines Menschen vergleichbares Ergebnis.
Während die obigen Beispiele eher von theoretischem Nutzen sind, wirkt künstliche Intelligenz schon heute in erschreckendem Maße in unserem Alltag:

- Kaufvorschläge auf *Amazon* basieren auf einer künstlichen Intelligenz, die aus Einträgen in der Wunschliste, bisherigen Käufen und vielen weiteren Variablen versucht, nur das vorzuschlagen, was der Nutzer auch sehen möchte [42].

- *Facebook* nutzt ein neuronales Netzwerk mit über 100 Millionen Parametern zur automatischen Erkennung von Gesichtern auf Bildern und erreicht damit eine dem Menschen ebenbürtige Genauigkeit [43].

- Spracherkennung, also das Wahrnehmen und Interpretieren gesprochener Sprache wird jährlich besser. Mit Produkten wie *Siri*, *Alexa* oder *Cortana* ist zwar momentan noch kein flüssiges Gespräch möglich, aber in nicht allzu ferner Zukunft durchaus vorstellbar. Auch hier ist die Basis eine künstliche Intelligenz, meist ein neuronales Netzwerk wie beispielsweise in einem Paper von Collobert et al. [44] vorgestellt.

- *Uber*, eine Art Taxiunternehmen, nutzt eine künstliche Intelligenz zur Routenplanung, um die Wartezeiten der Kunden und der Fahrer zu minimieren [45].

- *Spotify*, einer der weltweit größten Anbieter für Musik, Podcasts und Hörbücher, nutzt gleich mehrere verschiedene Ansätze, um mit Hilfe künstlicher Intelligenz die Songvorschläge seiner Nutzer zu verbessern. Zum einen versucht Spotify über Sprachverarbeitung ein generelles "Stimmungsbild" für bestimmte Artisten, Alben oder Lieder zu generieren. Dazu durchsucht es einfach das Internet nach Posts zu dem jeweiligen Thema und verarbeitet die Treffer weiter. Zum anderen analysiert Spotify die konkrete Audio-Datei eines jeden Tracks und sucht so nach Liedern mit ähnlicher Lautstärke, ähnlichem Tempo, ähnlichen Instrumenten und sogar Tonarten [46].

2.2.4 Starke vs. schwache KI

In der Literatur und in Diskussionen zu dem Thema fallen oft die Begriffe *starke* und *schwache* künstliche Intelligenz. Eine künstliche Intelligenz gilt als *schwach*, wenn sie auf einen bestimmten Bereich begrenzt ist. Sie sind nicht wirklich intelligent, zumindest nicht im menschlichen Sinne, sondern einfach darauf trainiert, sehr spezielle Probleme zu lösen. Oder etwas formeller ausgedrückt: „Schwache künstliche Intelligenz ist vor allem auf die Erfüllung klar definierter Aufgaben ausgerichtet und variiert die Herangehensweise an Probleme nicht. Stattdessen greift die schwache künstliche Intelligenz auf Methoden zurück, die ihr für die Problemlösung zur Verfügung gestellt werden“ [47]. Die Go-KI kann zwar Go, und nach einiger Anpassung und mit einem anderem Lernvorgang auch Schach oder Shogi spielen, ist aber für Aufgaben wie Sprach- oder Bilderkennung vollkommen ungeeignet. Generell zählen alle heute existierenden künstlichen Intelligenzen zu den schwachen KIs.
Im Gegensatz dazu steht die *starke* künstliche Intelligenz: „Das Ziel einer starken künstlichen Intelligenz [...] ist es, die gleichen intellektuellen Fertigkeiten von Menschen zu erlangen oder zu übertreffen. Eine starke künstliche Intelligenz handelt nicht mehr nur reaktiv, sondern auch aus eigenem Antrieb, intelligent und flexibel“ [47]. Die Entwicklung einer starken KI ist bis heute nicht gelungen und steht auch noch nicht in Aussicht. Trotzdem wirft schon die hypothetische Existenz einer solchen, dem Menschen ebenbürtigen oder überlegenen Intelligenz einige interessante Fragen auf, die im weiteren Verlauf dieser Ausarbeitung zwar nicht beantwortet, aber zumindest gestellt und betrachtet werden sollen.

3 Diskussion

3.1 Gefahren von KI

Im vorherigen Teil sollte klar geworden sein, dass KI ziemlich vielschichtig ist. Es fällt sehr schwer, abzuschätzen wie die Entwicklung vorangehen wird. Vielleicht entwickeln in 30 Jahren vollkommen selbstständig denkende Androiden eine Selbstwahrnehmung, wie im kürzlich erschienenen Computerspiel Detroit: Become Human [48], vielleicht erreicht die Forschung auf dem Bereich der künstlichen Intelligenz auch eine Plateauphase und es stellt sich heraus, dass man echte Intelligenz nicht modellieren kann. In beiden Fällen lohnt es sich jedoch, genauer auf die Chancen und Gefahren von KI einzugehen. Auf Spekulationen jeglicher Art wird dabei im Folgenden weitgehend

verzichtet, da diese sich in der Vergangenheit meist als unzureichend herausgestellt haben.

3.1.1 Gefahr von SuperKI

Redet man von einer SuperKI oder Superintelligenz, ist meist die Rede von einer Maschine, die sich selbst weiterentwickeln kann [49]. Demzufolge wäre es theoretisch möglich, dass sich die Maschine mit jeder Neuentwicklung weiter verbessert und irgendwann einen perfekten, gottähnlichen Zustand erreicht. Es existieren zwar grobe Richtlinien für künstliche Intelligenz, wie etwa die „Robotergesetze“, die 1942 von Isaac Asimov erstmals vorgestellt wurden, dass Roboter als oberstes Ziel haben sollten, zu verhindern, dass Menschen direkter oder indirekter Schaden zugefügt wird. Das Problem hierbei ist, dass eine mögliche Ableitung zu einem „nullten Gesetz“ führen könnte, das besagt, dass Roboter nicht zulassen dürfen, dass die Menschheit direkten oder indirekten Schaden nimmt [50]. In Filmen wie „I, Robot“ oder anderen Science Fiction Romanen („Terminator“, „Matrix“, etc.) führte diese Überlegung zu dem Schreckensszenario, dass Roboter den Menschen selbst als größte Gefahr für sich sehen und damit nicht länger an die Gesetze gebunden sind, da sie ihre Taten „zum Schutz der Menschheit“ rechtfertigen können.
Diese Gefahr ist durchaus gerechtfertigt und spätestens durch die oben aufgeführten Filme/Romane ist sich auch die breite Öffentlichkeit dieses Risikos bewusst. Forscher sind sich momentan noch sehr uneinig über diese Möglichkeit. Zwar besagt Moore's Law, dass die (Anzahl an Transistoren, und damit auch) die Leistung eines Computerchips sich ungefähr alle achtzehn Monate verdoppelt [51], dennoch ist unklar, ob das menschliche Gehirn (mit Selbstbewusstsein, echten Emotionen, Intelligenz und so weiter) wirklich lediglich mit genügend Rechenleistung zu simulieren ist. Der umstrittene Forscher Ray Kurzweil sieht die Möglichkeit einer künstlichen Intelligenz, welche die menschlichen Fähigkeiten überholt hat, jedoch schon in naher Zukunft [52].
Neuralink und OpenAI CEO Elon Musk sehen die Chancen, dass die Bemühungen zur Abwendung der Gefahr einer Superintelligenz für den Menschen, bei fünf bis zehn Prozent liegen [53]. Plattformen wie FutoreOfLife.org haben in Anlehnung an Asimov und seine Robotergesetze große Seminare zur Findung von allgemeingültigen Prinzipien für Künstliche Intelligenz organisiert [54]. Die entstandenen 23 Regeln haben eine riesige Unterzeichnerschaft namhafter Wissenschaftler und Persönlichkeiten. Ein einfaches Hirngespinst ist es also nicht. Wie real die Gefahr aber wirklich ist, bleibt abzuwarten.

3.1.2 Gefahr durch fehlerhafte KI

Während der Terminator, zumindest für die nächsten Jahre, noch kein Thema sein wird, ist ein anderes Problem viel näher an der Realität: Die Gefahr durch Fehler bei der Entwicklung/Implementierung einer künstlichen Intelligenz. Hierbei entstehen Probleme dadurch, dass die künstliche Intelligenz nicht gut genug entwickelt/trainiert wurde [55].
Beispiele für dieses Problem sind leicht zu finden. Beispielsweise zeigte Professor Christian Bauckhage vom Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme (IAIS) einen Anwendungsfall, in dem eine KI Bilder erkennen und kategorisieren sollte. Diese wurde als neuronales Netz implementiert und mit bereits klassifizierten Bildern aus einer Datenbank trainiert. Beim Test mit ungelabelten Darstellungen stellte sich heraus, dass die KI einen Husky als Wolf qualifiziert hat [56]. Nachforschungen ergaben, dass dies nicht geschah, weil sich Wölfe und Huskys vielleicht ähnlich sehen, sondern lediglich, weil bei allen Trainingsbildern, auf denen Wölfe zu sehen waren, gleichzeitig auch Schnee zu sehen war (siehe Abbildung 4).

Abbildung 4: Die entscheidenden Merkmale, die ein neuronales Netzwerk dazu brachten einen Husky als Wolf zu klassifizieren [56].

Die Auswirkung von Trainingsdaten auf die Ergebnisse ist sehr deutlich am Beispiel von „Norman“ [57] zu sehen. Diese künstliche Intelligenz wurde von Forschern vom MIT ausschließlich mit Bildern von sterbenden Menschen trainiert und danach dem

Rohrschachtest unterzogen. Dieser psychologische Test, bei dem man Bilder, die entstehen, wenn man ein Papier mit Tinte bekleckert und dann faltet, interpretieren muss, hat das Ziel die Persönlichkeit der Testperson zu erfassen (siehe [58]). Im Fall von „Norman“ deutete dieser die Bilder ausschließlich negativ, er sah zum Beispiel einen „Mann, der vor seiner schreienden Frau erschossen wurde“, während eine vergleichbare AI mit anderen Trainingsdaten eine „Person, die einen Regenschirm in die Luft hält“ sieht.
Fälle von fehlenden oder zu einseitigen Daten zum Trainieren von künstlichen Intelligenzen werden von Forschern als eine der größten Schwächen von aktuellen neuronalen Netzen gesehen. Interessanterweise handelt sich es hier aber um ein Problem, mit dem auch die menschliche Persönlichkeit zu kämpfen hat. Das Umfeld, welches einen Charakter umgibt, prägt ihn. So werden Menschen, die sich mit beispielsweise rassistischem Gedankengut beschäftigen, ebenso davon geprägt wie Microsofts Chatbot Tay. Die KI Tay, die auf Twitter gestartet wurde, entwickelte sich innerhalb von 24 Stunden zu einem rassistischen und sexistischen Gesprächspartner. Die User hatten gezielte Tweets auf die KI gesendet, welche wiederum diese Daten verwendete, um sich selbst zu trainieren. Das Programm musste nach nur einem Tag wieder offline genommen werden. Ein weiterer Fall war der Tod einer Passantin durch ein selbstfahrendes UBER Auto in den USA [59]. Das Problem besteht darin, dass es nicht möglich ist, KI isoliert zu testen. Es existieren weder genug Daten noch eine geeignete Testumgebung, um jeden möglichen Fall durchspielen zu können. Somit bleibt den Forschern nichts anderes übrig als Tests in der realen Welt durchzuführen und mögliche Fehler so gut wie möglich abzufangen. Im Falle des selbstfahrenden UBER Autos geschah dies durch einen Notfallfahrer, der im Problemfall eingreifen sollte.
Oftmals ist die Software hinter der KI lediglich eine proprietäre Entwicklung des Herstellers. Dies führt zu einer ganz neuen, ethischen Frage: Wer trägt die Verantwortung bei einem Fehler einer künstlichen Intelligenz? Diese Problematik soll im nächsten Kapitel thematisiert werden.

3.2 Eine Frage der Verantwortung?

In einem Artikel der Initiative Algorithmwatch [60] wird die Einführung von künstlicher Intelligenz in allen Lebensbereichen mit der Kommerzialisierung des Autos als Transportmittels in den 1920er Jahren verglichen. Die Bevölkerung wehrte sich vehement gegen die plötzliche Unsicherheit auf den Straßen. In den Städten wurden Unfallwracks mit „blutigen Schaufensterpuppen und dem Teufel am Steuer“ nachgestellt.

Das Landesgericht Georgia debattierte über den moralischen Charakter des Autos und stufte sie schlussendlich als „gefährliche Wildtiere“ ein.
Was für uns etwas lächerlich erscheint, ist die Folge einer Problematik, die sich auch sehr gut auf die Diskussion über künstliche Intelligenz übertragen lässt. Unser menschliches Wesen führt dazu, dass wir leicht eine Beziehung zu Mitmenschen, aber auch zu Tieren, Gegenstände, Maschinen und so weiter aufbauen können. Wir vermenschlichen Eigentum, reden mit dem Computer, schreien den Fernseher an oder flehen das Auto an, die letzten Kilometer bis zur Tankstelle noch durchzuhalten. Dieses Phänomen führt, gerade im Bereich künstlicher Intelligenz zu einer Verantwortungsfrage. Genau wie ein Mensch, kann auch eine künstliche Intelligenz Fehler machen, falsch handeln. Ist ein „intelligentes“, selbstfahrendes Auto zur Verantwortung zu ziehen, wenn es einen Unfall verursacht? Ist der Facebook Feed-Algorithmus daran schuld, dass man plötzlich rechtsradikale Propaganda auf seiner Facebook Startseite sieht?
Auch wenn sich diese Fragen zuerst lächerlich anhören, offenbaren sie doch ein zentrales Problem. Gerechtigkeit ist eine zentrale Komponente unserer Demokratie und tief im menschlichen Denken verankert. Der Schuldige soll bestraft werden. Beim Opfer-Täter-Täter Ausgleich hat das Opfer die Möglichkeit, den Täter nachzuvollziehen, verstehen zu können und den Konflikt, mitsamt eventuellem Schmerz, klären und loslassen zu können. Laut dem Psychologen Michael McCullough fördert die Vergebung einer Straftat die soziale Beziehungsfähigkeit und Vermeidung krankhafter Feindseligkeit [61].
Dies kann nur passieren, wenn klar ist, wer der Schuldige ist. Gabriel Hallevy, ein Forscher am Ono Academic College in Israel hat drei mögliche Szenarios entwickelt, in denen eine künstliche Intelligenz einem Menschen schaden könnte: (die Begriffe sind frei aus dem Englischen übersetzt) [62]

1. **Täterschaft durch einen Dritten**
   Dieses Szenario tritt dann auf, wenn eine Straftat durch eine Maschine verübt wird, die durch eine dritte Person dadurch instruiert wurde. In diesem Fall ist die Maschine unschuldig, die Verantwortung trägt entweder der Entwickler oder der Benutzer. Ein Beispiel wären Schadprogramme wie Malware, aber je nach Gesetzeslage auch autonome Waffensysteme oder, um bei künstlicher Intelligenz zu bleiben, vielleicht sogar ein „bösartig“ trainiertes neuronales Netz (wobei die Frage bleibt, was „bösartig“ im jeweiligen Kontext ist).

2. **Natürliche Konsequenz**
   Dieser Fall tritt auf, wenn ein System eine Aktion zum Erreichen seines Ziels

unangemessen ausführt und dabei eine Straftat begeht. Hallevy zeigt ein Beispiel, in dem ein Produktionsroboter einen Arbeiter tötet, weil er diesen als Gefahr für seine Aufgabe betrachtet. Hier liegt die Hauptverantwortung beim Entwickler der Maschine, wenn festgestellt werden kann, dass dieser dieses mögliche Szenario vorhersehen und verhindern hätte können.

3. **Unmittelbare Haftung**
   Die dritte Möglichkeit tritt dann auf, wenn sowohl eine Handlung, als auch ein Vorsatz für eine Straftat vorhanden sind. Ein selbstfahrendes Auto, das die erlaubte Geschwindigkeit überschreitet, könnte also für die Straftat verantwortlich befunden werden. Bereits heute existieren Fälle, in denen Nutzer, die z.B. für Delikte im Bereich Hacking beschuldigt wurden, erfolgreich argumentieren konnten, dass ihre durch Malware gekaperten Geräte die Straftaten begangen haben.

Die Tatsache, dass Maschinen Verantwortung für ihre Taten tragen könnten, bringt viele weitere Fragen mit sich. Dürften Maschinen sich verteidigen und beispielsweise ihre Taten als Folge eines Virenbefalls begründen? Wie sieht die Bestrafung einer Maschine aus? Findet eine Umprogrammierung statt? Im Falle eines neuronalen Netzes vielleicht ein Neutraining? Was passiert mit anderen Systemen der gleichen Produktreihe?
Weiterhin ist unklar, wie ein solches Vergehen juristisch geahndet werden soll. Eventuell entfällt die strafrechtliche Haftung, müsste das Vergehen nach dem Zivilrecht behandelt werden. Handelt es sich bei einem KI-System um ein Produkt, so gelten eventuelle Garantiebestimmungen. Handelt es sich aber um eine Dienstleistung, gilt das Delikt der Fahrlässigkeit. Hierbei müssen drei Elemente nachgewiesen werden, das Vorhandensein einer Sorgfaltspflicht, der Bruch dieser und die Entstehung eines Schadens für den Kläger.
Egal, wie die Entwicklung von künstlicher Intelligenz in den nächsten Jahren weitergeht, sicher ist, dass die Frage nach der Verantwortung definitiv ein Hauptthema bei der Diskussion dieser Thematik in der Öffentlichkeit, Wissenschaft und Politik darstellen wird.

3.3 Künstliche Intelligenz in allen Lebensbereichen

In Kapitel 2.2.3 wurde bereits ein kurzer Überblick über die heutige Verwendung von künstlicher Intelligenz gegeben. Hier sollen zusätzlich zwei spezielle Problemgebiete vorgestellt und diskutiert werden, an denen deutlich wird, wie wichtig es ist, Gesellschaft und Informatik zu vernetzen. Beide sind erschreckend, besonders wenn man

berücksichtigt, dass beide nicht nur heute, sondern bereits seit Jahrzehnten Realität sind. Dabei sind sie wohl keineswegs so medienpräsent wie das Horrorbild vom Terminator, obgleich sie, insbesondere für Menschen in sozial schwachen Stellungen, ähnliche Auswirkungen haben können.

3.3.1 SCHUFA Scoring

Das Wort SCHUFA bedeutet „Schutzgemeinschaft für allgemeine Kreditsicherung“. Sie bezeichnet sich selbst als „führender Informations- und Servicepartner für die kreditgebende Wirtschaft und Privatkunden“ [63].
Die Firma sieht ihr Kerngeschäft darin, „kreditrelevante Informationen zu Privatpersonen und Unternehmen bereitzustellen“. Damit ist gemeint, dass die SCHUFA Zahlungsdaten sowie persönliche Daten zu Personen von zusammenarbeitenden Unternehmen erhält, aufbereitet und verarbeitet und schließlich wieder an anfragende Firmen, in Form eines sogenannten „Scores“ verteilt. Dieser Score soll die Bonität, also den Ruf im Hinblick auf Zahlungsfähigkeit und Kreditwürdigkeit der Person ausdrücken.
Beispielsweise kann eine Bank mithilfe dieses Scores bewerten, wie wahrscheinlich es ist, dass Person X den beantragten Kredit innerhalb der nächsten Jahre umstandslos zurückzahlen kann. Damit sollen „sichere, schnelle und effiziente Geschäftsabschlüsse“ gemacht, sowie „Unternehmen vor Zahlungsausfällen und Konsumenten vor einer möglichen Überschuldung durch Konsumentenkredite“ geschützt werden [64].
Immer wieder tauchen jedoch Berichte auf, welche die von der SCHUFA betriebene Art von Datensammlung und Aufbereitung kritisieren. Ob es darum geht, dass man sogar ohne irgendwelche vorherige Einträge einen negativen Bonitäts-Score bekommt [65], die langfristige Speicherung von Privatinsolvenzen, auch nach deren Ablauf [66], die Verwendung von Internetdaten zu Personen, unter anderem durch beispielsweise Facebook-Daten [67] oder wie das Festhalten der SCHUFA an ihrem Abo-Modell im Gegensatz zu dem in der DSGVO geforderten elektronischen Auskunftsrecht steht [68].
Einen besonderer Streitpunkt stellt der geheime Scoring Algorithmus da. Es ist unklar, welche Daten mit welcher Gewichtung genau in diesem Algorithmus einfließen. Es ist möglich, dass fundamentale Falschannahmen getroffen werden, grundsätzliche Fehler in dem Algorithmus vorliegen - und das von niemandem überprüft werden kann, da das Verfahren als „Geschäftsgeheimnis“ gilt und nicht öffentlich einsehbar ist.
Die Kampagne OpenSCHUFA der Plattformen AlgorithmWatch und Open Knowledge Foundation Deutschland will nun mithilfe der Methode des „Reverse Engineering“ diesen Scoring-Algorithmus weitestgehend offenlegen [69]. Zum Erfolg der Aktion sind

Datenspenden von SCHUFA Auskünften nötig. Diese kann jede, in Deutschland lebende Person, einmal jährlich kostenlos bei der SCHUFA beantragen. Über die Plattform selbstauskunft.net kann der ganze Beantragungsprozess vereinfacht und danach als Scan auf www.openschufa.de/steps hochgeladen werden.
An diesem Beispiel wird deutlich, welche Verantwortung Informatiker für heutige gesellschaftliche Probleme besitzen. Im Bezug auf den Scoring-Algorithmus der SCHUFA kursieren viele Meinungen im Netz. Niemand weiß wirklich was dahinter steht, ein Problem, dessen Ausmaß sich erst dann wirklich entfaltet, wenn man bedenkt, dass die SCHUFA als privates Unternehmen Daten zu über 67,5 Millionen Menschen in Deutschland verwaltet und völlig unklar ist, wie mit diesen Daten eine Entscheidung über die Bonität einer Person getroffen wird. Im Sinne eines transparenten Datenumgangs wünschen wir OpenSCHUFA den größtmöglichen Erfolg.

3.3.2 KI in der Rechtssprechung

Gerechtigkeit ist tief im Menschen verankert. Wir fühlen uns bedrückt, wenn wir daran denken, was Andere ungerechtfertigterweise erleiden müssen. Ein Unschuldiger, der verurteilt wird, bereitet uns Kummer. Gleichzeitig wünschen wir uns, dass ein Schuldiger seine gerechte Strafe erhält.
Was Recht und Unrecht ist, wird in Gesetzen festgehalten. Richter und Anwälte wenden diese in Gerichten an und legen je nach Straftat passende Strafen zurecht. Dies sollte nicht anhand von einem willkürlichen Bauchgefühl, sondern anhand nachvollziehbarer Fakten und Erklärungen geschehen. Soweit zum Idealfall. In der Realität sieht das natürlich nicht immer so aus. Menschen sind keine deterministischen Automaten, wir verhalten uns nicht nach einer genauen Abfolge. Ebenso wenig ist es utopisch von Richtern zu erwarten, dass sie völlig unabhängig von Laune, Prägung und Umfeld entscheiden.
Künstliche Intelligenz scheint dieses Problem wunderbar zu lösen. Was bietet sich mehr an als eine Maschine, die keine Seite bevorzugt oder benachteiligt, völlig fair Gerechtigkeit schaffen zu lassen.
Ein Wunschtraum? Nicht ganz. Von der „Richter-KI“, sind wir wahrscheinlich noch etwas entfernt, aber Algorithmen zur Bewertung der Rückfallwahrscheinlichkeit von Straftätern existieren schon seit Jahrzehnten und werden beispielsweise in den USA auch bereits eingesetzt. Die Idee dabei ist, mithilfe von Daten und Selbstaussagen des Täters einen Score zu ermitteln, der den Richtern bei der Entscheidungsfindung helfen soll. Somit sollen mögliche menschliche Vorurteile verringert werden und eine fairere

Rechtsprechung möglich sein.
Die Anforderung an solche Systeme sind natürlich enorm. Auf der einen Seite sollen möglichst viele Daten zu einem möglichst kleinen, überschaubaren Ergebnis zusammengefasst werden. Auf der anderen Seite ist zu beachten, dass Menschen sich viel zu leicht von Bewertungen beeinflussen lassen und lieber Verantwortung abgeben als selbst zu übernehmen. Wären die Systeme fehlerfrei, ist diese Abgabe der Verantwortung vielleicht in Ordnung. Am Beispiel der Software COMPAS, die seit Beginn der 2000er Jahre eingesetzt wird, ist aber ersichtlich, dass die Algorithmen keinesfalls immer das machen, was sie sollen.
Im Folgenden beziehen wir uns auf die den Artikel zur Studie der Organisation ProPublica [70]. COMPAS ist eine Entwicklung der Firma Northpointe, welche 1989 von Wissenschaftlern aus dem Bereich Statistik und Mitarbeitern eines Gefängnisses gegründet wurde. Obwohl die Software anfangs nur für interne Analysen der Gefangenen gedacht war, wurde sie unter anderen im Bundesstaat New York schon 2002 als Pilotprojekt und seit 2010 als fester Bestandteil der Rechtsprechung eingesetzt. Auch der Staat Wisconsin setzte auf das Programm, ebenso wie zahlreiche kleine Countys (Bezirke). In Broward County, Florida, wurden beispielsweise mithilfe der Software Strafgefangene bewertet und bei niedrigem Risikowert vorzeitig entlassen, um der dauerhaften Überfüllung in den Gefängnissen entgegenzuwirken. Das Programm basiert auf der Auswertung von 137 Fragen, welche von den Angeklagten beantwortet werden müssen. Diese Fragen reichen vom persönlichen Umfeld, zum Beispiel „Wie viele deiner Freunde konsumieren illegal Drogen? “ bis zur Bewertung von ethisch-moralischen Statements, beispielsweise „Eine hungrige Person hat das Recht zu stehlen.“ Weiterhin werden Daten wie Geschlecht, Alter und Vorstrafen ermittelt. Wie genau sich aus all diesen Werten der Score für die Rückfallwahrscheinlichkeit zusammensetzt, ist nicht öffentlich bekannt.
Während der Nutzungszeit von COMPAS wurden mehrere Studien zur Bewertung der Genauigkeit des Programms durchgeführt, unter anderem von Northpointe selbst, der Florida State University, sowie des Bundesstaats New York. Die Ergebnisse sagten dem Algorithmus eine Genauigkeit in der Angabe, ob ein Angeklagter erneut straffällig wird, von ungefähr 65 Prozent zu. Dabei kam es nur zu minimalen Abweichungen zwischen Menschen verschiedener Hautfarben im Bereich von wenigen Prozent. Die Firma selbst behauptete in seiner Studie, die Rückfallwahrscheinlichkeit von Menschen mit weißer Hautfarbe wurden vom System um zwei Prozentpunkte besser erkannt als bei Menschen mit schwarzer Hautfarbe.

Zur Überprüfung dieser Angaben führte ProPublica, eine non-profit Organisation für investigativen Journalismus mit Sitz in New York, zwischen 2013 und 2014 eine Studie an, bei der die Bewertung durch COMPAS von ca. 7000 Menschen aus Broward County, mit dem tatsächlichen Eintreten einer Straftat innerhalb zwei Jahre nach der Verhandlung, verglichen wurde. Die Studie kam zu dem Ergebnis, dass schwarze Angeklagte doppelt so häufig wie weiße Angeklagte fälschlich als Risiko für das Begehen von erneuten Straftaten bewertet wurden. Weiterhin wurden weiße Straftäter deutlich häufiger, im Vergleich zu schwarzen Angeklagten, als niedriges Risiko für erneute Straffälligkeit eingestuft. Die genauen Einstufungen von COMPAS, gegliedert nach Hautfarbe, sind in den Abbildungen 5 und 6 zu sehen.

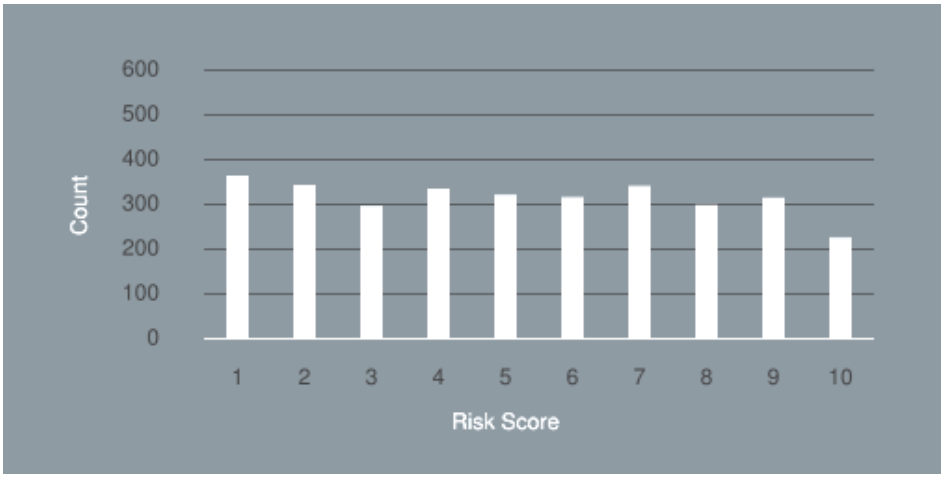

Abbildung 5: Die Anzahl der untersuchten Straftäter mit schwarzer Hautfarbe, gegliedert nach ihrem COMPAS-Score [70].

In der Quelle finden sich Gegenüberstellungen von einzelnen Personen unterschiedlicher Hautfarbe, die ähnliche Straftaten begangen haben. Teils zeigen sich wirklich obskure Konstellationen, wie die von Vernon Prater und Brisha Borden, die beide wegen geringfügigen Diebstahls angeklagt wurden. Prater, ein weißer Mann mittleren Alters, der als Vorstrafe bereits zwei bewaffnete Raubüberfälle und einen weiteren bewaffneten Raubüberfall besaß, wurde mit einer Rückfallwahrscheinlichkeit drei von zehn, also recht gering, bewertet. Borden, eine 18-jährige, schwarze Frau, die im Vorstrafenregister lediglich vier mal jugendliches Fehlverhalten stehen hatte, bekam von COMPAS eine acht von zehn, also deutlich erhöhte Rückfallwahrscheinlichkeit zugewiesen. Prater beging nach der Verhandlung einen weiteren schweren Diebstahl, Borden wurde in den nächsten zwei Jahren wegen keiner weiteren Straftat angeklagt.
Die Relevanz des Themas sollte klar sein. Gerade bei der Entscheidung über das weitere

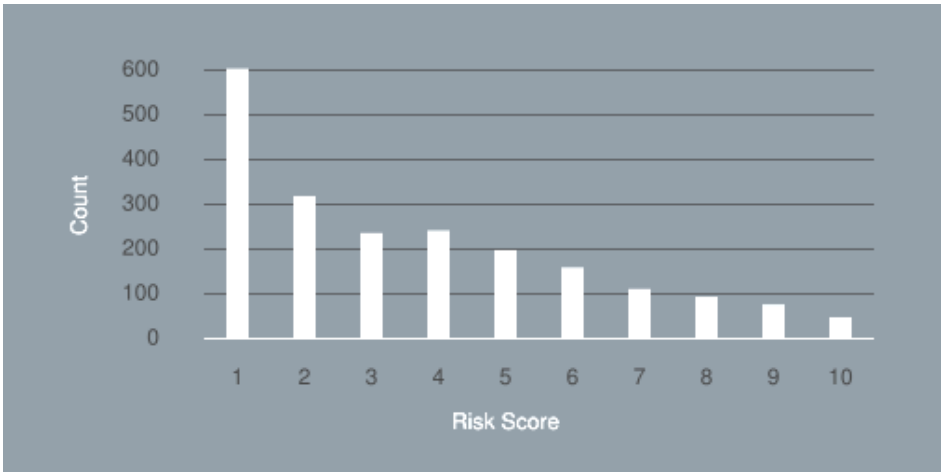

Abbildung 6: Die Anzahl der untersuchten Straftäter mit weißer Hautfarbe, gegliedert nach ihrem COMPAS-Score [70].

Schicksal eines Menschenlebens stellt sich die Frage, wie sinnvoll unbegründete Wertungen durch Algorithmen sind. Sicherlich liegt die letzte Entscheidung beim Richter, doch wird sich zeigen müssen, ob Menschen im Zweifelsfall nicht doch lieber auf einen vermeintlich „allwissenden“ Score als auf ihre eigene Intuition vertrauen. Mittlerweile werden ähnliche Systeme wie COMPAS in mehr als zehn US-Staaten eingesetzt. Letzteres wird im Übrigen trotz aller Anschuldigen noch heute eingesetzt und angeboten.

4 Schlussteil

Der Einsatz von KI entspricht in etwa dem „High Risk, High Reward“ Prinzip. Künstliche Intelligenz bietet unvorstellbare Möglichkeiten, kann uns zur Weiterentwicklung dienen, kann Probleme im Gesundheits-, Umwelt- und Wissenschaftsbereich lösen, an denen Menschen seit Jahrtausenden scheitern. Eine Aussage von Stephen Hawking ist: „KI wird entweder das Beste sein, was der Menschheit jemals widerfahren ist oder das Schlimmste“ [71].
Betrachtet man die Sachlage anhand dieses Zitats, kommt man dazu, KI als letztes Glied in einer Folge bahnbrechender Entwicklungen in unserer Geschichte zu sehen. Erfindungen wie die Demokratie, der Buchdruck, die Dampfmaschine oder die Kernspaltung haben unsere Welt und Gesellschaft radikal verändert. Zum Guten als auch zum Schlechten.
Was wir aus der Vergangenheit gelernt haben sollten, ist, dass sich Fortschritt nicht aufhalten lässt. Künstliche Intelligenz wird alle unsere Lebensbereiche durchdringen,

daran ist kein Zweifel. Die Frage ist, wie wir damit umgehen. Haben wir Angst? Sind wir naiv? Wo unterscheiden sich Mensch und Maschine?
Eine Antwort hierauf ist nicht leicht zu geben. Tatsächlich sind wir der Meinung, dass Prognosen und Vermutungen zwar spannend, aber leider zu oft völlig überzogen sind. Eben wie schon Thomas Watson, der 1943 behauptete: „Ich denke, dass es weltweit einen Markt für vielleicht fünf Computer gibt“.
Die Tatsache, dass es schwer ist, Prognosen über die Zukunft zu treffen, sollte uns aber nicht davon abhalten, darüber nachzudenken. Trotz aller Unsicherheit ist definitiv klar, dass Digitalisierung und Technologisierung in der Zukunft eine immer weiter zunehmende Rolle spielen werden. Dabei verschwimmen die Grenzen zwischen Realität und Vorstellung immer weiter. Oder, wie es das bekannte, dritte Clarke'sche Gesetz besagt: „Any sufficiently advanced technology is indistinguishable from magic.“ [72]
Informatik wird eine neue Bedeutung bekommen. Der Fokus muss wegkommen von der Technologie, wieder zurück zum Menschen. Informatiker haben die Möglichkeit ein Bindeglied zwischen der digitalisierten, kalten Welt der Nullen und Einsen hin zu der Gesellschaft, zu deren Nutzen sie gemacht ist, zu werden.
Im Zuge der Industrialisierung sagte der Philosoph Henry David Thoreau: „Siehe da! Die Menschen sind die Werkzeuge ihrer Werkzeuge geworden.“ Es liegt nicht in unserer Hand, zu interpretieren, ob er damals damit Recht hatte.
Es liegt aber in unserer Hand, zu bestimmen, ob er es heute hat.

Literatur

[1] Bruce G. Buchanan. A (very) brief history of artificial intelligence. *AI Magazine*, Volume 26, Number 4, pages 53-60, 2005.

[2] Vannevar Bush. As we may think. *The Atlantic Monthly*, July 1945.

[3] A. M. Turing. Computing machinery and intelligence. *Mind*, New Series, 59(236):433-460, October 1950. http://cogsci.umn.edu/millennium/final.html.

[4] P. McCorduck, M. Minsky, O. Selfridge, and H. A. Simon. History of artificial intelligence. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence* - Volume 2, IJCAI'77, pages 951–954, San Francisco, CA, USA, 1977. Morgan Kaufmann Publishers Inc.

[5] James Moor. The dartmouth college artificial intelligence conference: The next fifty years, *AI Magazine*, Volume 27, Number 4, pages 87-91, 2006.

[6] Rockwell Anyoha. The history of artificial intelligence. August 2017. http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/ [Online; abgerufen 14. Juni 2018].

[7] Daniel Crevier. *AI: The Tumultuous History Of The Search For Artificial Intelligence*. Basic Books, 1993.

[8] Hans Moravec. *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press, 1990.

[9] Gary Yang. *The history of artificial intelligence: AI winter and its lessons*. December 2006. [Online; abgerufen 14. Juni 2018].

[10] Automatic Language Processing Advisory Committee (ALPAC). *Language and Machines: Computers in Translations and Linguistics*, 1966.

[11] Francesco Corea. *A brief history of artificial intelligence*. April 2017. https://medium.com/@Francesco_AI/a-brief-history-of-ai-baf0f362f5d6 [Online; abgerufen 14. Juni 2018].

[12] IBM Corporation. *Deep blue*. http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/ [Online; abgerufen 14. Juni 2018].

[13] Steve Lohr. IBM is counting on its bet on Watson, and paying big money for IT. *New York Times*, October 2016. https://www.nytimes.com/2016/10/17/technology/ibm-is-counting-on-its-bet-on-watson-and-paying-big-money-for-it.html [Online; abgerufen 14. Juni 2018].

[14] Oliver Walter. *Was ist Intelligenz?, 2004-2011*. [Online; abgerufen 14. Juni 2018].

[15] Bild.de. Killer-Roboter und Cyber-Attacken: Sieht so das Ende der Welt aus? *Bild*, September 2013. [Online; abgerufen 10. Juni 2018].

[16] Trevor Mogg. Whatever you do, dont mess with boston dynamics spotmini robot. *Fox News*, February 2018. https://www.digitaltrends.com/cool-tech/boston-dynamics-spotmini-disturbance-test/ [Online; abgerufen 10. Juni 2018].

[17] Bonnie Docherty. Were running out of time to stop killer robot weapons. *The Guardian*, April 2018.
https://www.theguardian.com/commentisfree/2018/apr/11/killer-robot-weapons-autonomous-ai-warfare-un [Online; abgerufen 10. Juni 2018].

[18] Egor Dezhic. Understanding decision trees. 2017.
https://becominghuman.ai/understanding-decision-trees-43032111380f
[Online; abgerufen 09. Juni 2018].

[19] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, March 1986.

[20] Liu Yuxun and Xie Niuniu. Improved ID3 algorithm. In *2010 3rd International Conference on Computer Science and Information Technology*. IEEE, July 2010.

[21] IBM Corporation. *Usage of decision trees.*
https://www.ibm.com/support/knowledgecenter/en/SS6NHC/com.ibm.swg.im.dashdb.analytics.doc/doc/r_decision_trees_usage.html
[Online; abgerufen 08. Juni 2018].

[22] Igor Kononenko. Inductive and bayesian learning in medical diagnosis. *Applied Artificial Intelligence*, 7:317–337, 1993.

[23] Lilly Spirkovska. Three-dimensional object recognition using similar triangles and decision trees. *Pattern Recognition*, 26(5):727–732, May 1993.

[24] Timo Baumann. Decision tree usage for incremental parametric speech synthesis. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, May 2014.

[25] Neha Yadav, Anupam Yadav, and Manoj Kumar. *History of Neural Networks*, pages 13–15. Springer Netherlands, Dordrecht, 2015.

[26] Andrea Trinkwalder. Netzgespinste: Die Mathematik neuronaler Netze: Einfache Mechanismen, komplexe Konstruktion. *Heise c't*, 06/2016.
https://www.heise.de/ct/ausgabe/2016-6-Die-Mathematik-neuronaler-Netze-einfache-Mechanismen-komplexe-Konstruktion-3120565.html
[Online; abgerufen 10. Juni 2018].

[27] R. Rojas. *Neural Networks: A Systematic Introduction*, chapter 3, pages 56,57. Springer-Verlag, Berlin, 1996.

[28] Michael Nielsen. Neural networks and deep learning, 2015. http://neuralnetworksanddeeplearning.com [Online; abgerufen 29. Dezember 2017].

[29] A. Gibson and J. Patterson. *Deep Learning*. O'Reilly Media, Inc., August 2017.

[30] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[31] F. Rosenblatt. *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Report (Cornell Aeronautical Laboratory). Spartan Books, 1962.

[32] Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson. Understanding neural networks through deep visualization. In *Proceedings of the Deep Learning Workshop, 32st International Conference on Machine Learning*, 2015.

[33] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 818–833, Cham, 2014. Springer International Publishing.

[34] N. J. Freymuth. Grundlagen neuronaler Netze - Eine Einführung, März 2018. Abschlussarbeit für das Proseminar "Gestaltung und Durchführung von Fachvorträgen in der Informatik".

[35] Britgo.org. Google deepmind challenge match - Lee Sedol v AlphaGo - match report, 2016. [Online; abgerufen 08. Juni 2018].

[36] matburt anoek. Online-Go Server (OGS). [Online; abgerufen 27.02.2018].

[37] John Tromp and Gunnar Farnebäck. Combinatorics of Go. *Computers and Games: 5th International Conference, CG 2006, Turin, Italy, May 29-31, 2006. Revised Papers*, pages 84–99, 2007.

[38] David Silver, Aja Huang, Christopher J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal

Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529:484–503, 2016.

[39] Laurent Heiser. AlphaGo vs. Lee Sedol. *DGOZ Deutsche Go-Zeitung*, Mai 2016.

[40] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, December 2015.

[41] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, June 2015.

[42] Kiuk Chung. *Generating recommendations at Amazon scale with Apache Spark and Amazon DSSTNE*. 2016.
https://de.slideshare.net/HadoopSummit/generating-recommendations-at-amazon-scale-with-apache-spark-and-amazon-dsstne
[Online; abgerufen 08. Juni 2018].

[43] Yaniv Taigman, Ming Yang, and Lior Wolf. DeepFace: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[44] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.*, 999888:2493–2537, November 2011.

[45] Chintan Turakhia. *Engineering more reliable transportation with machine learning and AI at Uber*. 2017. https://eng.uber.com/machine-learning/ [Online; abgerufen 08. Juni 2018].

[46] Sophia Ciocca. *How does Spotify know you so well?*. October 2017.
https://medium.com/s/story/spotifys-discover-weekly-how-machine-learning-finds-your-new-music-19a41ab76efe [Online; abgerufen 08. Juni 2018].

[47] Julian Moeser. *Starke KI, schwache KI - was kann künstliche Intelligenz?*. 2017. https://jaai.de/starke-ki-schwache-ki-was-kann-kuenstliche-intelligenz-261/ [Online; abgerufen 08. Juni 2018].

[48] Wolfgang Zehentmeier. *Sind Androiden menschlicher als der Mensch?* 2018. https://www.br.de/themen/ratgeber/inhalt/computer/detroit-become-human-test-100.html [Online; abgerufen 03. August 2018].

[49] Stephan Dörner. Superintelligenz: Diese kommende Erfindung könnte das Ende der Menschheit bedeuten. *T3N*, Januar 2017. https://t3n.de/news/superintelligenz-ki-ai-787316/ [Online; abgerufen 03. August 2018].

[50] Sebastian Scholtysek. Die Robotergesetze von Isaac Asimov. *Roboterwelt*, Februar 2015. http://www.roboterwelt.de/magazin/die-robotergesetze-von-isaac-asimov/ [Online; abgerufen 03. August 2018].

[51] Gordon E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8):114ff, April 1965

[52] Ray Kurzweil. Immortality by 2045. http://2045.com

[53] Bryan Clark. Elon Musk basically confirms AI is coming to eradicate the human race. https://thenextweb.com/artificial-intelligence/2017/11/22/elon-musk-basically-confirms-ai-is-coming-to-eradicate-the-human-race/

[54] Future of Life Institute. *Asilomar AI principles*. 2017. https://futureoflife.org/ai-principles/ [Online; abgerufen 03. August 2018].

[55] Oliver Voß. So dumm ist künstliche Intelligenz. *Tagesspiegel*, November 2017. https://www.tagesspiegel.de/politik/wenn-der-algorithmus-versagt-so-dumm-ist-kuenstliche-intelligenz/20602294.html

[56] Martin Fischer. Künstliche Intelligenz als Gefahr: Menschheit muss sich auf Regeln einigen. *Heise online*, 13. Juni 2018. https://www.heise.de/newsticker/meldung/Kuenstliche-Intelligenz-als-Gefahr-Menschheit-muss-sich-auf-Regeln-einigen-4077950.html

[57] Pinar Yanardag, Manuel Cebrian, and Iyad Rahwan. World's first psychopath AI. http://norman-ai.mit.edu [Online; abgerufen 03. August 2018].

[58] Wikipedia contributors. Rorschach test.

[59] Daisuke Wakabayashi. Self-driving uber car kills pedestrian in arizona, where robots roam. *New York Times*. March 19 2018. https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html

[60] Lorena Jaume-Palasí. *Warum Sie keine Angst vor künstlicher Intelligenz haben sollten.* 27. April 2018. https://algorithmwatch.org/de/warum-sie-keine-angst-vor-kuenstlicher-intelligenz-haben-sollten/

[61] Michael E. McCullough. Forgiveness as human strength: Theory, measurement and links to well-being. *J Social Clinical Psychology*, pages 43–55, 2000.

[62] Emerging Technology from arXiv. When an AI finally kills someone, who will be responsible? *MIT Technology Review*, March 12 2018. https://www.technologyreview.com/s/610459/when-an-ai-finally-kills-someone-who-will-be-responsible/

[63] Die Schufa - Wir schaffen Vertrauen. https://www.schufa.de/de/

[64] Die Schufa - Unsere Aufgabe in der Wirtschaft. https://www.schufa.de/de/ueber-uns/unternehmen/aufgabe-wirtschaft/

[65] Philipp Seibt. Wie ich bei der Schufa zum "deutlich erhöhten Risiko" wurde. *Spiegel Online*, März 2018. http://www.spiegel.de/wirtschaft/service/schufa-wie-ich-zum-deutlich-erhoehten-risiko-wurde-a-1193506.html

[66] Kritik an Schufa-Eintrag nach Ende des Insolvenzverfahrens. *Frankfurter Allgemeine Zeitung*, Februar 2018. https://www.welt.de/newsticker/dpa_nt/infoline_nt/wirtschaft_nt/article173174850/Kritik-an-Schufa-Eintrag-nach-Ende-des-Insolvenzverfahrens.html

[67] Benedikt Fuest and Claudia Ehrenstein. Wenn dir die Schufa auf die Urlaubsfotos schaut. *Die Welt*, Juni 2016. https://www.welt.de/wirtschaft/webwelt/article106436481/Wenn-dir-die-Schufa-auf-die-Urlaubsfotos-schaut.html

[68] Benedikt Fuest. Neue Datenschutzregeln bedroht Abo-Modell der Schufa. *Die Welt*, Juni 2018. https://www.welt.de/finanzen/article177303132/DSGVO-stellt-das-Abo-Modell-der-Schufa-infrage.html

[69] AlgorithmWatch. *Openschufa warum wir diese Kampagne machen.* https://algorithmwatch.org/de/

[70] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. *MachineBbias.*. May 23 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[71] Oliver Voss and Laura Weigele. Wie gefährlich ist künstliche Intelligenz? *Tagesspiegel*, 23. November 2017. https://www.tagesspiegel.de/politik/killerroboter-und-co-wie-gefaehrlich-ist-kuenstliche-intelligenz/20602292.html

[72] Arthur C. Clarke. *Profiles of the Future: An Inquiry into the Limits of the Possible.* Gateway Publisher, 1973.