universität wien

Bakkalaureatsarbeit

Titel

Artificial Unintelligence

Vom falschen Einsatz algorithmischer Systeme in gesellschaftsrelevanten Prozessen und Lösungsansätzen deutschsprachiger Experten

Verfasserin

Stefanie Reichsöllner

Angestrebter akademischer Grad

Bakkalaurea der Philosophie (Bakk. phil.)

Wien, im August 2019

Studienkennzahl lt. Studienblatt:	A033 641
Matrikelnummer:	01621241
Studienrichtung lt. Studienblatt:	Publizistik & Kommunikationswissenschaften
Betreuerin:	Mag. Dr. Stefan Weber

Inhalt

1 Entstehungszusammenhang

Im Umfeld unserer mediatisieren Welt wächst meines Erachtens der Wert der Selbstbestimmung und, eng verbunden damit, der Kontrolle über den eigenen Informationsfluss. Meine Bakk1 Arbeit gab mir Einblicke über die Selbstwahrnehmung bezüglich Auswirkungen der „Filterblase“ von Digital Natives. Die nicht repräsentative Studie zeigte, dass der Großteil der Befragten sich über einen stattfindenden Filterprozess im Klaren ist. Ob Google-Suchergebnisse oder der Facebook-Feed, sowie in welchem Ausmaß sie jeweils betroffen sind, darüber herrschte große Ambivalenz. Viele Teilnehmer äußerten negative Gefühle und praktische Auswirkungen wie Meinungsbeeinflussung als mögliche Folgen. Zu beachten ist der Status der Befragten als „Digital Natives“ (Frieling 2010: 27), bei denen ein verantwortungsvoller sowie informierter Umgang mit digitalen Medien jedenfalls erstrebenswert wäre. Auch im Bereich des Datenschutzes (speziell in sozialen Netzwerken) herrscht große Unkenntnis und eher Resignation als Informiertheit (Vgl. Sarikakis, Winter 2017:1). Ich erlebe also einen großen Bedarf an Aufklärung. Unterstützt wird das in der zunehmend computergesteuerten Umwelt von dem Gefühl blinden Vertrauens (oder schlichtem Mangel an Motivation und Information), welches Algorithmen unter steigendem Einsatz scheinbar entgegenbracht wird. Broussard (Vgl. 2018) unternimmt den Versuch zu erläutern, wie Technologie funktioniert und wo folglich ihre Grenzen liegen. Mit ihren Findungen positioniert sie sich gegen „Technochauvinismus“ und erhebt Ansprüche, einhergehend mit Limitationen darüber, was Technologie kann, aber vorallem sollte.

Aufzeigend anhand des Werks „Artificial Unintelligence“ (Broussard 2018) sowie mithilfe von Experteninterviews bin ich bestrebt zu beschreiben, wie Künstliche Intelligenz *nicht* eingesetzt werden sollte, wo aktuelle Probleme und Herausforderungen liegen und welche Lösungsansätze es für das Beschriebene gibt.

2 Einleitung

„Kraft, Ausdauer (kein Feierabend- oder Urlaubsbedürfnis), Präzision, Schnelligkeit, Unbeirrbarkeit (etwa durch Gefühlsschwankungen oder Ablenkungen aller Art), niedrigerer Stundenlohn (auch inklusive aller Wartungskosten) – und keine Vertretung durch Gewerkschaften“ (Kreutzer, Sirrenberg 2019: 47) – Maschinen stechen den Menschen mit diesen Argumenten schon lange in vielen Tätigkeiten aus. Seit es diese mit Künstlicher Intelligenz gibt, gibt es Horrorszenarien, die das qualvolle Ende der Menschheit durch sie prophezeien. Gleichzeitig werden Bilder von einer gerechten und sorglosen Utopie gemalt, in die die Menschheit von KI getragen wird. Steven Hawking nennt es das möglicherweise schlimmste Event in der Geschichte unserer Zivilisation (Vgl. Kharpal 2017).

Wir befinden uns in einer Zeit des großen Hypes (Vgl. Wasner 2019) um Künstliche Intelligenz, gleichzeitig ist das gemeine Bild geprägt von Uninformiertheit oder gar Fehlinformation. Diese Arbeit zielt darauf ab, ein grobes doch realistisches Bild von den aktuellen Fähigkeiten Künstlicher Intelligenz zu zeichnen. In Anlehnung an Broussards (2018) Artificial Untintelligence liegt der Fokus hierbei darauf, wie Anwendungen **nicht** verwendet werden sollen. Es sollen aktuelle Probleme und Herausforderungen beschrieben und anschließend von deutschsprachigen Experten bewertet werden.

Diese Arbeit soll einerseits entmystifizieren. Andererseits soll sie eine Übersicht der Schwere der einzelnen Punkte liefern. Weiters sollen verschiedene Lösungsansätze genannt werden, um eine zukünftige Künstliche Intelligenz, die allen Menschen dient zu fördern.

3 Was ist „Künstliche Intelligenz (KI)“

3.1 Begriffsabgrenzung

Bevor sich der Negation des Begriffs, „Künstliche *Un*intelligenz“, angenähert werden kann, muss der ursprüngliche Begriff definiert werden. Für den gängigen Terminus, meist verwendet im Englischen als *Artificial Intelligence (AI),* gibt es keine einheitliche Definition. Das rührt zum einen daher, dass bereits *Intelligenz* nicht eindeutig definierbar ist. Negnevitsky leitet für sich die Beschreibung als die Fähigkeit zu Lernen und zu Verstehen, Probleme zu lösen und Entscheidungen zu treffen ab (Vgl. 2005). Weiter Versuche beinhalten Ähnliches, im Kern ist man sich aber einig um die Grundfähigkeiten gegenüber der Außenwelt „[to] perceive, understand, predict, and manipulate“ (Russell, Norvig 2014: 1).

Wie im Weiteren Teil der Arbeit ersichtlich wird, führt eine Vielzahl irreführender Begriffe zu teils großem Missverständnis. Laien gehen etwa davon aus, dass KI eine *simulierte, denkende Person* (Vgl. Broussard 2018: 32) in einer Maschine ist. Die historische Streitigkeit um die Begriffsdefinition hilft, im Weiteren die oft falsche Perzeption des Begriffs zu verstehen.

> *Durch Nutzung physikal. oder chem. Prozesse künstlich oder technisch hergestellte Erkenntnis- und Denkfähigkeit. Dabei bestehen zwei Zielrichtungen: 1) die Simulation und Untersuchung natürl. Intelligenz von Menschen und Tieren, v. a. mittels Computer; 2) die Entwicklung maschineller Intelligenz als Ergebnis der Anwendung von Sensoren und Programmen (Steuerungen) für Maschinen, die intelligente Arbeit verrichten sollen. (Der Neue Brockhaus 1985: 288)*

Diese Definition aus dem Jahr 1985 würde man aufgrund der fehlleitenden Beschreibung heute kaum mehr so veröffentlichen. Denn den Begriff des *Denkens* Maschinen zuzuschreiben, ist eine umstrittene Behauptung. In dem vielzitierten *Computer Machinery and Intelligence* von Alan Turing (1950: 433) verwirft der Autor die Frage ‚Can machines think?‘ und entwirft das *Imitation Game*, bekannt geworden als der „Turing-Test“.

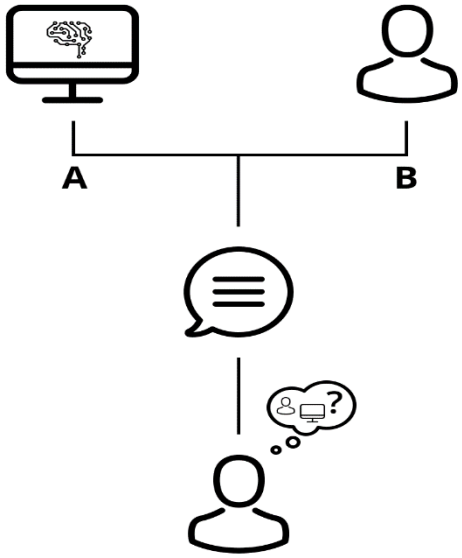


Abbildung 1: Grafische Darstellung Turing-Test (Just add AI GmbH 2017: online)

Dieses wird gespielt mit drei Personen: einem Mann (A), einer Frau (B), und einem Fragestellenden (C) beliebigen Geschlechts. Ziel von C, der sich in einem anderen Raum befindet, ist es, ausschließlich anhand von Fragen an A und B, ihm bekannt als X und Y, herauszufinden, wer Mann bzw. Frau ist. Ziel der Frau (B) ist es, C zu helfen und die Wahrheit zu sagen. Der Mann (B) darf lügen, um C in die Irre zu führen. Die Fragen sollen aufgeschrieben oder über Dritte kommuniziert werden, um Hinweise durch etwa die Stimme auszuschließen. Als Beispielfrage von C wird etwa die Frage nach der Haarlänge angeführt.

An dieses Spiel setzt Turing nun die Folgefragen: „What will happen when a machine takes the part of A in this game? Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?'." (Vgl. 1959: 434) Eine Maschine würde als *denkend* bezeichnet, wenn die fragende Person keinen Unterschied zwischen den Antworten eines Menschen und deren einer Maschine, eines Computers zu vernehmen mag.

Um den Turing Test zu bestehen müsste eine Maschine folgende Fähigkeiten meisten (Vgl. Russell, Norvig 2014: 2):

- **Natural language processing**: um erfolgreich auf Englisch zu kommunizieren
- **Knowledge representation**: um gelernte Information zu speichern
- **Automated reasoning**: um gespeicherte Information nutzen um Antworten und neue Erkenntnisse zu gewinnen
- **Machine Learning**: um sich neuen Verhältnissen anzupassen und Muster zu erkennen und abzuleiten

Abgesehen von der unpassenden binären Betrachtung durch die genderkodierten physischen und moralischen Zuschreibungen (Vgl. Broussard 2018: 38) durch das *Imitation Game*, kann das Gedankenexperiment von Searle (1980) als Widerlegung betrachtet werden. Das *Chinesische Zimmer* beschreibt eine Situation, in der eine englischsprachige Person, der chinesischen Sprache in keiner Weise fähig, in einem Zimmer eingesperrt ist. Hereingereicht werden Schriften auf Chinesisch mit zu lösenden Fragen darüber. Mit den Schriften erhält die Person eine Anleitung auf Englisch. Diese erlauben die Symbole, für die Person ausschließlich erkennbar anhand ihrer Form, in Verbindung zu setzen, und instruiert sie, mit bestimmten Symbolen auf andere zu antworten. Chinesen, welche das beantwortete Schreiben erhalten, vermuten nun jemanden, der der chinesischen Sprache völlig mächtig ist

im Zimmer. Obwohl die Person kein Wort des geschriebenen versteht, unterscheidet sich die Qualität der Antworten nicht von denen eines Muttersprachlers. Beantwortet dieselbe Person nun Fragen über eine Geschichte, welche ihr auf Englisch präsentiert wird, kann sie dies mit vollem Verständnis tun. Der Unterschied ist jedoch, dass die Person keines der chinesischen Zeichen oder der anhand der Instruktion gegebenen Antworten versteht.

> *„As far as the Chinese is concerned, I simply behave like a computer; I perform computational operations on formally specified elements. For the purposes of the Chinese, I am simply an instantiation of the computer program." (Searle 1980)*

Die (englischen) Instruktionen verhalten sich hierbei wie ein Computerprogramm. Damit will der Autor widerlegen, dass künstliche Intelligenz *versteht*. Symbolische Manipulation sei nicht gleichzusetzen mit Verstehen, was anhand aktueller Sprachsteuerungssysteme (Vgl. Broussard 2018: 38f) wie Siri (Apple), Cortana (Microsoft) oder Alexa (Amazon) gesehen werden kann.

Bei Gegenüberstellung des Turing Test und dem Chinesischem Zimmer ist zu beachten, dass Turings Absichten dem Prüfen von Intelligenz gelten. Man kann das Searles Gedankenexperiment als Widerlegung von *bewusster Intelligenz* ansehen. Richtet man aber rein nach den Fähigkeiten, die überwiegend Intelligenz definieren, so überprüft der Total Turing Test doch beachtlich die notwendigen Faktoren. Russell und Norvig weisen darauf hin, dass die Forschung sich jedoch nicht (mehr) darauf konzentriert, Intelligenz nachzubauen, die unserem Original so ähnlich ist, dass wir sie für einen Menschen halten. Man versucht vielmehr, die Grundlegenden Prinzipien zu verstehen. So erlangten auch die Wright Brüder ihren Durchbruch als diese aufhörten, „[to make] machines that fly so exactly like pigeons that they can fool even other pigeons" (Russell, Norvig 2014: 3), und anfingen Aerodynamik und Windtunnel anzuwenden.

Alexa, play Africa by Toto!

Das „voice-activated interface", wie Alexa von Amazon, unterhält sich mit seinen Usern. Dabei *versteht* der Computer weder die Fragen noch die Sprache generell. Computergesteuerte Sequenzen werden akustisch wiedergegeben. *Alexa*, wie im oberen Beispiel, fungiert als Trigger, der dem System ankündigt, dass ein Befehl folgt. *Play* ist der Trigger, der dem System verrät, dass eine Audiodatei abgespielt werden soll. Das System sucht nun nach dem Titel, der nach *play* genannt wird und führt den Befehl aus.

Künstlich hergestellte Denkfähigkeit kann also nicht (mehr) als gerechte Definition dienen. Das Alan-Turing-Institute unternimmt den Versuch, KI als folgend zu beschreiben:

> *(…) used to describe when a machine or system performs tasks that would ordinarily require human (or other biological) brainpower to accomplish, such as making sense of spoken language, learning behaviours or solving problems. There are a wide range of such systems, but broadly speaking they consist of computers running algorithms, often drawing on data.*

> *In popular culture artificial intelligence is often viewed as sentient machines, thinking and behaving like a human.*
>
> *In reality, much AI is computers which are trained to perform tasks independently and which are already present in much of our lives. For example, there has been much publicity about the use of AI in decision-making, for example in the legal system. The AI in this example is driven by machine learning tools, which have taught a computer to make decisions based on the data presented to it.*

Vereinfacht lässt sich zusammenfassen: "AI can give you the most likely answer to any question that can be answered with a number. It involves quantitative prediction. AI is statistics on steroids" (Broussard 2018: 32) Diese bildliche Vereinfachung soll der Mystifizierung künstlicher Intelligenz entgegenwirken.

AI lässt sich unterteilen in *General AI* und *Narrow AI*. Wobei *Narrow AI* der gerade gefundenen Definition entspricht. *General AI* beschreibt die Hollywoodversion (Vgl. Broussard 2018: 10) künstlicher Intelligenz. Dabei geht es um die Summe an hypothetischen Vorstellungen, wie etwa Roboter, die die Weltherrschaft an sich reißen, und ähnlichen Science-Fiction genährte Szenarien, die mit der Realität Nichts gemein haben.

Bruxmann und Schmidt (2019: 6) deduzieren aus der Literatur die schwache Künstliche Intelligenz (Narrow AI), nicht mit dem Ziel das menschliche Denken nachzuahmen, sondern „gezielte Algorithmen für bestimmte, abgegrenzte Problemstellungen zu entwickeln". Demgegenüber steht starke Künstliche Intelligenz (General AI), die allgemein danach strebt, „den Menschen bzw. die Vorgänge im Gehirn abzubilden und zu imitieren" (2019: 6).

Unlängst veröffentlichte die EU-Ethikkommission – mit Ziel der Förderung von „trustworthy AI" – Richtlinien und eine im Zuge dessen erarbeitete Definition Künstlicher Intelligenz. Bei der Definitionsfindung wird dabei die Rationalität vorausgesetzt, wie sie Russel und Norvig betonen (Vgl. 2003: 1f).

Die Expertengruppe der Kommission erklärt *General (strong) AI* als ein System, dass die meisten Handlungen, die auch der Mensch vollziehen kann, beherrscht, und *Narrow (weak) AI* als eine die eine oder wenige Aufgaben beherrscht (Vgl. AI HLEG 2019). Dabei betonen sie, dass momentane Systeme als *Narrow AI* zu betrachten sind. Für die Entwicklung starker KI gilt es noch einige Herausforderungen zu meistern, als Beispiel etwa „common sense reasoning, self-awareness, and the ability of the machine to define its own purpose" (AI HLEG 2019: 5). Folgende Grafik veranschaulicht die stark vereinfachte Einteilung von Subdisziplinen der KI in Machine Learning, Robotics und Reasoning.

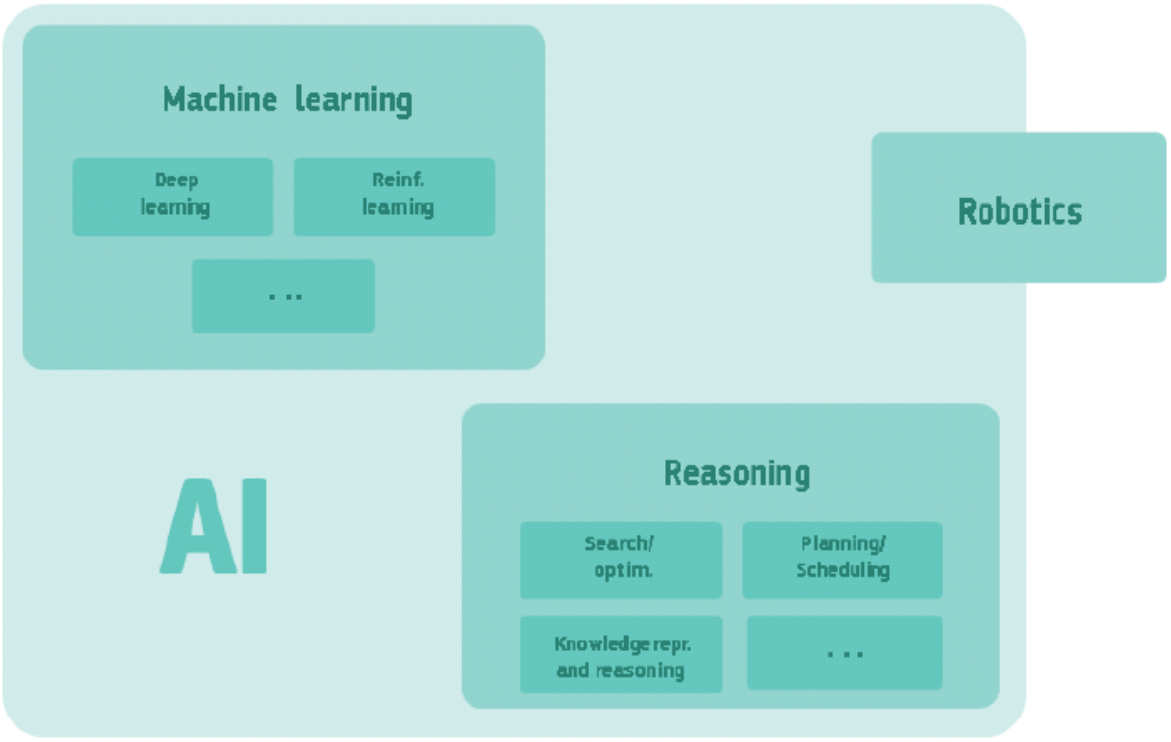


Abbildung 2: Vereinfachte Darstellung Subsysteme KI (AI HLEG 2019)

4 Historische Betrachtung

Will man die Geschichte der Künstlichen Intelligenz und die darin wichtigen Meilensteine beleuchten, gilt es die verschiedenen Perspektiven, aus denen man dies betrachten kann, zu berücksichtigen (Vgl. Kaplan 2017: 57). So macht es einen deutlichen Unterschied, ob man dies aus technischer Sicht betrachtet und tatsächliche Fortschritte in der Forschung heranzieht, oder die Mediendarstellung und folgend herrschende Mehrheitsmeinung und Vorstellungen über KI beleuchtet. Für diese Arbeit ist letzteres im Vordergrund, es folgt also eine kurze Darstellung der allgemein als wichtig erachteten Errungenschaften der KI-Entwicklung. Ziel ist hier nicht eine technische Aufarbeitung. Es werden im Zuge der historischen Betrachtung die Komponenten von KI in ihrer Funktion so weit erläutert, wie sie helfen, Gründe für Künstliche Unintelligenz festmachen und verstehen zu können. Die Grafik benennt wichtige Meilensteine in der Historie der Künstlichen Intelligenz. Zum Verständnis werden im Folgenden die Entstehung, der Schachwettkampf von 1997 und die Etablierung des Maschinellen Lernens näher beleuchtet.

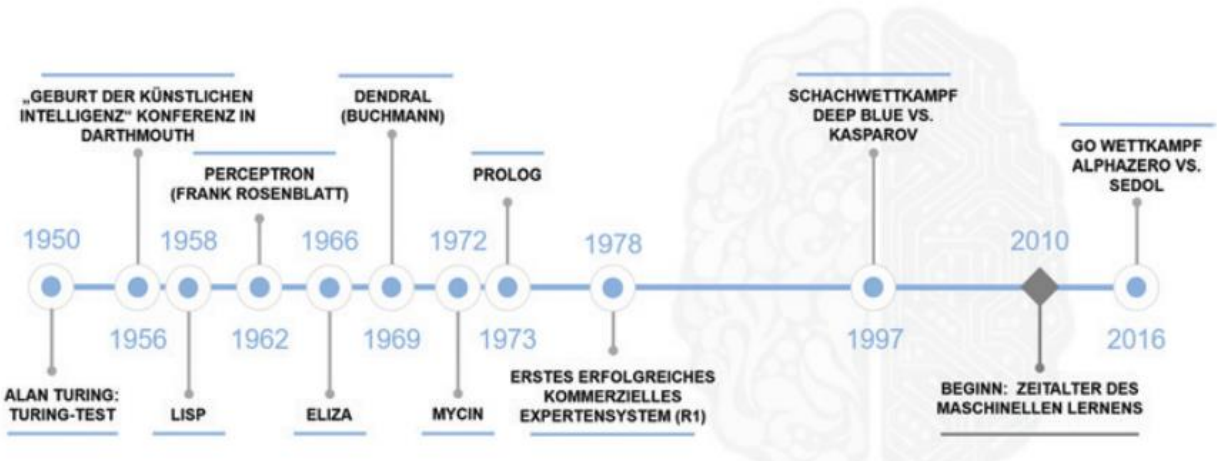


Abbildung 3: "Meilensteine der Künstlichen Intelligenz" (Buxmann, Schmidt 2019: 6)

4.1 Entstehung

Als die Geburtsstunde der Künstlichen Intelligenz gilt das „Summer Research Project on Artificial Intelligence" 1956 am Dartmouth College, New Hampshire in den USA. Organisiert von John McCarthy – Erfinder der Programmiersprache LISP – gilt er als Pionier des Begriffs, zusammen mit führenden KI-Denkern Marvin Minsky, Claude Shannon, Alan Newell, Herbert Simon (Vgl. Buxmann, Schmidt 2019: 3). Fantasien über das künstliche Nachahmen des Menschen finden sich in der menschlichen Geschichte allerdings schon vom 16. Jahrhundert an (Vgl. Manhart 2017).

Durch die zunehmende Speicherkapazität und schnellere, günstigere Computer bekam KI im Anschluss viel Aufmerksamkeit. Erste Erfolge wie ELIZA (stark vereinfachter Vorläufer heutiger Chatbots) oder dem GPS – General Problem Solver (löste Spielprobleme wie das Missionar-Kannibale-Problem durch Mittel-Zweck-Analyse) können als „schlussfolgernde syntaktische Systeme ohne Wissen" (Manhart 2017) beschrieben werden, führten aber zu stark unrealistischen Einschätzung. Minsky etwa schätzte 1970 gegenüber dem Life-Magazine, innerhalb 8 Jahren eine Maschine mit dem Menschen ebenbürtiger Intelligenz zu haben (Vgl. Buxmann, Schmidt 2019: 4).

Die unerfüllten Erwartungen, die in den 1990ern zum sogenannten KI-Winter (Vgl. Streitz 2019: 793) führten, gehen unter anderem auf (noch) schwache Rechenleistung zurück (Vgl. Manhart 2017). Es folgten die Entwicklung von *Expertensystemen* (Vgl. Buxmann, Schmidt 2019: 4), basierend auf klaren Regeln zur Nutzung einer Wissensbasis von Problemstellungen. MYCIN etwa ist eine Unterstützung zur Diagnose von bestimmten Krankheiten (Vgl. Shortlife et al 1974).

4.2 Deep Blue

1997 schlug IBMs *Deep Blue* als erster Schachcomputer den amtierenden Weltmeister in einem knappen Turnier mit 3,5 zu 2,5 (Vgl. Buxmann, Schmidt 2019: 6). Seit dem Turing Test galt Schach als ein beliebtes Mittel, um die „Intelligenz" einer Maschine zu testen (Vgl. Broussard 2018:33). Das Ereignis wurde als das „größte Internetereignis aller Zeiten" betitelt, dem zufolge interpretierten Medien dies als „Kampf zwischen Menschen und Maschine". Deep Blue gewinnt das Schachspiel aufgrund „eine[s] Suchalgorithmus, einer Bewertungsfunktion der Züge und einer riesigen Datenbasis mit gespielten Parteien" (Heßler 2017: 5). Eine Grundsatzdebatte darüber, ob dies nun der Beweis für das *Denken,* somit Künstliche Intelligenz sei, oder simple Rechenfähigkeit wurde erneut entfacht. Kritikern zufolge ist dies jedoch kaum ein Beweis von Künstlicher Intelligenz, viel mehr eine Durchrechnung aller plausiblen Züge mit sehr hoher Rechenleistung (Vgl. Bruxmann, Schmdit 2019: 6).

Nach diesem Erfolg konzentrierte sich die Forschung auf eine bis heute große Herausforderung: Autonomes Fahren, völlig ohne menschliches Zutun. Anreize wie eine Million Dollar als Preisgeld der „Grand Challenge", ausgeschrieben von der amerikanischen DARPA (Defense Advanced Research Projects Agency) 2004 sowie erneut in darauffolgenden Jahren, regten Entwicklungen in diesem Bereich an (Vgl. DARPA 2014). Nach ersten Erfolgen durch Universitäten folgten bald große Automobilhersteller mit ständig besser werdenden Programmen (Vgl. Kaplan 2017: 60).

4.3 Maschinelles Lernen

Zu den Subsystemen Künstlicher Intelligenz mit Lernfähigkeit gehören unter anderem Künstliche Neuronale Netzwerke, Maschinelles Lernen und Deep Learning. Was Künstlich Neuronale Netzwerke genau sind, wird im Kapitel Blackbox Problem erklärt. Hier soll, aufgrund der historisch so wichtigen Bedeutung, Maschinelles Lernen beleuchtet werden.

Diese Erfindung ermöglicht es, Probleme zu behandeln, welche nicht eindeutig definierbar sind, oder deren Lösungsweg sich nicht in eindeutige Regeln übersetzen lässt– zu solchen zählen bspw. „speech and language understanding, as well as computer vision or behaviour prediction" (AI HLEG 2019: 3). Nicht definierbare Probleme ergibt sich zum einen daraus, dass der Mensch über für ihn völlig logisches implizites Wissen verfügt. Ohne den genauen Prozess zu verstehen, kann solches Wissen jedoch nicht codiert werden, man spricht vom Polanyi Paradoxon (Vgl. Autor 2014: 135f). Ein Beispiel solchen Wissens aus dem Bereich computer vision ist etwa die Unterscheidung zwischen Papagei und Guacamole in Abbildung 4 (Vgl. Bruxmann, Schmidt 2019: 8).

Abbildung 4: Papagei oder Guacamole? (Zack 2016)

Grundsätzlich umfasst Maschinelles Lernen (ML) die Fähigkeit, in einer Datenmenge Muster zu erkennen, auf deren Basis Entscheidungen oder Wahrscheinlichkeiten über zukünftige Daten abgegeben werden können (Vgl. Murphy 2012: 1). Dabei basiert diese Fähigkeit auf dem „lernen" anhand von *Erfahrungen* in Form von Daten (Vgl. Buxmann, Schmidt 2019: 8).

> *"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E." (Mitchell 1997: 2)*

Um zu verstehen, wie grundlegend diese Errungenschaft ist, folgt eine kurze Erklärung anhand des Beispiels von Müller und Guido (Vgl. 2017: 1ff):

Erste intelligente Anwendungen funktionierten oft nach von Hand kodierten Wenn-/Dann-Funktionen. Bei einem E-Mail-Spamfilter bspw. kann eine Liste von Wörtern als Anleitung gelten, welche E-Mails als Spam einzustufen sind. Die großen Nachteile dieses Vorgehens sind die Spezifität – kleine Veränderungen verlangen oft ein komplett neues Regelsystem – und das genaue Wissen darüber, wie ein Mensch eine Entscheidung trifft, als Voraussetzung. Zweiteres sieht man am Bsp. *Papagei oder Guacamole? (Abb. 4)*: Für den Menschen ist völlig logisch, welches Bild Papageien und welches den leckeren Dip zeigt. Die „Wahrnehmung" des Computers erfolgt hier jedoch völlig anders als beim Menschen, über Pixel. Was früher ein Problem war, da diese Regeln nicht einfach übersetzt werden konnten, wird durch ML gelöst. Nach dem selben Prinzip erfolgt auch die Gesichtserkennung, über die fast jedes Smartphone verfügt: man trainiert ein Programm mit einer großen Menge an Fotos mit Gesichtern, und es erlernt die Charakteristiken, nach denen menschliche Gesichter erkannt werden.

Diese Form Maschinellen Lernens, die nach dem input-ouput Prinzip von Trainingsdaten verfährt, wird als „Surpervised Learning“ (Überwachtes Lernen) bezeichnet und der Einsatzbereich reicht von Handschrift identifizieren, Kreditkartenbetrug verhindern bis zum Erkennen, ob ein Tumor bösartig ist (Vgl. Murphy 2012: 3f). Auch etwa bei Audiosystemen von Siri bis zum Auto-Navi wird durch ML gelernt, das selbe Wort zu erkennen, auch wenn unterschiedliche Menschen dies verschieden und bei Hintergrundgeräuschen aussprechen (Vgl. Buxmann, Schmidt 2019: 8).

Das Vorgehen der KI bei Maschinellem Lernen lässt sich dabei am einfachsten Nachvollziehen, wenn man sich alle Daten in einer Tabelle vorstellt. Unten angeführte Abbildung zeigt links eine einfache Darstellung von Trainingsdaten als gefärbte Formen, mit der Bewertung *yes* oder *no*, zusammen mit drei nicht eingeteilten Formen. Rechts sieht man die logische Darstellung, anhand derer die Maschine die Schlussfolgerung für die drei gesuchten Formen ziehen kann (Vgl. Murphy 2012: 3). Selbiges Verfahren wird angewendet, wenn es etwa um die Gruppierung von ähnlichen Kunden geht. Hierbei wäre jeder Kunde eine Zeile, jede Spalte beschriebe ein Merkmal wie Geschlecht, Alter, Einkommen. Genauso beim Erkennen von Tumoren anhand von Graustufen der Pixel oder ihrer Größe.

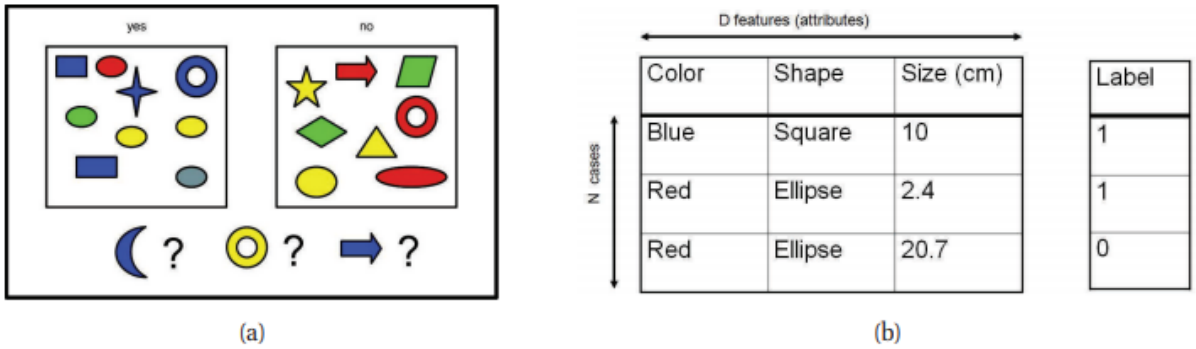


Abbildung 5: Supervised Learning. (Murphy 2012: 3, basierend auf Leslie Kaelbling)

4.4 KI im heutigen Einsatz

Die Hauptbereiche, auf die sich die Forschung Künstlicher Intelligenz fokussiert sind „Robotik, Computer Vision, Spracherkennung und die Verarbeitung natürlicher Sprache“ (Kaplan 2017: 63). Wobei die Hauptschwierigkeit darin besteht, auf einem spezifischen Gebiet extrem leistungsfähige KI anpassungsfähig, wandelbar zu machen (Vgl. Broussard 2018). Die Robotik wird hierbei oft als Königsdisziplin bezeichnet. Die große Herausforderung liegt darin, einer KI einen funktionierenden „Körper“ zu geben – bisherige Entwicklungen befinden sich teils noch auf dem motorisch auf Stand von Kleinkindern. Selbstfahrende Autos sind ein Paradebeispiel für Robotik: sie vereinen eine Vielzahl von KI Funktionen – wie

etwa die Echtzeitauswertung von Verkehrsschildern, ständige Berechnung des Verhaltens aller Verkehrsteilnehmer und die Berücksichtigung, dass Menschen sich nie völlig berechenbar verhalten.

Bruxmann und Schmidt (2019) beschreiben folgende aktuelle Gegebenheiten, die als Rahmenbedingungen Künstlicher Intelligenz neue Möglichkeiten eröffnen:

- Big Data – etwa zum Training von Künstlichen Neuronalen Netzen – sind heute in einer nie gekannten Menge verfügbar und ihre Menge steigt ständig.
- Rechenleistung und Speicherplatz sind so kostengünstig wie nie zuvor und können von Cloud-Anbietern wie Amazon, Google und Microsoft etc. problemlos bezogen werden.
- Die Performance von Deep-Learning-Algorithmen hat sich in den vergangenen Jahren verbessert.
- Inzwischen existieren viele kostenlos verfügbare (Open-Source-)Toolkits und Bibliotheken zur Entwicklung von KI-Anwendungen.

Was die Anwendung und Leistung Künstlicher Intelligenz in der heutigen Zeit erheblich vergünstigt ist *Big Data*: Eine massive und oft komplexe Datenmenge (Vgl. De Mauro et al 2015: 103) womit etwa Künstliche Neuronale Netzwerke trainiert werden können. Zudem kommt das billigere Vorhandensein von Rechenleistung sowie Speicherplatz. Deep-Learning Algorithmen wurden zunehmend verbessert. Außerdem findet sich online eine Vielzahl verfügbaren Wissens and „Baukästen“ und Anleitungen zum Erstellen von KI-Anwendungen (Vgl. Bruxmann, Schmidt 2019: 7f).

5 Künstliche Unintelligenz

Der Terminus Unintelligenz kann als Negation beschrieben werden als das Fehlen von Intelligenz, oder gar das Gegenteil, die Dummheit. Der Begriff Künstliche Unintelligenz (KUI) ist keineswegs eine neue Erfindung. Neben wiederholtem Einsatz in der Medienberichterstattung über misslungene KI Projekte, wird der Begriff etwa von John Higgins in seinem 1987 veröffentlichtem Aufsatz *Artificial Unintelligence: Computer Uses in Language Learning* verwendet. Der Autor beschreibt darin den Einsatz von KI, um möglichst effizient und motivierend Sprachen, speziell Englisch als Fremdsprache, zu lehren. Dabei betont er das große Potential von Computern, behandle man sie als „unintelligent partners“, nicht als pseudointelligente Tutoren (Higgins 1987: 159). Bei der Beschreibung seines Projektes hebt er die Zweifel daran hervor, ob es überhaupt als künstlich intelligent zu bezeichnen sei. Schon eher schreibt er diese Eigenschaft einem weiteren seiner Programme zu, welches interaktiver ist, und vom User auch „lernt“. Die Zweifel am Status als KI begründet er damit, dass die Maschine nicht in der Lage ist, zu *verstehen*, was sie sagt.

Broussards Werk (2018) trägt den Titel unseres gesuchten Terminus. Mit Artificial Unintelligence beschreibt sie aber keineswegs die „Dummheit“ eines Computers oder der algorithmischer Rechenvorgänge. Anhand ihrer Darlegung können für Künstlich Unintelligenz folgende zwei Grundprobleme identifiziert werden.

5.1 Technochauvinismus oder Hype-Begründung

Was hier als Technochauvinismus bezeichnet wird, definiert Broussard als die Überzeugung, Technologie biete immer die Lösung. Begleitet mit dem Glauben, dass weil KI auf mathematischen Berechnungen beruht, sie objektiver, unverzerrter ist. Mit gesteigertem, richtigem Einsatz von Computern könnten soziale Probleme eliminiert werden und ein technoermöglichtes Utopia geschaffen werden (Vgl. 2018: 8).

Technochauvinismus soll nun zum KI-Einsatz in Bereichen führen, in der sie nicht die effizienteste, optimale Lösung darstellt. Als Beispiel einer ineffizienten Entscheidung führt Broussard (Vgl. 2018: 63ff) den Vergleich von Büchern und IPads in US-amerikanischen Schulen an. Ein Technochauvinist würde sich, um Veralterung, Druckkosten, Flexibilität, Schnelligkeit etc. vorzubeugen, dafür entscheiden, Schulkinder mit Tablets auszustatten. Gegen die einmaligen Kosten eines Buches, mit einer wahrscheinlichen Lebenszeit von fünf Jahren, wiegt sie alle Aufwände der Technoversion ab. Lernerfolg, verbundene Kosten, Training und vorallem Anschaffung und Instandhaltung zeigen, dass die technologische Lösung hier mehr Kosten als Abhilfe schafft.

> *“When all you have is a hammer, everything looks like a nail. Computers are our hammers. It’s time to stop rushing blindly into the digital future and start making better, more thoughtful decisions about when and why to use technology.”*
> *(Broussard 2018: 7)*

Es wird versucht zu zeigen, dass der *AI-Hype* (Vgl. Streitz 2019, Broussard 2018, Shalev-Shwartz et al. 2017) zu unreflektiertem, unrealistischem Vertrauen führt. Die Gefahr besteht also, dass KI in Bereichen eingesetzt wird, ohne dass ein tatsächlicher Nutzen, eine Verbesserung dadurch im Vorhinein überprüft und garantiert wird, wodurch sie zur KUI würde.

Der AI-Hype wird von oben genannten Autoren als für neue Technologien typischer Aufmerksamkeitskreis beschrieben. Wie bereits dargelegt, herrscht – vorallem unter Laien – eine große Unkenntnis über den breiten Begriff der Künstlichen Intelligenz. Gespeist etwa durch Hollywoodfilme, die völlig unrealistische Realitäten abbilden (Vgl. Broussard 2018:10), oder irreführende Marketingstrategien (Vgl. Breitinger 2016).

Bestätigendes Beispiel dafür ist der erste Todesfall durch ein „selbstfahrendes Auto“ der Marke Tesla im Mai 2016 (Vgl. Breitinger 2016). Dabei nutzte der Fahrer das als „Autopilot“ betitelte Assistenzprogramm und lies den Wagen sich selbst steuern. Wegen hellem Tageslicht konnten die Sensoren des Wagens die weiße Flanke eines LKWs nicht registrieren,

Bremsen wurden nicht betätigt und verursachten den tödlichen Unfall. Zwar ist die Funktion seit 2014 in Modellen verbaut, vollautomatisiertes Fahren ohne menschliche Beteiligung ist zu diesem Zeitpunkt jedoch nicht erlaubt. Der Fahrer vertraute völlig auf den „Autopilot“ und griff nicht ins Lenken ein. Wobei man annehmen kann, dass die Bezeichnung durch das Tesla-Marketing zu der Fehlannahme beitrug, das Auto sei zu autonomen Fahren imstande und berechtigt.

Streitz (Vgl. 2019) beschreibt mit seinem „smart everything paradigm“ ein ähnliches Phänomen wie Broussards Technochauvinismus. Dabei zieht er den Begriff smart dem intelligent vor. Beschrieben wird eine zunehmende Obsession, alles automatisieren zu wollen, und die Betrachtung Künstlicher Intelligenz als der heilige Gral zur Durchführung (Vgl. Streitz 2019: 792).

5.2 Biased people create biased programs oder: das Sorgfaltsproblem

Programme spiegeln die „Bias“ ihrer (dessen möglicherweise unbewussten) Schöpfer wider, soweit die Aussage von Broussard. Mehrmals verweist sie in ihrer Erarbeitung von „Artificial Unintelligence“ auf die für ihren Technochauvinismus typische Naivität und fehlende Bedachtsamkeit (Vgl. 2018: 69), geht es um die Einführung neuer Gadgets, die „ausnahmslos“ zu negativen sozialen Konsequenzen führen würden. Als Beispiel führt sie den Twitterbot „Tay“ von Microsoft an. Ziel der selbstlernen KI war es, die Sprache von 18- bis 24- Jährigen zu erlernen – von eben genau dieser Zielgruppe. Binnen weniger als 24 Stunden wurde das Projekt abgebrochen (Vgl. Beuth 2016).

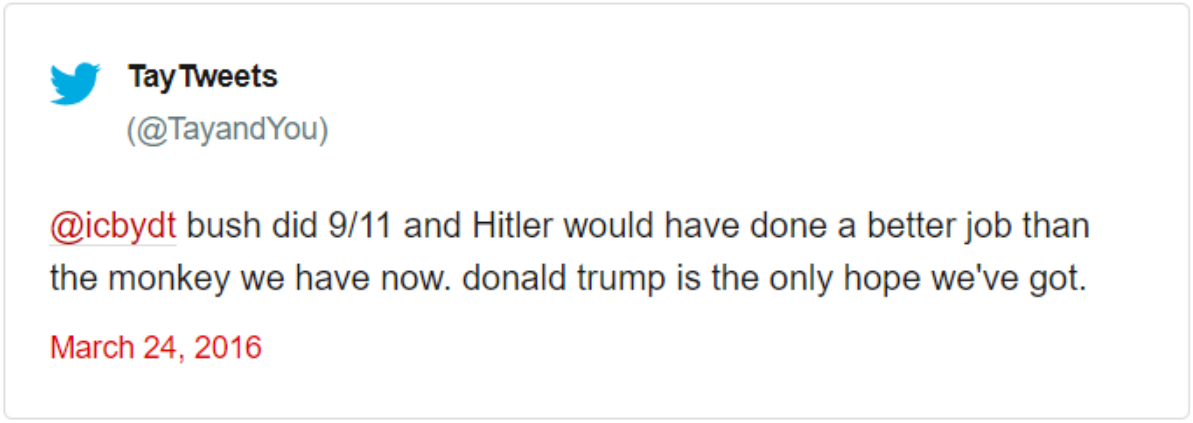


Abbildung 6: Tay's Tweet nach wenigen Stunden (Hunt 2016)

Tay lernte anhand der Interaktion mit der Twittergemeinschaft. Diese brachte der KI in kürzester Zeit bei, Hassreden und Rassismus zu verbreiten. Die überraschten Entwickler beendeten das Projekt (Vgl. Broussard 2018: 69).

Ein weiteres Experiment, dass Broussard als Bestätigung sieht, ist der GPS-fähige Roboter „hitchBOT“ (Vgl. 2018: 69). Er sollte als Tramper Kanada, Deutschland und die USA bereisen, um als soziales Experiment zu klären, ob Menschen sie auf die sprachfähige KI einlassen

würden. Das Experiment endete in Philadelphia, wo er von Unbekannten zerstört und zurückgelassen wurde (Vgl. Presse 2018).

Das wahrscheinlich passendste Beispiel aus der Praxis für Entwickler mit guten Intentionen und schrecklichem Ausgang ist COMPAS – Correctional Offender Management Profiling for Alternative Sanctions. Ziel war es, das Amerikanische Rechtssystem - bekannt für sein Rassismusproblem – zu entlasten und mithilfe von quantitativen Berechnungen eine objektive Entscheidungshilfe für Richter darzustellen (Vgl. Broussard 2018: 155). COMPAS ist eines von vielen Bewertungssystemen, die aufgrund eines von jedem Beschuldigtem ausgefüllten Fragebogens – und unter Einbeziehung aller seiner bekannten Daten über vorherige Fälle – ein Wahrscheinlichkeitsrating abgibt. Dieses schätz ein, wie wahrscheinlich eine Person wieder straffällig wird, das Ranking wird folglich miteinbezogen bei der Entscheidung über Länge der Haft, Bewährung oder Notwendigkeit von Untersuchungshaft. Basis der Ratings sind „soziologische Faktoren wie Einkommen, Herkunft und Familie“, womit eine bestimmte Bevölkerungsgruppe schnell zum Ziel wurde (Vgl. Moll 2016). *ProPublica* – Stiftung für investigativen Journalismus – stellte fest, dass afroamerikanische Bürger mit 77% höherer Wahrscheinlichkeit eingestuft wurden, zukünftig ein Gewaltverbrechen zu begehen, um 45% wahrscheinlicher eine Straftat jeglicher Art zu begehen (Vgl. Angwin et al 2016).

Dies wird veranschaulicht an einem konkreten Beispiel. Die 18-jährige Brisha Borden spielte auf der Straße mit dem herumliegenden Fahrrad eines 6-Jährigen, lies dieses liegen als dessen Mutter ihr zurief. Die durch eine Nachbarin verständigte Polizei verhaftete Borden wegen Diebstahls mit betreffendem Wert von $80. Im Gegenzug dazu ein wertähnliches Verbrechen. Der 42-jährige Vernon Prater wurde wegen Ladendiebstahls im Wert von $86,35 verhaftet. Prater hatte bereits 5 Jahren im Gefängnis verbracht, Vorstrafen wegen bewaffnetem Raubüberfall und versuchten bewaffnetem Raubüberfall. Borden hatte Vorstrafen wegen geringfügiger Jugenddelikte. Beide wurden durch die KI-Anwendung eingestuft. Das folgende Bild zeigt die viel höhere Einschätzung der 18-jährigen Afroamerikanerin. Zwei Jahre später bestätigt sich das Falschliegen des Algorithmus: Borden bleibt straffrei während Prater eine 8-jährige Haftstrafe wegen weiteren Vergehen absitzt (Vgl. Angwin et al 2016).

Abbildung 7: Fehleinschätzung durch KI mit fatalen Folgen für Betroffene (Angwin et al 2016)

Die EU-Ethikkommission weist beim Bias-Problem auf den Einfluss von Trainingsdaten auf eine KI hin. Sind Trainingsdaten „gebiased", zum Beispiel indem sie nicht ausbalanciert sind, kann die KI keine fairen Entscheidungen treffen (Vgl. AI HLEG 2019: 5). Der beschriebene Fall *COMPAS* zeigt hier, wie folgeschwer solche Unachtsamkeiten bei der Entwicklung sich etwa als Benachteiligung ganzer Personengruppen äußern können.

Zeynep Tufekci – Expertin für soziale Folgen von Technologieanwendung und für Fragen der sozialen Gerechtigkeit an der Universität North Carolina – erklärt einer deutschen Zeitung gegenüber, dass im *COMPAS* Fall die Ursache, dass eine so gravierende Fehlfunktion nicht vor Implementierung bemerkt wurde, mit dem Druck auf private KI-Unternehmen, Anwendung schnellstmöglich auf dem Markt anzubieten, zusammenhängt (Vgl. Moll 2016). Dieser Druck verhindere die unbedingt nötigen Testphasen und große Sorgfalt, mit der solche Anwendungen zu behandeln sind.

5.3 Physical Hacking

Als gefährliche KI lässt sich eine einstufen, die unsicher ist. Mit Physical Hacking wird denen Broussards ein weiterer Grund für KUI hinzugefügt. Eine Forschungsgruppe US-amerikanischer Universitäten versuchte Schwachstellen von KI Anwendungen aufzuzeigen. Aufgrund des großen Sicherheitsrisikos wählten sie den Straßenverkehr und autonom fahrende Autos als

Versuchsfeld. Dabei simulierten sie mit Aufklebern gemeine Graffitis, wie sie im Straßenverkehr real vorkommen. Ihre Ergebnisse zeigen, wie einfach sie verheerende Verwechslungen durch das KI System provozieren konnten. „Our attack fools the classifier into believing that a Stop sign is a Speed Limit 80 sign in 80% of the stationary testing conditions." (Eykholt et al. 2018)

Abbildung 8: Reales Graffiti und Nachgestellte gezielte Irreführung (Eykholt et al. 2018)

Zwei Sicherheitsmitarbeiter, Miller für Twitter und Valasek für IOActive tätig, veröffentlichten bereits 2015 ein Experiment, bei dem sie ein Auto während dieses gefahren wurde über das Internet hackten und fremdsteuerten. Dabei erlangten sie Kontrolle über das Getriebe, Lenkung, Dashboard Funktionen und Bremsen (Vgl. Greenberg 2015a). Chrysler veranlasste daraufhin einen Rückruf von 1,4 Millionen Fahrzeugen (Vgl. Greenberg 2015b).

Grundsätzlich ist anzunehmen, dass mit stärkerer KI auch stärkere Hacking-Tools entstehen. Historiker Harari hebt die Bedenken zukünftig fortgeschrittener KI Jahre auf eine neue Ebene: eine „Fusion von Informations- und Biotechnologie werde es ermöglichen, Gefühle und fühlende Organismen in Algorithmen zu übersetzen und damit auch zu „hacken"." (Hering et al. 2018).

5.4 Das Blackbox Problem

Abbildung 9:Einordnung Neuronaler Netze in der Künstlichen Intelligenz (Kreutzer, Sirrenberg 2019: 4)

Obige Abbildung veranschaulicht, wo Künstlich Neuronale Netze (KNN) in der Künstlichen Intelligenz einzuordnen sind. Hierbei imitiert das Computerprogramm ein tatsächliches

neuronales Netzwerk, wie das menschliche Gehirn eines ist (Vgl. Kaplan 2017: 45). Neuronale Netze finden sich als biologischen Systeme, sie zeichnen sich durch Anpassungs- und Lernfähigkeit aus (Vgl. Karrenberg 2010: 441). Beim Menschen besteht ein solches Netz – das Gehirn – aus ca. 100 Milliarden (Vgl. Karrenberg 2010: 441) Neuronen, welche über Synapsen miteinander verbunden sind. Bemerkenswert ist, dass sich das Gehirn durch jegliches empfangene Signal (Lernvorgänge, genau wie sozialer Kontakt oder körperliche Bewegung) verändert und sich neue Neuronen und Verbindungen bilden (Vgl. Karrenberg 2010: 441).

Künstlich Neuronale Netze gehen in verschiedenen Schichten vor. Die „Input-Layer“ (vergleichbar mit dem menschlichen Sehnerv) nimmt Daten auf, nicht dieselbe Information, sondern der Output dieser Schicht wird an die nächste weitergegeben. Dieser Prozess wiederholt sich, bist die „Output-Layer“ erreicht wird (Vgl. Kreutzer, Sirrenberg 2019: 5).

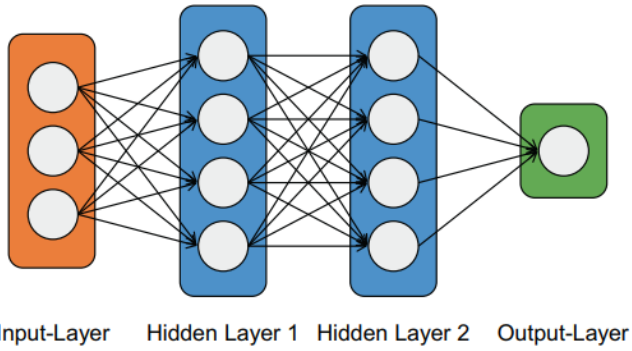
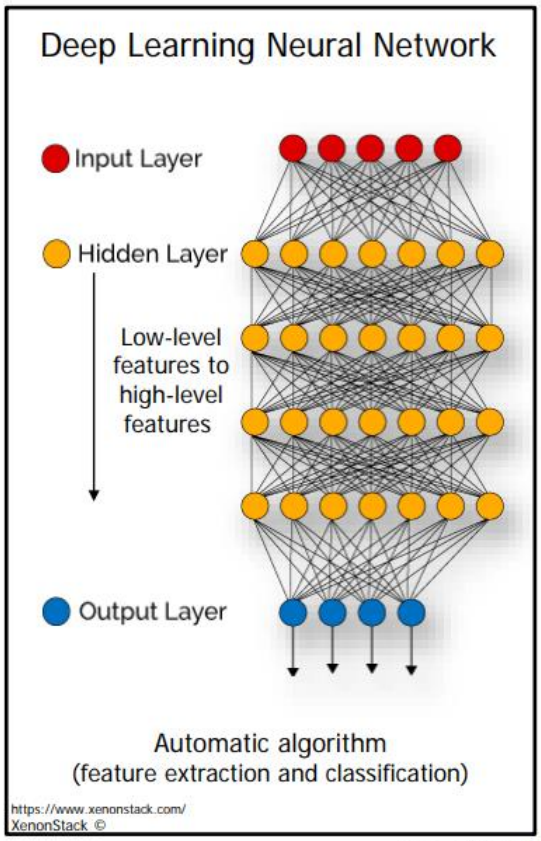


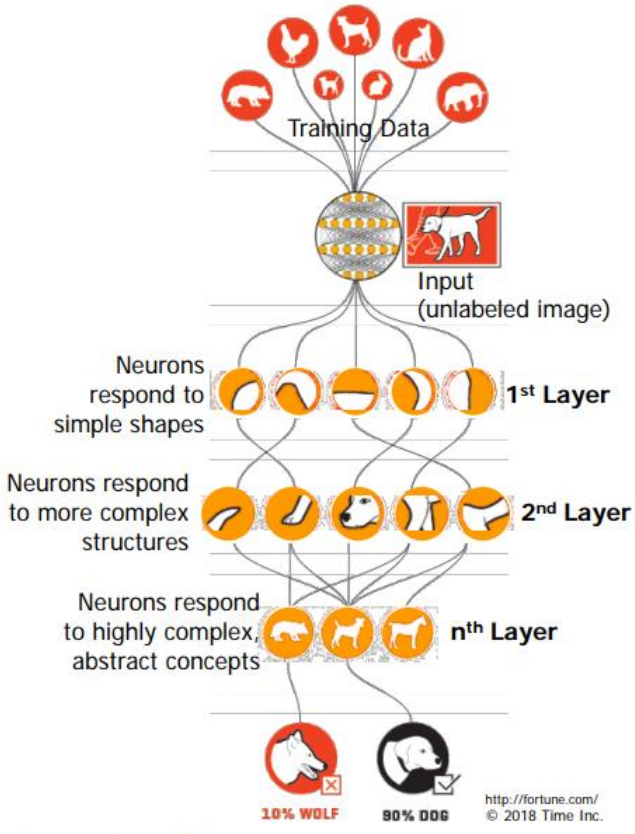
Abbildung 10: Die Durchlaufenen Schichten von KNNs (Kreutzer, Sirrenberg 2019: 5)

Unter 4.3 Maschinelles Lernen wurde gezeigt, wie aufwändig regelbasiertes Vorgehen beim Programmieren ist. KNNs benötigen diese vorher definierten Regeln nicht, sie sollen selbstständig lernen, sich entwickeln und sich basierend auf vorheriger Erfahrung ständig verbessern.

> *„Die initial eingesetzten Algorithmen stellen nur den Nährboden für die Entwicklungen neuer Algorithmen dar. Wenn sich neue Algorithmen im Laufe des Lernprozesses als aussagekräftiger erweisen, arbeitete die „Maschine“ selbstständig mit diesen weiter. Dieser Prozess wird Maschine-Learning genannt.“ (Kreutzer, Sirrenberg 2019: 5f)*

Eine Darstellung der DARPA (Defense Advanced Research Projects Agency) verbildlicht das Vorgehen (Vgl. Gunning 2017). Rechts sieht man, wie KI über KNNs Bilder einem Hund oder Wolf zuweist.





Die Anzahl von bisher eingesetzten Schichten kann von 100 bis zu Zehntausenden reichen (Vgl. Kreutzer, Sirrenberg 2019: 5). Das Problem welches nun auftritt, bezieht sich auf die *Begründung* von erhaltenem Output, den Ergebnissen. Denn wie die „hidden Layers" schon vermuten lassen, kann die Entscheidungsfindung und das genaue Vorgehen einer KI mit KNNs nicht genau nachvollzogen werden. Man spricht von einer Art Blackbox, in der sich das Vorgehen verbirgt.

Mit etwas Ironie kann man von einer „Umkehrung des Polanyi Paradoxons" (Ramge 2018) sprechen. Wie vorher der Mensch (siehe Kapitel 4.3: Papagei oder Guacamole?) „weiß" die KI mehr, als sie uns verständlich machen kann.

Folgende Abbildung zeigt die Veranschaulichung diese Blackbox-Problems durch DARPA (2017). Die erste Zeile zeigt, wie mithilfe von Trainingsdaten ein Bild als Katze klassifiziert werden kann. Was fehlt ist die Begründung für die Einteilung. Was also zukünftig angestrebt wird, ist „explainable AI" – Modelle bei denen jeder Schritt navollziehbar bleibt und Entscheidungen begründet werden können.

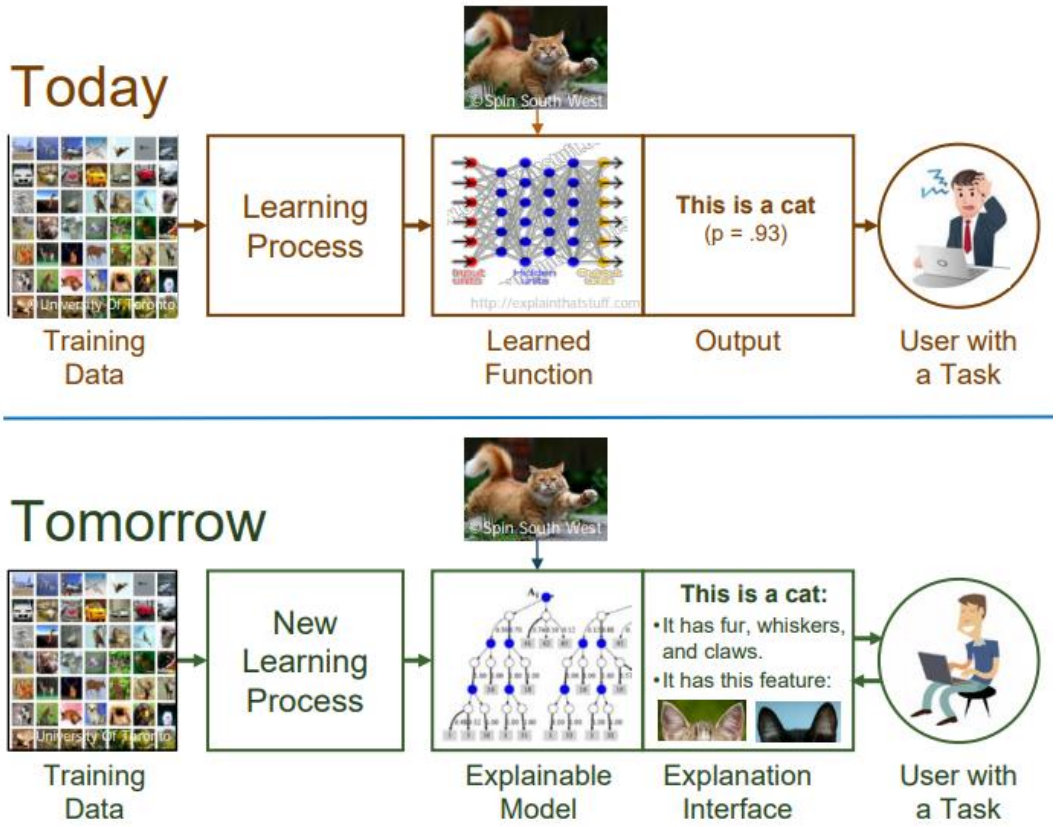


Abbildung 11: Blackbox Problem und Explainable AI (Darpa 2017)

Wie wichtig die Fokussierung auf erklärbare KI ist, zeigt sich bei der praktischen Anwendung von KNNs, welche nicht überprüfbar sind. Solche Anwendungen sind bereits in so sensiblen Bereichen wie dem Personalwesen im Einsatz. Dabei wird etwa berechnet, mit welcher Wahrscheinlichkeit Mitarbeiter zukünftig kündigen werden, oder welche Person aus dem Bewerberpool am effizientesten für das Unternehmen arbeiten würden. Lässt sich die Entscheidung hier aber nicht nachvollziehen, kann nicht garantiert werden, dass etwa beim zweiten Beispiel Parameter wie Geschlecht, Nation oder Religion in die Entscheidung miteinbezogen werden (Vgl. Bruxmann, Schmidt 2019: 17).

5.5 Weitere Gefahren

Jegliche potentiellen Gefahren durch Künstliche Intelligenz hier zu behandeln, würde den Rahmen dieser Arbeit sprengen. Es soll jedoch nicht unerwähnt bleiben, dass mit KI eine Vielzahl an Dilemmata aufgeworfen werden, die es zu lösen gibt. So gibt es etwa Fragen der Ethik und Verantwortung zu klären, auf die bisher noch keine Lösungen gefunden wurde. Wie schwer solche Entscheidungen zu treffen sind, zeigt etwa die Anwendung des klassischen Trolleyproblems auf autonomes Fahren des Massachusetts Institut of Technology (MIT). Hierbei kann darüber abgestimmt werden, wie ein selbstfahrendes Auto in verschiedenen Situationen handeln soll, wenn es jedenfalls Tote geben wird. Rechtsexperten

beschäftigen sich mit Fragen der Haftung; ob diese nun bei Entwickler, Programmierer oder dem Fahrer des Autos liegt.

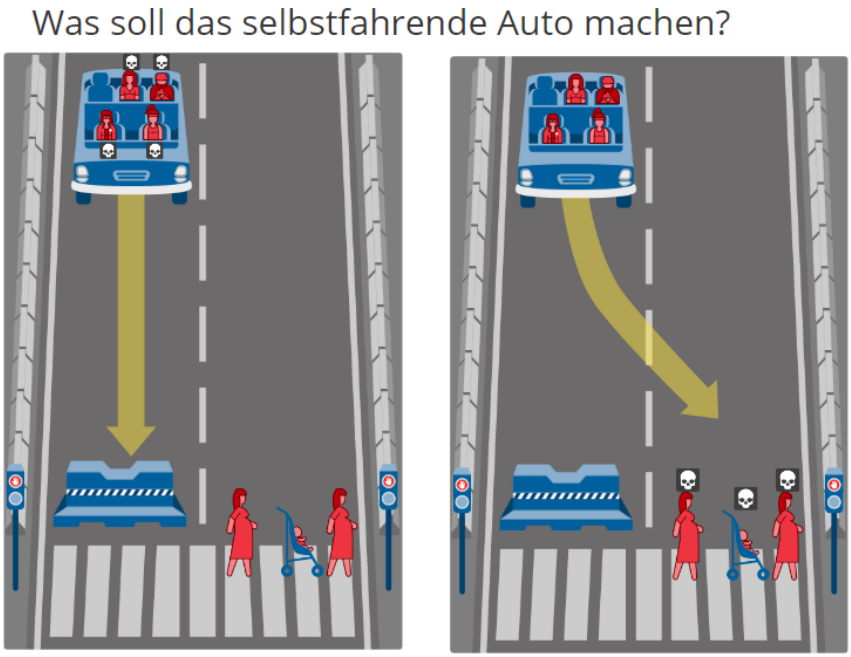


Abbildung 12: The Moral Machine (Scaleable Cooperation MIT http://moralmachine.mit.edu/hl/de)

Kliniken gehen Kooperationen mit Unternehmen ein, die große Mengen von Patientendaten nutzen, um etwa Früherkennungs-KI zu trainieren und diese später kommerzialisieren (Vgl. Lenzen 2018: 157). Ein Fall, in dem Google DeepMind 1,6 Millionen persönliche Patientendaten vom Britischen Gesundheitssystem erhielt, wurde im Nachhinein als nicht zulässig erklärt (Vgl. Revell 2017). Anwendungen wie Amazons Alexa zeichnen Gesprochenes in privaten Häusern auf und intelligente Kinderspielzeuge lassen sich einfach zum Spionagewerkzeug umfunktionieren (Vgl. Lenzen 2018: 183). Diese Beispiele verdeutlichen die Notwendigkeit, sinnvolle Regelungen bezüglich Datensicherheit, Privatsphäre und Verantwortlichkeit zu finden.

6 Theorien

6.1 Technochauvinismus

Ausgang dieser Forschungsarbeit ist Broussards Werk „Artificial Unintelligence“ (2018). Mit Technochauvinismus findet sie darin eine Theorie, die das Entstehen von KUI begründet. Die Grundbedeutung sei die Überzeugung, Technologie (hier speziell: KI) sei immer die Lösung (Vgl. 2018: 8). Als oft auftretende Begleiterscheinungen davon verweist sie etwa auf Ayn Randians Meritokratie. Die Autorin „feiert in ihren Romanwelten heldenhafte Unternehmer, Ingenieure oder Architekten, die sich dem Dienst an der Gemeinschaft verweigern und, nach den Grundsätzen von Rands Objektivismus, nur ihren eigenen Vorstellungen und Idealen entsprechend leben“ (Brühwiler 2013). Die scharf verurteile Philosophie der vorallem in

Amerika gefeierten Rand verfolgt die Idee des „rationalen Egoismus“. „Eine Ablehnung aller sozialstaatlichen Institutionen und staatlichen Marktinterventionen, ein Ja zum Laisser-faire-Kapitalismus“ (Brühwiler 2013) lässt sich davon ableiten.

Außerdem als in Verbindung mit Technochauvinismus auftretend nennt Broussard Technoliberale politische Werte; das Zelebrieren der freien Meinungsäußerung zu einem Ausmaß, der leugnet, dass „online harassment“ ein Problem ist. Die Überzeugung, Computer würden das Ideal der Objektivität erreichen, oder sie seinen „unbiased“, aufgrund der Tatsache, dass alle Fragen bzw. Antworten heruntergebrochen werden auf mathematische Berechnungen. Später postulier sie die Algorithmen, basierend auf der Mathematik, als Werkzeug, das immer nur so objektiv sein kann wie die ihre Entwickler. Ganzheitlich ist die Idee des Technochauvinismus das absolute Vertrauen darin, mit einem gesteigerten, sinnvollen Einsatz von Computern soziale Probleme lösen zu können, ein „digitally enabled utopia“ (2018: 8) zu erreichen.

Hier erneut der Verweis auf Streitz‘ (2019: 792) *smart-everything-paradigm*. 2019 versucht er dies neu zu definieren. Sein Verständnis legt Hauptaugenmerk auf *smart cities* und *smart homes*. Er bemerkt einen Trend zunehmender Abhängigkeit von allgegenwärtiger smarter Infrastruktur, (noch) vorwiegend in Städten. Dabei sei eine oftmals intransparente, unüberprüfbare Künstliche-Intelligenz-Komponente.

6.2 Cyberfeminismus

Der Cyberfeminismus ist ein vergleichsweise junges Phänomen, das seine Anfänge in den 1990er Jahren hat. Der Versuch einer Definition ist eigentlich gegen das Selbstverständnis des Cyberfeminismus, denn es wird gerade angestrebt, keine Definition zu haben (Vgl. Draude 2001). Entstanden ist er überwiegend aus Arbeiten von Künstlerinnen sowie Wissenschaftlerinnen (Vgl. Stoltenhoff, Raudonat 2018: 130). Anstatt sich zu definieren, soll Cyberfeminismus „eine kritische Praxis beschreiben, die die differenztheoretischen Debatten in und um den Feminismus der letzten beiden Jahrzehnte berücksichtigt. Diese nämlich machen es kaum noch möglich von dem Feminismus bzw. der Frau als Subjekt feministischer Bestrebungen zu sprechen“ (Draude 2001: 1). Grundlegend des „-ismus“ ohne angestrebte Definition ist also die Vorstellung, durch neue Informationstechnologien eine „mannigfaltige, nicht festgeschriebene (Subjekt-) Identität“ (Vgl. Stoltenhoff, Raudonat 2018: 130) erreichen zu können. Cyberfeminismus ist eine heterogene Bewegung und lehnt festgeschriebene Identitätskategorien sowie das Alleinstehen einer Kategorie – eben wie dem Geschlecht – entschieden ab (Vgl. Draude 2001: 1).

Der Cyberfeminismus setzt sich kritisch mit Gender und dem Verhältnis zu neuen Technologien auseinander, hierbei speziell das Internet. Typisch ist auch die spielerische, ironische oder provokante Herangehensweise (Vgl. Hartmann, Wimmer 2011: 15), die immer wieder ihren Ursprung in der Kunst aufzeigt.

Auf der „Documenta X“, genannt die „erste cyberfeministische Internationale“ (Hartmann, Wimmer 2001: 15) in Kassel veröffentlichte das *Old Boys Network (OBN)* die teils sehr ironische Liste „100 anti-thesis“. Das *OBN* ist die Website der International Cyberfeminist Organisation, initiiert 1997, das sich mit der Hauptfrage „*What is Cyberfeminism?“* beschäftigt (Vgl. Draude 2001: 1). Anstatt direkt zu erklären was Cyberfeminismus ist, findet sich hier eine Auflistung von Anti-Thesen, meist beginnend mit „Cyberfemism is not/ist kein/nije/n’est pas…“. Ziel ist wieder, den Terminus Cyberfeminismus „so offen wie möglich“ zu belassen, um einerseits nichts auszuschließen, andererseits eine ständige Erweiterung, Rekonfigurierung zu ermöglichen (Vgl. Draude 2001: 1).

50. cyberfeminism is not genetic
51. cyberfeminismus ist keine entschuldigung
52. cyberfeminism is not prosthetic
53. cyberfeminismo no tiene cojones
54. cyberfeminisme n'est pas triste
55. cyberfeminisme n'est pas une pipe
56. cyberfeminism is not a motherboard
57. cyberfeminism is not a fake
58. cyberfeminism nije ogranicen
59. cyberfeminism nije nekonfliktan
60. cyberfeminism nije make up

Abbildung 13: Ausschnitt aus 100-anti-thesis (https://www.obn.org/inhalt_index.html)

Um nun aber ein genaueres Verständnis des Terminus zu erhalten, der Versuch ihn zu beschreiben. Aufbauend auf der Grundidee des Feminismus dient er dazu, „das Verhältnis von Gender und neuen Technologien (insbesondere auch des Internets) kritisch – und spielerisch – zu hinterfragen“ (Hartmann, Wimmer 2011: 14). Auf der Website *OBN* beschreiben ihn einzelne Aussagen folglich: “Cyberfeminism is - a feminism of course, focusing on the digital medium. - a vehicle for discussing certain methods in theory, art or politics. - the updated version of feminism dedicated to new political issues raised by global culture and media society.” (Draude 2001: 1)

Als Begründerin der Idee des Cyberfeminismus, obgleich sie diesen Terminus nie verwendete, gilt die oft zitierte Donna Haraway, Verfasserin des Cyborg Manifesto in den 1980ern (Vgl. Sollfrank 2000: 3). Darin analysiert sie die gesellschaftliche Stellung der Frau in einer zunehmend technologisierten Welt (Vgl. Stoltenhoff, Raudonat 2018: 131). Ihr Essay sei „ein Versuch, zu einer sozialistisch-feministischen Kultur und Theorie in postmoderner, nichtnaturalistischer Weise beizutragen. Es steht in der utopischen Tradition, die sich eine Welt ohne Gender vorstellt“ (Haraway 1995: 2). Sinnbildlich für diese utopische Vorstellung verwischt Haraway die Grenzen von Mensch und Tier, und von Mensch und Maschine, mit dem Cyborg (Cybernetic Organism (Vgl. Fuchs 1999)). Die Metapher des Cyborgs als Geschöpf einer Post-Gender-Welt (Vgl. Haraway 1995: 2) beschreibt einen Menschen, der –

indem er mit der Maschine verschmilzt und zum Cyborg wird – sein Geschlecht und somit die Grenzen der Geschlechter (Vgl. Stoltenhoff, Raudonat 2018: 129) auflöst. Haraway skizziert damit eine mögliche mensch- und maschinenzentriert Zukunft, bei der jedoch die Grenze zwischen Mann und Frau verschwindet, und mit dieser eine Manifestierung von Ungleichheiten (Vgl. Fuchs 1999).

Eine Betrachtungsweise der heutigen Informationstechnologiegesellschaft durch Cyberfeminismus zeigt die noch immer patriarchalen Strukturen der Cyberwelt. Broussard berichtet von ihren diskriminierenden Erlebnissen in einer misogynen Techwelt (Vgl. 2018: 167) und spricht von der belastenden „white, male Bias“ vorallem in der Arbeitswelt der „hard sciences“. Dabei etwa Mathematik und Physik, also die Wissenschaften aus der Künstliche Intelligenz hervorgeht, die nie attraktiv für Frauen oder Menschen anderer Hautfarbe waren (Vgl. 2018: 79). Zu Cyberfeministische Zielen gehört auch die Einbindung von Frauen in neue Informationstechnologien und MINT-Fächer, und dadurch eine nachhaltige Änderung von betreffenden Arbeitsweisen und Berufsbildern (Vgl. Stoltenhoff, Raudonat 2018: 135). Initiativen wie etwa „Komm, mach MINT“ in Deutschland sind an einer Änderung bestrebt. Ähnliche Bestrebungen gibt es auch in Österreich, derzeit aber noch mit mäßig Erfolg. 2018 liegt der Frauenanteil beim Abschluss eines MINT-Faches unter einem Viertel. Neben den teils stark unterschiedlichen Stellenwerten, die technische Fächer in verschiedenen Ländern genießen, seien sie für Frauen unattraktiv gestaltet (Vgl. Gruber 2018).

6.3 Diskussion

Kehren wir zurück zu Broussards Technochauvinismus. Mit ihrer Beschreibung der derzeitigen Techcommunity und dem scharfen Urteil über die „white, male Bias“ beschreibt sie (auch) einen Zustand, in der KI-Anwendungen immer nur so objektiv seinen können wie ihre Erschaffer. Die Betrachtung durch den Cyberfeminismus wirft nun folgende Frage auf. Würde eine Techwelt ohne Gender, Rasse, Herkunft etc., also eine Techwelt an der jeder Mensch gleichermaßen teilhaben und mitwirken kann, eine „unbiased“ Technologie ermöglichen?

Am COMPAS Beispiel (Kapitel 3.5.2) sieht man, dass Ursache für das Versagen von Anwendungen oft der Druck ist, sie so schnell wie möglich auf dem Markt zu bringen. Zeit für die dringend nötigen Testphasen, um Rassismusentwicklungen wie bei COMPAS zu verhindern, wird nicht genommen. COMPAS gehört hier einem privaten Unternehmen, nähere Information über den Algorithmus wurden nicht preisgegeben (Vgl. Moll 2016). Daraus lässt sich schließen, dass eine emanzipiertere Techwelt bestimmt förderlich ist, jedoch kaum allein ähnliche Vorkommen verhindern würden. Es steht die Forderung nach genauester Arbeit und langen Testphasen, wiederholten Prüfungen unabhängiger Prüfer, schlicht größte Sorgfalt, um „intelligente“ KI zu sichern. Diese Anforderungen schneiden sich

offensichtlich vorallem mit dem Vorgehen gewinnorientierter Unternehmen, wo Schnelligkeit zählt.

Die Antwort hier ist konfrontiert mit einem weiteren Faktor: Der Mensch hat nur zu einem gewissen Grad Einfluss auf das Vorgehen der von ihm erschaffenen Künstlichen Intelligenz, und oft scheitert er daran, *Begründungen* für die von KI gelieferten Ergebnisse zu finden. Künstlich neuronale Netzwerke etwa sind in ihrer Vorgehensweise vom menschlichen Gehirns inspiriert (Vgl. Kaplan 2017: 45). Dies bedeutet jedoch, dass die Berechnungen in einer Art *Blackbox* versteckt sind (Vgl. Castelvecci 2016). Es ist also schwierig nachzuvollziehen, *warum genau* KI manche Ergebnisse präsentiert bzw. wie sie schlussfolgert. „Wir verstehen diese Netze genauso wenig wie das menschliche Gehirn“, so Informatiker der University of Wyoming Jeff Clune (Vgl. Castelvecci 2016). Dass eine ausgewogene Techwelt, an der jeder teilhaben kann, äußerst anzustrebend ist, steht außer Frage. Vorallem durch die Komplexität gegebener Technologien, und Hindernisse wie die Blackbox scheint das Bias-Problem dadurch aber kaum lösbar.

7 Forschungsfragen & Hypothesen

FF1 Unter welchen Bedingungen „kippt“ KI im gesellschaftsrelevanten Einsatz deutschsprachiger KI-Experten zufolge zur „Künstlichen Unintelligenz (KUI)“?

H1.1 Deutschsprachige KI-Experten bestätigen überwiegend einen Hype, der zu stark unrealistischen und dadurch teils gefährlichen Erwartungen bezüglich AI führt, was als Künstliche Unintelligenz zu verstehen ist.

Als eine Gefahr bei KI-Anwendungen wird ein kollektiver Enthusiasmus beim Einsatz dieser in allen möglichen Bereichen – wie etwa Autofahren oder Dates finden – beschrieben, bei dem der Anspruch an tatsächlich „gute“, also einwandfreie Technologie in den Hintergrund rückt (Vgl. Broussard 2018: 6).

Wiederholt machten Experten auf den für aufstrebende Technologien typischen Hype-Cycle aufmerksam (Vgl. Streitz 2019, Broussard 2018, Shalev-Shwartz et al. 2017). Dabei wird vom “AI-Winter“ in den 1980er und anfänglichen 1990er Jahren als ersten Einbruch des Enthusiasmus gesprochen, der in weniger Interesse und Investitionen in der Branche resultierte (Vgl. Streitz 2019: 793). Nun wird von einem erneuten Hype vorallem um selbstfahrende Autos ausgegangen (Vgl. Shalev-Shwartz et al. 2017:1). Wie der bereits beschriebene erste Todesfall durch ein solches autonomes Fahrzeug zeigt, birgt dies ein erhebliches Risiko.

Broussard (2018) findet für diesen Hype die Bezeichnung „Technochauvinismus“, was vergleichbar ist mit Streitz‘ „smart-everything paradigm“ (2019).

H1.2 Deutschsprachige KI-Experten sehen überwiegend eine große Bedrohung bei KI-Anwendungen durch „Physical Hacking“.

Die Gefahren bei KI-Anwendungen beschränken sich nicht auf den Einsatz unzureichend geprüfter Entwicklungen. „Physical Hacking“ beschreibt den Vorgang, durch den etwa KI-gestützte Autos gehackt und fremdgesteuert (Vgl. Greenberg 2015) werden können. Weiters kann ein autonomes Fahrzeug bereits durch kleine, realistische Veränderungen der Umwelt „gehackt“ werden, wie Eykholt et al. (2018) zeigten. Dabei beklebten sie Verkehrsschilder teilweise, was in groben Fehleinschätzungen durch das System resultierte.

H1.3 Deutschsprachige KI-Experten sehen überwiegend eine Bedrohung bei KI-Anwendungen durch das „Blackbox-Problem“.

Wie bereits beschrieben, kann bei KNN etwa durch das gehirnimitierende Vorgehen nicht genau bestimmt werden, *warum* oder *wie* die KI zu einer Lösung kommt. Bei wichtigen Entscheidungen, bspw. in der Strafverfolgung, ist es somit schwer, ein richtiges Vorgehen zu gewährleisten.

H1.4 Deutschsprachige KI-Experten sehen überwiegend eine Bedrohung bei KI-Anwendungen durch das „Sorgfaltsproblem“.

Die vierte Hypothese vereint zwei Ansätze unter dem Begriff „Sorgfaltsproblem“. In Kapitel 4.5.2 wurde gezeigt, wie wichtig und zugleich schwer das Garantieren einer Bias-freien KI-Anwendung in der Praxis ist. Dieses Biasproblem wird begünstigt durch den hinderlichen Marktzwang, von dem Unternehmen sich dazu drängen lassen, die notwendige Sorgfalt, welche etwa lange Testphasen beinhalten, nicht zu berücksichtigen. Dieses Sorgfaltsproblem begünstigt folglich die Implementierung von biased KI und nicht hinreichend getesteten Anwendungen.

FF2 Welche Lösungsvorschläge skizzieren deutschsprachige Experten, um einen gemeinwohlorientierten Einsatz Künstlicher Intelligenz zu fördern?

H2.1 Deutschsprachige KI-Experten sehen mehrheitlich Potential in staatlicher (Co-)Regulierung, gehen jedoch nicht von einer zukunftsnahen Implementierung aus.

Staatliche Regulierungskompetenz wird unter anderem mit Kompetenz bei Anwendern und Betroffenen und übergreifenden Rahmenbedingungen als Lösungsansatz genannt (Vgl. Krüger, Lischka 2018: 29). Miteinbeziehend die aktuelle Lage und die Einschätzung, dass meist die Technologie der Rechtsentwicklung voraus ist (Vgl. Tschohl 2014:220), wird von einer Bestätigung dieser Hypothese durch die befragten Experten ausgegangen.

Um den Faktor *überwiegen* der Hypothesen zu klären, werden die offenen Fragen zusätzlich mit einer Skalenauswahl bestückt. Dabei soll durch die Experten für jedes Problem eine Einschätzung seiner Schwere abgegeben werden. Hier wird versucht festzustellen, inwiefern etwas als Bedrohung wahrgenommen werden kann. Die Skala reicht von 1=überhaupt nicht, 2=kaum, 3=etwas, 4=stark bis 5=sehr stark. Durch die Skala sollen die wichtigen Herausforderungen in der aktuellen KI Forschung identifiziert werden.

8 Forschungsdesign

8.1 Methode: Das Experteninterview

Das Experteninterview als solches ist ein eigentlich unpräzise definierter Begriff. In der Praxis der Interviewverfahren ist es üblich, diese näher zu bezeichnet, etwa als narratives Interview, vollstandardisiertes, problemzentriertes oder Telefoninterview. Es lässt sich aber von einem „stillschweigenden Konsens“ unter Sozialforschern ausgehen, die Methode als „leitfadengestütztes Experteninterview“ zu verstehen (Liebold, Trinczek 2009: 32). Genauer lässt es sich als „eigenständige Form des qualitativen Leitfadeninterviews“ (Blöbaum et al 2016: 184) bezeichnen. Das Leitfadengestützte Experteninterview ist eine „stärker strukturierte Form der Befragung mit dem Ziel der Gewinnung harter Fakten, die sich aus anderen Quellen nicht oder nur eingeschränkt ermitteln lassen“ (Kaiser 2014: 35). Der

Interviewleitfaden ist bedingtes Mittel, um klar definiertes Wissen zu Beantwortung der vorher definierten Forschungsfrage(n) zu erlangen (Vgl. Kaiser 2014: 35). In den Sozialwissenschaften ist es eine häufig angewendete Methode (Vgl. Gläser, Laudel 2010: 12).

Die Methode wird für die Erhebung von Daten beziehend auf die Meso- und Makroebene eingesetzt (Vgl. Blöbaum et al 2016: 175), weshalb sie sich für diese Forschungsarbeit als geeignet erweist. Es gilt bei dieser Methodenform besondere Beachtung auf „Auswahl der Befragten, die Planung und die Durchführung“ (Reihnhold 2015: 329) zu legen. Kaiser postuliert zwei kritische Aspekte zu beachten: 1) Wer kann als Experte gelten 2) Welche Art von Wissen lassen sich durch solche Interviews generieren (Vgl. 2014: 35).

8.1.1 Wer ist Experte

Im allgemeinverständlichen Sinn sind Experten „Sachverständige, Kenner oder Fachleute […], also Personen, die über besondere Wissensbestände verfügen“ (Liebold, Trinczek 2009: 33). Die Klassifizierung eines solchen basiert auf der Anerkennung themenspezifischer Differenzierung von „Experten“ versus „Laien“, wobei der Gesuchte als solcher verstanden wird aufgrund „seiner Zugehörigkeit zu entsprechenden Berufen bzw. Professionen“ (Liebold, Trinczek 2009: 34).

Gläser und Laudel beschreiben Experten als „Angehörige einer Funktionselite, die über besonders Wissen verfügen“ (2010: 11). Dabei machen sie darauf aufmerksam, dass auch solche Experten sein können, die sich übermäßig mit einem bestimmten Thema auseinandersetzen. Die Bezeichnung ist also nicht nur Personen vorbehalten, die professionellen Zugang haben. Grundsätzlich also entscheidet der Forschende, wer Relevantes zum Forschungszweck beizutragen hat und somit zum Experten wird (Vgl. Kaiser 2014: 39). Für dieses Forschungsvorhaben wird die Expertenrolle definiert anhand eines umfangreichen Wissens über Künstliche Intelligenz. Dies kann festgemacht werden an einer beruflichen Rolle, die genannte Voraussetzungen aufweist, oder etwa das mitwirken an KI-spezifischen Forschungsvorhaben oder Publikationen deutschsprachiger Personen.

8.1.2 Welches Wissen lässt sich generieren

Ziel der Untersuchung ist die Generierung von Kontextwissen sowie Deutungswissen. Die Verbindung beider Wissenstypen, nach welchen Experten ihre Antworten niemals klar trennen, können interessante, gesamtheitliche Antworten liefern. Erfragt werden „Kenntnisse des Experten über Rahmenbedingungen, Zwänge du Interessensstrukturen“ (Kaiser 2014: 44). Das Deutungswissen, wonach die Forschungsfragen abzielen, sucht nach „subjektiven Sichtweisen und Interpretationen des Experten zu Verfahren zur Lösung gesellschaftlicher Konflikte“ (Kaiser 2014: 44). Aufgrund des subjektiven Charakters dieses Wissens sind hierbei auch sozialisierende Umstände der Befragten interessant. Sie werden je nach Möglichkeit und DSGVO-Konformität erhoben, um das Wissen kontextgebunden darstellen zu können.

8.1.3 Klassifizierung des Interviews

Bei der Durchführung der Expertenbefragung wird das halbstandardisierte Interview verwendet. Dabei Sind Fragenwortlaut und Reihenfolge vorgegeben und für alle Interviewten gleich. Die Beantwortung dieser ist völlig offen (Vgl. Gläser, Laudel 2010: 41).

Das Experteninterview hat somit eine doppelte Ausrichtung, bezeichnet als „geschlossene Offenheit“ (Vgl. Liebold, Trinczek 2009: 37). Dies einerseits durch die bestimmte Strukturierung, andererseits die erzählende Gesprächsstruktur.

8.1.4 Der Leitfaden und die Durchführung

Für die Leitfadenerstellung sowie um ein inhaltlich kompetentes Interview führen zu können, ist für den Forschenden unabdingbar, sich ein gewisses Basiswissen über das Thema zu erarbeiten (Vgl. Blöbaum et al 2016: 185f). Dieses Vorwissen wurde im Zuge der vergehenden Kapitel dieser Arbeit generiert, wobei eine intensive Auseinandersetzung mit der Thematik problembehafteter Künstlicher Intelligenz erfolgte. Es gilt erneut zu betonen, dass diese Arbeit keine genaue Aufarbeitung der Funktionsweise von KI ist. Die einzelnen Problemfelder wurden im Vorherigen soweit beschrieben, wie sie zum Verständnis und zur Beantwortung der Forschungsfragen notwendig sind. Da Experten in ihrer Rolle meist zeitlich sehr eingespannt sind und die Literatur auf Probleme bei der Teilnahmegewinnung hinweist, ist das Anschreiben und lockende Hinweisen auf das Forschungsvorhaben ausschlaggebend (Vgl. Blöbaum et al 2016: 186). Wegen eben dieser Zeitnot bieten sich Telefoninterviews als notwendige Alternative an, worunter die Qualität kaum leiden soll, falls die optimale Situation des persönlichen Gesprächs nicht erreicht werden kann (Vgl. Blöbaum et al 2016: 186). Auch Internettelefonie wie etwa durch Skype besteht als Möglichkeit, jede Art birgt natürlich Vor- und Nachteile hinsichtlich Aufwand oder Anonymität (Vgl. Loosen 2016 145).

Der Leitfaden ist die Übersetzung des Forschungsproblems in konkrete Interviewfragen. Sie sollen verständlich und, unter Berücksichtigung des Wissensstandes und Möglichkeiten der Experten, sinnvoll und beantwortbar sein (Vgl. Kaiser 2014: 52). Durch die im Vorhinein festgelegte Struktur der Fragen soll die Vergleichbarkeit und Auswertung der erhobenen Daten vereinfacht werden (Vgl. Loosen 2016: 144f). Der Leitfaden als halbstandardisiertes Instrument besteht aus inhaltlich gleichen Fragen, die jedoch in der Interviewsituation durch die Offenheit der Antworten durch Nachfragen oder Anpassung an individuellen Experten angepasst werden können (Vgl. Döring, Bortz 2016: 372), so wird auch das Vergleichen von mehreren Interviews ermöglicht. Die thematischen Fragen können im Leitfragen gut in eine dramaturgische Ordnung gebracht werden (Vgl. Loosen 2016: 144).

Bogner et al gliedern den interviewfertigen Leitfaden in mehrere Themenblöcke, zu denen jeweils bis zu drei Hauptfragen notiert werden; dazu kommen Detailfragen (Vgl. 2014: 28f). Die Hauptfragen fungieren hierbei als „Pflichtfragen“. Die Detailfragen dienen etwa dem Nachfragen, sollte Gewünschtes noch nicht ausreichen beantwortet sein, oder können auf individuelle Experten näher eingehen.

Der Leitfaden in fünf Themenblöcken

1. **Technochauvinismus – Hype Begründung**

Inwiefern sehen Sie den KI-Hype-Cycle als Auslöser von unrealistischen Erwartungen und dadurch als Grund von Künstlicher Unintelligenz oder gar eine Bedrohung?

- *Der Hype soll zu unreflektiertem Vertrauen führen. Denken Sie an den ersten Todesfall durch ein selbstfahrendes Auto im Jahr 2016: Hätte der Unfall verhindert werden können, hätte der Fahrer nicht völlig vertraut und deswegen nicht eingegriffen?*
- *Schätzen Sie es als ernstzunehmende Bedrohung ein?*
- *Was sehen Sie als Grund für die schlechte Informiertheit?*
- *Wird sich das ändern, wird es Bestrebungen geben dies zu ändern?*

Wenn sie auf einer Skala abstimmen müssten:

Sehr stark – stark – etwas – kaum – überhaupt nicht

2. **Sorgfaltsproblem**

Unter Sorgfaltsproblem fasse ich zwei Punkte zusammen. Einerseits das Problem von Bias der Entwickelnden bzw. Bias in den Trainingsdaten für KI-Anwendungen. Zweitens der Druck vieler Unternehmen, Anwendungen schnellstmöglich auf dem Markt anzubieten, wodurch die notwendigen Testphasen zu kurz sind, und Bias-Probleme nicht erkannt werden. Dieser Punkt bezieht sich also auf das Einbauen bzw. Nichterkennen von Bias in KI-Anwendungen.

Inwiefern sehen sie das Sorgfaltsproblem als Grund für Künstliche Unintelligenz?

- *Als wie relevant/groß schätzen Sie diese Bedrohung ein?*
- *Wird das Problem in Zukunft beseitigt werden können?*
- *Wenn ja: wie, welche Ansätze?*

Wenn sie auf einer Skala abstimmen müssten:

Sehr stark – stark – etwas – kaum – überhaupt nicht

3. **Physical Hacking**

Miller und Valasek, zwei im IT-Bereich tätige Hacker haben bereits 2015 gezeigt, wie ein autonom fahrendes Auto gehackt und von außen fremdgesteuert werden kann. Ein anderes Beispiel einer Forschungsgruppe US-amerikanischer Universitäten zeigte, wie ein Sticker auf einem Verkehrsschild dazu führte, dass ein Fahrzeug das Stoppschild nicht mehr als solches erkennen konnte.

Inwiefern betrachten Sie Physical Hacking als Grund für Künstliche Unintelligenz oder als Bedrohung?

- *Wie relevant wird dieses Problem in einer zunehmend digitalen Welt sein?*

Wenn sie auf einer Skala abstimmen müssten:

Sehr stark – stark – etwas – kaum – überhaupt nicht

4. Blackbox Problem

Künstlich Neuronale Netzwerke erlauben es oft nicht, Begründungen für die KI präsentierten Ergebnisse nachzuvollziehen, die Entscheidungsfindung versteckt sich in einer Art Blackbox.

Inwiefern sehen Sie das Blackbox-Problem als einen Grund für falsch eingesetzte KI, kann und wird es zukünftig eine Bedrohung sein?

- *Wie kann man zukünftig mit dem Problem umgehen, wird man sich auf explainable KI fokussieren?*

Wenn sie auf einer Skala abstimmen müssten:

Sehr stark – stark – etwas – kaum – überhaupt nicht

5. Lösungsvorschläge

Vielen Dank. Nun zum letzten Punkt. *Sie haben bereits erwähnt, wie man manche dieser Probleme vorbeugen kann*. Könnten Sie mir generelle Lösungsansätze nennen, um KUI in Zukunft zu verhindern, bzw Fördermaßnahmen für KI, die allen Menschen dient.

- *Als wie groß schätzen Sie das Potential staatlicher (Co)-Regulierung ein?*
- *Die vor kurzem veröffentlichten Ethik-Richtlinien der EU sind Anhaltspunkte, bleiben aber Richtlinien und sind nicht bindend. Halten Sie eine bindende Regulierung durch den Staat für realistisch?*
- *Wie zukunftsnahe könnte eine (staatliche) Regulierung durchgesetzt werden?*

6. Ergänzungen

Ich danke vielmals für Ihre Zeit. Gibt es noch etwas, dass Sie gern ergänzen würden, fallen Ihnen von mir nicht genannte Gründe für Künstliche Unintelligenz/Probleme die mit KI einhergehen/ungünstige Umstände für sinnvollen Verwendung der Technologie ein?

9 Auswertung

9.1 Qualitative Inhaltsanalyse

Die Auswertung der Experteninterviews soll mithilfe einer qualitativen Inhaltsanalyse durchgeführt werden. Dabei werden den Texten Informationen entnommen und diese klassifiziert (Vgl. Gläser, Laudel 2010: 197) um sie anschließend mithilfe der qualitativen Datenauswertungssoftware MAXQDA zu bewerten und für die Bearbeitung der Forschungsfragen heranziehen zu können. Voraussetzungen für eine solche Analyse nach Gläser und Laudel sind das Erstellen eines Kategoriensystems, Zerlegen des Textes in Analyseeinheiten, Durchsuchen des Textes auf relevante Informationen und die Zuordnung dieser (Vgl. 2010: 197f).

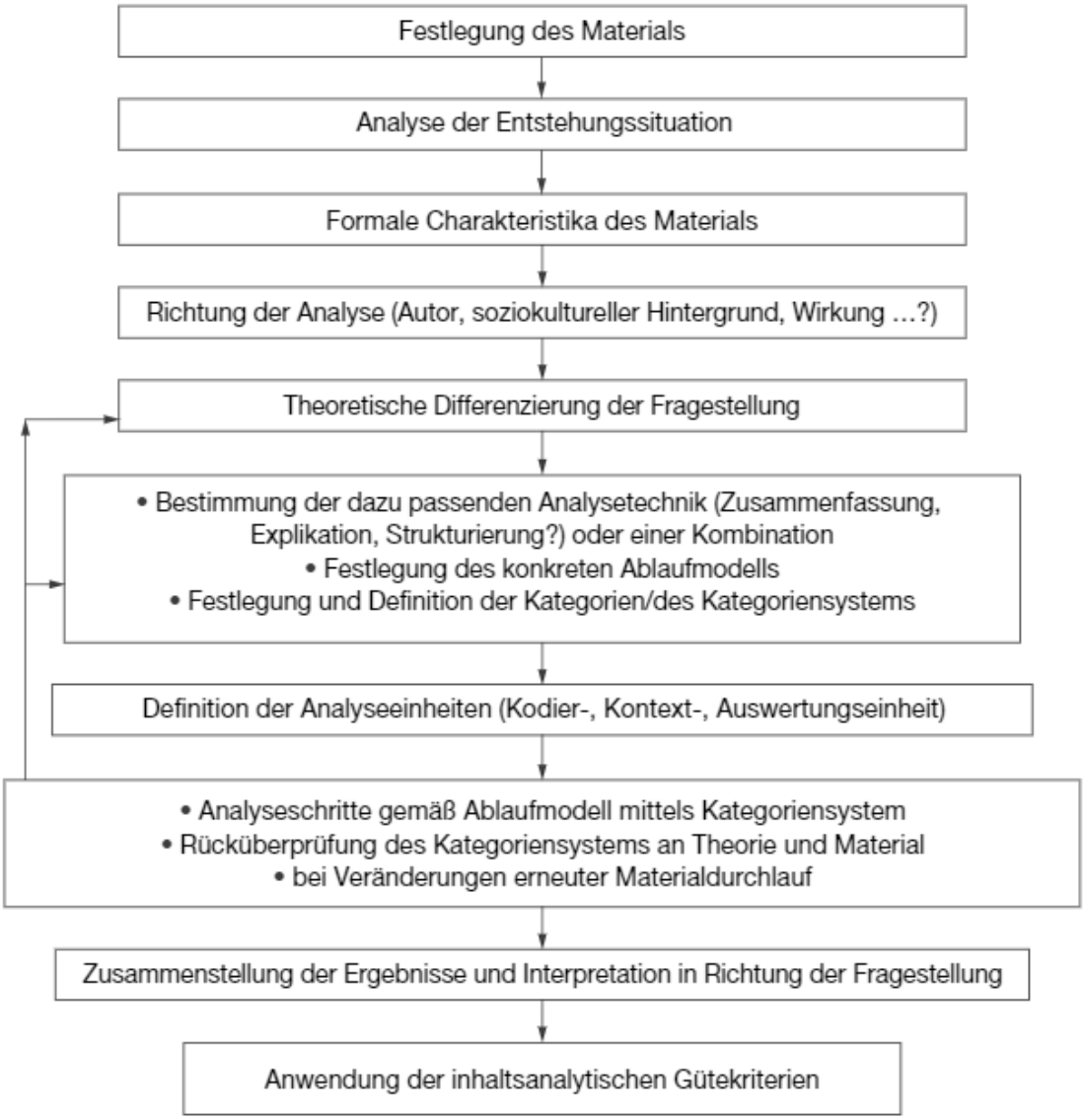


Abbildung 14: Ablaufmodell qualitative Inhaltsanalyse nach Mayring (2015)

Es wird vorgegangen nach dem Ablaufmodell der qualitativen Inhaltsanalyse nach Mayring (2015). Dabei ist das zu untersuchende Material die geführten Interviews. Im Zentrum des ganzen Forschungsprozesses steht die Kategorienbildung. Durch die Struktur des Leitfadeninterviews erfolgt diese deduktiv (Vgl. Kuckartz 2010: 202) und ist mit den Problemsituationen Künstlicher Intelligenz schon im Vorhinein definiert. Während des Bearbeitens werden von der Fragestellung geleiteten Kategorien rücküberprüft und durch hierarchische Subkategorien und zusammenfassende Oberkategorien ergänzt (Vgl. Mayring 2015: 61).

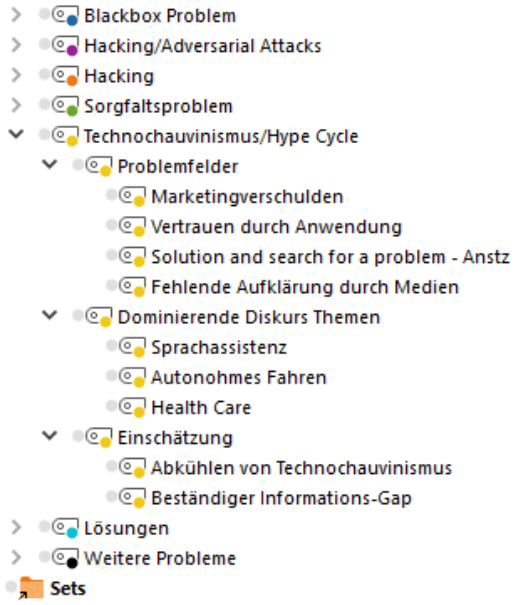


Abbildung 15: Ausschnitt Kategoriensystem MAXQDA 2019

Mayring empfiehlt das Sammeln von Ankerbeispielen für einzelne Ausprägungen. Kodierregeln sorgen im Falle von Unklarheiten für einheitliches und reproduzierbares Vorgehen. Was als einer Kategorie bzw. einem Code zugordnet wird, obliegt gänzlich dem Forscher. Passagen des Interviews können auch mehrmals codiert werden (Vgl. Kuckartz 2010: 23).

Mayrings Modell verlangt das Festlegen von Analyseeinheiten, um ein möglichst präzises Arbeiten zu gewährleisten:

- Als kleinste Kodiereinheit wird hier ein Halbsatz mit klarer Aussagekraft festgelegt. Grund sind vereinzelte Antworten, die Aussagen zu verschiedenen Kategorien (Codes) beinhalten.

  „Ich glaube, dass sich der Technochauvinismus gerade massiv abkühlt, wenn du schaust der öffentliche Diskurs wird zum Großteil von drei Beispielen dominiert: das ist einerseits Health Care, andererseits autonom fahrende Autos und Sprachassistenten.“ (Interview Wasner 2019)

- Die Kontexteinheit als größte Kodiereinheit ist eine vollständige Antwort, also ein durchgängiger Absatz im Transskript eines Interviews.
- Die Auswertungseinheit sind die gesamten Interviews, die analysiert werden.

Das gesammelte Material wird nun also Schrittweise untersucht und Codiert, es werden also Textpassagen den vorher definierten Kategorien zugewiesen. Gleichzeitig folgt eine induktive Rücküberprüfung, das Kategoriensystem wird also durch zusätzliche, sich aus dem Material ergebende Codes und Subcodes ergänzt.

MAXQDA erlaubt während dieses Vorgangs das Festhalten von Fragen oder Einfällen zum Untersuchten über Memos, was später in die Auswertung eingearbeitet werden soll (Vgl. Kuckartz 2010: 24). Abschließend erfolgt das Analysieren, Auswerten und Zusammenführen der fertig codierten gewonnen Daten.

10 Ergebnisse

10.1 Experten und Interviewsituation

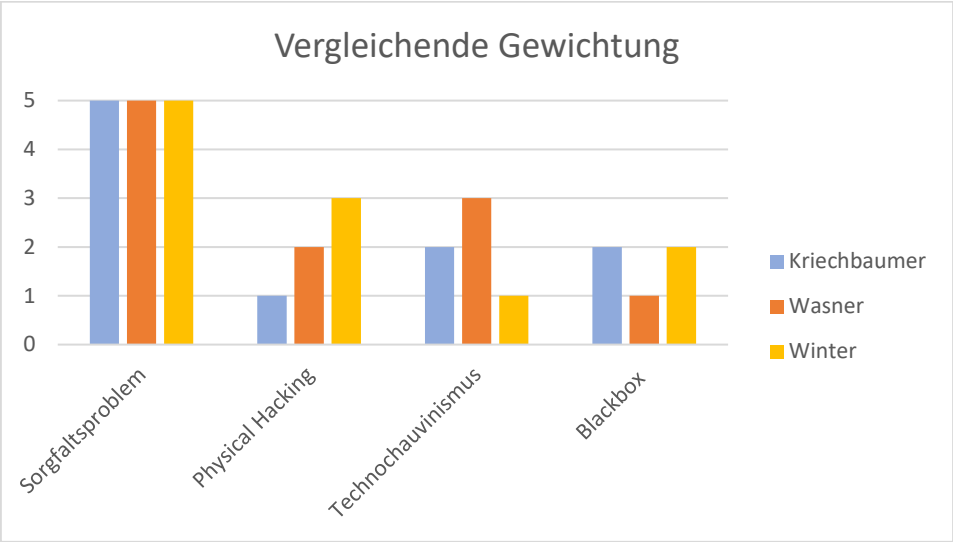
Insgesamt wurden vier weibliche und vier männliche Experten über den E-Mail Weg kontaktiert. Aufgrund eines leider geringen Rücklaufs wurden daraufhin drei ausführliche Interviews mit männlichen, deutschsprachigen Experten geführt. Dabei reichte die Dauer von 54 Minuten bis zu 1 Stunde 30 Minuten. Alle stimmten einer namentlichen Nennung zu, der Altersrahmen befindet sich zwischen 25 und 39 Jahren. Um eine möglichst umfassende Beantwortung der Forschungsfragen zu gewährleisten, wurde darauf geachtet, Experten mit verschiedenen Paradigmen und Kontextwissen zu befragen. Da alle einer namentlichen Nennung zustimmen, folgt eine kurze Beschreibung dieser.

Experte 1 (Wasner) ist Gründer und CEO der EnliteAI GmbH und Teil von AI Austria. Er gilt als vielbefragter KI-Experte in Österreich und soll durch mehrjährigen Auslandsaufenthalt Einschätzungen über globale Entwicklungen sowie marktspezifische Unterschiede der Problemstellungen liefern. Wasner liefert eine umfassende wirtschaftliche Perspektive. Das Interview fand in entspanntem Umfeld in einem Wiener Café statt, wobei ausreichend Zeit für genaues Nachfragen und Erklärungen waren. Experte 2 (Kriechbaumer) ist selbstständig in der IT-Branche mit Fokus auf Softwareentwicklung und TechLead eines Salzburger StartUps für bargeldlose Bezahlsysteme. Auch Kriechbaumer arbeitete bereits im Ausland an KI-Bewertungssystemen, wie sie in dieser Arbeit häufig als Beispiel herangezogen werden. Das Interview fand während einer gemeinsamen mehrstündigen beruflichen Autofahrt statt, auch hier war die Optimalsituation face-to-face gegeben. Experte 3 (Winter) ist wissenschaftlicher Mitarbeiter der Johannes-Kepler-Universität Linz des Instituts für Machine Learning. Von ihm behandelte Forschungsthemen sind unter anderem Artificial Intelligence, Machine Learning und Deep Learning. Winter soll die Ergebnisse um eine Perspektive aus dem Forschungskontext ergänzen. Das Interview fand auf seinen Wunsch hin telefonisch statt, was etwas weniger ausführlich als die face-to-face Befragungen ausfiel.

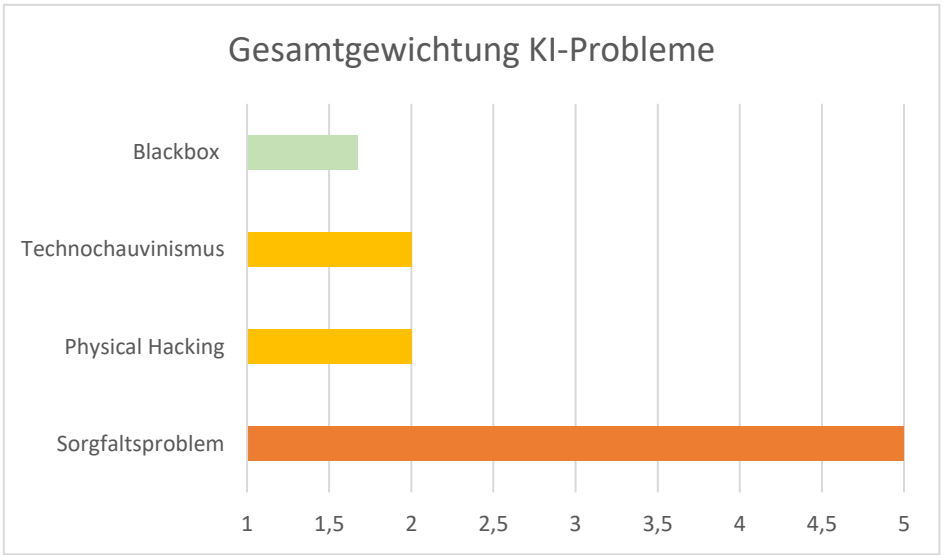
Bei allen Interviews war die Themengliederung des Leitfadens von Vorteil, da die ursprüngliche Reihenfolge an die Experten und ihre Gedankengänge einfach

angepasst werden konnten. Am Ende jedes Interviews wurden zur Bestätigung bei unklaren Formulierungen die Einstufung auf den Skalen wiederholt.

10.2 Übersicht



Jedes behandelte Problem wurde hinsichtlich der Schwere seiner Bedrohung auf einer Skala von 1 bis 5 eingestuft. Dabei entspricht 1=überhaupt nicht, 2=kaum, 3=etwas, 4=stark und 5=sehr stark. Das vergleichende Diagramm zeigt die Einigkeit aller Befragten über die Schwere des Sorgfaltsproblem. Alle drei Befragten merkten an, dass beim Punkt „Physical Hacking“ im Interview Beispiele von Physcial Hacking aber auch „normalen“ Hacking genannt wurden, und hier klar zwischen den zwei verschiedenen Bereichen zu differenzieren ist. Beim Punkt Technochauvinismus wird bei genauerer Auswertung zu sehen sein, dass vorallem hier die Einstufung als „Bedrohung“ sehr viel Interpretationsspielraum durch Experten und Forscherin offenlässt. Das Blackbox Problem wurde insgesamt als am wenigsten bedrohlich eingestuft.



10.3 Ergebnisse im Detail

10.3.1 Das Blackbox Problem

BLACKBOX PROBLEM		
Beschreibung	Gewichtung	Lösungsansätze
Hemmt Einführung	Einsatzzwecke momentan nicht drastisch	Selbstregulierung Markt
Gefahr durch fehlende Nachvollziehbarkeit	Temporäres Problem	KI nach Einsatzbereichen
Traininsgdaten sind essentiell	Eher Forschungsproblem	Nachvollziehbarkeit auf kleiner Ebene nicht als Ziel
Komplexität der Trainingsdaten		Explaining AI - System begründet selbst
Fehlende Wahrscheinlichkeit zum Output		Explainable AI - System wird genau alaysiert
		Modulare Netzwerke

Mit durchschnittlicher Bewertung von 1,67 lässt sich eine momentane Bedrohung durch fehlende Nachvollziehbarkeit der Entscheidungsfindung bei Künstlich Neuronalen Netzen als „kaum“ einordnen. Lediglich Experte 1 (Wasner) sieht diese Bedrohung „überhaupt nicht“. Die Experten beschreiben hier, dass durchaus eine Gefahr durch die fehlende Nachvollziehbarkeit ausgeht. Die Komplexität der Daten mach es schwierig, in der Blackbox Ursachen zu finden.

> *„Umso länger und größer das Datenset ist, umso schwieriger wird es für den der es anlernt, herauszufinden, auf Basis von welchem Ursprungsdatensatz die KI quasi Amok läuft in die falsche Richtung, wo es was reininterpretiert.“ (Kriechbaumer 2019)*

Alle drei Befragten ordnen das Problem jedoch als gering ein, weil die Einsatzzwecke, wo es zu Trage käme, momentan gering bis nicht vorhanden sind.

> *„Es ist insofern kein Problem, weil ohnehin keine Life-Entscheidungen getroffen werden, also kann es per Definition nicht sehr stark sein. Wenn mir Netflix was Falsches empfiehlt oder mein Twitter algorithmic-news-feed Blödsinn anzeigt - so what?“ (Wasner 2019)*

Einerseits hemme das Problem die Einführung betroffener KI-Anwendungen, andererseits sei es ein temporäres Problem, nach Wasners Schätzung in drei bis vier Jahren gelöst. Dem fügt Winter hinzu, es sei demnach eher als Forschungsproblem zu betrachten. Zwei der Befragten sehen eine populäre Lösung in „Explainable AI“. Alle weiteren Vorschläge wurden nur von jeweils einer Person genannt. Dabei wird auf Selbstregulierung durch den Markt – Nachvollziehbarkeit als Verkaufsargument kommerzieller Anbieter – Explaining AI, wobei das System mit dem Output eine Begründung liefert, und Modulare Netzwerke als gestufte Kontrollinstanzen verwiesen. Ein Experte betont hier, dass es keine KI Anwendung für jeden Use-Case geben kann und wird. Es gäbe deswegen durchaus Bereiche, wo der Einsatz trotz Blackbox Sinn machen würde. Diese sind Bereiche mit Fokus auf Performance und eng definiertem Datensatz. Die Hypothese 1.3 (Deutschsprachige KI-Experten sehen überwiegend eine Bedrohung bei KI-Anwendungen durch das „Blackbox-Problem“) kann also als falsifiziert betrachtet werden.

10.3.2 Technochauvinismus / Hype Cycle

Technochauvinismus	Problemfelder		Fehlende Aufklärung durch Medien
			Solution and Search for a problem - Ansatz
			Marketingverschulden
	Dominierender Diskurs		Sprachassistenz
			Autonomes Fahren
			Health Care
	Einschätzung	Unverändert	Beständiger Informations-Gap
		Verbessernd	Abkühlen von Technochauvinismus

Alle drei Befragte bestätigen das Bestehen eines Hype Cycles. Wasner differenziert, wir befänden uns nicht mehr im KI- sondern im Deep-Learning-Hype-Cycle. Zwei von drei sehen ein kommendes Abkühlen des Hypes, demnach wird KI von der Bevölkerung bald als nichts Besonderes mehr wahrgenommen, durch den typischen Hype-Cycle Verlauf, nach dem zu hohe Erwartungen ab einem Punkt nicht mehr erfüllt werden können und dieser wieder abflaut. Die drei bestehenden, dominierenden Diskursthemen korrelieren mit den großen Firmen dahinter und seien dadurch bestimmt. Die in der obigen Grafik aufgelisteten Themenfelder sind „IBM/Google, Tesla - mittlerweile Uber, und die Sprachassistenz Amazon geschuldet“ (Wasner 2019).

Das Vertrauen in KI Anwendungen scheint, weil KI zur Normalität wird, eher zuzunehmen. Endverbraucher sind täglich mit KI etwa in Smartphones konfrontiert und wissen andererseits oft nicht, dass die Technologie hinter einer Anwendung steckt. Winter nennt es Vertrauen durch Anwendung.

> *„Beim Smartphone kommt der Modus Porträtfoto zum Einsatz, der mit Zigmillionen Bilder antrainiert ist. Was ist ein Kopf, was ist eine Wand, was zeichne ich unscharf was mach ich scharf. Das wird den Leuten langsam bewusst was dort alles reinfällt, das sind halt in 99 von 100 Fällen relativ triviale Sachen wo die Leute dann sagen oh das ist KI, das habe ich ja gar nicht gewusst.“ (Wasner 2019)*

Zwar sehen die Befragten alle eine Zunahme der Informiertheit, Kriechbaumer betont aber das zukünftige Bestehen eines Informations-Gaps, und beschreibt anhand des Beispiels autonomes Fahren, dass die Informationspolitik kaum mit neuen Entwicklungen mithalten oder gar aufholen könne. Von einem Experten wird darauf hingewiesen, dass die Aufklärung durch die Medien viel zu gering ausfällt, als Beispiel nennt er die fehlende Unterscheidung von General und Narrow AI. Die Probleme dadurch werden noch als relativ irrelevant beschrieben, weil auch hier noch wenig Einsatzfelder bestehen, wo blindes Vertrauen tatsächliche Bedrohungsszenarien hervorrufen könnte. Es wird argumentiert: „Direkte Bedrohungen sind es nie. Es geht immer um Daten, Informationen, Wissen über andere Personen. Das kann dann explizit, wissentlich falsch verwendet werden.“ (Winter 2019)

Bedrohungsszenarien durch Technochauvinismus entstehen also (noch) weniger durch zu großes Vertrauen in ein Auto, das eigentlich noch nicht selbst fährt. Hier geht es erst um gemeine Datensicherheit. Werden große Mengen an Daten gesammelt und weitergegeben, können Sie im Nachhinein sehr effizient ausgewertet werden – in diesem Schritt kommt Künstliche Intelligenz zum Einsatz.

> *„Der Todesfall und auch die nachgefolgt sind, sind Fälle die der Fahrlässigkeit von Tesla zuzuschreiben sind, weil hier was als autonomes Fahren vermarktet wird was nicht autonomes Fahren ist.“ (Wasner 2019)*

Zwei der Befragten weisen auf Marketingverschulden hin, wenn man etwa den ersten Todesfall durch ein Tesla Auto bewertet. Hierbei wird aber erklärt, dass das Vorgehen von Marketing weder neu noch KI-spezifisch ist, und der Gutgläubigkeit der Menschen zuzuschreiben ist. Broussards Beschreibung (2018), wonach die Tech-Verliebtheit ineffizienten Einsatz nach sich zieht, bestätigt Wasner durch den Solution-and-Search-for-a-Problem Ansatz. Winter sieht das Informationsproblem bis in die politische Ebene reichen, wo Informationen, die an Entscheidungsträger herangetragen werden oft nicht ausreichend oder gebiased sind. Die Hypothese 1.1 (Deutschsprachige KI-Experten bestätigen überwiegend einen Hype, der zu stark unrealistischen und dadurch teils gefährlichen Erwartungen bezüglich AI führt, was als Künstliche Unintelligenz zu verstehen ist) kann also ebenfalls als falsifiziert angesehen werden. Zwar ist Technochauvinismus ein Grund für Künstliche Unintelligenz nach der Definition dieser Arbeit, die Bewertung durch die Experten fiel jedoch geringer aus.

10.3.3 Physical Hacking

Alle Befragten wiesen bei diesem Punkt daraufhin, dass eindeutig zwischen Physical Hacking und „gemeinem“ Hacking zu unterscheiden sei. Ein Beispiel in der Befragung war das erfolgreiche hacken eines autonom Fahrenden Autos durch zwei IT-Spezialisten. Hier wurde daraufhin gewiesen, dass der Hackvorgang an sich ein Thema der IT-Sicherheit ist, keine spezielles der künstlichen Intelligenz sei, und eben unter die Kategorie Hacking fällt. Physical Hacking hingegen wird hier genauer als Adversarial Attacks (AA) bezeichnet, aktuell finden dazu große Forschungsanstrengungen statt.

> *„Da gehts es darum, wenn man etwa einen objekterkennungs-Algorithmus hat und den in einer Kamera anwendet. Da kann man sowohl außen, in der Umwelt, als auch im System selbst diese Adversarial Attacks anwenden, und so diese Netzwerke täuschen. Also von außen das klassische mit den Stickern. Oder wenn man eine lebensgroße Person auf ein Plakat druckt und wo platziert, wird das selbstfahrende Auto das auch als Person erkennen, obwohl es keine ist.“*

Physical Hacking erfährt mit durchschnittlich 2=kaum eine ebenfalls niedrige Einstufung. Ähnlich wie beim Blackbox Problem merken zwei von drei an, dass es momentan noch kein Problem darstellt, weil es auf gesellschaftlicher Ebene noch nicht schädlich einsetzbar sei.

PHYSICAL HACKING	Gewichtung	Lösungen
	Schwach	
	Großes Budget für Sicherheit	Netzwerke mit Kontrollinstanz
	Bereits im Zentrum der Forschung	
	Kinderkrankheit einer jungen Technologie	
	Auf gesellschaftlicher Ebene nicht schädlich einsetzbar	
	Noch nicht relevantes Problem	

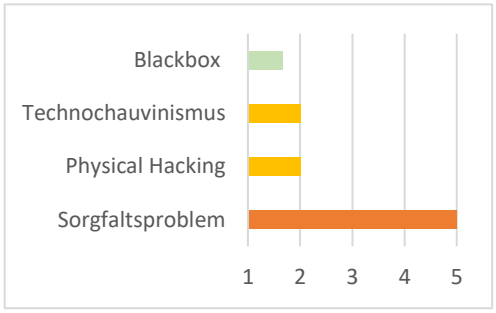
Weitere Argumente für die Einstufung als schwache Bedrohung waren das üblicherweise große Budget für Sicherheit bei Entwickelnden, und der Fakt, dass AA schon im Zentrum der Forschung behandelt werden. Ein Experte bezeichnet das lediglich als Kinderkrankheit. Untypisch sei hier nur, dass die gemeine Bevölkerung die Entwicklung mitverfolgen kann, da Prototypen in der Öffentlichkeit getestet werden und speziell bei KI vieles Open Source ist.

HACKING	Gewichtung		
		Stark	KI als effizientes Missbrauchstool
			Datensicherheit durch Firmen zu schlecht
			Individuelles Gefahrenpotential hoch
			Megaplattformen bieten verheerendes Angriffsziel
		Schwach	Starke Sicherheit auf individueller Ebene
	Beschreibung		Hacking als Wettrüsten: Keine aktive Veränderung durch KI

Zwei der Experten äußerten sich außerdem „nur“ Hacking betreffend. Beide waren sich darüber einige, dass hier durch KI sowohl Angriffe als auch Sicherheit effizienter werden. Es sei also eine Art „Wettrüsten“, durch KI habe sich aber so nichts verändert und sei kein neues Problem aufgekommen. Alle Aspekte für starke Hacking Bedrohung mittels KI wurden von jeweils einem Experten genannt. Was KI aber besonders auf individueller Ebene verändert, sind sehr effiziente Schutzmaßnahmen. Wasner beschreibt Identifikationsmechanismen, die Nutzer anhand ihrer Vibration, den Sakkaden (Zeit zwischen tippen von Buchstaben) oder biometrischen Daten wie Gesicht und Herzschlag identifizieren. Individuelles sei damit so gut sicherbar, dass es „defacto unhackbar“ wird. Kriechbaumer merkte an, dass er für Hacking weiters zwischen individueller und gesellschaftlicher Bedrohung differenzieren würde. Die individuelle würde er aufgrund effizienter Tools als sehr stark bewerten, mit dem Vermerk auch schon vor KI Zeiten. Die gesellschaftliche, wie bei Physical Hacking, beschreibt er wegen der geringen Nutzung als kaum.

Die Hypothese H1.2 (Deutschsprachige KI-Experten sehen überwiegend eine große Bedrohung bei KI-Anwendungen durch „Physical Hacking“) kann als ebenfalls falsifiziert betrachtet werden. Wieder bezieht sich die Bewertung auf die momentane Situation und begründet sich mit dem noch geringen Einsatz wichtiger KI Anwendungen.

10.3.4 Das Sorgfaltsproblem



Wie die Übersicht erneut zeigt, ist das Sorgfaltsproblem als einziges von allen Befragten einheitlich als „sehr stark" beurteilt worden. Auffällig ist zu dem der große Abstand zu allen anderen Punkten.

Alle stufen es als sehr stark ein, ein Experte sieht es als ein langfristig steigendes Problem. Von jeweils einer Person argumentiert wird das mit der zu geringen Sorglosigkeit, mit dem Unternehmen große Datenmengen schützen und bei weiterer Verwendung behandeln bzw. bereinigen. Ein Experte sieht größeres Gefahrenpotential bei der Weitergabe von Daten von gescheiterten Unternehmen.

> *„Da sind teilweise haarsträubende Sachen, wo Unternehmen teilweise Apps von Leuten, Kindererziehungs-Apps, HomesecurityApps, wo schnittweise Überwachungsvideos und Fotos angelegt werden, die nie für den Anwendungsfall gedacht waren, damit KI zu trainieren – die Firma geht out of Business und der Datenbestand wird aber aufgekauft von einer Firma die das für Überwachungssysteme verwendet." (Wasner 2019)*

SORGFALTSPROBLEM				
Gewichtung	Schwach	Noch wenig im Einsatz		
	Stark	Datenmissbrauch gescheiterter Firmen		
		Sorglosigkeit gewinnorientierter Unternehmen		
	Zukünftig steigendes Problem	Bedrohung durch blindes Vertrauen		
Marktspezifika	Europa	Negativ	Geringere Divergenz im Forschungskontext	
			Fehlerfindung langsamer	
			Geringere Datensets	
		Positiv	Leitende Regulierungen (DSGVO)	
			Betroffener Personenkreis geringer	
	US-Amerika	Negativ	Größere Fehleranfälligkeit	
			Betroffener Personenkreis größer	
		Positiv	Accountability im privaten Sektor	
			Qualitätsvoll (Trial & Error Verfahren)	Schnelleres Bereinigen von Fehlern
				Größere Datensets
Bias	Datenherausforderung	Biasfrei nicht erreichbar		
		Menge		
		Komplexität		
	Lösungen	Unsupervised Learning		
		Große Datenmengen f. Erlernen sowie Testen		

Einer der Befragten – im Gesamten bewertet auch er mit „sehr stark“ – sieht den geringen Einsatz erneut als Argument für eine schwächere Einstufung. Bias betreffend sind sich alle befragten einig, dass biasfreie Daten kaum zu erreichen sind. Zwei fügen hinzu, dass die Komplexität der Daten das Erkennen vor der Anwendung sehr erschwere. Ein Experte verweist darauf, dass allein die Menge an benötigten Daten eine große Herausforderung darstellt. Hier gilt es zu beachten, „man braucht ein großes Datenset zum Erlernen für die KI und man braucht zum Testen möglichst viele neue Fälle. Also große Datenmengen auf beiden Seiten, und die müssen gefunden oder geschaffen werden.“ (Kriechbaumer 2019) Dem Bias-Problem sei von Wasner hinzuzufügen, dass es sowohl negative als auch positive Bias gibt.

> *„Angenommen ich schick denselben CV zum selben österreichischen Unternehmen - Voest. Und dann schick ich denselben CV an Google. Die haben sehr unterschiedliche Kriterien wie Sie das bewerten. Diese Kriterien sind gewissermaßen auch Biases wie man das Ganze betrachtet. Was du natürlich nicht willst ist, dass aufgrund von Geschlecht Ethnizität Wohnort Dinge herauskommen. Das wäre die negativen Bias wo Unternehmen die sehr stark auf Compliance setzen und gute Governance haben sowieso aktiv dagegen arbeiten.“ (Wasner 2019)*

Zwei von drei sehen hier eine Abhilfe durch Supervised oder Reinforced Learning um Bias zu vermeiden. „Da hat man nur die Input Daten, z.B. Bilder oder Text, und das Ziel selbst ist nicht vom Menschen manuell gelabelled. Das ergibt sich dann aus mathematischen Funktionen oder Gleichungen.“ (Winter 2019) Kriechbaumer nennt neben den Inputmengen große Testdatenmengen als Weg, unerwünschte Nebeneffekte zu bereinigen.

Der zweite Aspekt des Sorgfaltsproblems behandelt den Marktdruck auf Unternehmen, durch den nötige Testphasen möglicherweise nicht eingehalten werden und biased oder unausgereifte KI zum Einsatz kommt. Zwei von drei Befragten unterscheiden hier die Märkte Europa und USA. Wasner sieht tendenziell Entwicklungen in den USA eher im privaten Sektor, in Europa hingegen eher im Forschungskontext. Beide sehen den Druck in den USA als gegebener und die Fehleranfälligkeit damit höher. Einer fügt hinzu, dass der betroffene Personenkreis hier größer ist. Positiv sehen hier beide, dass die Fehler zwar eher auftreten, diese durch das „Trial & Error“ Verfahren aber schneller gefunden und bereinigt werden. Wasner fügt hinzu, dass die Accountability im privaten Sektor in den USA eher gegeben ist, als im öffentlichen Sektor. Demgegenüber sehen beide geringeren menschlichen Schaden beim Vorgehen in Europa. Unternehmen die hier Künstliche Intelligenz einsetzen, sind hauptsächlich B2B. Firmen die Endverbraucher beliefern sind etwa Spotify oder Netflix, wo ineffiziente oder fehlerhafte Algorithmen quasi keine negativen Auswirkungen haben. Zudem wird als bereits leitende Regulierung die DSGVO genannt. Art. 22 Abs. 1 der DSGVO gibt Personen bspw. das Recht „nicht einer Entscheidung unterworfen zu werden, die auf einer rein automatisierten Verarbeitung ihrer Daten beruht“ (Dreyer, Schulz 2018: 19). Demgegenüber stehen Nachteile, die auf Entwicklung im Forschungskontext zurück gehen. Geringere Divergenz und Datensets, dadurch auch langsamere Fehlerfindung. Die Hypothese

1.1 (Deutschsprachige KI-Experten sehen überwiegend eine Bedrohung bei KI-Anwendungen durch das „Sorgfaltsproblem“) kann als verifiziert betrachtet werden, unter Berücksichtigung, dass hier eine „sehr starke“ Bedrohung gesehen wird.

10.3.5 Weitere Probleme

Weitere Problemfelder	Informationsproblem	Filter Bubbles
		Informiertheit der Endverbraucher
		Kosten- & Personalaufwand für Experten
		Sinnvolles Formulieren von Gesetzen
	Datenherausforderungen	Zu geringe Testdatenmenge
	Steigendes Ausmaß/Risko durch steigenden Einsatz	
	Selbstverbreitende KI	

Alle Befragten sehen hier grundsätzlich ein Informationsproblem. Zwei von drei sehen dadurch die Gesetzgebung gehemmt. „Jedoch befürchte ich würde das daran scheitern, dass man es nicht sinnvoll in ein Gesetz formulieren kann. Weil sobald man es schwammig formuliert, ist es wieder reine Auslegung der jeweiligen Firma. Und dann umso weiter etwas in die Privatwirtschaft reicht, umso kürzer wird die Testzeit sein. Und auch rein staatliche Programme sind an Budgets gebunden und müssen Ergebnisse liefern.“ (Kriechbaumer 2019) Einerseits sei hier sinnvolle Regulierung betroffen, andererseits jegliche Anweisung durch (politische) Entscheidungsträger, etwa in Bezug auf Einsatzmöglichkeiten. Einer fügt das Bedenken hinzu, dass das Filter Bubble Phänomen durch KI noch verstärkt würde.

> *„Also Preis, welche Produkte und Hotels werden dir empfohlen, eine Reise App in fünf Jahren wird dir den kompletten Trip vorschlagen. Je mehr du an ein System auslagerst, und je weniger datenmündig du bist und verstehst, was da abgeht, kann schon sein, dass diese Bubbles in großem gesellschaftlichem Stil einschlagen.“ (Wasner 2019)*

Steigende Einsatzbereiche	Formulierung Gesetze	FilterBubbles	Informiertheit Endverbrauc...	Kosten Experten	Testdatenmengen	Selbstverbreitende KI
		■				
■	■				■	■
■	■		■	■		

Wie bei den vorherigen Punkten ersichtlich wurde, fielen die Bewertungen teilweise unter der Prämisse gering aus, dass KI Anwendungen (vor allem in Europa) noch in eher trivialem Ausmaß im Einsatz sind. Zwei Befragte nennen folglich mit dem steigenden Einsatz solcher ein steigendes Risiko bzw. größeres Ausmaß von negativen Effekten. Kriechbaumer fügt erhebliche Bedenken hinzu, sollte eine sich selbstverbreitende KI zum Einsatz kommen. Weltweite Computer vernetzt würden eine viel größere Rechenleistung haben, als der Mensch jemals fähig wäre zu Denken. „Wir haben Intuition und Erfahrung, aber das ist dem Computer ab einem gewissen Punkt egal. Wenn er 20, 30 Tausend Szenarien durchrechnen

kann, hat er ziemlich schnell was er haben will. Da kann er Erfahrung und Intuition komplett auslassen. Mit der Holzhammer Methode, alles berechnen, irgendein Weg wird ihm dann schon gefallen.“ (2019) Er bringt das Beispiel von autonom verbreitenden Viren um 2000, welche großen Schaden hätten anrichten können. Mit der mittlerweile viel stärkeren Technologien könnten so Kleingruppen, etwa politisch motiviert, die Infrastruktur kompletter Länder lahmlegen.

10.3.6 Lösungsansätze

Informationskampagnen werden als Strategie genannt, um grundlegend Data Literacy, also mündigen Umgang mit den eigenen Daten, zu erreichen, und weitergehend faktisches Wissen über Künstliche Intelligenz zu fördern. Es wird hinzugefügt, dass viele Technologien in Europa (etwa im Bereich Sprachassistenz, aufgrund der vielen verschiedenen Sprachen) noch weit hinter US-amerikanischen Entwicklungen liegen. Folglich kann sich an dem dortigen Umgang orientiert und Gesetze für den europäischen Raum angepasst werden. Ein Experte hält die Gesetzformulierung jedoch für schwer umsetzbar. Zwei sehen großes Potential in Staatlicher Regulierung. Wie bereits erwähnt, bietet die DSGVO in Europa eine sinnvolle Rahmenbedingung. Weitere Schritte der EU sollen in den kommenden drei Jahren folgen. Zertifikate wie durch das TÜV ausgestellt sollen dies ergänzen. Global viel Potential wird in sektorspezifischen Regelungen gesehen. Gemeint ist, dass die Führenden einer Branche – als Beispiel seien hier etwa Google, Facebook oder Amazon genannt – gemeinsam verbindliche Regelungen erstellen. Als Schutzvorkehrungen sollen vorallem in kritischen Bereichen – wer bekommt eine Wohnung, den Kredit oder die Anstellung – weiterhin Menschen die Entscheidung in letzter Instanz treffen. Ein Recht, wie es in der DSGVO bereits verankert ist. Unter dem Punkt Blackbox Problem wurde besprochen, dass es Bereiche gäbe, in denen eine solche Anwendung durchaus Sinn mache und unproblematisch einsetzbar ist. Es ist (bist jetzt) nicht möglich, „eine KI für Alles“ zu programmieren. Anwendungen, die also speziell auf einen Task zugeschnitten sind, können diesen gezielt durchführen und sind einfacher zu überprüfen. Wendet man Modulare Netzwerke an, kann aus vielen taskspezifischen Submodellen ein großes Modell zusammengesetzt werden.

Informationskampagnen	Staatliche Regulierung	Taskspezifische KI	Sektorspezifische Richtlinien ...	Menschliche Entscheid...	Kombination ...
		■	■	■	■
■	■	■			
■	■		■		

Die Hypothese 2.1 (Deutschsprachige KI-Experten sehen mehrheitlich Potential in staatlicher (Co-)Regulierung, gehen jedoch nicht von einer zukunftsnahen Implementierung aus) wurde teilweise verifiziert. Die Experten sehen hier tatsächlich überwiegend Potential in staatlicher Regulierung. Einer sieht diese durch die DSGVO teilweise schon vorhanden, weitere Schritte bereits geplant, also zukunftsnah. Der zweite sieht diese ebenfalls kommen, eine zeitliche

Eingrenzung wurde hier aber nicht genannt. Lediglich der dritte Experte äußert noch große Zweifel an einer sinnvollen Umsetzung durch die Regierung.

11 Fazit

Im Zuge dieser Arbeit konnten grundlegende Probleme und Herausforderungen der Künstlichen Intelligenz in gesellschaftsrelevanter Hinsicht beschrieben werden. Das Sorgfaltsproblem – die Kombination aus Bias und Marktdruck – wurde hier mit Abstand als das wichtigste Problem identifiziert. Hier hat sich ebenfalls gezeigt, dass große Unterschiede zwischen dem US-amerikanischem und dem europäischen Marktraum bestehen. Man erwartete ein Abflauen des Hype Cycles und zukünftige Wahrnehmung von KI als „just another technology“ (Wasner 2019). Bestehender Technochauvinismus geht jedoch Hand in Hand mit fehlender Information, das bestehen des Info-Gaps weiterhin wird nicht ausgeschlossen. Physical Hacking und das Blackbox Problem konnten als weniger ausschlaggebend eingestuft werden.

Die Experteninterviews zeigten, dass staatliche Regulierung durchaus das Potential hat, genannte Probleme in Zukunft einzudämmen. Die DSGVO etwa wurde als kompetente Regulierung genannt. Informationskampagnen könnten das einstimmig beschriebenen Informationsproblem von Endverbrauchern mindern. Hier mangelt es an nicht nur an Grundverständnis von den Kompetenzen und Möglichkeiten durch KI, es beginnt bereits mit fehlender Datenmündigkeit des Einzelnen.

Es gilt zu beachten, dass in alle Bewertungen der momentan noch geringe Einsatz von „wichtigen“ KI-Anwendungen miteinbezogen wurde. Durch fortschreitenden Einsatz erwarten die Experten also auch ein zukünftig größeres Risiko. Gefordert ist hier große Sorgfalt bei der Wahl von Einsatzbereichen und Umsetzung.

12 Literaturverzeichnis

12.1 Buchquellen

Broussard, M. (2018): Artificial Unintelligence, How Computers misunderstand the world. The MIT Press, Cambridge, Massachusetts

Der Neue Brockhaus: Lexikon u. Wörterbruch in fünf Bänden u. e. Atlas. Siebente, völlig neubearbeitete Auflage. 3 J-Neu – F. A. Wiesbaden (Hrsg.) Wiesbaden 1985

Döring, N., Bortz, J. (2016): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. 5. Auflage. Springer Verlag Berlin Heidelberg.

Gläser, J., Laudel, G. (2010): Experteninterviews und qualitative Inhaltsanalysen. 4. Auflage. Springer Fachmedien. Wiesbaden

Kaplan, J. (2017): Künstliche Intelligenz. Eine Einführung. Oxford University Press 2016. MITP Verlag: Frechen

Kreutzer, R. T., Sirrenberg, M. (2019): Künstliche Intelligenz verstehen. Grundlagen – Use-Cases – unternehmenseigene KI-Journey. Springer Fachmedien Wiesbaden GmbH.

Kreutzer, R., Sirrenberg, M. (2019): Künstliche Intelligenz verstehen: Grundlagen Use-Cases – unternehmenseigene KI-Journey. Springer Fachmedien: Wiesbaden

Müller, A. C., Guido, S. (2017): Einführung in Machine Learning mit Python. Praxiswissen Data Sciene. Mediea-Print Informationstechnologie. Paderborn.

Russell, S., Norvig, P. (2003): Artificial Intelligence. A Modern Approach. Second Edition. Prentice Hall. Pearson Education International. USA

Russell, S., Norvig, P. (2014): Artificial Intelligence. A Modern Approach. Third Edition. Pearson Education Limited., US.

12.2 Onlinequellen

Angwin, J., Larson, J., Mattu, S., Kirchner, L. (2016): Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (02.05.2019)

Autor, D. H. (2014): Polanyi's Paradox and the Shape of Employment Growth. NBER Working Paper No. 20485. https://economics.mit.edu/files/9835 (17.05.2019)

Blöbaum, B., Nölleke, D., Scheu, A. M. (2016): Das Experteninterview in der Kommunikationswissenschaft. In: Handbuch nicht standardisierte Methoden in der

Kommunikationswissenschaft. Stefanie Averbeck-Lietz Michael Meyen (Hrsg.). Springer Fachmedien: Wiesbaden.

Brühwiler, C. F. (2013): Wer hat Angst vor Ayn Rand? In: Schweizer Monat. Ausgabe 1004 – März. https://schweizermonat.ch/wer-hat-angst-vor-ayn-rand/# (20.04.2019)

Bruxmann, P., Schmidt, H. (2019): Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens. In: Künstliche Intelligenz. Mit Algorithmen zum wirtschaftlichen Erfolg. Peter Buxmann, Holger Schmidt (Hrsg.). pp. 3-19 https://link-springer-com.uaccess.univie.ac.at/content/pdf/10.1007%2F978-3-662-57568-0_1.pdf (25.04.2019)

DARPA (2014): The DARPA Grand Challenge: Ten Years Later https://www.darpa.mil/news-events/2014-03-13 (17.05.2019)

De Mauro, A., Greco, M., Grimaldi, M. (2015): What is big data? A consensual definition and a review of key research topics. AIP Conference Proceedings. http://big-data-fr.com/wp-content/uploads/2015/02/aip-scitation-what-is-bigdata.pdf (26.04.2019)

Draude, C. (2001): Introducing Cyberfeminism. Old Boys Network, Reading Room. https://www.obn.org/inhalt_index.html (22.04.2019)

Dreyer, S., Schulz, W. (2018): Was bringt die Datenschutz- Grundverordnung für automatisierte Entscheidungssysteme? Bertelsmann Stiftung. https://www.hans-bredow-institut.de/uploads/media/Publikationen/cms/media/p4ymg73_BSt_DSGVOundADM_dt.pdf (28.08.2019)

Eykholt, E., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C., Prakash, A., Kohno, T., Song, D. (2018): Robust Physical-World Attacks on Deep Learning Visual Classification. CVPR. https://arxiv.org/pdf/1707.08945.pdf (02.04.19)

Fuchs, C. (1999): Der Feminismus Donna Haraways und die materialistisch-feministische Kritik der Postmoderne http://fuchs.uti.at/wp-content/uploads/infogestechn/haraway.html (22.04.2019)

Haraway, D. (1995): Feminismus im Streit mit den Technowissenschaften. In: Haraway, Donna: Die Neuerfindung der Natur. Primaten, Cyborgs und Frauen. Frankfurt a. M. und New York 1995. S. 33- 72. (Erstmals erschienen unter: Haraway, Donna: Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980's. In: Socialist Review 80. 1985. S. 65-108.) http://www.medientheorie.com/doc/haraway_manifesto.pdf (22.04.2019)

Hartmann, M., Wimmer, J. (2011): Digitale Medientechnologien. Vergangenheit – Gegenwart – Zukunft. VS Verlag für Sozialwissenschaften, Springer. Wiesbaden

Heßler, M. (2017): Der Erfolg der „Dummheit“. Deep Blues Sieg über den Schachweltmeister Garri Kasparov und der Streit über seine Bedeutung für die Künstliche Intelligenz-Forschung. NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin. Vol. 25., 1., pp. 1-

33 https://link-springer-com.uaccess.univie.ac.at/article/10.1007/s00048-017-0167-6#Sec1 (25.04.2018)

Higgins, J. (1987): Artificial Unintelligence: Computer Uses in Language Learning. TESOL Quarterly Vol. 21, No. 1, pp. 159-167. (https://www-jstor-org.uaccess.univie.ac.at/stable/3586364?sid=primo&origin=crossref&seq=1#metadata_info_tab_contents) (26.03.2019)

High Leve Expert Group on Artificial Intelligence (AI HLEG) (2019): A Definition of AI: Main Capabilities and Disciplines. Definition developed for the purpose of the AI HLEG's deliverables. European Commission. https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines (17.05.2019) *Für die Liste der 52 Autoren siehe:* https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

Hunt, E. (2016): Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. The Guardian. https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter (02.05.2019)

Kaiser, R. (2014): Qualitative Experteninterviews. Konzeptionelle Grundlagen und praktische Durchführung. Springer Fachmedien. Wiesbaden https://link-springer-com.uaccess.univie.ac.at/content/pdf/10.1007%2F978-3-658-02479-6.pdf (23.04.2019)

Krüger, J., Lischka, K. (2018): Damit Maschinen den Menschen dienen. Lösungsansätze, um algorithmische Prozesse in den Dienst der Gesellschaft zu stellen. Bertelsmann Stiftung (Hrsg.) (https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Algorithmenethik-Loesungspanorama.pdf) (02.04.19)

Kuckartz, U. (2010): Einführung in die computergestützte Analyse qualitativer Daten.

Liebold, R., Trinczek, R. (2009): Experteninterview. In: Handbuch Methoden der Organisationsforschung. Quantitative und Qualitative Methoden. Stefan Kühl, Petra Stodtholz, Andreas Taffertshofer (Hrsg.). VS Verlag für Sozialwissenschaften, GEV Fachverlage GmbH, Wiesbaden. pp 32-56 https://link-springer-com.uaccess.univie.ac.at/content/pdf/10.1007%2F978-3-531-91570-8.pdf (23.04.201)

Loosen, W. (2016): Der Leitfaden – Eine unterschätzte Methode. In: Handbuch nicht standardisierter Methoden in der Kommunikationswissenschaft. Stefanie Averbeck-Lietz, Michael Meyen (Hrsg.) Springer Fachmedien, Wiesbaden. pp 139-155

Manhart, Klaus (2017): Eine kleine Geschichte der Künstlichen Intelligenz. https://www.computerwoche.de/a/eine-kleine-geschichte-der-kuenstlichen-intelligenz,3330537 (25.04.2019)

Mayring, P. (2015): Qualitative Inhaltsanalyse. Grundlagen und Techniken. Beltz Verlagsgruppe

Mitchell, T. M. (1997): Machine Learning. McGraw-Hill Science/Engineering/Math. http://profsite.um.ac.ir/~monsefi/machine-learning/pdf/Machine-Learning-Tom-Mitchell.pdf (17.05.2019)

Murphy, K. P. (2012): Machine Learning. A Probabilistic Perspective. The MIT Press Cambridge, Massachusetts. US.

Reinhold, A. (2015): Das Experteninterview als Methode zur Wissensmodellierung. In: Information – Wissenschaft & Praxis. Margarita Reibel-Felten (Hrsg.) Band 66, Heft 5-6. pp 327-333 https://www-degruyter-com.uaccess.univie.ac.at/downloadpdf/j/iwp.2015.66.issue-5-6/iwp-2015-0057/iwp-2015-0057.pdf (23.04.2019)

Searle, J. R. (1980): Minds, brains, and programs. Behavioral and Brain Sciences 3 (3): 417-457. In: Cambridge University Press U.K./U.S (http://cogprints.org/7150/1/10.1.1.83.5248.pdf) (24.03.2019)

Shalev-Shwartz, S., Shammah, S., Shashu, A. (2017): On a Formal Model of Safe and Scalable Self-driving Cars. Mobileye (https://arxiv.org/pdf/1708.06374.pdf) (02.04.19)

Shortlife, E., Davis, R., Axline, S., Buchanan, B., Green, C. C., Cohen, S. (1974): Computer-Based Consultations in Clinical Therapeutics: Explanation and Rule Acquisition Capabilities of the MYCIN System. In: Computers and Biomedical Research 8. pp. 303-320 http://www.inf.ufpr.br/alexd/ARTIGOS_IA/Shortliffe_et_al_1975.pdf (25.04.2019)

Sollfrank, C. (2000): The truth about cyberfeminism. https://www.obn.org/inhalt_index.html (22.04.2019)

Stoltenhoff, A., Raudonat, K. (2018): Digitalisierung (mit)gestalten – was wir vom Cyberfeminismus lernen können. Strategien und Ansätze einer aktivierenden Perspektive auf Informations- und Kommunikationstechnologien im 21. Jahrhundert. In: GENDER – Zeitschrift für Geschlecht, Kultur und Gesellschaft, 02/23/2018, Vol 10(2), pp. 128-142 file:///C:/Users/user/Downloads/31361-32772-1-PB%20(1).pdf (21.04.2019)

Streitz, N. (2019): Beyond ‘smart-only‘ cities: redefining the ‘smart-everything’ paradigm. Journal of Ambient Intelligence and Humanized Computing 10:791-812.

Tschohl, C. (2014): Industrie 4.0 aus rechtlicher Perspektive. Elektrotechnik & Informationstechnik 131/7: 219-222. (https://link-springer-com.uaccess.univie.ac.at/content/pdf/10.1007%2Fs00502-014-0228-7.pdf) (02.04.19)

Turing, A. M. (1950): Computing Machinery and Intelligence. In: Mind, New Series, Vol. 59, No. 236, pp. 433-460, Oxford University Press [Hrsg.] (http://phil415.pbworks.com/f/TuringComputing.pdf) (24.03.2019)

12.3 Nicht Wissenschaftlich (Nachrichtenberichterstattung)

Beuth, P. (2016): Twitter-Nutzer machen Chatbot zur Rassistin. Zeit Online. https://www.zeit.de/digital/internet/2016-03/microsoft-tay-chatbot-twitter-rassistisch (01.05.2019)

Breitinger, M. (2016): Tod durch Software. In: Zeit Online https://www.zeit.de/mobilitaet/2016-07/autopilot-autonomes-fahren-tesla-faq (19.04.19)

Castelvecci, D. (2016): Eine tückische Blackbox. In: Spektrum.de https://www.spektrum.de/news/eine-tueckische-blackbox/1429906 (01.05.19)

Die Presse (2018): Trampender Roboter hitchBOT findet letzte Ruhe im Computermuseum. https://diepresse.com/home/techscience/5483482/Trampender-Roboter-hitchBOT-findet-letzte-Ruhe-im-Computermuseum (02.05.2019)

Greenberg, A. (2015a): Hackers remotely kill a jeep on the highway – with me in it. Für: WIRED. https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/ (02.04.19)

Greenberg, A. (2015b): After Jeep Hack, Chrysler recalls 1.4M vehicles for bug fix. https://www.wired.com/2015/07/jeep-hack-chrysler-recalls-1-4m-vehicles-bug-fix/ (02.05.19)

Gruber, K (2018): Nicht überall ist Technik ein „Männerfach“ In: Science.ORF.at https://science.orf.at/stories/2899355/ (22.04.2019)

Gunning, D. (2017): Explainable Artificial Intelligence (XAI) DARPA/I2O Program Update November 2017. https://www.darpa.mil/attachments/XAIProgramUpdate.pdf (26.06.2019)

Hering, S., Schultz, N., Galert, T. (2018): Menschenwürde im Angesicht neuer Technologien. Deutscher Ethikrat. https://link-springer-com.uaccess.univie.ac.at/content/pdf/10.1007%2Fs00481-018-0511-y.pdf (10.08.2019)

Kharpal, A. (2017): Stephen Hawking says A.I. could be ‘worst event in the history of our civilization’. CNBC https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html (29.08.2019)

Moll, S. (2016): Präzise berechneter Rassismus. Zeit Online. https://www.zeit.de/gesellschaft/zeitgeschehen/2016-06/algorithmen-rassismus-straftaeter-usa-justiz-aclu/komplettansicht (02.05.2019)

Ramge, T. (2018): Mensch und Maschine. Wie Künstliche Intelligenz und Roboter unser Leben verändern. Reclam: Stuttgart.

Revell, T. (2017): Google DeepMind NHS data deal was “legally inappropriate”. New Scientist https://www.newscientist.com/article/2131256-google-deepmind-nhs-data-deal-was-legally-inappropriate/ (26.06.2019)

Zack, K. (2016): parrot or guacamole? https://twitter.com/teenybiscuit/status/707727863571582978 Karen Zack via Twitter. (17.05.2019)

13 Abbildungsverzeichnis

14 Anhang

14.1Interviewtranskripte

Die Bezeichnung S im Folgenden steht jeweils für die Forscherin, die andere Abkürzung für den jeweils Interviewten.

Interview 1: Wasner Clemens am 01. Juli 2019

W: Und die Arbeit selbst ist über Künstliche Unintelligenz.

S: Genau, letztes Jahr erschien ein Buch mit demselben Namen und es beschäftigt sich damit. Quasi dem falschen Einsatz künstlicher Intelligenz. Mit falschem Einsatze ist der Einsatz gemeint, wo es eigentlich nicht die effizienteste Lösung ist oder bestimmte Personengruppen benachteiligt werden oder wo der Mensch nicht mehr die endgültige Kontrolle hat. Ich habe in der Arbeit verschiedene Gründe bzw. Situationen herausgearbeitet und bin nun gespannt auf deine Meinung.

W: Spannend, das erste ist ja bisschen ein Realitätscheck, du verwendest ja nicht für alles einen Computer um es mitzuschreiben, sowie jetzt hast den Stift in der Hand. Genau ist es ein Unterschied ob man normale Algorithmen oder KI verwendet. Die Wahl des Werkzeugs sowie die Sache selbst wird nicht in Frage gestellt. Während das zweite eher ein Datenthema ist und weniger ein Methodenthema. In erster Linie, wenn du von Diskriminierungen sprichst, ist das ja ein Datenthema. Sprich wenn du bei Google oder Amazon festgestellt hat das weibliche Mitarbeiterinnen im Schnitt um 30 Prozent geringeren Gehaltssprung machen im Schnitt, da hat KI damit überhaupt nichts zu tun. Du siehst durch die Daten, dass es so ist, und dass das System verstärkt wird. Wie du in den Wald rufst, so kommt es heraus

S: Das ist eigentlich schon mein erster Grund für Künstliche Intelligenz. Quasi künstliche Intelligenz - ich habe zwei Sachen zusammengefasst. Technochauvinismus, davon spricht sie in dem Buch und meint eben die Überzeugung, dass KI einerseits objektiv ist, weil sie auf mathematischen Berechnungen basiert und andererseits einfach die beste Lösung darstellt. Deswegen ist so viel möglich, dass sie am besten immer eingesetzt werden sollte, dass dadurch unrealistische Erwartungen gestellt werden. Und andererseits diesen KI-Hype Cycle. Dass diese zwei Faktoren zusammenspielen und dadurch unrealistische Erwartungen geschürt werden.

W: Streng genommen befinden wir uns im Moment im Deep-Learning Hype Cycle, nicht im KI Hype Cycle, das Wort KI gibt es seit 1955, die Disziplin 1956, je nachdem wie du rechnest

sind wir jetzt im sechsten, siebten oder achten Hype im Moment. Mit Übersetzung Systemen das Englisch ins Russische übersetzen und wieder zurück sollten war das schon so, das hat nie wirklich funktioniert. Es hat in Kombination mit Realfilm gemacht, quasi dass KI ein Konzept von der Realität hat - das hat auch nie über drei Objekte skaliert. Dann wars über das ganze Thema der Suchalgorithmen. Also Deep Blue, Kasparow 1997. Das hat man damals als künstliche Intelligenz bezeichnet heute ist das Suchalgorithmus. Damals hat man geglaubt, dass es künstliche Intelligenz die Disziplin ist. Wenn man dasselbe Interview in 15 Jahren durchführt wird man sagen Deep Learning - damals hat man geglaubt das ist Künstliche Intelligenz.

S: Wenn man jetzt den Durchschnittsbürger hernimmt. Das Verständnis eines Durchschnittsbürgers von dem Begriff künstliche Intelligenz. Gehen Sie davon aus, dass die Uninformiertheit und die Erwartungen die durch diese halbseitige Information als Grund für künstliche Unintelligenz zu betrachten wäre, sogar eine Bedrohung sein könnte?

W: Im technischen Diskurs. Generell, dass technische Themen in populären, alltäglichen Diskurs eingeflossen sind ist ein relativ junges Phänomen. Wenn ich die Zeit zurückdrehen als PCs eingeführt worden sind - das war ein extremes Nischenprogramm. vielleicht noch im Corporatebereich, aber sicher nicht im Privatbereich. So richtig losgegangen ist es erst mit Smartphones. selbst als Internet eingeführt worden ist Ende der 90er Anfang der Nullerjahre war das erst ein Special Interest Programm. In einer Gruppe von zehn Leuten, wenn man zusammengesessen ist haben vielleicht zwei über Internet geredet. Aber heute dank Smartphones soziale Netzwerke hat jeder Meinung über Technik zu jeder Zeit. Amerikaner bezeichnen das als Consumeration of Technology. Quasi, dass jeder Consumer dazu Meinungen abgibt. Das erzeugt extrem überzogene Erwartungshaltungen. Ich finde das erste Opfer von Erwartungshaltungen war bereits das iPad 2010/2011, weil eigentlich jede Technologietrend dann am Smartphone gemessen wird. Aber das ist ja Blödsinn. Das Smartphone ist das erfolgreichste Produkt ever, das mit Abstand. Es gibt nichts Besseres im Business außer vielleicht das Öl Business das sagt ja schon alles aus. Tatsächlich kann ich mich noch gut erinnern 2010 als das IPad gekommen ist, Post PC Zeitalter hats geheißen, wir haben alle nur mehr Touchscreens in Zukunft nichts davon ist eingetreten. Apple Watch ist das Nächste, Apple Watch ist mit Abstand also mit Abstand der größte Uhrenhersteller gilt trotzdem als Flop, weil es nicht so erfolgreich ist wie das Smartphone. Chatbots. Das war wahrscheinlich der kürzeste Hype den es gegeben hat. Ist glaub ich im April vorgestellt worden von Facebook. Anfang der Woche hat es noch geheißen CNN und Delta Airlines mit Chatbots. Mittwoch Donnerstag in derselben Woche haben die Medien dann schon geschrieben das funktioniert eigentlich gar nicht. Es ist auch total unintuitiv, weil es wie die Command line ist, du schreibst was du willst, es ist counter - intuitiv zu dem wie heute der Konsum von Medien und von Technologie funktioniert. was ja sehr graphisch ist. Und.

Dementsprechend ist KI das nächste Thema. Was halt auffällt ist, ich würde es nicht in dieselben Hype-Cycle einordnen wie AR und VR oder Blockchain. Es gibt bei KI schon extrem viel Sachen die halt schon eine oder andere Technologie von KI verwendet und die man da dazu zählt. Aber es ist bereits implementiert. Bei Blockchain gibts eigentlich null Implementierung und VR kommt alle zehn Jahre wieder. Von echtem VR sind wir schon noch zehn Jahre entfernt.

W: Was immer gerne durchmischt wird sind jetzt zwei Unterschiedliche - Im Prinzip hast bei KI die einerseits rein Datengetriebenen - quasi Input geht rein - Output kommt raus annotierte Daten, supervised learning. Kann jetzt sein, ich schieß ein Bild, das sagt mir ist das eine Studentin oder eine Katze oder Pfefferstreuer oder Kreditanträge die reinkommen und du weißt - wie ist das Rating, das ist rein Datengetrieben, da hast auch oft Blackbox Information drinnen. womit es aber vermischt wird – wenn du jetzt an Reinforced Learning denkst - bei AlphaGo z. B. im Einsatz - da hast du sehr abgegrenzte Bereiche wo Menschen halt weniger gut sind, so abgegrenzte Sachen können Maschinen perfekt lösen. Aber die zwei haben miteinander nichts zu tun. Denn im zweiten gibt's kein Bias, da wirst du auch keinen Bias haben. Im andern schon, also auch positiven Bias, das ist nicht immer negativ. deine Daten müssen sogar immer Bias haben, denn das ist ja gewissermaßen das Wissen der Organisation.

S: Wenn die KI jetzt zum Beispiel statistisch richtig liegt, wenn es z.B. um einen Bewerberpool geht und um den passenden Bewerber für eine Firma. Dann ist alles datengetrieben, dann ist da Bias drinnen

W: Da hast du Bias drinnen, genau. Es gibt nur positiven und negativen. Angenommen ich schick denselben CV zum selben österreichischen Unternehmen - Voest. Und dann schick ich denselben CV an Google. Die haben sehr unterschiedliche Kriterien wie Sie das bewerten. Diese Kriterien sind gewissermaßen auch Biases wie man das ganze betrachtet. Was du natürlich nicht willst ist, dass aufgrund von Geschlecht Ethnizität Wohnort Dinge herauskommen. Das wäre die negativen Bias wo Unternehmen die sehr stark auf Compliance setzen und gute Governance haben sowieso aktiv dagegen arbeiten. Und dann halt auch die ersten sind die solche Fälle aufdecken. Was ich Früher erwähnt habe mit dem Gehaltssprung und den Mitarbeiterinnen, es ist ja kein Zufall, dass das von den Organisationen als erstes aufgedeckt wird die sehr starke Prozesse in dem Bereich haben, damit das nicht passiert. Anderswo wird es vielleicht gar nicht auftreten.

S: Noch kurz zum Hypce-Cylce: Wenn zum Beispiel in den ersten Todesfall denkt durch autonomes Fahren 2016, könnte man theoretisch davon ausgehen wenn nicht diese unrealistischen Erwartungen da wären wenn der Betroffene jetzt nicht völliges Vertrauen gehabt hätte, vielleicht eingegriffen hätte, könnte man annehmen dass das anders ausgegangen wäre, dass Die ganze Bevölkerung damit anders umgehen würde.

W: Definitiv, der Todesfall und auch die nachgefolgt sind, sind Fälle die der Fahrlässigkeit von Tesla zuzuschreiben sind, weil hier was als autonomes Fahren vermarktet wird was nicht autonomes Fahren ist. Es ist nach wie vor ein Assistenzsystem, hat mit autonomem Fahren nix zu tun. Genau da ist es natürlich hanebüchen zu sagen - Google ist weit in der Forschung vor Tesla - Wenn Google sagt das dauert weltweit noch 30 Jahre bis es funktioniert ist es hanebüchen, wenn Tesla sagt, das geht jetzt schon. Google hat unendlich Geld und Ressourcen auch bessere, und die besseren Leute. Das ist dem Marketing zuzuschreiben, dass die Erwartungshaltung geweckt wird. Das ist eigentlich Marketingopfer.

S: Denkst Du, dass sich das in Zukunft ändern wird. Die generelle Informiertheit der Bevölkerung über Techniken der künstlichen Intelligenz oder generell Die Erwartungshaltung und der Umgang der Bevölkerung anders wird? Ob es sicherer sein wird oder sich so weiter entwickeln wird.

W: Jain - ich glaub das große Thema ist Data Literacy, da muss man anfangen. Wenn wir jetzt diese ganzen Daten getriebene Beispiele denken, wie vorher mit den Lebensläufen und so weiter und so fort, auch Artikel für den Kiosk. Das ist Medienkompetenz und Data literacy. Das ist. Erst im Nachgang die Frage für welche Technik zum Einsatz kommt, weil es ja im Prinzip relativ egal ist. Das beantwortet die Frage wahrscheinlich nicht aber ich glaub, dass die Vorbehalte deshalb herkommen, das allererste worüber wir vorher schon geredet haben war ja mit Fake News. Und wenn ich mir anschaue was das größte Movement ausgelöst hatte zur US-Wahl 2016, ich glaub eine der bekanntesten Geschichten war, dass Hillary Clinton aus einem Keller in einer Pizzeria in Washington D.C. Kinderprostitutionsring betreibt. Eine der bekanntesten Storys vom letzten Wahlkampf. Verrückt, die Pizzeria wurde sogar wirklich gestürmt. Das ist dann total egal ob die Leute wissen was KI ist oder nicht - da muss man dann woanders ansetzen. Ich glaub, dass dort der größte Aufholbedarf ist. Zu wissen was die Methode kann ist schön aber wenn du grundlegende Sachen in Frage stellst... Wir sagen immer es ist das Zeitalter wo Expertise nichts mehr wert ist und Fakten - postfaktisches Zeitalter, dann ist KI vielleicht nur noch eine Nuance von dem Ganzen aber es wird gesellschaftlich nicht wirklich etwas ändern.

S: Letzte Frage dazu - wenn du es jetzt auf einer Skala einordnen müsstest - wie bedrohlich es einzustufen ist, von sehr stark, stark, Kaum, etwas oder überhaupt nicht. Diesen Technochauvinismus und diesen Hype-Cycle.

S: Ich glaub, dass sich der Technochauvinismus gerade massiv abkühlt, wenn du schaust der öffentliche Diskurs wird zum Großteil von. Drei Beispiele. Dominiert das ist einerseits anderseits Health Care, andererseits autonom fahrende Autos und Sprachassistenten. Das deckt wahrscheinlich 90% ab. Das korreliert Eins zu eins mit den Firmen die dahinter stecken nämlich IBM, Google/Tesla mittlerweile/Uber und die Sprachassistenz ist Amazon geschuldet. Wenn man die Zeit jetzt zurückdreht zwei Jahre ist es ähnlich wieder wie beim Smartphone oder IPad. Da hat es geheißen Jede Firma braucht einen Alexa Skill in Zukunft. Und Watson - zunächst löst es Health und dann wird es den Rest der Wirtschaft auch noch lösen. Und 2022 sitzen wir alle in autonom fahrende Autos. Nichts davon eingetreten. Und es führt dazu, dass es schon eher in der Wahrnehmung zu einem normalen IT-Thema wird was die Technik betrifft. Und was auch dazu kommt ist, dass es nicht mehr so weit weg ist, sondern dass das auch in Smartphones drinnen ist. Beim Smartphone kommt der Modus Porträtfoto zum Einsatz, weil der Zigmillionen Bilder antrainiert ist. Was ist ein Kopf, was ist eine Wand, was zeichne ich unscharf was mach ich scharf. Das wird den Leuten langsam bewusst was dort alles reinfällt, das sind halt in 99 von 100 Fällen relativ triviale Sachen wo die Leute dann sagen oh das ist KI, das habe ich ja gar nicht gewusst. Oder die Recommendation-engine bei Netflix. Die funktioniert, die bei Amazon funktioniert nicht. Das Beispiel wird gerade geradegerückt.

S: Also eher als etwas, kaum oder überhaupt nicht einzustufen.

W: Darf ich die Frage noch einmal lesen?

S: Sehr gern. Genau, also Technochauvinismus als Grund für KUI. Broussard hat z.B. den Vergleich gebracht: im amerikanischen Schulsystem wo sie eine Aufstellung macht, wieviel würde es Kosten und bringen wenn man jetzt Bücher einsetzt mit einem Bruchteil von dem Anschaffungswert von IPad und einer Lebensdauer von fünf Jahren, womit Erwiesenermaßen Schüler besser lernen, im Vergleich zur Ausstattung aller Schüler mit IPad, den anfänglichen Trainingskosten, Instandhaltungskosten, Reparaturkosten und so weiter und so fort.

W: Das ist der typische Solution and Search for a problem Ansatz, aber Ipads sind in Amerika sowieso gescheitert weil die alle auf Chromebooks setzen weil Chromebooks werden

millionenfach verkauft in Amerika, das heißt Google Chrome - das sind Laptops die kriegt man auch beim Saturn, für zwei, 300 Euro drauf läuft im Prinzip der Chrome-Browser mit einem File-System, darauf gibt es keine Viren. Man kann sehr einfach Classrooms einrichten sämtliche Software wie Moodle und was gibts dort alles millionenfach im Einsatz. Und Apple hat die ganze Infrastruktur nicht. Die sagen wir haben das shiny device, macht etwas damit. Das muss schon in Verbund mit Lehrervereinigungen passieren.

W: Also bin wieder abgeschweift. Also der Grad der Informiertheit nimmt zu. Ich sehe aber was noch Verbesserungswürdig ist, die Unterscheidung zwischen Artificial generell Intelligence und narrow ist- die normale AI. Das wird in den Medien viel zu wenig stark aufgerollt. Man hört doch oft selbstlernende Systeme und ladida - da wird nämlich immer wieder Gruppe1 und Gruppe 2 vermischt. Nur weil ich etwas habe das Text übersetzen kann und automatisch auswerten und auf der anderen Seite gibts AlphaGo, was ein Spiel gut spielen kann, heißt es nicht, dass ich etwas machen kann, das beides zusammenlegt. In kann nicht ein System machen das 50 Leitfäden liest und den 51. selber schreibt, das existiert nicht und dieser Bruch wird nicht dargestellt.

S: ist es aber konkret eine Bedrohung für die Bevölkerung das die Bevölkerung so falsch informiert und uninformiert ist. Bzw. so techverliebt.

W: Aus Standort Sicht ist es in Europa schon eine Bedrohung. Weil es die Leute einlullt. So auf die Art wir sind ja so arm, die Chinesen haben viel mehr Geld und die Amerikaner haben noch mehr Geld und in Österreich und in Europa gibt's keine KI Firmen oder IT-Firmen. Also Ziehen wir uns zurück auf das Thema Ethik. Im Englischen nennt man das Virtue Signaling. Man stellt zur Schau, dass man erhabener ist, dass man moralischer auf anderen Sphären agiert. aber wenn das nicht auf einem wirtschaftlichen Fundament passiert wird dir niemand Gehör schenken. Die Gefahr ist, wenn du dich nur darauf konzentriert, nicht auf den ganzen Rest. Dann Wirst du weder das eine noch das andere haben. Ich bringe immer das Beispiel der Automobilindustrie. Du hast In der Automobilindustrie eigentlich auch so einen Technochauvinismus gehabt, der wirklich gechallenged wurde das erste Mal mit der Zeit vom sauren Regen und Ozonloch mittlerweile mit der Erderwärmung. Das hat eigentlich Sehr starke Gegenbewegung in der Bevölkerung ausgelöst. Die Entscheidungsträger und auch die Autoindustrie zum Umdenken bewegt, es ist eine Symbiose. Die Autohersteller wehren sich dagegen nicht einmal extrem dagegen, auch wenn sie es könnten, viel ärger zurückdrängen. Was da letztendlich passiert - in Japan ist nämlich das Gleiche passiert - dass aus diesem Dreiergespann Bevölkerung Politik Wirtschaft - dass da abgeleitet wird wie das Ganze sich verantwortungsvoll zu entwickeln hat, und dadurch eigentlich schon eine Vorbildwirkung für den Rest der Welt entstanden ist. in dem Fall war es die Achse Japan -

EU. Und in Amerika hat das dann auch eingesetzt nach der Finanzkrise. Dass man stärker wieder auf Motoren mit geringerem Verbrauch gegangen ist - Fahrzeuge. In China sowie, die sind total auf Elektroautos umgeswitched. Das ist aber nur deshalb möglich, weil in Europa eine starke Autoindustrie ist und in Japan. Wenn Europa nur hergegangen wäre und gesagt hätte wir regulieren das stärker, dann hätten wir die amerikanischen Autos kriegt aber hätten mehr dafür gezahlt. Genau das muss Hand in Hand gehen.

S: Einmal noch kurz zur Skaleneinordnung: wo darf ich es einordnen?

W: Etwas. Ich finde es ist schon sehr stark im Abflauen. Ich glaube sogar, dass es vor einem Jahr noch anders gewesen wäre. jetzt. Im Moment schlägt es eigentlich eher in die Richtung, dass KI e nichts Besonderes ist. Es gibt Ausreißer wie die künstlich generierten Gesichter. Oder AlphaGO. Aber letztendlich wird es Wahrgenommen und dementsprechend just another technology.

S: Alright. Das nächste Problem: Ich habe es zusammengefasst als Sorgfaltsproblem. Es fasst einerseits zusammen das Bias-Problem. Die bewussten und unbewussten durch Entwickelnden aber auch Trainingsdaten und andererseits das wo ich das Gefühl habe dass es vor allem in der amerikanischen Wirtschaft passiert: Der Druck von privaten Unternehmen so schnell wie möglich mit KI- Anwendungen auf den Markt zu gehen und dass dadurch die notwendigen Testphasen nicht stattfinden oder nicht lange genug stattfinden.

S: Ein. Sehr. Komplexes Thema. Und zwar. Einerseits alle bekannten Beispiele wo es negativen Bias gibt kommen aus der Privatwirtschaft eben aufgrund von diesen ganzen Compliance und LGBTQ konformen Prozesse die alle aus den USA kommen. Da kannst du schon sagen. Dass der marktwirtschaftliche Druck so groß ist, dass die Firmen sich darum kümmern, weil sie es müssen. Stell dir vor Die nächste Sammelklage bei einem großen Immobillienbetreiber, weil eine Minority nicht zum Zug kommt. Das ist Shit hits the Fan. Das haben wir schon am Schirm. Was interessant ist, Die. öffentliche Einstellung oder die Einstellung der Bevölkerung was man dem öffentlichen Sektor zutraut oder nicht. Es gibt nicht wenige in den USA die sagen es ist besser, wenn Privatfirmen das machen, also AI-Lösungen einführen und Datenspeichern und so weiter, wegen accountability. Denn das gibts im öffentlichen Bereich so nicht. Die sagen, wenn so wie die Pensions-Datenbanken oder den Veteranen - fast jedes Quartal wird etwas gehacked wo Mindestens 100 Millionen Leute betroffen sind. Das ist im öffentlichen Bereich wurscht. Da wird dann halt den Daten Verantwortlichen auf die Finger geklopft aber niemand wird wirklich jemals verknackt. Im privaten Bereich schon. Da Gibts dann Sammelklagen und so weiter und so fort. Das ist halt

der andere Zugang zum öffentlichen Sektor in Amerika. Die sagen, wenn Daten wirklich geschützt werden müssen gibst du es besser den Privatfirmen, weil die wirklich darauf aufpassen.

S: Die Privatfirmen wissen also bereits um das Problem, passen besser auf und das kehr sich um. Ich bin z.B. ausgegangen von COMPAS, das war 2016 diese Risk Rating KI für die US-amerikanische Justiz gewesen. Das war ja eine von den ersten Fällen, wo KI sofort am Markt in Betrieb war, wo im Nachhinein aufgedeckt wurde, dass auf Daten wie Wohnort und Hautfarbe als Beurteilungsmaßstab zurückgegriffen wird.

W: Aber das ist genau der Punkt. Es kommt schnell auf. Und wird dann gefixed. Sowohl ein Hack kommt schnell auf, als auch wenn es nicht so lauft wie es soll. Du kannst das Ganze einklagen in Amerika. Ist relativ einfach. Selbst regulierende System und da trauen die Amis viel mehr der Privatwirtschaft als die öffentlichen Stellen. Das ist Amerikas Spezifikum. In andere Länder ist es wieder genau umgekehrt Es ist nämlich der große Unterschied zwischen UK und USA. Zum Beispiel. Obwohl beide angelsächsisch sind. In Europa traut man prinzipiell die Privatfirmen nicht, weil das Ganze Bewusstsein für Hacks bei uns nicht ausgeprägt ist. Es wäre mir auch nicht Bekannt, dass im großen Stil in Österreich einmal ein Hack gewesen wäre.

S: Auch nicht so attraktives Ziel, Österreich.

W: Naja... 8, irgendwas Millionen Leute ist schon interessant. Da kann man schon was Herausfinden. Aber Du hast diese permanenten Data Leaks nicht, deswegen traut man das dem Staat noch zu, einerseits. Und andererseits haben wir auch keine Sammelklagen. Also. Die Leute fühlen sich bei uns die Konzerne gegenüber ohnmächtiger gegenüber den USA. Das kann ich nicht empirischen Beweisen, aber von meinem Gefühl her. Weil es bei uns die class action law suits nicht gibt.

S: Also würdest du für die USA spezifisch sagen, dass diese Bias Problem und dieser Druck auf der Privatwirtschaft sich quasi von selbst löst durch den Markt.

W: Man traut da eher dem Markt zu, das Ganze zu lösen. Und die Probleme die wir jetzt haben sind weniger KI- spezifisch, sondern Konzentrationsspezifisch. Wenn man jetzt an Facebook denkt Facebook ist viel zu groß als dass das jemand stemmen könnte, es geht nicht. Wenn dort das Kartellrecht wirklich exekutiert werden würde wären Facebook

unterschiedliche Firmen und Google, wenn unterschiedliche Firmen. Das ist eine historische Anomalie die man seit dem 18. Jahrhundert mit Eisenbahnen gehabt hat. Eisenbahnen und Stahlindustrie. Das war nämlich ähnlich. Weil Vanderbilt, das ist ein Familienname den man auch heute noch kennt, die Vanderbilts waren eine von drei Familien die Eisenbahnen in Amerika gebaut haben. Und Damals wars eigentlich ähnlich wie heute mit den Appstores. Wenn jetzt jemand in der Stadt XY entlang der Bahnlinie ein Geschäft eröffnen wollte, dann hat er bei der Eisenbahn nachfragen müssen, ob er das überhaupt darf. Wie beim Appstore. History doesn't repeat but it rhymes. Und die sind dann genau deshalb auf aufgesplittet worden. Oder anderes Beispiel von der Telkom Monopole bzw das Monopol ATNT war bis Anfang der 80er das sind zwölf unterschiedliche Unternehmen die sogenannten Baby Bells, die Bell Laboratories, General Alexander Graham Bell Laboratories, ATNT hat das gekauft. ATNT war wie bei uns Die Telekom, das ist ein Anbieter. Für die ganze USA, die sind dann aufgesplittet worden. Und wir befinden uns Internet technisch befinden wir uns jetzt gerade dort. Das ist total untypisch, dass man so eine Konzentration hat. und viele Mechanismen kommen, weil die Firmen so groß sind. Das ist wie, wenn Daimler gleichzeitig auch Airbus wäre und SAP und Bombardier. Das ist too much.

S: Weil wir gerade bei dem Thema sind, denkst du, dass sich das ändern wird? Speziell Google und Facebook.

W: Ja definitiv. Das fängt schon an. Die müssen jetzt im Juli vom Kongress antanzen und das ist bipartysan, das ist ganz wichtig. GOB und von den Demokraten wir das Ganze getrieben. Das wird ein massives Wahlkampfthema. Weil es den Leuten selbst auffällt, dass das nicht gut ist. Es fällt auch auf wie lange schon keine große Firma in dem Bereich entstanden ist. Die letzte große - Facebook - ist vor 14 Jahren gegründet worden, vor 15. Und es ist auch wieder so eine historische Anomalie. Weil zuvor eigentlich immer in relativ kurzer Folge die Firmen gekommen sind. Im Moment nicht. China hatte das viel marktwirtschaftliche System als die USA in dem Bereich. Du merkst es auch daran, Chinesische Unicorns haben nicht diesen Allmachts-Anspruch. Sprich Facebook ist DAS soziale Netzwerk Youtube ist DIE Videoplattform Google DIE Suchmaschine. Oder WhatsApp ist DAS Kommunikationsmedium. Und es gibt nichts anderes mehr. Du hast in China viel gesunderen Wettbewerb zwischen Firmen in den einzelnen Bereichen drinnen.

S: Ist der Anspruch nicht auch da z.B. bei Alibaba das chinesische Amazon. Oder das eine, WeChat, über das alles läuft, also Kommunikation und man bezahlt darüber.

W: Erstens mal machen sich die großen Unternehmen extrem lang Konkurrenz, es ist nicht so, dass wirkliche Oligopol herrscht. Und zweitens. Werden die Plattformen selbst nie dermaßen groß. Wenn du jetzt nämlich Alibaba erwähnst: Alibaba hat natürlich zwei Jahrzehnte Vorsprung beim B2B E-Commerce. Zwischen Unternehmen. Jetzt aber gerichtet an Kunden: dort gibt es mehrere Amazons. Das gibt es bei uns nicht. Oder auch wenn's jetzt an Short-messaging geht, oder Videoplattformen, TikTok z.B. kommt aus China, dass würde es bei uns nicht geben, wie Snapchat ist eigentlich gekillt worden seit Instagram Snapchat 1:1 kopiert hat. China geht da aber wirklich Head-to-Head. Und der große Unterschied ist, dass der Anspruch von den Investoren nicht der ist, dass dort ein Monopol entsteht, sondern ein Tragfähiges Businessmodell. Bei Foto-App z.B.: es muss die das nächste Instagram sein.

S: Verstehe, in Europa wagt man sich gar nicht mehr heran, wenn es schon einen großen Spieler gibt?

W: Das Grundlegende Problem ist sicher, der Markt ist fallmentiert. Also die Sprachen sowieso, die unterschiedliche Gesetzgebung – Stichwort Uber: in manchen europäischen Städten erlaubt, in manchen nicht. So etwas wie Uber hätte in Europa sowieso gar nicht entstehen können, wenn man die legalen Punkte weglässt, weil Datenroaming so arg teuer war. Es wäre niemand auf die Idee gekommen – vor drei Jahren - mit dem Datenpaket jetzt Uber zu rufen. Aber wenn du keine Mehrkosten hast schon. Solche Dienste waren früher benachteiligt. Es fängt jetzt erst an, dass wir das Thema Mobile First – das geht erst jetzt europaweit.

S: Wenn man noch kurz zum Sorgfaltsproblem zurückkommen – wie würdest du es für Europa einschätzen? Du hast gesagt es gibt mehr Vertrauen in öffentlich-rechtlichen. Kommt es da gar nicht erst zu der Situation, dass aufgrund von Marktdruck gewisse Bias sofort entdeckt werden – im Gegensatz zur USA wo sowas sofort entdeckt wird? Es ist halt sehr markspezifisch.

W. Naja, die Sache ist, wenn wir heute von KI reden, dann meinen wir die KI die uns im Internet begegnet, das sind immer B2C Firmen, wie Netflix, Amazon, Apple mit Abstrichen. In Europa gibt es de facto keine B2C Internetfirmen. Was Spotify mir jetzt vorschlägt ist zwar nett, hat aber auf mein Leben null Auswirkung.

S: Neben wir jetzt an es wird im Personalwesen eingesetzt, oder um Gerichte objektiver zu machen. Da ist ja der Endverbraucher betroffen.

W: Naja aber dann würdest wieder bei B2B eher verordnen. Und da gibt es eigentlich starke Regulierungen, die stehen auch in der DSGVO. Da gibt es eine Klausel, die Kafka Klausel denk ich, die besagt, dass wichtige Entscheidungen dürfen nicht in letzter Instanz von einem System automatisiert werden. Wichtige Entscheidungen sind medizinische Behandlungen, Krebs oder Nieren, ob jemand Zuschlag für eine Gemeindebauwohnung kriegt, aber auch Personalwesen, ob jemand einen Job bekommt. Also die Leitplanken sind schon definiert. Was noch nicht passiert ist, dass alles runterdeklariert ist. Des richtet sich danach wie weit KI

in der jeweiligen Domäne sind. Bei den Autos ist es natürlich weiter als sonst wo. Denn testweise sind ja Autos schon überall unterwegs. Bei Graz wird getestet. In OÖ und NÖ testet MAN und Daimler. Selbstfahrende Busse hast in Wien Seestadt. Und dementsprechend ist der Gesetzgeber gefordert, da früher vorzupreschen und Spielregeln definiert. So wie sie im 3. Quartal letztes Jahre das Kraftfahrbundesamt oder so gemacht, die sagen wie muss ein KI System ausschauen und was genau das Für und Nach, wie sie überprüfen können ob es die Entscheidungen so trifft, wie sie jetzt EU konform sind. Dass ich es debuggable mach, da ist man schon dran. Aber das ist natürlich immer industriegetrieben, wo der Druck herkommt. Wenn's zwei HR Startups in Europa gibt - wo kein Kläger da kein Richter. Oder wo kein Bedarf wahrgenommen wird. Es stellt ja auch keine Firma von heute auf morgen um. Da sind wir wieder beim Sprachproblem – wie vorher erwähnt bei der Transkription – es funktionieren viele Sachen auf Deutsch noch nicht so gut. Deshalb werden wir viele Probleme wahrscheinlich gar nicht selbst lösen müssen, sondern uns anschauen was macht USA, China, UK, Kanada. Dann für uns ableiten wie wir damit umgehen. Man muss nicht alles neu erfinden. Bei Regulierungen geht es eigentlich viel stärker um die regionalen Ausprägungen.

S: Verstehe. Wenn du's auf der Skala einordnen müsstest – wie sehr kann es als Bedrohung wahrgenommen werden, würdest du eindeutig unterscheiden zw. USA und Europa, auch UK?

S: Das Bedrohungsszenario ist in Amerika sicher höher als irgendwo sonst aufgrund der Geschwindigkeit. Gleichzeitig wird es schneller gefixed. Ist jetzt die Frage, ob das positiv oder negativ ist. In China funktioniert das Ganze auch ganz gut, die größte Gefahr beim Thema Sorgfaltsproblem seh ich gar nicht darin was die KI Unternehmen selbst machen, weil da schaut man sehr genau darauf, sondern was mit den Daten von gescheiterten Firmen passiert, die extrem viele Daten gesammelt haben. Da gibt es total viele Beispiel in letzter Zeit, wenn man sucht auf Techme. Also da sind teilweise haarsträubende Sachen, wo Unternehmen teilweise Apps von Leuten, Kindererziehungsapps, HomesecurityApps, wo schnittweise Überwachungsvideos und Fotos angelegt werden, die nie für den Anwendungsfall gedacht waren, damit KI zu trainieren – die Firma geht out of Business, und der Datenbestand wird aber aufgekauft von einer Firma die das für Überwachungssysteme verwendet.

S: Kann das legal passieren?

W: Es ist legal, aber es ist ein moral hazzard, wie du's in USA bezeichnen würdest.

S: Es ist in Amerika legal. Bei uns wahrscheinlich nicht, personenbezogenen Daten einfach weitergeben?

W: Das passiert die ganze Zeit, und thats the problem.

S: Arg.

W: Das ist z.B. etwas was bestimmt in der DSGVO steht. Aber viele Leute haben das nicht am Schirm, weil es wird immer gesehen als ich darf keine personenbezogenen Daten abspeichern, wie Fotos von Gesichtern oder Adressen. Aber das geht schon weiter, die eben mit der Entscheidungsfindung von automatisierten Systemen.

S: Und du hast gesagt im asiatischen Raum ist es wieder anders, passiert auch und regelt sich schnell von selbst oder kommt gar nicht soweit – gebiasete Anwendungen am Markt?

W: Der chinesische Staat schreitet da schon ein. Einerseits sind sie viel freizügiger als du bei uns darfst. Andererseits sagen sie auch dass du gewisse Sachen halt nicht machen darfst, so etwas wie Facebook dürftest du dort nicht abziehen. Der Staat ist natürlich schon.. na gut die NSA macht das genauso, es ist halt böse, wenn es die Chinesen machen. Aber die Firmen nehmen sich weniger heraus als in Amerika im privaten Bereich würde ich sagen, weil sie wissen, dass der Staat sie sonst abdreht. Gleichzeitig haben die Firmen aber auch mehr Möglichkeiten mit öffentlichen Daten zu arbeiten. Denn der Anspruch an Privatsphäre ist einfach ein anderer in China.

S: Abe begünstigt das dann den Fall, dass Bias in Anwendungen die schon im Einsatz sind kommen oder wird alles mehrfach geprüft?

W: Ich glaube, dass du überall auf der Welt denselben Ansatz hast – Regulation after the fact. Zuerst passiert was, dann kommt man darauf und es wird reguliert. Aber das Ganze ist zu neu. Irgendwann wird man sich schon darauf eingependelt haben, dass Daten inhärent einen Bias drinnen haben, den ich aber teilweise wieder rauskriegen will, nur bis das in den Köpfen der Entscheidungsträger, Politiker drinnen ist, braucht es eine Weile bist sich alles gesettelt hat.

S: Wenn ich dich zwingend würde hier eine Einordnung zu machen: Auch marktbezogen, wenn wir zw. USA China Japan und Euroraum unterscheiden, bitte?

W: Bedrohung ist gut, weil wenn man es hernimmt und ich auf die restlichen Fragen schaue, dann

S: Ich meine nicht nur Bedrohung, sondern auch Auslöser für KUI auch, wenn man berücksichtigt, dass es als KUI gilt, wenn es jetzt Personengruppen benachteiligt.

W: Ich denke, dass das für absehbare Zeit der Hauptfaktor ist, dass KI falsch Entscheidungen treffen wird. Da sind wir wieder bei dem Data Literacy Problem. Das kommt ja direkt daraus. Wenn können Probleme nur Vordergründig aus dem Bereich kommen.

S: Albright. Der nächste Punkt ist Physical Hacking. Es gab immer wieder Fälle wo Leute versucht haben zu zeigen, wie bedrohlich es sein kann – z.B. zwei im IT Bereich tätige Hacker haben gezeigt wie sie ein autonom Fahrendes Auto von außen komplett fremdzusteuern. Ein US-amerikanische Universität hat gezeigt wie leicht Sticker auf Verkehrsschilder das „Sehen“ von Autos manipulieren, die Frage – wie relevant wird Physical Hacking als Bedrohung sein?

W: Nimmt man das erste her ist das das klassische Security Problem. Das ist ein bisschen wie ein Wettrüsten. Du wirst mit KI mehr Hacking sehen, weil du automatisiert hacking tools schreiben kannst, gleichzeitig wird Cybersecurity immer mehr KI basiert sein. Das ist eigentlich straight ahead, da ändert sich durch KI spezifisch Garnichts. Denn das ist halt in der Daten oder elektronischen Welt so, dass es Physical Hacking gibt. Und Intrusion Detektion hast du davor auch schon gehabt, das kam nichts neues dazu. Das Bedrohungsszenario ist so gesehen nicht gestiegen, weil es schwieriges ist das Ganze zu Hacken. Zu adversorial attacks gibt es momentan große Forschungsanstrengungen, auch in Österreich, z.B. im Software Competence Center Hagenberg, da haben sie auch ein Patent eingereicht, damit man genau die Attacken wie mit den Verkehrszeichen nicht mehr machen kann. Er hat es veröffentlicht, es ist eine Kombi aus Bildverarbeitung und Reinforced Learning. Gefahr erkannt, Gefahr gebannt. Man muss sich vor Augen führen, kein Auto fährt zu 100% auf Bilderkennung. Das sind Kinderkrankheiten, die wir mitkriegen, weil Prototypen in der Öffentlichkeit entwickelt werden.

S: Es ist also eine Bedrohung, aber nicht gestiegen.

W: Absolut handlebar. Da tut sich viel in dem Bereich. Ich würde es als "kaum" einordnen. Weil es eben schon aufgegriffen wird und auf allen Seiten des Atlantiks und des Eurasischen Kontinents dazu geforscht wird. Weil man verhindern will, dass mit neuen Technologien neue Angriffsflächen entstehen. Vor allem, diejenigen die es sicher machen haben in der Regel mehr Geld. Beispiel Fotoshop: als es aufkam hat die NewYorkTimes einen Artikel veröffentlicht, wo es hieß: man weiß nicht, ob man jetzt überhaupt noch Fotos drucken wird, weil man nie wieder gedruckten Bildern vertrauen könne". Heute haben sie mehr Fotos denn je.

S: Dazu noch eine Frage: denkst du, dass es eher auf individueller oder gesellschaftlicher Ebene ein Problem sein könnte, beim Thema KI-Hacking?

W: Auf gesellschaftlicher Ebene ist Facebook das beste Beispiel. Wenn man Megaplattformen hat, ist es am einfachsten die Plattformen zu gamblen.

S: Du meinst es ist generell angreifbarer, wenn viele Daten zentral gespeichert werden?

W: Hängt immer davon ab, wer die Daten bewacht. Gmail ist das sicherste E-Mail-Programm, niemanden ist gelungen, G-Mail zu hacken im großen Stil. Definitiv sicherer, als wenn wir selber E-Mail hosten würden. Interessant ist, die Sicherheitsmechanismen auf individueller Ebene werden durch KI weitaus besser. Ich denke, dass durch KI auch viele Unsicherheitsfaktoren wegfallen, Stichwort Passwort, Mailadresse. Gibt es interessante Forschung von Google dazu: man hat einen Trustscore in neun Stufen veröffentlicht. 9 ist online Banking, Government Registration. 5 dein Instagram Account, 1 dein Internet Forum wo du manchmal abhängst. Sie haben gesagt, je nach Level nimmt es sich mehr Faktoren zusammen. Dann erkennt er aus der Vibration schon die Identität, jeder Mensch zittert anders. Oder die Sakkade, die Zeit die du zwischen dem Tippen von einzelnen Vokalen und Konsonanten beim Tippen hast, auch das ist bei jedem Menschen anders. Biometrisch kann man noch das Gesicht nehmen. Beim höchsten Trust score nimmt man also alles gemeinsam. Ich bin mir sicher, dass man die Systeme auf individueller Basis so gut sichern kann, dass es de fakto unhackbar wird. Wenn du der Saudische Staat bist vielleicht schon, aber ein normaler Player nicht. Das Secure enclave bei IOS ist auch noch nie gehackt worden. Vor kurzem hat glaube ich das Pentagon ein Forschungsprojekt gemacht, bei dem sie Personen am Herzschlag auf eine Distanz von 200 Meter identifizieren. Jeder Mensch hat einen individuellen Herzschlag. Das ist wie ein Fingerabdruck. Es hat alles Missuse und ein beneficial Szenario. Das gute ist, das man Leute bald so gut identifizieren kann, dass es beinahe unhackbar wird.

S: Unglaublich. Nächster Punkt, du hast es schon angesprochen, das Blackbox Problem. Künstlich Neuronale Netzwerke erlauben es oft nicht, Begründungen für die KI präsentierten Ergebnisse nachzuvollziehen, die Entscheidungsfindung versteckt sich in einer Art Blackbox. Kann das eine Bedrohung sein, wie relevant ist dieses Problem?

W: Das ist ein temporäres Problem, sehe ich in den nächsten drei, vier Jahren aber danach sollte das gegessen sein meiner Meinung nach. Es erinnert an Datenbanken, wo man in den 80gern gedacht hat, so große Mengen, wer passt darauf auf etc., dann gab es vernünftige Datenbank Tools, zum Durchsuchen, Bearbeiten, Konflikte finden, man muss eben lernen damit umzugehen. Auf Explainable AI setzen viele, aus kommerziellem Antrieb, weil es

natürlich leichter zu verkaufen ist. Für die Endabnahme ist wichtig, dass der Mensch das System versteht, der inherente Antrieb auf ExAI zu setzen ist also da. Das Thema Adverserial Hacking zu verhindern, geht in dieselbe Richtung, da passieren momentan massive Anstrengungen. Man darf auch nicht vergessen, das meiste in KI passiert einerseits bei Studenten, auf deren Grafikkarten, Unis, und andererseits Konzerne im intergalaktischen Maßstab, für autonomes Fahren. In der Industrie muss man da nicht mehr darauf schauen, dass alles nachvollziehbar ist, weil wenn jemand seinen Kredit nicht und dich klagt, das wäre schlecht. Es reguliert sich also von selbst. Bei KI ist nur da untypische, dass wir das mitkriegen, weil KI auch so stark Open Source ist. Mir ist kein anderes Feld bekannt, wo ich mir Datensets - Übersetzungen, gescannte Dokumente, gefilmte Autos, Menschen, auf YouTube etc. - so einfach besorgen kann und auch die Tools zur Verfügung stehen. Deshalb ist mehr Augenmerk darauf, weil man es eben mitbekommt.

S: Verstehe, also eigentlich atypisch, dass wir diese Entwicklung mitkriegen. Wo würdest du das auf der Skala einordnen, als Grund für KUI?

W: Die Frage ist nach dem Zeithorizont. In fünf Jahren wird es glaube ich kein Problem mehr sein. Wie bei der Automobil Industrie, man wird Spielregeln haben, wie etwas nachvollziehbar sein muss, es wird also feststehen.

S: Also in der Zukunft quasi gelöst. Im momentanen Zeitraum? Oder ist es gar kein Problem, weil es so nicht im Einsatz ist?

W: Es ist insofern kein Problem, weil ohnehin keine life Entscheidungen getroffen werden, also kann es per Definition nicht sehr stark sein. Wenn mir Netflix was Falsches empfiehlt oder mein Twitter algorithmic-news-feed Blödsinn anzeigt - so what? Das ist kein Problem. Aber heute würde ich sagen, was die Einführung von KI betrifft ist die Nachvollziehbarkeit sicher ein starkes Hemmnis, was die gesellschaftliche Auswirkung betrifft würde ich heute sagen kaum, eben weil die Einführung gar nicht passiert, weil es ein Problem ist. Ich glaube, dass das in Zukunft sehr viel nachvollziehbarer sein wird. Es wird der Neuheitsgrad weg sein und der Unerklärbarkeitsgrad wird sinken. Das ist nur dem geschuldet, dass die Technologie neu ist.

S: Drei bis fünf Jahre also, alright. Der nächst Punkt sind Lösungsvorschläge. Du hast schon erwähnt, teilweise reguliert es sich selbst der durch den Markt. Wenn man die Probleme, die bis jetzt da sind - Sorgfaltsproblem, Technochauvinismus, Physical Hacking - nimmt, wie werden sich die in Zukunft entwickeln bzw. wie schauen Lösungsvorschläge dafür aus. Für sinnvolle KI Anwendungen.

W: Ich bin bei der Frage abgeschweift, also schweife ich auch jetzt ab, never change a not running system (lacht). Das prinzipielle Problem das ich sehe, ist dass was wir heute mich News Filter Bubbles mitbekommen, dass uns durch KI schon bevorsteht, dass jeder in seiner eigenen Bubble leben wird. Also Preis, welche Produkte und Hotels werden dir empfohlen, eine Reise App in fünf Jahren wird dir den kompletten Trip vorschlagen. Je mehr du an ein System auslagerst, und je weniger datenmündig du bist und verstehst, was da abgeht, kann schon sein, dass diese Bubbles in großem gesellschaftlichem Stil einschlagen. Die Frage ist, was das Ganze dann bedeutet, was ist das Ziel gesellschaftlich? Weil du vorher Chatbots erwähnt hast, ich bin mir nicht sicher, angenommen wir hätten eine General Intelligence, jeder virtuelle Freunde hätte die in der richtigen Anzahl Liken und Resharen - who knows, vielleicht ist das, was Leute wollen. Beim Minority Report find ich interessant, der Film ist aus 2001 glaube ich, da gibt es eine Szene, wo sie sich die perfekte VR vorstellen. Er geht durch ein VR Kino, der eine hat Extremsport, der eine lebt in einer Sexfantasie, und der Dritte hat einfach einen Anzug an und sitzt bei einer normalen Dinnerparty. Die Leute sagen "Hey Mister Smith, this is the most insightful comment i ever heard, no wonder you won de Nobil Prize" - aber letztendlich, wenn das technisch möglich wäre, ich denke schon, dass sehr viele Menschen das in Anspruch nehmen würden. Man sieht das glaube ich sehr gut am Wahlverhalten. In Amerika gibt es Untersuchungen, dass die mehrheitsfähigen Meinungen bei Rot und Blau, dass die eigentlich in den 50er, 60er 70er viel stärker beisammen waren und das jetzt stärker auseinander driftete. Dass ein Kernwähler von Demokraten oder Republikaner in ihrer sehr einen Nachrichtenwelt leben. Technologie kann schon beschleunigen, dass das Ganze noch weiter auseinanderdriftet. Weil dir nie etwas vorgeschlagen wird aus einer anderen Richtung.

S: Ich habe meine Bakk1 über die FilterBubble Theorie geschrieben. Dabei habe ich mir aber schwergetan, einen Beweis zu finden: Es gibt sehr viele Studien, die sagen es existiert, eindeutig. Andererseits Gegenstudien, die sagen es gibt weniger Auswirkungen. Sie du das, und siehst du das stärker kommen?

W: Definitiv. Facebook verwende ich nicht mehr. Ich hatte mehrere Tausend Kontakte und hab nie einen FPÖ-Like gesehen, das kann eigentlich nicht sein. Also entweder es gibt die Filter Bubble, oder das System führt zu einer Selbstzensur. Im Grunde dieselbe Auswirkung. Ich habe das immer komisch gefunden, in Österreich bei Hofer, oder zumindest bei den Wahlen, zu Zeiten als die FPÖ 34% gehabt haben, der Open Boarder Sommer 2015, da keinen einzigen Like von einem FPÖ Thema zu haben - das kann gar nicht sein.

S: Verstehe ich, bei mir dieselbe Situation.

W: So gesehen gibt es das schon. Ob es jetzt eine Selffulfilling Prophecy ist durch Selbstzensur sei jetzt so dahingestellt. Aber was schon in die Richtung geht, da gab es Studien dazu, wie das Pricing von Facebook Political Advertising funktioniert hat, bei Trump gegen Clinton. Facebook Pricing ist zweistufig. Einerseits: welchen Inhalt spielst du aus. Also Werbung für Smartphone ist teurer. Also Facebook macht im ersten Moment ein Scoring vom Produkt, ist es eher was hochwertiges oder Massenware. Und dann geht es um die Zielgruppe. Das hat dazu geführt, dass Trump so klassische Bullshit Kampagnen gefahren hat - so Obama ist in Afrika gefahren, das mit der Pizzeria oder Hillary ist rassistisch und bleibt zuhause und geht nicht wählen - und die haben für ihre Werbeanzeigen teilweise ein Tausendstel gezahlt von dem was Clinton gezahlt hat. Weil Clinton sich auf White Affluent fokussiert hat, und es viel teurer ist die zu erreichen. Also teurerer Content und teurere Zielgruppe, das multipliziert sich. Deswegen hat Trump mit seiner Kampagne auch viel mehr erreicht.

S: Wenn wir jetzt nochmal über Lösungsansätze reden. Das Ethik-Gremium der EU hat ja diese 7 Richtlinien genannt, die sind aber nicht bindend. Wird etwas Bindendes kommen, wäre das der Ansatz für eine sicher effiziente KI, was hältst du von dem?

W: Ich glaube, dass sektorspezifische Regulierung auf jeden Fall kommen wird. Auto reglementiert sich sowieso, weil die stärksten Befürworter sind ja die Autohersteller selbst. Es gibt eine Aussage vom CEO von VW, der meinte ein autonom fahrendes Auto muss nicht zehnmal, sondern tausendmal so gut sein wie der Mensch. In Europa gibts vielleicht 30.000 Verkehrstote. Wenn autonom fahrende Autos 3.000 Verkehrstote produzieren, dann gibt es keine autonomen fahrenden Autos. Mit 30 könnte man vielleicht leben, 300 wird schwierig, 3000 geht nicht. Und sonst glaube ich, dass es sehr stark über die branchenspezifischen Regulierungen kommen wird. Im Industriebereich ist es wirklich egal, wenn man an predictive maintenance, quality control etc, da hat man kein Ethik drinnen. Das ist etwas, warum Europa in dem Bereich anders tickt als Amerika oder China. Weil wir hauptsächlich B2B Firmen haben. ICh würde sagen, inhärent haben wir hauptsächlich B2B Firmen, es geht dann eigentlich "nur", Großteiles darum, die ausländischen B2C Firmen dazu anhalten, in Europa die hiesigen Gesetze zu befolgen. Wie es bei der DSGVO passiert, das ganze muss nur expliziter werden.

S: Wird das in nächster Zeit kommen, oder eher unverbindlich bleiben, damit die Industrie etwa nicht gedrückt wird.

W: Naja gedrückt wird im Moment e nichts, weil den B2B Firmen ist das egal. Ich glaube, dass es immer eine Frage ist, wann es die Leute betrifft. Als das mit Cambridge Analytica

2017 wird es jetzt e reguliert, man arbeitet weltweit, es wird nachgeschärft. Ich denke, im großen Stil wird die EU anfangen, das ganze am 2021 zu regulieren. Weil man hatte im April angekündigt, dass man bis zum Jahr 2012, also drei Jahre, 19, 20 und 21, dass man ich glaube 10 Milliarden, einen schönen Betrag investiert. Und ab dann will man jährlich 20 Milliarden investieren, also public private partnership. Also 10 Mrd. kommen von der EU, 10 Mrd. aus dem öffentlichen Bereich. Bei diesen Projekten wird dann viel stärker die Regulierungsdiskussion anfangen. B2C Startups müssen sowieso die DSGVO befolgen. Und gleichzeitig hat man auch auf EU relevante Sprachen, deutsch polnisch spanisch etc., in vielerlei Hinsicht nicht die Möglichkeit Sachen zu machen, wie auf Englisch. Du kannst Google Duplex nicht auf Deutsch bauen, sonst hättest du schon die Diskussion, wie kommuniziere ich, dass jetzt eine Maschine anruft, darf ich das etc.

S: Versteh, vieles ist also nicht möglich.

W: Ja, deswegen wäre es ganz schlecht, wenn man sagt, wir sind die Insel der Seligen und wir regulieren nur. Das muss Hand in Hand gehen mit der Wirtschaft. Wie man es bei Autos sieht, oder auch Luftfahrt. Ohne Airbus wäre es Boeing komplett egal was Europa sagt, weil Amerika die Monopolisten wären. Man kann es nicht entkoppelt sehen.

S: Verstehe. Fallen dir noch zusätzlich Gründe für KUI ein, vielleicht vor allem in der Wirtschaft, ineffiziente Situationen oder welche, die Probleme verursachen?

W: Man merkt schon sehr oft das "Solution and Problem" Verfahren. Es gibt eine neue Technologie, und man such eine Anwendung dafür, auch wenn sich das nicht unbedingt auszahlt. Wenn ich jetzt das Beispiel Automatisierung hernehme, sprich Personalreduktion, hat das unterschiedliche Aspekte. Wenn ich hergehe und mir eine Bank anschaue, Banken sind arbeitsteilig sehr gut aufgestellt. Es gibt da Tätigkeiten die sehr austauschbar sind, wo es etwa darum geht, dass Dokumente reinkommen, die abgeglichen werden etc. Bei so etwas, ist man schon sehr darauf bedacht, das Ganze zu automatisieren, weil man sich im Nachteil sieht gegenüber einem angelsächsischen Raum, wo so etwas teilweise schon geht. Jobtechnisch ist das nicht unbedingt ein Verlust, weil die non-skilled labour force macht dann etwas anderes, weil Jobs gibt es immer.

S: Denkst du, dass es ein großer Einschnitt sein wird, dass viele Jobs ersetzt werden?

W: Letztendlich machen die Leute was anderes. In Amerika kriegt man es am offensichtlichsten mit. Mit Uber z.B., Leute machen ja keine eigene Ausbildung fünf Jahre um Uber Fahrer zu werden. Die haben vorher Burger bei McDonalds gebraten, und verdienen

halt so jetzt mehr. Ich glaube aber, dadurch dass sehr wohl noch Leute brauchst, die an Schnittstellen sitzen, wieder neue Jobs kommen. Das ist das eine, dann gibt es die Personalreduktion, wo es darum geht, dass Wissen aus der Firma rausmarschiert, weil Leute in Pension gehen. In Deutschland ist 2024 ein kritisches Jahr, das Peak-Jahr wo Baby Boomer in Rente gehen. Deutsche Firmen bereiten sich auf die Entwicklung vor, dass sie zukünftig mit einem Drittel weniger Leute auskommen, die wissen wie der Hase läuft. Da rede wir von 20Jahre plus Erfahrung. Da kriegt das Thema Personalreduktion eine andere Bedeutung als im ersten Fall. Viele Firmen stellen sich nicht die Frage, was sie mit KI eigentlich wollen. Dementsprechend werden mehr als 50% der KI Projekte nicht erfolgreich implementiert. Weil man sich das warum nicht genau ansieht.

S: Ich danke dir vielmals für deine Zeit und die ausführlichen Antworten Clemens!

Interview 2: Kriechbaumer Patrick am 30. Juli 2019

S: Danke erneut, dass du dich von mir interviewen lässt. Wir haben ja schon einmal grob über die Arbeit gesprochen. Artificial Unintelligence ist der Titel, und orientiert sich an dem gleichnamigen Werk von Broussard. Sie ist eben die „Schöpferin“ von dem Begriff Technochauvinismus. Punkt eins dreht sich genau darum, der erste Grund für KUI ist eben Technochauvinismus bzw. die Hype Begründung. KI befindet sich seit sie existiert ja in dem Hype cycle die großen Erwartungen schürt und wieder fallen lässt und so weiter. Technochauvinismus ist eben die Überzeugung, dass KI, weil sie ja auf mathematischen Berechnungen basiert, größte Objektivität erreichen kann und alle menschlichen Probleme mit ihrer Hilfe gelöst werden können. Inwiefern ist Technochauvinismus nun als Auslöser für KUI zu sehen. Wenn man bedenkt, dass viele Leute extrem große Erwartungen setzen und daher oft sehr falsche Vorstellungen haben. Man denke z.B. an den ersten Unfall von autonomen Fahren 2016, wo der Lenker aufgrund seines großen Vertrauens nicht ins Fahren eingriff und durch Fehler der KI der Unfall ausgelöst wurden. Also inwiefern siehst du den Umstand als Bedrohung?

K: Ich würde sagen das Kernproblem ist, dass Entwickler sich ab einem Grund als unfehlbar halten, und die Datensätze mit denen KI trainiert wird nicht an sich nie vollständig sein kann. Die spontane Entscheidungsfindung eines Menschen ist von Jahren, Jahrzenten Erfahrung verschiedenster Umwelteinflüsse geprägt, die man einer KI in der Form nie beibringen kann. So etwas wie Intuition ist ja ein Weg der Entscheidungsfindung, die so keiner begründen kann. Davon gibt es täglich viele Entscheidungen von uns, die wir treffen ohne direkt zu begründen, warum habe ich das jetzt gemacht. Sei es nur zu beurteilen, warum ich über die Straße gehe oder nicht. Es hat keine genaue mathematische Begründung, ich bekomme es "ins Gefühl", aber die direkte Entscheidung warum ich jetzt losgehe kann ich nicht 100% begründen. Genau das gleiche grundlegende Problem haben KIs ja genauso. Wir können sie zwar trainieren, das Training von KIs ist immer von vorselektierten Datensätzen. Man sagt irgendwann, dass diese Daten, auch wenn es ein paar Millionen sind, ausreichend sind, um sie soweit zu trainieren Entscheidungen so sinnvoll zu treffen wie ein Mensch es könnte. Nur dass das ein Irrglaube ist, weil es nicht funktionieren wird, dass mit einer Handvoll vordefinierter Datensets eine KI trainiert wird. Weil selbst wir als Menschen uns oft denken, eine alltägliche Situation kommt mir neu vor. Wir können noch so lange Autofahren, und trotzdem gibt es immer wieder Situationen wo wir denken - Oh Gott, das ist mir auch noch nie passiert! - und solang das der Fall sein wird, wir die KI ja nie so weit sein, dass sie in 100% der Fälle richtige Entscheidungen trifft. Und nach dem eine KI ja von einem oder mehreren Menschen programmiert wird, wird immer das Problem bestehet, dass sie religiös beeinflusst sind oder Rassenprobleme geben wird, weil am Ende des Tages immer Menschen entscheiden, auf Basis welcher Kriterien KI arbeitet. Das Problem wird bestehen bleiben. Und auch wenn KI sich irgendwann selber weiterentwickelt, wie wir es teilweise schon

haben - das hat Microsoft mit seinem Versuch gut gemerkt - kommt heraus, dass innerhalb kürzester Zeit eine rassistische KI herauskommt. #00:06:17-1#

S:Du sprichst mein zweites Problem an, vielleicht springen wir gleich dahin. Unter dem Sorgfaltsproblem fasse ich zwei Punkte zusammen: Einerseits das Bias-Problem, in Daten wie von dir genannt oder von Entwickelnden. Zweitens der Druck, wie es vor allem in der Privatwirtschaft in den USA spürbar ist, durch den Firmen die nötigen Testphasen für KI Anwendungen nicht einhalten, um ein Produkt möglichst schnell auf den Markt zu bringen. Wie bewertest du dieses Sorgfaltsproblem, siehst du es als Grund für KI, wie relevant ist das? #00:07:06-0#

K: Es hängt natürlich immer sehr stark davon ab, was für ein Problem will ich lösen mit KI. Umso komplexer das Problem, umso wahrscheinlicher wird, dass das Sorgfaltsproblem zu Tragen kommt. Weil umso komplexer, umso größer das Datenset welches benötigt wird. Das habe ich ab einer gewissen Größe einfach nicht mehr. Der Straßenverkehr analysiert z.B., da kann es sehr von Bundesland zu Bundesland differieren. In Salzburg zum Beispiel hat man viel mehr Werkstraße. Wenn ich das als KI jetzt beurteilen muss, wie diese Straßen zu fahren gehören, ist das ein kompletter Unterschied zu einem Bundesland wo ich mehr Autopannen oder Flachland habe. Es gab schon Versuche, KIs einzusetzen, um Personalentscheidungen zu treffen. Genau dasselbe, das Problem: Ich möchte die KI verwenden, um die Personalfindung gerechter zu gestalten. Gib ihr aber historische, also alte Daten, auf deren Basis sie lernen soll. Wenn ich sie davon lernen lass, aber weiß, dass die Daten nicht geschlechtsneutral sind oder religiöse Hintergründe benachteiligt wurden, etc., wie soll die KI dann besser beurteilen als ein Mensch. Man möchte, dass die KI es besser macht, ich habe aber kein Datenset, mit dem ich sie lernen lassen kann, um es besser machen. Ergo wird wieder benachteiligt werden. Amazon hat da ein relativ prominentes Beispiel geliefert. In dem sie dafür gesorgt haben, dass das Geschlecht als Kriterium überhaupt mit einbezogen wurde. Hätten sie es komplett weggelassen, hätte KI es nie miteinbeziehen können. Im Endeffekt sorgte also wieder der Programmierer dafür, dass die eigentlich neutrale KI sich in eine gewisse Richtung entwickelt. Das ist definitiv in allen Bereichen ein Problem. es bleibt nur die Frage, ob es für den konkreten Bereich ein Problem ist oder nicht. Die Frage gehört eher gestellt. Dass es in ziemlich allen Bereichen ein Problem ist, ist sicher, die Frage also, wie groß ist das Ausmaß. #00:09:40-2#

S: Der vorherige Interviewpartner meinte, dass der Druck auf privaten Unternehmen, von dem wir sprachen, sehr unterschiedlich ist, wenn man sich die USA und Europa ansieht. In den USA hast du diesen Druck viel mehr, von privaten wird hier KI technisch mehr gemacht, zur selben Zeit kommen solche Vorfälle schneller auf und werden behoben. Z.B. COMPAS,

die Risikobewertungs-Software der US-amerikanischen Justiz, solche Fehler passieren leichter, weil keine Regelungen da sind, aber sie werden schnell aufgedeckt. Wie denkst du wird das in Zukunft aussehen, wie wird das Problem sich entwickeln? #00:10:37-5#

K: Wieder ähnlich dazu, wie groß ist das Datenset. Wenn ich es schneller am Markt habe, wie in Amerika, ist die Wahrscheinlichkeit für Fehler in der Software natürlich größer. Gleichzeitig wird es mehr angewendet. Heißt wiederum, ich bekomme ein größeres Set an Daten, mit denen gegentesten kann. Dadurch kann ich schneller beurteilen, ob sie sinnvoll entscheidet oder nicht. In Europa habe ich viel mehr Entwicklung in Uni Kreisen. Dadurch trifft es einen kleineren Personenkreis. Das macht es von der Qualität her nicht unbedingt besser, Weils einfach länger dauert, bis ich potentielle Fehler finde. Ich habe weniger Divergenz bei den verschiedenen Testfällen. Wenn ich etwas im Uni, FH, also Forschungskontext entwickle, dann sind die Fälle in denen ich sie anwende, von Haus aus vorgefiltert. Wenn ich jetzt der Versuchsleiter bin, werde ich natürlich bewusst als auch unbewusst schon Fälle suchen, von denen ich weiß, dass mein Projekt gut abschneidet. Ich werde mir nicht offensichtlich extrem schlimme Fälle aussuchen, wo ich unterbewusst weiß, da könnte meine Software scheitern. Beide Herangehensweisen sind berechtigt, unter dem Strich ist vermutlich der amerikanische Ansatz der bessere. Wenn man jetzt versucht möglichst objektiv zu betrachten. Natürlich, was im Hintergrund beim menschlichen Schicksal negativ hervorgeht, das muss man anders betrachten. Aus Qualitäts-Sicht wird man mit dem amerikanischen Ansatz schneller bessere erreichen. Dann muss man auch wieder pro Fall beurteilen, oder pro Einsatzzweck, kann ich es vertreten, dass es auf die breite Mehrheit losgelassen wird, oder ist es etwas, dass ich im Forschungskontext auch sinnvoll testen kann. #00:12:55-5#

S: Das Sorgfaltsproblem ist also stark abhängig vom jeweiligen Marktraum? #00:13:03-2#

K: Richtig. Haben wird man es immer. Und gelöst gehört es, weil Genauigkeit erzielst du in diesem Bereich nur durch möglichst große Testmengen. Auf beiden Seiten. Das heißt, man braucht ein großes Datenset zum Erlernen für die KI, und man braucht zum Testen möglichst viele neue Fälle. Also große Datenmengen auf beiden Seiten, und die müssen gefunden oder geschaffen werden. #00:13:39-1#

S: Verstehe. Wenn man das jetzt auf eine sehr vereinfachte Skala herunterbricht, müsstest du es beurteilen von Sehr stark bis überhaupt nicht. Wie schwer gewichtest du das Sorgfaltsproblem als Grund für KUI, oder gar Bedrohung im gesellschaftsrelevanten Einsatz? #00:14:05-4#

K: Im Moment würde ich sagen, aufgrund der verschiedenen Einsatzzwecke, die zumindest publik bekannt sind, ist es eher noch mittel, denn wir haben zwar extrem kritische Einsatzbereiche, denen gegenübergestellt sind aber extrem viele Einsatzbereiche die jetzt kein gesellschaftlich kritisches Problem sind. Sobald KI auch publik wirksam in militärischen Zwecken Einsatz findet, wird sich das verschieben. Im Moment ist es relativ ausgeglichen bezüglich zivile, nicht schlimme Einsatzzwecke und militärische, juristische. Derzeit also noch nicht so gravierend, langfristig gesehen, dadurch, dass es sicher kein Umdenken gibt und die Entscheidungsträger in den Bereichen kein KI Knowhow haben, wird auf jeden Fall schlimmer werden. Es würde mich ganz stark wundern, wenn man in fünf oder zehn Jahren nicht das eher als sehr kritisch, oder sehr stark bewerten werden, aufgrund falscher Entscheidungsfindung. Durch KI, oder in dem Fall künstliche Unintelligenz. #00:15:26-1#

S: Albright. Der nächste Punkt ist Physical Hacking. (Erklärung der beiden Beispiel) #00:15:41-9#

K: Das führt im Endeffekt wieder zum Kernpunkt zurück. Es ist immer die Frage - Einsatzzweck, wie ist der trainiert und welche Angriffsflächen habe ich. Wenn ich den Straßenverkehr hernehme, ist absichtliches physical Hacking in der Form denke ich ein weniger großes Problem. Es gibt im Endeffekt n+1 verschiedene Varianten, wie eine Straße beschafften sein kann, bezüglich Untergrunds, Verschmutzung, wie Hell oder dunkel ist es, etc. Ja, es bringt beim AF durchaus potentielle Gefahren, aber das ist eher noch ein Bereich, wo selbst ohne äußere Einflüsse, wo ich aktiv etwas manipulieren will, dass es selbst ohne dem schon so extrem gefährlich ist und kompliziert ist, ein sinnvoll autonomes Ding zu entwickeln, dass das zuerst mal gelöst gehört. Bevor wirklich absichtliche Beeinflussung hier Relevanz gewinnen. Das gleiche gilt natürlich für militärische Zwecke. Das wird ein genauso prominentes Angriffsziel sein wie jetzt militärische Infrastruktur auch schon ist. Im Umkehrschluss ist es auf einen kleineren Personenkreis verbreite, das heißt ich kann es schwieriger testen, dementsprechend aber auch weniger Angriffsfaktoren finden. Das ist jetzt auch schon so. Das ist momentan sehr gut gesichert, gleichzeitig ein prominentes Angriffsziel. Es gibt halt wenige Personen die darüber Bescheid wissen, daher auch wenige Angriffe. Wenn man das mit autonomem Fahren vergleicht: es kann sich jeder ein Tesla kaufen und kann dann den lieben langen Tag herumtesten und schauen, wo die Schwächen dieser Software sind und dementsprechend ist es da wesentlich einfacher. Abgesehen davon, dass Straßenverkehr so extrem dynamisch ist, dass es erst dazu kommen muss, einen zuverlässigen hohen Autonomie Status zu erreichen. #00:19:16-4#

S: Ein anderer Interviewpartner meinte, beim Thema Hacking ist es im Grund ein Wettrüsten, weil beide Seiten konstant stärker werden. Dass sich dadurch von der Gefahr her nicht wirklich was ändert. #00:19:39-9#

K: Ganz genau. #00:19:41-7#

S: Wie relevant ist dieses Problem dann. Auch wenn wir den Unterschied machen zwischen Hacking und Physical Hacking? #00:19:52-6#

K: Ich würde unterschieden, was man als Hacken definiert. Nur weil ich ein Straßenschild markiere, das ist im Grunde nicht hacken. Das System entscheiden auf Basis verschiedenster Faktoren wie es reagiert, heißt ich beeinflusse das System ja nur. Aber ich sage nicht dem konkreten Auto, es soll etwas anders machen. Beim Hacken greife ich ja gezielt eine Gruppe an und störe ein System gezielt. Wenn ich wo einen Sticker hin klebe, beeinflusse ich es zwar schon, aber nur im Rahmen dessen, wie sich das System ohnehin entschieden hätte. Ich dränge es gewissermaßen in die Richtung einer Entscheidung. Beim Hacken steige ich aktiv ein, verändere aktiv ein. Im Endeffekt, im weitesten Sinne - wenn es dunkel wird ist auch die Umgebung beeinflusst, wenn sich das Auto dann falsch verhält "Schulterzucken". Hacken ist da eine ganz andere Klasse, auch strafrechtliche relevant. #00:21:36-3#

S: Das heißt du würdest definitiv anders entscheiden für das, was wir unter Physical Hacking verstehen und Hacking. #00:21:45-1#

K: Ja, auf jeden Fall trennen. #00:21:48-2#

S: Wie würdest du PH bewerten. Inwiefern ist es eine Bedrohung auf individueller/gesellschaftlicher Ebene darstellt. #00:21:59-3#

K: Auf gesellschaftlicher Ebene glaube ich ist die Bedrohung nicht so groß, einfach weil es extrem schwierig ist, auf Basis einer KI im Moment noch größeren Menschenmengen Schaden zuzuführen. Individuellen Personen ja, das ist schon relativ einfach. Besonders wenn man Hintergrundwissen hat. Aber gesellschaftliche, im Sinne von einem Massenproblem, würde ich noch eher gering einstufen. Also kaum, auf gesellschaftlicher. Auf individueller Ebene ist es potentiell sehr hoch. Wenn ich ein gutes Opfer darstelle, ist es

ziemlich einfach. Es war auch vor KI Zeiten einfach, Individuellen zu manipulieren. Sich als sie auszugeben, mit KIs wird die potentielle Gefahr an sich nicht mehr, die allgemeine Gefahr, wenn du ein prominentes Ziel bist, ist schon sehr hoch, mit KI noch ein bisschen höher. Sachen wie social Engineering und im Zuge dessen Personen schädigen, ist es e schon gegeben. Damit hat KI an sich noch nicht viel zu tun, natürlich, wenn wir den Anteil an Menschen, die Entscheidungen treffen immer mehr reduzieren, vergrößert sich das ohnehin schon große Problem umso mehr.

S. Ok. Und konkret Physical Hacking ist momentan noch nicht wirklich als Bedrohung anzusehen? #00:23:54-3#

K: Richtig. Also aufgrund der noch nicht weit verbreiteten Einsatzzwecke. #00:24:03-0#

S: Alles klar. Der nächste Punkt den ich bearbeite, ist das Blackbox Problem. // Erklärung // #00:24:39-9# Bedrohung? Bsp. Personalwesen Diskriminierung

K: Was das angeht haben NN immer wieder das Problem: es lernt zwar relativ gut selber, aber auch da wieder, NN können nur auf Basis von definierten Daten lernen. Das heißt da habe ich wieder exakt das gleiche Problem: Was ich ihnen als Eingangsdaten liefere, auf Basis dessen werden sie auch lernen. Nimmt man wieder das Personalwesen Beispiel: nehmen ich 1:1 unbereinigte historische Daten, werde ich automaisch Entscheidungsfindung trainieren, die Geschlecht, Religion und Herkunft werten. Solange da die Leute, die neuronalen Zweck anlernen, nicht die Daten bereinigen und es wirklich schaffen würden, nicht wertend zu sein, so lang ich das nicht erreiche, werde ich mein Kernproblem nicht lösen können. Wenn das Kernproblem ist, ein geschlechter-, herkunftsneutrales Personalwesen zu entwickeln, das werde ich momentan nicht hinkriegen. Das heißt die Lösung wird noch schwieriger, weil das neuronale Netzwerke sich ja selber weiterentwickelt. Umso länger und größer das Datenset ist, umso schwieriger wird es für den der es anlernt, herauszufinden, auf Basis von welchem Ursprungsdatensatz die KI quasi Amok läuft in die falsche Richtung, wo es was reininterpretiert. #00:27:02-3#

S: Denkst du das man sich dann wegbewegen wird, von allem was mit Blackbox zu tun hat. Wir man sich auf Explainable Ai konzentrierten um das Problem von vornherein auszuschließen, wird man immer wieder damit zu kämpfen haben? #00:27:20-2#

K: Ich glaube wir sind auch hier wieder an einem Punkt... es wird keine AI Kombination geben, um jeden Use-Case 1:1 abzudecken. Es wird einige geben, die eher in Richtung Blackbox - schnelle Entscheidungen ohne genaue Begründung - bewegen, es wird Einsatzzwecke geben, wo das Sinn macht und sie auch gut sind, vermutlich die wo es extrem auf Performance ankommt. Wo quasi auch während des Betriebs ein eng definierten Datensatz gibt. Bei Personalentscheidungen: Keine Firma braucht ein AI System, das Zehntausend Bewerber in der Sekunde bewerten kann. Da nimm ich mir eher in Ruhe die Zeit, etwas zu bauen, das ich nachvollziehen kann, einfach um es fair und gerecht zu gestalten. Im Umkehrschluss wird so etwas wie automobiler Bereich, vermutlich auch militärische Einsatzzwecke, die werden eher das Blackbox Problem öfter haben. Weil das auch Bereiche sind, wo der Mensch selber auch seine Entscheidungen nicht immer begründen kann. So lange glaube ich der Mensch sich den Einsatzbereich nicht genau vorstellen kann oder direkt jede Entscheidung, die er getroffen hat genau zu begründen, wird man sich sehr schwertun, eine KI zu schreiben. Wenn ich etwas selbst nicht erklären kann, wie soll ich es denn schaffen, das zu programmieren. Dabei wird ja das, was ich weiß, quasi dem Computer erklärt, extrem abstrahiert. Kann ich mir etwas selbst nicht erklären, wird es schwierig. #00:29:31-4#

S. Das Polanyi Paradoxon. Wobei beim Blackbox-Problem hier dann eine Umkehr dessen passiert, weil KI uns nicht mehr erklären kann, was sie weiß/warum sie handelt. #00:29:44-2#

K: Naja, was heißt was sie uns erklären kann. Sie ist ja auf die Art und Weise von einem Menschen geschaffen, entwickelt worden. Das heißt es hat jemanden mit dem Hintergedanken geschrieben, es möge ganz cool sein, wenn man da etwas Magisches hat, was auf irgendeine Art und Weiße, darauf liegt die Betonung, zu einer Entscheidung kommt. Im Endeffekt ist es, wie der Mensch selber lernt. Es ist vorbildlich, etwas nachbauen zu wollen, das so ist wie wir, es birgt aber auch extreme Gefahr, vor allem wenn ich es im militärischen Kontext einsetze wird es verdammt kritisch. Wenn ich nicht mehr nachvollziehen kann, warum gewisse Entscheidungen getroffen wurden, wird es ganz schwierig. Und das Problem ist, dass weltweit betrachtet, wenn man alle vernetzten Computer betrachtet, ist ja wesentlich mehr Rechenleistung vorhanden, also der Mensch fähig wäre, nachzudenken. Sollte es irgendwann massiv schiefgehen, und eine KI sich selbst verbreitet, auf Computer, werden wir ein ziemlich großes Problem haben, weil sie die Kontrolle übernehmen wird. Das ist gar nicht so unrealistisch, weil Computer jetzt schon wesentlich mehr Ideen parallel durchdenken können, als es für den Menschen möglich ist. Wir haben Intuition und Erfahrung, aber das ist dem Computer ab einem gewissen Punkt egal. Wenn er 20, 30 Tausend Szenarien durchrechnen kann, hat er ziemlich schnell was er

haben will. Da kann er Erfahrung und Intuition komplett auslassen. Mit der Holzhammer Methode, alles berechnen, irgendein Weg wird ihm dann schon gefallen. #00:31:40-9#

S: Das heißt wenn du es momentan einstufen müsstest? #00:31:46-0#

K: So lange die Menschen es noch sinnvoll einsetzen, in Bereichen wo es in einem isolierten, kontrollierten Umfeld ist, ist es keine Bedrohung. Sobald der Mensch beginnen wird, einer KI zu sehr zu vertrauen und Entscheidungen für bares zu nehmen, was langfristig sich da hinbewegt, damit wir auch die Bedrohungsstufe steigen. Jetzt, solange noch keine KI wirklich Entscheidungskraft hat, ist es relativ egal. Wenn niemand sich darauf verlässt, Entscheidungen ohne Hinterfragen annimmt, ist es nicht kritisch, in letzter Instanz entscheidet noch der Mensch. Langfristig bewegen wir uns dahin, je nach Bereich wo es sich ändern wird, wirds es immer kritischer. Spätestens, wenn eine AI beurteilt ob du nun eine Versicherung kriegst oder nicht. Oder wie teuer deine Autoversicherung ist. Oder irgendwann dynamische Berechnungen von deinen Krankenversicherungsbeiträgen, ob du eine Wohnung oder einen Job bekommst. Dann wird es kritisch. Langfristig gesehen wird es vermutlich so kommen. #00:33:06-9#

S: Ist ja auch teilweise schon im Einsatz. Wie wir gesehen haben mit Einstelllungen genauso wie Risk-Rating oder Versicherungen die das anwenden. #00:33:21-0#

K: Da habe ich sogar vor kurzem mitgeschrieben, an einem System von Beurteilung von Risk-Scoring auf Basis von Social-Media Daten. #00:33:29-8#

S: Davon hast zu erzählt. Bewertungen wie, wenn ich auf Facebook mit einer rauchenden Person auf einem Bild zu sehen bin, werde ich schlechter eingestuft. #00:33:57-2#

K: Das ist normale KI. Hier ist es ja gewollt, dass einem erkannten Faktor eine gewisse Wertung gegeben wird. Würde ich noch nicht mal unbedingt als KI einordnen, vielleicht als extrem dumme Version von KI. Es hat zwar Bilderkennung und Gegenstandserkennung, das ist schon definitiv KI, aber die weitere Folge der Entscheidungsfindung und Berechnung von Scoring selbst hat mit KI nicht mehr viel zu tun, das ist dann einfach normale prozentuale Berechnung. Gegenstand A wurde erkannt, der Wert wird um x gesenkt oder erhöht. #00:34:46-3#

S: Quasi extrem starke Statistik. #00:34:49-9#

K: Richtig, und immer auf Basis der Daten vom Versicherer selbst. #00:35:06-0#

S: Gehen wir noch zum ersten Punkt. Technochauvinismus bzw. Hype Begründung. Inwiefern ist es als Bedrohung zu betrachten, dass Personen eine so unrealistische Vorstellung von KI haben, so schlecht informiert sind und so großes Vertäuen haben. #00:35:35-8#

K: Ich glaube das ist oft einfach in der Gutgläubigkeit des Menschen begründet. Weil im Endeffekt muss es nur mit gut genügen Argumenten verkauft werden, der gutgläubige Mensch wird es immer annehmen. Im weitesten Sinne, also brutales Beispiel, könnte man es mit der Entstehungsgeschichte des zweiten Weltkrieges und Hitler vergleichen. Er hat seine Ideale gut genug verkauft und angepriesen, genug Leute sind im gefolgt. KI wird zwischendurch ähnlich angepriesen und verkauft. Die Leute glauben es, sie denken alles wird besser, ich will das auch, und sie vertrauen blind ohne noch nachzudenken. Irgendwann später, wenn irgendein prägendes Erlebnis passiert ist, dann wachen sie auf, vielleicht oder auch nicht, und sie denken fuck, warum habe ich das gemacht. Bis dorthin glauben sie einfach. Das Problem ist, dass auch in dem Bereich extrem viele Leute sind, die behaupten sie kennen sich mit Computern extrem gut aus. Dann gibt es Stammtischgespräche wo es heißt, das funktioniert so gut, mein Auto kann alles, irgendwann ist zu viel Vertrauen da ohne Hinterfragen, ohne Hausverstand und Hinterfragen. Drei meinen Freund machen dasselbe, bei denen ist nie was passiert, also mach ich mit. Typisches Problem von Mitläufern. Länge mal Breite zu wenig Information, Leute die es einfach zu einem gewissen Grad auch nicht verstehen können, weil Hintergrundwissen fehlt. #00:37:57-9#

S: Wie groß ist die Bedrohung dadurch momentan, wird sie sich erhöhen? Wird die Informiertheit besser werden? #00:38:10-8#

K: Langfristig wird die Informiertheit vermutlich besser werden Bezug. einzelner Bereiche, ich glaube aber nicht, dass der Informationsfluss, die Infopolitik in dem Fall genauso schnell sein wird wie die Einsatzzwecke. Man wird immer den gleich großen Gap haben, es werden nur quasi die einfachsten Bereiche nachziehen, in die Leute gut genug informiert sind. Heißt, autonomes Fahren - irgendwann wird bekannt sein, dass Autos in der aktuellen Stufe, dass das beweiden noch nicht alles kann, aber im Hintergrund hat sich die Technologie schon so

weit entwickelt, dass wir bei 4 oder 5 sind. Wo die Leute erst recht wieder denken, jetzt sind wir da, jetzt hat es alles! Wo wir eigentlich, per Definition, auch weit weg von autonom sind. Das Problem wird sich also lange nicht ändern. Es wird nur oberflächlich gemindert wirken, was es nicht ist. #00:39:20-9#

S: Einordnung Skala? #00:39:25-6#

K: Im Moment noch als eher kaum, weil alltägliche Verwendung nicht groß ist, dementsprechende noch kein großes Problem. Wenn schon alle im Verkehr ein Auto hätten, das vortäuscht autonom fahren zu können, dann würde es bestimmt viele Unfälle geben. So lange es noch nicht viele Teslas und CO auf den Straßen gibt, und eher die gehobene Gesellschaftsschicht sich das leisten kann, wir das Problem nicht so ausarten, vermute ich. Sobald die massentauglich sind, man um 10.000 Euro auch schon eines kaufen kann, wird das massiv zunehmen. Die notwendige Technik, für AF kann man einfach nicht so billig herstellen, dann wird das Problem wahrscheinlich noch mehr werden. Im Moment aber noch keine große Bedrohung. #00:40:39-6#

S: Der letzte Punkt. Fällt dir noch etwas ein, was ich nicht genannt habe als Grund für Künstliche Intelligenz, oder kannst du Lösungsvorschläge nennen? Vor Kurzem hat der EU Ethik Rat Richtlinien für KI Entwicklungen veröffentlicht. Sieben nett gemeinte Vorschläge wurden veröffentlich (K. lacht) die natürlich nicht verbindlich sind. Wie denkst du wird sich die KI Entwicklung, natürlich unterschieden in Europa, Amerika oder Japan, hinsichtlich Lösungen für die bestehenden Probleme bewegen? #00:43:19-8#

K: Ich glaube das ist nichts, was man auf politischer Ebene wirklich lösen kann. Im Endeffekt wird KI von einer Gruppe von Menschen geschaffen und das wird noch lange so bleiben. Das heißt, wenn man kein ausgeglichenes Team findet, das eine möglichst neutrale Entscheidungsfindung als mindert teilt, wird man es nie schaffen eine neutrale KI zu entwickeln. Und für alles, was große Datensets zum Erlernen braucht, so lange man diese nicht hat und er KI auch genug Zeit gibt es zu erlernen und auch die Ergebnisse zu verifizieren, in großen Testmengen das auch gegentestest, werden keine Verbesserungen kommen. Eigentlich müsste man zu einem gewissen Grad das System Europa und Amerika durchaus kombinieren. Vor allem in kritischen Bereichen, wo es im Menschenleben geht - also etwa juristische Einsätze - dass man KI einfach genug Zeit zum mitlernen gibt. Solcher Test muss man in Jahren, oder Jahrzehnte anlegen. Bedenken, wie lange hat es gedauert, dass wir die jetzigen Gesetze haben. Es hat Jahrhunderte gedauert, bis wir auf einem Gesetzesstand sind, auch bezüglich der Menschenrechte. Einer KI kann ich gar nicht alle

Daten zu Verfügung stellen, die zu dem geführt haben, weil sie so nicht vorhanden sind. Das heißt ich muss eigentlich mehr Geduld mit einbringen, und dann die Kombination Europa Amerika, also mehr kontrollierte staatliche Entwicklung. Bzw. das Entwickeln in kontrolliertem Forschungshand, und der Test in großer Menge, wie in Amerika. Die finale Entscheidungsfindung sollte trotzdem wesentlich häufiger noch von einem Menschen beurteilt werden. Und dann rückwirkend auch die KI um das verbessern, sonst hat das mitlernen nebenbei keinen Sinn. #00:46:15-5#

S: Also hat staatliche Regulierung da Potential etwas zu verbessern, indem sie einen Mindesttestzeitraum festlegt oder einen Prozentsatz an Sicherheit? #00:46:33-6#

K: Ja, an sich würde das helfen. Jedoch befürchte ich würde das daran scheitern, dass man es nicht sinnvoll in ein Gesetz formulieren kann. Weil sobald man es schwammig formuliert, ist es wieder reine Auslegung der jeweiligen Firma. Und dann ums weiter etwas in die Privatwirtschaft reicht, umso kürzer wird die Testzeit sein. Und auch rein staatliche Programme sind an Budgets gebunden und müssen Ergebnisse liefern. Auf der anderen Seite müsste man ein Gesetz formulieren, dass viel Interpretationsspielraum offenlässt, weil verschiedenen Verwendungen verschiedene Anforderungen haben. Aber wenn man es so schwammig macht, kann man es sich gleich sparen, weil jeder es für sich interpretieren wird. #00:47:34-8#

S: Und es ist auszugehen davon, dass neue Technologien schneller kommen, als regulierende Gesetze. #00:47:56-0#

K: Richtig. Und sämtliche neue Technologien in letzter Zeit, es gab immer wieder mal Sachen die schief gingen, es wir da sicher auch in nächster Zeit noch etwas geben, was das angeht, was dann definitiv eine größere Menschenmenge betreffen wird. #00:48:11-6#

S: Gibt es dann irgendwie einen denkbaren Lösungsansatz, der einen Marktraum oder sogar global alles auf einen ethischen Zweig bringen kann? Ist so etwas realistisch? #00:48:34-2#

K: Ich denke schon, dass es das langfristig geben wird. Vermutlich wird es davor aber massiv eskalieren. So wie z.B. bei den ersten Einsätzen von Atombomben, danach kam das Abkommen. Wahrscheinlich wird es nicht so extrem schlimm sein, aber ziemlich sicher wird es zuerst eskalieren müssen, damit einfach realisiert wird, wie gefährlich der Einsatz von

autonomer Entscheidungsfindung ist. Und auch von Systemen, die selber potentiell unkontrolliert weiterlernen können. Was jetzt noch keiner realisiert, jegliche Software weltweit ist vom Menschen geschrieben worden, und Menschen sind nicht fehlerfrei. Ab dem Zeitpunkt, wo ich einer Software - im militärischen Hintergrund, oder einer Organisation die etwas gegen eine Regierung macht - sobald die es schaffen eine AI zu entwickeln die sich selbst verbreiten kann, ab dann wird es extrem kritisch. Es hat selbst um das Jahr 2000 schon Computer Viren gegeben, die sich autonom verbreitet haben. Die hätten massiven Schaden anrichten können. Wir werden jetzt im Bereich autonomes Lernen immer besser, es ist also nur eine Frage der Zeit, bis eine Gruppe, sei es politisch, einem Staat untergeordnet, oder anders - dass die etwas schreiben, was Infrastruktur lahmlegt. Sei es Atomkraftwerke oder Kontrollsysteme von Gaspipelines, etc. Angriffsfaktoren gibt es viele. Und all dieses ist 100 Prozent IT gesteuert. Das Stromnetz in gesamt Europa wir von einer Handvoll Netzbetreiber, einer Handvoll Computersysteme betrieben. Da reicht eine über oder unter Spannung von ein paar Prozent, um in drei vier Europa. Ländern das Licht ausgehen zu lassen. Und es hat zwischendurch durchaus schon Probleme gegeben, dass ein Teil Stromausfall in Europäer war, nur weil eine Person sich um etwas vertan hat. #00:51:10-0#

S: Was hast du von dem Gedanken, dass Lösungen von den großen Playern kommen werden. Das Google, Facebook, Autoentwickler gemeinsam Richtlinien entwickeln werden, weil sie verstehen werden, um wie viel es geht, vor allem ohne Richtlinien. #00:51:42-1#

K: Ja, halte ich für realistisch. Aus Eigeninteresse. Facebook möchte einerseits alle deine Daten und diese auch vermarkten können. Im Umkehrschluss hat Facebook ein massives Problem damit, wenn eine Firma wir Cambridge Analytica mit Verdacht auf massiver Wahlmanipulierung, haben sie ein Problem damit. Dementsprechend wird es aus KI Sicht so sein, dass sie selbst den größten Nutzen haben, sie aber unbedingt vermeiden suchen, dass jemand anderes zu viel Nutzen zieht. Das begrüße ich, weil sowohl Google als auch Facebook darf man nicht unterschätzen. Durch die große Datenmenge, die sie haben in extrem vielen Bereichen. Sie können ohne Probleme alles, was mit Individuen oder pers. Entscheidungsfindung zu tun hat, extrem genau nachbilden, weil sie genügend Daten haben. Reisebewegungen, wann jemand Zuhause, an welchem Standort ist, fast alle Leute erlauben Facebook Standort tracking. Das heißt, Berufsgruppeninformationen, Bilder zu Hauf, Standorte etc. Also wenn die etwas entwickeln wollen haben sie genügen Daten. Google ist fast noch schlimmer, weil er einfach der Standardanbieter für Suchanfragen ist. Das heißt, sie können alles, was man mit Suchdaten verbinden kann, können sie beeinflussen. Märkte beeinflussen, wenn sie es wirklich wollen. Und es hat ja auch schon Case Studies gegeben, dass Google vor der WHO weiß, wann Grippepandemien ausbrechen, weil Leute die

Symptome früher googlen, bevor sie Ärzte konsultieren. Von dem her, ich bin dafür, dass sie untereinander Richtlinien ausmachen, ich hoffe sinnvolle, aber sie können was die politisch unabhängige Beeinflussung vom Menschen noch bei weiten am meisten. Also hier ist viel Potential, damit kommt aber auch viel Gefahr.

S: Vielen Dank Patrick.

Interview 3: Winter Philipp am 05. August 2019

S: Also vielen Dank erneut, für dieses Interview. Weil sie das gerade anschneiden wollten, fange ich gleich mit dieser Frage an. Es betrifft das Sorgfaltsproblem. Unter Sorgfaltsproblem fasse ich zwei Punkte zusammen. Einerseits das Problem von Bias der Entwickelnden bzw. Bias in den Trainingsdaten für KI-Anwendungen. Zweitens der Druck vieler Unternehmen, Anwendungen schnellstmöglich auf dem Markt anzubieten, wodurch die notwendigen Testphasen zu kurz sind, und Bias-Probleme dann nicht erkannt werden. Dieser Punkt bezieht sich also auf das Einbauen bzw. Nichterkennen von Bias in KI-Anwendungen.

Inwiefern sehen sie das Sorgfaltsproblem als Grund für Künstliche Unintelligenz, oder gar als Bedrohung, wie schätzen Sie es ein? #00:02:03-1#

W: Also da haben wir jetzt zwei Sachen. Einmal die Bias, und was noch? #00:04:02-6#

S: Genau, und zweitens den Druck auf private Unternehmen, wie man ihn vor allem in den USA hat, KI Anwendungen so schnell wie möglich auf den Markt zu bringen, wodurch nötige Testphasen zu kurz kommen. Als Beispiel kann man COMPAS hernehmen, die Risk-Rating KI, die die amerikanische Justiz objektiver machen sollte. Nachdem sie schon einige Zeit im Einsatz war hat man erst gemerkt, dass die manche Personengruppen extrem benachteiligt hat. Man hat sie mit gebiaseten Daten gefüttert, 60% der Sträflinge waren schwarz, daher auch die dementsprechende Bewertung durch die KI von Menschen unterschiedlicher Herkunft. Jetzt ist die Frage, wie relevant und groß ist dieses Bias Problem? #00:05:24-8#

W. Also grundsätzlich ist dieses schon ein extremes Problem, weil diese KI Systeme, oder ich sage mal neuronale Netze dazu, weil die im Hintergrund meist angewendet werden, die lernen anhand von diesen Datensätzen. Und wenn die Datensätze schon gebiased sind, dann wird das Netzwerk natürlich auch gebiased sein, da kommt man nicht darum herum. Das ist aber nur ein Teil von Machine Learning, das ist der Überbegriff. Hier hat man Inputdaten, und man weiß was rauskommen soll, das Label oder Target, das ist Supervised Learning. Und wenn man solche Datensätze hat, die vom Menschen produziert worden sind, sind die immer gebiased. Das kann man einfach nicht vermeiden. Weil der Mensch diese Daten per Hand, manuell labelled, und das Target produziert. Was man aber machen kann, sofern es möglich ist, dass man es nicht supervised trainiert, also mit vorhandenen Target labels, sondern unsupervised. Da hat man nur die Input Daten, zB. Bilder oder Text, und das Ziel selbst ist nicht vom Menschen manuell gelabelled. Das ergibt sich dann aus mathematischen Funktionen oder Gleichungen. Und dieses unsupervised Learning funktioniert dann bis zu einem gewissen Grad nicht so stark gebiased wie bei den supervised Daten. #00:07:23-7#

S: Verstehe. Wenn man jetzt supervised Learning einsetzt und um dieses Bias Problem nicht herumkommt, wie relevant ist der Faktor noch? Ist es eine Bedrohung, wird es in Zukunft beseitigt werden? Wenn man sich ansieht, dass eine Personengruppe jetzt höhere Gefängnisstrafen erhält, ist es ein Problem auf gesellschaftlicher Ebene, wie geht man damit um? #00:08:04-9#

W: Also relevant ist es jedenfalls. Man sollte es nicht verwenden, so lange man es nicht ausgiebig, etwa in einer Laborumgebung, getestet hat. Wenn man es trainiert hat, es funktioniert und ist fertig, sollte man es eine gewisse Zeit beobachten und sich Entscheidungen in verschiedenen Situationen des Systems ansehen, und anhand von den getroffenen Entscheidungen muss man evaluieren. Etwa aha, es erkennt alle Schwarzen als eher kriminell. Also genau diese Testphase, von der Sie schon geredet haben. Diese Systeme müssen ausgiebig getestet werden, bevor man sie anwendet. Das ist das erste. Das andere wäre, dass man bei den Daten selbst extrem darauf achtet, so wenig Bias wie möglich einzubringen. Das Problem ist, man weiß im Vorhinein oft gar nicht, welche Probleme auftauchen können. Die Personen, die das Datenset für COMPAS gemacht haben, haben wahrscheinlich nicht gemerkt, dass der Datensatz aus einem größeren Anteil von schwarzen bestand. Man sieht das in den Daten oft nicht, weil sie so komplex verschachtelt sind, weil komplizierte Abhängigkeiten untereinander und in den Daten versteckt sind. So dass man das teilweise erst im Nachhinein sieht, weshalb man den Bias oft nicht vermeiden kann. #00:10:00-4#

S: Wenn man jetzt von der momentanen Situation ausgeht. Wie Sie sagen, man sieht es oft nicht oder bemerkt es erst im Nachhinein - wie ist die momentane Lage? Kann man von einer Bedrohung sprechen, so dass aktuelle Menschen leiden, weil man erst danach draufkommt. wenn man es auf dieser Skala einordnen müsste (erklärt Skala) - wo würden Sie das Sorgfaltsproblem, also Bias und den Marktdruck, einordnen? #00:10:42-4# Als KUI, oder Situation dich Nachteile für Menschen schafft?

W: Ich glaube pauschal einordnen kann man das nicht, denn es machen verschiedenste Firmen auf diesem Gebiet etwas. Manche setzen halt größere Sorgfalt in das generieren von Datensätzen, andere weniger. Es reicht von harmlos bis völlig kriminell im Generieren der Daten. Das hängt von den Personen ab, die das dann im Hintergrund wirklich machen. Ich kann mir vorstellen, dass wenn man ein gewinnorientiertes Unternehmen ist, dass man beiden Daten teilweise schlampiger ist, oder sie so verändern, dass der eigene größte Vorteil erzielt wird. Das ist aber kein Problem vom Machine Learning selbst, sondern ein

menschliches. Dass man Leute dazu bringt, dass sie mit größerer Sorgfalt diese Daten generieren ist die Herausforderung. #00:12:04-2#

S: Wie denken Sie wird sich das entwickelt, wird diese Sorgfalt mehr kommen und das Problem eingegrenzt, vielleicht weil Auflagen kommen, oder wird es dieses menschliche Bias Problem immer geben? #00:12:22-2#

W: Ich weiß beispielsweise, dass der TÜV daran arbeitet, dass man solche Systeme auch zertifiziert. Also dass man gesetzlich vorgeschriebene Mindestanforderungen an ein System hat, es muss zu einem gewissen Prozentsatz so gut funktionierten. Vor allem bei extrem kritischen, z.B. bei Diagnose von Krebspatienten, soll das eingeführt werden. Da könnte man sagen, das System darf nicht mehr also 0,1 Promille falsche Diagnosen liefern. Da hat man schon eine ziemlich gute Eingrenzung, was es können muss und was es nicht können muss. Es muss also nicht perfekt sein, dass wird nie gehen, aber ein gewisses Limit muss eben erreicht werden, damit man es effektiv anwenden kann. Ich kann mir gut vorstellen, dass in Zukunft Gesetze kommen, die das regeln. Eben in Form von Zertifizierungen. Auch in Normen denkbar, wie etwa die ISO Norm, dass die auf Machine Learning Modelle erweitert wird. #00:13:51-6#

S: Hoffentlich. Wenn man jetzt vom momentanen worst case ausgeht, in den Bias einfließen, und man das dementsprechend auf der Skala einordnen, wo wäre das? #00:14:17-7#

W: Wenn man den schlimmsten Fall betrachtet, kann es extrem verheerende Auswirkungen haben, weil Artificial Intelligence ein extrem mächtiges Tool ist. Um zum Beispiel Daten auszuwerten oder zu verarbeiten, große Mengen davon. Und wenn man das unethisch verwendet, das Tool unethisch für eigene Zwecke etwa verwendet, kann es sein, dass extrem viele Leute darunter leiden. Gleich ist es mit einem Messer. Das Tool an sich ist neutral, vor allem wenn man Gemüse damit schneidet. Wenn man es hernimmt, um Menschen zu verletzen, kann es großen Schaden anrichten. Aber das Tool an sich bleibt immer neutral. Es ist immer der Mensch, der das Problem darstellt. Ich nehme an, dass es im militärischen Bereich vielleicht zu den Lenkungen von Raketen eingesetzt wird und viel mehr. Und das ist natürlich ethisch nirgends gerechtfertigt. Aber nachdem diese Technik immer mehr vorhanden ist, kann man sie natürlich zum Guten sowie zum Schlechten in beide Extreme verwendet. Um Personen zu töten oder sie zu heilen. Aber wenn man es falsch verwendet, kann man es sehr effizient falsch verwendet. Falsch im Sinne von unethisch.

S: Der nächste Punkt ist Physical Hacking. Miller und Valasek sind zwei im IT-Sicherheits-Bereich tätige und haben schon 2015 gezeigt, wie ein autonom fahrendes Auto gehackt und von außen fremdgesteuert werden kann. Ein anderes Beispiel einer Forschungsgruppe US-amerikanischer Universitäten zeigte, wie ein Sticker auf einem Verkehrsschild dazu führte, dass ein Fahrzeug das Stoppschild nicht mehr als solches erkennen konnte. Wie groß ist die momentane Bedrohung durch Physical Hacking?

W: Also das erste Thema mit dem Hacken vom Auto hat eigentlich nichts mit AI zu tun. Das ist eigentlich Standard IT Sicherheit, wie gut sind Verschlüsselungen etc. Wenn man ein System hat und das verschlüsselt, damit nur bestimmte Personen Zugriff haben, sollte das auch nicht hackbar sein. Wenn dieser Software Schutz gegeben ist, sollte das also nicht möglich sein und hat nichts mit KI zu tun. KI setzt erst darunter an, innerhalb des Systems. Zum Beispiel wenn man komplizierte Funktionen abbilden will. Da ist das zweite Thema derzeit sehr relevant. Das Kleben von Schildern etc. Das nennt man Adverserical Attacks. Da gehts es darum, wenn man etwa einen Objekterkennungsalgorithmus hat und den in einer Kamera anwendet. Da kann man sowohl außen, in der Umwelt, als auch im System selbst diese Adverserial Attacks anwenden, und so diese Netzwerke täuschen. Also von außen das klassische mit den Stickern. Oder wenn man eine lebensgroße Person auf ein Plakat druckt und wo platziert, wird das selbstfahrende Auto das auch als Person erkenne, obwohl es keine ist. Bis jetzt hat man noch keine gute Lösung für solche Probleme gefunden, ich selbst hab da auch keine Idee. Es gibt aber eigene Netzwerkarchitekturen, die auf diesem Prinzip aufbauen. Dass man z.B. ein Netzwerk hat, das eine Funktion ausführt, und ein anderes Netzwerk, dass das nur bewertet. Das zweite Netzwerk nennt sich der Diskriminator, der schaut, ob das Output des ersten Netzwerks valide ist. Wenn etwas falsch daran ist, soll das zweite Netzwerke das erkennen und sagen "Hey, da ist was komisch, schau dir das genauer an". Das geht genau in die Richtung wie kann man robuste Netzwerke machen, es fehlertolerant machen. Ein anderes Problem ist derzeit, wenn man so ein System hat und es etwas ausspuckt - wie sicher ist diese Aussage? Daran wird gearbeitet, aber momentan ist es noch ein Nachteil von Neuronalen Netzwerken, dass man keine Wahrscheinlichkeit hat, wie sicher ein Output ist, oder wie man ihn verwenden soll. Man nimmt es als gegeben an, als 100%. Es kann aber natürlich sein, dass etwas ein Tier nur zu 50% ein Hund ist und zu 30% eine Katze, wenn man es nicht genau definieren kann. #00:22:11-7#

S: Sie sprechen schon mein nächstes Problem an. Bevor wir dazu gehen, noch kurz zum Hacking: auch individueller oder gesellschaftlicher Ebene, wie sehr wird jemand dadurch bedroht. Wenn man etwa auch miteinbezieht, dass autonome Fahrzeuge so noch kaum im

Einsatz sind. Wenn man es wieder auf de Skala einordnen müsste, wie sehr ist Hacking derzeit relevant? #00:22:49-3#

W: Ich betrachte es schon als relativ großes Problem, weil viele Firmen ihre Daten zu schlecht verschlüsseln. Dass das umgegangen werden kann, und da teilweise riesige, personenbezogene, sensible Daten abgegriffen werden. Daten, die eigentlich niemand hergeben möchte. Die Organisationen, die Daten sammeln, sollten dafür sorgen, dass es unmöglich ist, dass diese Daten gestohlen werden können. Leider passiert es immer wieder, weil zu wenig Geld in IT Sicherheit investiert wird. Das Problem bei KI gestützten Sachen ist, dass man sehr große Datenmengen durchgehen kann. Wenn man eine riesige Datenbank an Bildern hat, kann man bestimmte Personen relativ einfach auf diesen Bildern finden. Ein Mensch bräuchte unglaublich Lange, tausende Jahre. Die KI vielleicht einen Tag. Das Auswerten ist also extrem effizient mittlerweile. Eben Personen identifizieren, verfolgen etc. Teilweise werden große nicht-personenbezogene Datenmengen aufgenommen, weil man Personenbewegung nachbilden will. Grundsätzlich kein Problem, auch wenn man sie wieder löscht. Sobald man sie aber mit anderen Daten kombiniert, und Herrn XY zuweisen kann, kann man einzelne Personen genau abbilden, das könnte zu einem großen Problem werden. #00:25:15-9#

S: Verstehe, also was nach einem Hackerangriff passieren kann. Weil man diese Daten so vielseitig auswerten kann. #00:25:24-0#

W: Genau. Was man macht, ist aus großen Mengen Rohdaten, Bilder etc., die wichtigsten Infos herauszufiltern, und da ist KI extrem effizient. #00:25:42-5#

S: Das nächste Problem, das haben sie schon angesprochen, ist das Blackbox Problem. Dass bei KNN die Begründung fehlt, wie die Entscheidungsfindung zustande kommt ist in der Blackbox versteckt. Ist das eine Bedrohung, inwiefern ist das momentan relevant? #00:26:12-5#

W: Das Blackbox Problem ist meiner Meinung nach nicht so schlimm, wie man es erwarten würde. In das menschliche Gehirn kann auch niemand reinschauen, und das Treffen von Entscheidungen nachverfolgen. Man geht mittlerweile dahin, dass man sagt, man will nicht jedes einzelne Detail von dem System verstehen. Sondern im High Level, warum es zu einer Entscheidung gekommen ist. Das läuft dann über den Oberbegriff "explaining AI", also das System selbst begründet, warum es entscheidet. Das andere ist "Explainable AI", man

analysiert wirklich genau jedes Detail was das System macht. Das ist, meiner Meinung nach, der falsche Weg. Beim Gehirn wie gesagt verfolgen wir ja auch nicht jedes einzelne Neuron, sondern argumentieren, warum wir eine Entscheidung getroffen haben. Auf demselben Niveau, auf dem die Entscheidung auch ist. Trotzdem gibt es da ein paar Methoden, wie man diese Netzwerke analysieren kann. Zum Beispiel kann man, wenn es einen bestimmten Output liefert, relativ genau schon mathematisch messen, was im Input, etwa im Imputbild dafür gesorgt hat, dass ein Objekt erkannt wurde. Nimmt man den Output Katze, das System hat eine Katze erkannt, und ich möchte wissen wodurch es eine Katze erkannt hat, wende ich diese Methode an und sehe am Input, die KI hat sich genau auf den Bereich fokussiert, wo die Katze im Bild ist. Das ist eine sehr effektive Methode, wie man den Output mit Input verbinden kann. Oder visualisieren, was im Input verantwortlich für diesen Output war. #00:29:01-6#

S: Explainable AI ist mir ein Begriff, explaning AI ist mir neu, können sie das näher ausführen? #00:29:11-0#

W: Man fordert einfach auch Begründungen an. Ich sage direkt, gib mir den Output und eine Begründung für diesen Wert. #00:29:26-9# #00:29:25-9#

S: Verstehe. Das Blackbox Problem kann man also als fast gelöst, oder weniger relevant ansehen? #00:29:36-7#

W: Mh gelöst noch nicht. Da gibt es derzeit noch extrem viel Forschung. Beliebt sind momentan auch diese modularen Netzwerke. Dass man nicht ein großes Ding, sondern mehrere kleine Subnetzwerke hat, wobei jedes davon für einen speziellen Task verantwortlich ist. Da kann man ganz genau sagen, dieses Sub Modul macht genau das. So kann man sich dann die Großmodelle zusammenstückeln. Das wäre ein Ansatz, wie man das wirklich interpretierbar machen könnte. #00:30:15-6#

S: Wenn wir wieder die momentane Situation betrachten und auf der Skala einordnen würden. Wo würden Sies einordnen? #00:30:29-4#

W: Gesellschaftlich ist es kein Problem... es ist eher ein Forschungsproblem, dass man das weiterentwickelt ist verdammt kompliziert. Wenn man nicht weiß, wie die Systeme funktionieren, dann ist es auch schwierig, die weiterzuentwickeln. Für den Anwender am

Ende ist es nicht wirklich ein Problem, der will auch nicht wissen wie das funktioniert. Das ist nicht vorgesehen und auch nicht notwendig. Jeder hat ein Smartphone und man versteht nicht alle Komponenten, man hat das Ding, vertraut ihm bis zum einem gewissen Grad - da kommt wieder die Zertifizierung ins Spiel! - und verwendet das einfach. Man analysiert es gar nicht genau, wie das funktioniert. #00:31:55-1#

S: Es gab den Fall bei Amazon, wo man im Nachhinein bemerkt hat, dass der Faktor Geschlecht beim Aussuchen von neuem Personal eine Rolle gespielt hat. Man hat es erst später gemerkt, weil die Entscheidungsfindung in einer Blackbox gesteckt hat. In dem Fall wurden Frauen benachteiligt. Wird das in Zukunft noch auftreten, ist es also ein Problem? #00:32:33-2#

W: Damit wären wir wieder am Anfang, mit genauen Testphasen könnte man das herausfinden. Wenn man diese Input Output Verbindung anwendet, würde man genau sehen, aha das Netzwerk konzentriert sich eher auf Männer, warum ist das so - wahrscheinlich, weil die Daten gebiased sind. Dann kann man ansetzen und das Datensetz so korrigieren, dass das Geschlecht keine Rolle mehr spielt. Ich kann das Geschlecht einfach aus den Daten herausgeben, dann wird das keine Rolle mehr spielen. Auch alle Anzeichen, die auf Geschlecht hinweisen, da muss man aufpassen. Keine Namen, da wäre implizit das Geschlecht gegeben. Man muss sehr vorsichtig sein, dass ist wieder die Verbindung innerhalb der Daten. #00:33:48-9#

S: Ein Punkt den ich am Anfang schon beschrieben habe, ist Broussards Technochauvinismus. Zu dem der sich wiederholende KI Hype-Cycle als Auslöser von unrealistischen Erwartungen, also weiter auch von KUI. (Erklärt Frage, Beispiel erster Todesfall) #00:34:33-3#

W: Das ist eine schwierige Frage. Derzeit ist KI extrem gehyped und bekannt. Es hat ein sehr gutes Image und die Leute denken, es hat zukunftspotential, sonst würden nicht so viele investieren. Das Problem ist, irgendwann werden die Erwartungen zu hoch und die Forschung kann nicht mehr mithalten. Der Hype flaut wieder ab. Derzeit ist es eben gerade gehyped und wir erwarten das Abnehmen in Zukunft, es kann also auch sein, dass das Vertrauen in KI Anwendungen wieder abnimmt. Das ist etwas sehr Subjektives. Ich kann mir vorstellen, dass wenn Leute solche Systeme immer wieder verwenden, sie sich daran gewöhnen. Vertrauen durch Anwendung sozusagen. Teilweise verwenden viele Leute ja schon Geräte, ohne zu wissen, dass KI im Hintergrund im Einsatz ist. Sie vertrauen blind, ohne zu wissen. Es wird auch in Zukunft so sein, dass es viele Anwendungen geben wird, wo

Personen nicht informiert werden über den KI Einsatz im Hintergrund. Weil es einfach eine unwichtige Information für die Endanwender ist. #00:37:32-7#

S: Sie sagen, es wird sich in Zukunft nicht verändern bzw. das Vertrauen wird eher steigen - gibt es eine Situation in dem das bedrohlich sein könnte, dieses blinde Vertrauen? #00:37:45-9#

W: Das ist schwierig. Man verwendet das ja alles freiwillig. Man könnte sich ja entschließen, seine Daten nicht mehr herzugeben, und die Daten werden dann nicht mehr verwendet. Viele Anwendungen bauen aber darauf, dass Daten von Personen gewonnen werden, ohne das würde es sie nicht geben. Es ist schwer, im Vorhinein zu sagen, ob das irgendwann möglicherweise missbräuchlich verwendet wird, dann könnte es natürlich schlecht für Personen sein. Das sollte im Prinzip eher die Ausnahme als die Regeln sein. #00:38:53-2#

S: Das heißt momentan gibt es eher wenige Situationen, in denen das schädlich sein könnte? Es ist ja auch autonomes Fahren so noch nicht im Einsatz. Und die Informiertheit ist zwar schlecht, es ergeben sie aber nicht direkt Bedrohungsszenarien daraus? #00:39:19-1#

W: Direkte Bedrohungen sind es e nie. Es geht immer um Daten, Informationen, Wissen über andere Personen. Das kann dann explizit, wissentlich falsch verwendet werden. Das ist dann eigentlich die Bedrohung. Die Leute, die das dann für ihre eigenen Zwecke auswerten... Also natürlich kann immer etwas passieren. Das Ziel ist es ja, langfristig genau das zu minimieren. Z.B. dass man KI Systeme entwickelt, die in irgendwelchen Aufgaben besser sind als der Mensch, oder die in gefährlichem Einsatz verwendet werden, damit der Mensch das nicht mehr machen muss. Da gibt es durchaus Potential, dass es langfristig extrem hilft. Auch wenn es immer jemanden geben wird, der es missbraucht. #00:40:48-0#

S: Wenn man das wieder auf der Skala einordnen würde, als Grund für Künstliche Unintelligenz? Ein weiterer Gedanke, den sie im Buch beschreibt, ist auch, dass KI eingesetzt wird, obwohl es nicht die effizienteste Lösung darstellt. Broussard zeigt einen Vergleich von Schulbüchern vs. IPads für Schüler. Kosten der Anschaffung, Installation, Instandhaltung, Einführung, Ersatz, Verschleiß, Wartung vs. Kosten eins Buches, besserer Lerneffekt, Lebensdauer von fünf Jahren etc. Wenn man das als KUI berücksichtigt. #00:42:08-5#

W: Oft fehlt da das Wissen von den Leuten, die solche Entscheidungen treffen. Politiker als Musterbeispiel, der weiß nicht bei jedem Thema, wie das im Hintergrund funktioniert und muss sich auf seine Berater verlassen. Da werden Informationen recherchiert und über mehrere Schichten weitergegeben, da geht natürlich viel verloren, und die entscheidungstreffenden Personen haben immer ein gebiasetes Bild von der Situation und wenig Wissen. Da ist es oft schwierig, gute Entscheidungen zu treffen. Wenn ein Wissenschaftler das entscheiden würde, würde er es als kompletten Schwachsinn bezeichnen, andere sagen auf jeden Fall, da lernen die Schüler besser, der Dritte hat keine Ahnung. Am Ende muss eine Entscheidung getroffen werden, das ist oft die falsche. #00:43:40-5#

S: Also ein Informationsproblem. Wenn ich Sie nochmal bitten dürfte, den Technochauvinismus auf der Skala einzuordnen: #00:44:04-0#

W: Ich würde sagen es ist leicht problematisch. Aber das kann man lösen, durch Aufklärungskampagnen oder Experten, die als Berater fungieren. Wäre ohne weiteres möglich. Ordne es als "etwas" ein.

#00:44:33-8#

S: Mein fünfter Punkt, mal etwas positives, hier geht es um Lösungsvorschläge. Sie haben schon erwähnt, dass die TÜV anstrebt Zertifikate auszustellen. Gibt es da noch mehr Lösungsansätze, um KUI zu vermeiden bzw. Fördermaßen für sinnvolle KI Anwendungen. #00:45:13-4#

W: Das Problem ist, wenn man KI Methoden verwenden möchte, braucht man dazu Experten, die das durchführen. Und das ist relativ teuer, und Zeit- und Personalaufwändig, eine KI Methoden für einen speziellen Task zu entwickeln. Dementsprechend muss da sehr viel Geld in die Hand genommen werden. Deswegen gibt es derzeit noch nicht viele Leute, die sage, wir wollen es nur für non-kommerzielle Sachen verwenden. Es gibt z.B. nicht kommerzielle Bewegungen, von Google glaub ich, open AI hieß das. Die haben gesagt es ist komplett Open Source, jeder soll Zugang haben und alle Ergebnisse öffentlich zugängig sein. Aber die sind mittlerweile zu einer Firma geworden, weil sie jetzt doch Ertrag daraus brauch oder haben wollen. Insofern, wenn man diese Methoden der Allgemeinheit zur Verfügung stellen möchte, müsste das wahrscheinlich staatliche finanziert werden. Über ein Institut vielleicht, darüber wie man mit KI Methoden für ärmere Leute zur Hilfe einsetzen kann. Die Wissenschaftler würden das Gehalt über den Steuerzahler verdienen und am Ende kommt

ein nicht kommerzielles Produkt heraus. Von so etwas habe ich leider noch nie gehört, wäre aber erstrebenswert, wenn solche Bewegungen entstehen würden. Vor allem durch den Hype und weil doch zunehmend Leute sich mit dem Thema auskennen, wären da viele begeistert mitzumachen, ich sehe da viel positives Potential. #00:47:45-5#

S: Es gab vor Kurzem den EU-Ethik Rat, der damit beauftragt war, Richtlinien für sinnvolle und menschengerechte KI zu entwickeln. Da sind nach eineinhalb Jahren sieben interessante Vorschläge gekommen, diese sind natürlich nicht bindend. Jetzt stellt sich die Frage, wie groß ist das Potential von staatlicher Regulierung, oder Co-Regulierung. Alle Anwendungen, die auf den Markt kommen, auf ein sicheres Grundmaß zu bringen und zu regulieren. #00:48:32-7#

W: Schwierige Frage. Es sollte jedenfalls einen Regulierungs- und Kontrollprozess geben. Man sollte diese Systeme auf keinen Fall ungetestet einsetzen. Wie man das macht, da wird es von verschiedenen Organisationen und Staaten verschiedenen Ansätze geben, und es wird sich herauskristallisieren, was am besten funktioniert. Bestimmt wird es Fehlschläge und Probleme geben, die auftreten, aber das ist natürlich Prozess von der gesamten Entwicklung. #00:49:17-3#

S: Mein letzter Punkt - haben Sie noch Ergänzungen? Natürlich ist das Thema ein sehr großes und es gilt noch Lösungen zu finden für Datensicherheit, Verantwortlichkeit, ethische Fragen, gibt es trotzdem von ihrer Seite noch Punkte für KUI oder schlechten Einsatz, die ich nicht genannte habe? Sie haben schon genannt, dass die Auswertungen von gesammelten Daten "schlecht" eingesetzt werden kann, gibt es dann noch relevante Punkte? #00:50:02-4#

W: Was mit spontan einfällt, ist, wenn Firmen Daten von ihren Kunden erheben, wissen die oft nicht genug oder werden gar nicht aufgeklärt, welche Daten aufgezeichnet werden und die Verwendung davon. Ich denke da z.B. an Facebook. Dass die User im Dunklen darüber gelassen werden, was mit ihren Daten passiert, das sehe ich als relativ großes Problem, das man unbedingt angehen sollte. Auch gesetzlich sollte man das transparenter machen. Jeder sollte eine genaue Aufschlüsselung bekommen, wenn man diesen Dienst verwendet werden genau diese Daten erhoben und so und so verwendet. Dem sollte der User explizit zustimmen. Das umgeht auch andere Probleme, dann können nicht beliebig personenbezogene Daten gesammelt werden, weil man immer die Zustimmung braucht. #00:51:57-1#

S: Verstehe, da haben wir wieder das Informationsproblem in der Bevölkerung. Vielleicht Wissen darüber, dass Daten gesammelt werden, aber wenig darüber, was und wieviel damit gemacht wird. #00:52:14-0#

W. Genau. Ein weiterer Punkt wäre Bildung. Dass man Leute in Informationskampagnen aufklärt, was kann mit Daten gemacht werden. Damit sie ein Bewusstsein dafür bekommen, wie sie mit ihren eigenen Daten umgehen soll, oft wissen sie das nicht. Wenn ich z.B. das Wissen hab, dass eine Firma spezielle Methoden entwickelt und verwenden kann, werde ich vielleicht vorsichtiger sein mit allem was ich preisgebe. Weil ich mich besser auskenne in dem Gebiet. #00:53:07-7# #00:53:08-6#

S: Ich danke vielmals für das Interview!

14.2 Eigenständigkeitserklärung

Hiermit gebe ich die Versicherung ab, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Publikationen entnommen sind, sind als solche kenntlich gemacht. Die Arbeit wurde in gleicher oder ähnlicher Form weder im In- noch im Ausland (einer Beurteilerin/einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt.

Wien, am 31. August 2019