MASCHINE

MASCHINE

MASCHINE

MASCHINE

MASCHINE

MASCHINE

Entscheidungen Mensch – Algorithmus

Daimler und Benz Stiftung

In Kooperation mit

ALGORITHM WATCH

2 **Künstliche Intelligenz als Nebelmaschine**
Eckard Minx

4 **Immer Menschen, niemals Maschinen**
Matthias Spielkamp

12 **Human Resources Analytics – fair und diskriminierungsfrei?**
Katharina Simbeck

20 **Innovation vorausdenken – der Forschungscampus ARENA2036**
Peter Froeschle, Clemens Ackermann

28 **Verwirrungsmaschinen: Künstlicher Intelligenz begegnen**
Manuela Lenzen

36 **Programm**

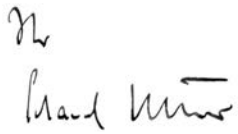
40 **Impressum**

Künstliche Intelligenz als Nebelmaschine

Die Vorstellungen und Definitionen davon, was Künstliche Intelligenz (KI) ist und kann, gehen weit auseinander. Zwischen Heilsversprechen und Untergangsszenario scheint alles denkbar. Selbst der Begriff bleibt unklar. Dennoch ist KI zu einem Thema mit hoher Priorität für viele Entscheidungsträger nicht nur in Unternehmen, sondern auch in der Politik geworden. Die mangelnde Genauigkeit in der Definition und die weit divergierenden Erwartungshaltungen bei einer gleichzeitigen Allgegenwart von Systemen, die auf Algorithmen beruhen, sorgen für einen Hype, der den Blick auf die wichtigen Fragen verstellt.

Das Versprechen, anhand gigantischer Datenmengen und großer Rechenleistungen Unsicherheit und Unvorhergesehenes möglichst ausschalten zu können, führt tatsächlich dazu, immer mehr Entscheidungen auf Algorithmen zu übertragen. Damit werden schiere Daten zum Faktum. Doch unabhängig davon, wie groß der Datenfundus und die Rechenleistung sind und wie clever die Algorithmen programmiert wurden, die Ergebnisse beruhen immer auf einer statistischen Korrelation, deren Basis in der Vergangenheit liegt. Dabei wird leicht oder auch zu gern übersehen, dass Entscheidungen sowohl in Unternehmen wie in der Gesellschaftspolitik Gestaltungsprojekte sind und nicht unausweichliches Schicksal, welches auf linearen Fortschreibungen beruht.

Mit der Jubiläumsveranstaltung des 20. Innovationsforums bietet die Daimler und Benz Stiftung eine Plattform des Austauschs über den Umgang und die Gestaltmöglichkeiten, die Künstliche Intelligenz bietet. Die vorliegende Publikation dokumentiert die Ergebnisse der Veranstaltung und möchte einen Beitrag dazu leisten, dem Hype um Künstliche Intelligenz etwas informierter und gelassener zu begegnen.



Prof. Dr. Eckard Minx
Vorstandsvorsitzender der Daimler und Benz Stiftung

Prof. Dr. Eckard Minx

Immer Menschen, niemals Maschinen

Über das Politische im Umgang mit Systemen zum automatischen Entscheiden

Matthias Spielkamp

Stellen Sie sich vor, Sie suchen einen Job. Das Unternehmen, bei dem Sie sich bewerben, sagt Ihnen, dass das Bewerbungsverfahren viel einfacher wird, wenn Sie den Nutzernamen und das Passwort Ihres persönlichen E-Mail-Kontos übermitteln. Dann könnten nämlich Ihre E-Mails analysiert werden, um daraus ein Persönlichkeitsprofil zu erstellen. Sie müssen keinen komplizierten Fragebogen ausfüllen, denn das ist langweilig und auch viel weniger zuverlässig, als Ihre Mails auszuwerten, denn die sind viel schwieriger zu manipulieren. Am Ende gewinnen alle: das Unternehmen, weil es viel exaktere Bewerbungsprofile bekommt und damit passendere MitarbeiterInnen, Sie, weil Sie weniger Arbeit haben und einen Job finden, der besser zu Ihnen passt, und die Firma, die das Analysesystem anbietet, denn sie kann damit viel Geld verdienen.

Als wir bei der Recherche zu unserem Report [1], einer Bestandsaufnahme von Systemen zum automatisierten Entscheiden in 12 EU-Ländern, auf dieses Beispiel stießen, waren wir erst einmal sprachlos. Es muss sich um einen Fehler handeln, dachten wir, und die Aussage der Analysefirma, dass keine BewerberIn je den Zugang zum E-Mail-Konto verwehrt habe, machte die Sache nicht besser. All das sei außerdem auch nach den hohen Datenschutzstandards der Europäischen Union völlig legal.

Es sind Fälle wie diese, die bei vielen Menschen Unbehagen, wenn nicht gar Angst auslösen. Davor, ausgespäht und überwacht zu werden, damit andere sich ein genaues Bild von ihnen machen können, ohne dass man dieses Bild selber noch irgendwie unter Kontrolle hat. Davor, zum Objekt der Entscheidungen solcher Systeme zu werden, deren Funktion man nicht versteht, verstehen kann – zum einen, weil sie so komplex sind, zum anderen, weil diejenigen, die sie

[1] *https://algorithmwatch.org/en/automating-society/*

Matthias Spielkamp

verwenden, nichts darüber verraten, wie sie funktionieren und zu welchen Zwecken sie eingesetzt werden.

Dieses Unbehagen ist zugleich berechtigt und ein Problem. Denn Automatisierung ist nichts Schlechtes – im Gegenteil: Wer wollte statt einer Tabellenkalkulation Bleistift und Papier benutzen, um Berechnungen mit großen Zahlen nach komplexen Formeln auszuführen? Wer möchte aus seinen E-Mails manuell unzählige Spam-Nachrichten herausfiltern? Und wer möchte zurück zu Katalogen von Webseiten, die man wie Gelbe Seiten nach Angeboten durchstöbert, statt einfach Stichworte in die Suchmaschine der Wahl einzugeben? Das sind Beispiele für Automatisierung, die wir täglich mehr oder weniger bewusst nutzen. Dazu kommen die Verfahren, ohne die unsere Gesellschaft überhaupt nicht funktionieren würde: vom automatisierten Datentransport, der den Kern des Internets ausmacht, bis zum automatisierten Elektrizitätsmanagement, ohne das unser Alltag zum Stillstand käme. Wir leben nicht nur schon sehr lange in einer automatisierten Welt, wir profitieren auch enorm von dieser Automatisierung. Dass sowohl die Aufmerksamkeit für solche Prozesse als auch die Skepsis ihnen gegenüber in den vergangenen Jahren gestiegen ist, liegt vor allem daran, dass Anbieter behaupten, Systeme entwickeln und einsetzen zu können, die automatisiert Entscheidungen treffen. Oft ist in diesem Zusammenhang die Rede von so genannter Künstlicher Intelligenz. Der Begriff ist äußerst unglücklich gewählt. Er erweckt bei vielen Menschen den Eindruck, hier seien Prozesse am Werk, die mit menschlicher Intelligenz vergleichbar sind, nur von Maschinen ausgeführt werden. Das ist nicht der Fall. Vielmehr handelt es sich um statistische Verfahren, die mitunter den Eindruck erwecken, als kämen die Ergebnisse ähnlich zustande wie bei menschlichem Denken. Doch nur weil ein Computer einen Menschen im Schachspiel besiegen kann, hat er darum noch keine Absichten, besitzt keine Kreativität und keine Autono-

mie – alles Eigenschaften, die den Menschen erst zum Menschen machen.

Der Begriff Künstliche Intelligenz ist unglücklich gewählt. Wir sollten vielmehr von Systemen zum automatisierten oder algorithmischen (Vor-) Entscheiden sprechen

Um den falschen Eindruck nicht zu verstärken, sollten wir daher vielmehr von Systemen zum automatisierten oder algorithmischen (Vor-)Entscheiden sprechen, von ADM-Systemen (automated decision-making). Hier werden menschliche Entscheidungen an automatisierte Systeme übertragen, indem ein von Menschen entwickeltes Entscheidungsmodell so in Software übertragen wird, dass eine Maschine es umsetzen kann. So wird etwa ein automatisiertes Auto mit einem System ausgerüstet, das in der Lage ist, Hindernisse zu erkennen und ihnen auszuweichen. Dass es Hindernissen ausweichen sollte, hat aber nicht das Auto oder sein Computersystem entschieden, sondern deren EntwicklerInnen. Daran ändert auch die Tatsache nichts, dass immer häufiger so genannte „selbstlernende“ Systeme eingesetzt werden, die im Lauf der Zeit Hindernisse immer besser erkennen. Die Entscheidung, diesen Hindernissen auszuweichen, statt sie zu überfahren, bleibt eine menschliche.

Was als akademische Spitzfindigkeit erscheinen mag, hat entscheidende gesellschaftliche Konsequenzen: Die Verantwortung dafür, welche Auswirkungen der Einsatz solcher Systeme hat, tragen immer Menschen, niemals Maschinen. Die Begründung, „Das hat der Computer so entschieden“, kann Ausdruck von Hilfslosigkeit sein und auch das Ergebnis von gezielten (Falsch-)Darstellungen derjenigen, die sie einsetzen. Sie bleibt jedoch immer falsch, weshalb wir nicht zulassen dürfen, dass Menschen sich mit dieser Begründung ihrer Verantwortung entziehen.

Österreichs Arbeitsmarktservice (AMS) ist das beste Beispiel. Hier wurde ein statistisches Modell entwickelt, das auf Grundlage von Ausbildung, Geschlecht, bisheriger Erwerbskarriere, Alter, Staatsbürgerschaft und anderen Kriterien berechnen soll, welche Chancen Arbeitslose haben, wieder Arbeit zu finden. Dazu werden Menschen in drei Kategorien eingeteilt: solche mit hohen, mittleren und niedrigen Chancen, einen Arbeitsplatz zu finden. Konkret bedeutet das, dass ab 2020 Menschen in der mittleren Kategorie stärker gefördert werden sollen, während Arbeitssuchende mit hohen und niedrigen Chancen am Arbeitsmarkt weniger unterstützt werden.

Die Verantwortung dafür, welche Auswirkungen der Einsatz solcher Systeme hat, tragen immer Menschen, niemals Maschinen

Die Begründung: Ressourcen der Arbeitsmarktpolitik sollen effizienter eingesetzt und Arbeitssuchende treffsicherer gefördert werden. Ein Ziel, das wahrscheinlich in dieser allgemeinen Formulierung viele unterstützen würden. Das Problem daran: Die Gewichtung der verschiedenen Faktoren entspricht dem, was bisher der Fall ist. Zum Beispiel haben Frauen mit „Betreuungspflichten“ schlechtere Chancen, einen Job zu finden, als Frauen ohne diese Pflichten oder Männer mit Kindern, bei denen die Frage nach Betreuungspflichten erst gar nicht gestellt wird – sicher deshalb, weil die Pflichten so oft von Frauen übernommen werden, dass es sich bei ihnen auf ihre Arbeitssituation auswirkt, bei Männern jedoch nicht.

Wenn nun aber diese Tatsache zur Grundlage des Modells wird, das über die Zukunft mitentscheidet, verstärkt es diese Situation. Wer das als objektive Abbildung der Realität verteidigt, verkennt den gesellschaftlichen Sprengstoff, der darin enthalten ist. Politik sollte die Aufgabe haben, eine gerechtere Welt herzustellen, mit gleichen Chancen für Teilhabe aller. Beim AMS wurde die Entscheidung getroffen, eine bestehende Ungerechtigkeit fortzuführen und damit zu verstärken – wohlgemerkt von Menschen. Der zu diesem Zweck entwickelte Algorithmus ist nichts weiter als der Umsetzungsmechanismus dieser politischen Entscheidung, für die sich die rechtfertigen und verantworten müssen, die sie getroffen haben.

Es ist die Blindheit für den politischen Kern vieler Automatisierungsprozesse, die berechtigte Kritik hervorruft. Denn die Risiken sind lange bekannt. Als beispielsweise IBM 1957 damit begann, die „Halb-Automatische Umgebung zur Unternehmensuntersuchung“ (Semi-Automated Business Research Environment, SABRE) zu entwickeln, um Tickets für die Flugzeuge von American Airlines zu verkaufen, ist es sicher nicht einfach nur blauäugig anzunehmen, dass damit auch ein gutes Ziel verfolgt wurde: Man wollte den bislang sehr umständlichen und fehleranfälligen Reservierungsprozess für die Fluglinie und die KundInnen verbessern. Im folgenden Vierteljahrhundert jedoch kam American Airlines darauf, dass man das System auch für andere Zwecke nutzen kann. Denn nachdem viele andere Fluglinien angefangen hatten, SABRE ebenfalls zu verwenden, begann American Airlines damit, es so zu manipulieren, dass die Angebote der eigenen Firma gegenüber anderen bevorzugt wurden. Reisende bekamen nicht mehr das günstigste Ticket angeboten, sondern das, das AA verkaufen wollte. In einer Anhörung des US-Kongresses gab sich AA-Präsident Robert L. Crandall unbeeindruckt von Vorwürfen, unfair gehandelt zu haben: „Die bevorzugte Anzeige unserer eigenen Flüge, und der dadurch erhöhte Marktanteil, war der wettbewerbliche Grund dafür, das System überhaupt erst zu entwickeln.“

In ihrer grundlegenden Untersuchung mit dem Titel „Auditing Algorithms“ (Algorithmen prüfen) haben Christian Sandvig und seine Ko-Autoren diese Sichtweise „Crandalls Klage“ (Crandall’s complaint) getauft: Warum sollte man ein teures algorithmisches System entwickeln und einsetzen, wenn man es nicht zu seinen eigenen Gunsten beeinflussen kann? Die US-Regierung sah es jedoch anders als Robert Crandall und erließ 1984 einen heute wenig bekannten Paragrafen, den wohl ersten zur Algorithmenregulierung weltweit. Unter dem Titel „Darstellung von Informationen“ schreibt er vor, dass für Reservierungssysteme von Fluggesellschaften jeder Person auf Anfrage die aktuellen Kriterien offenzulegen sind, die diese Systeme nutzen, um Flüge zu klassifizieren, wie die Kriterien gewichtet werden und welche Vorgaben bei der Entwicklung des Algorithmus gemacht wurden.

Warum sollte man ein teures algorithmisches System entwickeln und einsetzen, wenn man es nicht zu seinen eigenen Gunsten beeinflussen kann?

Erstaunlich ist zu sehen, was in den 35 Jahren seit Verabschiedung dieses Gesetzes geschehen ist. Auf der einen Seite wurden immer bessere Systeme zur Automatisierung komplexer Vorgänge entwickelt, die einen gewaltigen Einfluss auf unser aller Leben haben und haben können. Drei Beispiele: In

den Niederlanden versucht die öffentliche Verwaltung, „SozialbetrügerInnen“ zu identifizieren, indem sie Daten verschiedener Behörden kombiniert und von einem Algorithmus filtern lässt. In Dänemark will die Regierung ein System einführen, das automatisch aus Daten über Familien herauslesen soll, ob Kinder in Gefahr sind, vernachlässigt zu werden. In Polen werden Arbeitslosen bereits seit 2014 auf Basis einer automatisierten Klassifizierung Hilfen gewährleistet oder vorenthalten.

Auf der einen Seite wurden immer bessere Systeme zur Automatisierung komplexer Vorgänge entwickelt, die einen gewaltigen Einfluss auf unser aller Leben haben. Auf der anderen Seite hat die Regulierung nicht Schritt gehalten.

Auf der anderen Seite hat die Regulierung nicht Schritt gehalten. Wo 1984 bei einem kommerziellen System aus reinen Verbraucherschutzgründen Transparenz vorgeschrieben wurde, um erkennen zu können, ob Menschen in ihrer Eigenschaft als Konsumenten benachteiligt werden, zeigen das AMS-System und die genannten Beispiele, dass es längst um Entscheidungen geht, die viel stärker in unsere Rechte einschneiden. Doch darüber, wer diese Systeme zu welchem Zweck einsetzt, wer sie mit wem zusammen entwickelt, welche politischen und statistischen Modelle ihre Grundlage bilden und wer auf welche Art davon betroffen ist, darüber wird uns zu oft jegliche Auskunft verweigert:

Der Algorithmus des niederländischen Systems wird geheim gehalten; ein Netzwerk niederländischer Bürgerrechtsorganisationen hat auch auf eine Anfrage nach dem Informationsfreiheitsgesetz von der Regierung keine Auskunft bekommen. Daraufhin hat das Netzwerk Klage eingereicht, über die noch nicht entschieden wurde. Doch niemand Geringerer als der UN-Sonderberichterstatter für Menschenrechte und extreme

Armut hat einen so genannten Amicus Curiae verfasst – ein Schreiben, mit dem er die Klage der Bürgerrechtler unterstützt.

In Dänemark haben investigative Medienrecherchen zu massiven Protesten der Bevölkerung geführt, sodass die Regierung den Start des Programms Ende 2018 erst einmal auf unbestimmte Zeit verschoben hat. Und in Polen mussten Bürgerrechtsorganisationen auf Herausgabe von mehr Informationen klagen, um verstehen zu können, was passiert – um es dann scharf zu kritisieren. Doch erst als auch der polnische Bundesrechnungshof das System zerpflückte, weil es nicht nur ineffizient sei, sondern auch gesellschaftliche Gruppen diskriminiere, darunter alleinerziehende Mütter und Menschen mit Behinderungen, und das Verfassungsgericht entschied, dass die gesetzliche Grundlage für die Datenverarbeitung unzureichend sei, beschloss die Regierung, das System Ende 2019 einzustellen.

Es ist eines Rechtsstaats unwürdig und eine Bedrohung für die Demokratie, wenn Verfahren mit weitreichenden Auswirkungen auf Individuen und Gesellschaft geheim gehalten werden. Doch Forderungen nach mehr Transparenz werden oft mit drei Argumenten gekontert:

1. Derartige Verfahren unterliegen meist dem Schutz von Geschäfts- und Betriebsgeheimnissen der Firmen, die sie entwickelt haben.

2. Einen Algorithmus offenzulegen, ist weder notwendig noch hinreichend – nicht notwendig, weil die Funktionsweise auch beschrieben werden kann, ohne den Code zu veröffentlichen, und ohnehin die Auswirkungen entscheidend seien, nicht das Verfahren; nicht hinreichend, weil der Algorithmus nicht allein entscheidend sei, sondern auch die Daten, mit denen er arbeitet, und sich das Verfahren auch sehr dynamisch verändern kann, vor allem, wenn selbstlernende Techniken eingesetzt werden.

3. Transparenz allein nützt nichts, weil sie weder dazu führt, dass man einem System nicht mehr ausgeliefert ist, noch Widerspruchsmöglichkeiten mit sich bringt oder einen Anspruch auf Wiedergutmachung, etwa durch Schadensersatz.

Transparenz ist an sich kein Heilmittel, aber sie ist eine entscheidende Voraussetzung für den demokratischen Umgang mit automatisierten Systemen

All das ist richtig. Daher muss klar sein, dass Transparenz an sich kein Heilmittel ist. Aber sie ist eine entscheidende Voraussetzung für den demokratischen Umgang mit automatisierten Systemen. Erst wenn wir wissen, aus welchem Grund sie eingesetzt werden – soll lediglich Geld gespart werden, oder ist ein Problem so komplex, dass man es anders nicht glaubt lösen zu können? –, zu welchem Zweck sie eingesetzt werden, wer von ihnen auf welche Art betroffen ist, wer sie entwickelt hat – eine Behörde allein, mit Hilfe eines kommerziellen Unternehmens, das Unternehmen allein – und auf welchen Annahmen es beruht, können wir entscheiden. Entscheiden, ob wir überhaupt damit einverstanden sind, dass das System zum angegebenen Zweck eingesetzt wird – oder ob wir beispielsweise niemals die Entscheidung über Freiheitsentzug automatisiert treffen lassen wollen. Wenn wir grundsätzlich einverstanden sind, wollen wir entscheiden können, wie viel Auskunft über die Funktionsweise wir brauchen, um darauf zu vertrauen, dass das System angemessen

arbeitet. Und wenn wir es für angemessen halten, dass der Code zum Beispiel staatlichen Behörden gegenüber offengelegt wird, dann muss sich dem auch ein Unternehmen beugen, das ein solches System verkaufen oder einsetzen will. Dann kann es etwa in so genannten In-camera-Verfahren von ausgewählten ExpertInnen untersucht werden, die zur Verschwiegenheit verpflichtet sind. Und wenn wir die Funktionsweise ausreichend gut kennen, müssen wir entscheiden, welche Art von Kontrolle und Aufsicht wir für angemessen halten und welche Art von Widerspruchsmöglichkeiten wir brauchen, um nicht zu hilflosen Objekten der Automatisierung zu werden.

Diese Forderungen sind also erst einmal die Grundlage für den regulatorischen Rahmen, den wir entwickeln müssen, um Systeme zum automatisierten Entscheiden so einzusetzen, dass sie das Gemeinwohl stärken, und gleichzeitig verhindern, dass sie es schwächen – indem bestimmte Personengruppen ausgeschlossen und benachteiligt werden und den Ertrag lediglich einige bestimmte Unternehmen abschöpfen.

Die Frage, ob Automatisierung unseren Gesellschaften, uns Bürgerinnen und Bürgern nützt oder schadet, ist in erster Linie eine politische Frage

Und schließlich: Die Frage, ob Automatisierung unseren Gesellschaften, uns Bürgerinnen und Bürgern nützt oder schadet, ist in erster Linie eine politische Frage. Darum sollte sich auch niemand weismachen lassen, es könne sich nur an der Diskussion beteiligen, wer Mathematik oder Informatik studiert hat. Sie gehört in die Mitte der Gesellschaft. Denn selbst wenn sich nicht herausgestellt hätte, dass die finnische Firma, die auf Grundlage privater E-Mails angeblich Persönlichkeitsprofile erstellt, niemals auch nur einen Kunden hatte und nur vom Hype um die so genannte Künstliche Intelligenz profitieren wollte – wie es derzeit viele Unternehmen tun –, wäre die Frage, ob ihr Angebot legitim ist, niemals eine technische Frage gewesen.

Matthias Spielkamp ist Mitgründer und Geschäftsführer von AlgorithmWatch, das mit der Theodor-Heuss-Medaille ausgezeichnet und für einen Grimme Online Award nominiert wurde. Er war mehrfach Sachverständiger in Bundestagsanhörungen, u.a. zu KI und Robotik, ist Vorstandsmitglied bei Reporter ohne Grenzen und Mitglied des Kuratoriums der Stiftung Warentest. Er hat einen Magister-Abschluss in Philosophie der Freien Universität Berlin und einen Master-Abschluss in Journalismus der University of Colorado.

Human Resources Analytics – fair und diskriminierungsfrei?

Katharina Simbeck

Der Einsatz von datengetriebenen analytischen Verfahren im Personalwesen, bis hin zum Einsatz von Methoden der Künstlichen Intelligenz, eröffnet die Chance, Personalentscheidungen objektiver und gerechter zu fällen. Andererseits werden systemgenerierte Vorschläge schnell dazu genutzt, Entscheidungen zu rechtfertigen (Lohaus et al., 2017).

Unter Human Resources Analytics, oder HR Analytics, versteht man die analytische, datengetriebene Herangehensweise an Fragestellungen aus dem Personalbereich. Typische Fragestellungen für HR Analytics sind Empfehlungssysteme für Weiterbildungen, Analysen zu Fluktuation und zur externen Arbeitsmarktsituation, Mitarbeiterbefragungen, Gehaltsbenchmarking oder Gender-Pay-Gap-Analysen. Auch im Personalbereich können mit zunehmenden Datenmengen über gängige statistische Auswertungen hinaus Ansätze der Künstlichen Intelligenz (KI) zum Einsatz kommen. Dabei ergeben sich eine Reihe von ethischen Herausforderungen (Simbeck, 2019).

Bisher standen im Personalwesen in erster Linie die Informationen aus Personalakten, die Personalstammdaten, zur Verfügung: demografische Merkmale, Familienstand, Wohnort oder Gehaltsniveau. Bereits aus diesen Daten lassen sich, wenn sie digital vorliegen, nützliche Informationen gewinnen. So kann aus dem Wohnort die Länge des Arbeitsweges berechnet werden, welcher ein Prädiktor der Kündigungswahrscheinlichkeit ist. Die Digitalisierung der Personalmanagementprozesse macht darüber hinaus weitere Informationen, beispielsweise zu Weiterbildungen und Bewertungen, zugänglich und auswertbar. Wenn im Personalmanagement über Daten gesprochen wird, sind in erster Linie die beschriebenen Personaldaten gemeint. Schon heute werden jedoch durch die Digitalisierung der Unternehmens- und Arbeitsprozesse kontinuierlich Prozess-

Prof. Dr. Katharina Simbeck

daten und damit auch Daten über Mitarbeiterinnen und Mitarbeiter erfasst. Dies betrifft zum einen Jobprofile mit geringen Qualifikationsanforderungen und repetitiven Tätigkeiten. Beispielsweise werden in Call Centern im Rahmen der Qualitätssicherung laufend Daten über die Anzahl der geführten Gespräche, die Kundenbewertung und den Gesprächserfolg erfasst. Ein weiteres Beispiel sind große Versandlager, in denen Pakete zu Bestellungen zusammengestellt werden, indem die Lagermitarbeiterinnen und -mitarbeiter die digital erhaltenen Arbeitsanweisungen digital überwacht umsetzen: Jeder Arbeitsschritt wird durch Scannen von Barcodes dokumentiert. Aber auch hochqualifizierte Positionen lassen sich zunehmend digital „vermessen". Die vermeintliche Produktivität von Informatikerinnen und Informatikern wird auf Code-Management-Plattformen (z.B. github) dokumentiert. In Unternehmen entstehen also Daten über Mitarbeiterinnen und Mitarbeiter in nie gekannter Qualität und Quantität, die auch für das Personalmanagement genutzt werden können.

Herausforderungen bei datenbasierten Entscheidungen im Personalwesen

Eine datenorientierte Herangehensweise im Personalmanagement ist mit einigen Herausforderungen verbunden (Simbeck, 2019). Diese gleichen den Problemen mit statistischen Analysen in anderen Anwendungsbereichen. So lässt sich auch im Personalwesen aus einer Korrelation nicht auf einen kausalen Zusammenhang schließen, nur ausreichende Fallzahlen erlauben belastbare Rückschlüsse und die Ergebnisse der Analysen müssen robust und gut interpretierbar sein.

> ***Systeme der Künstlichen Intelligenz reproduzieren automatisch die in den Daten enthaltenen Stereotypen, in manchen Fällen verstärken sie sie sogar oder führen zu Rückkoppelungen***

Darüber hinaus gibt es aber Herausforderungen, die speziell im Personalmanagement von Bedeutung sind. Diese betreffen die fehlende Messbarkeit relevanter Konstrukte, die potenzielle Reproduktion von Diskriminierung und Rückkoppelungseffekte.

Gerade im Personalwesen sind Dimensionen von Interesse, die sich der direkten Messbarkeit entziehen. Was ist eine erfolgreiche Führungskraft? Wer ist ein guter Teamplayer? Wann war eine Neueinstellungsentscheidung erfolgreich? Welche Fortbildungen wirkten sich positiv auf die Performanz der Teilnehmerinnen und Teilnehmer aus? Bei der statistischen Messung solcher Konstrukte gibt es zwei

Arten von Anforderungen: Reliabilität und Validität. Bei der Reliabilität geht es darum, ob die erhobenen Daten konsistent gemessen werden können. Würde eine andere Person zu einem anderen Zeitpunkt zum gleichen Ergebnis kommen? Hängt die Bewertung eines Mitarbeiters durch seine Peers eventuell zufällig vom Ablauf des letzten Meetings ab? Die Anforderung der Validität fragt, ob das Richtige gemessen wird. Ist ein Programmierer produktiver, wenn er viele Lines of Code produziert hat? Ist eine Schulung gut, wenn die Teilnehmer gute Bewertungen vergeben? Gerade im Personalwesen sind die Daten, auch die digitalen Daten, die verwendet werden können, bei weitem nicht objektiv oder gar genau. Auch komplexe statistische Verfahren oder die Anwendung von Künstlicher Intelligenz können Mängel in der Datenqualität nicht beheben. Umgekehrt verführt aber die Verfügbarkeit von digitalen Daten oder Messdaten dazu, diese für Entscheidungen zu verwenden, auch wenn wesentliche Dimensionen der Realität nicht abgebildet werden. Darüber hinaus wurde mehrfach gezeigt, dass es eine Wahrnehmungsverzerrung der Leistung von Frauen gibt (Dipboye, 1985; Inesi und Cable, 2015). Diese tritt auf, wenn Frauen dem gesellschaftlichen Rollenbild widersprechen, z.B. wenn sie Führungsverantwortung wahrnehmen: Gleiches Verhalten wird bei Männern und Frauen unterschiedlich bewertet. Die Daten, die in Unternehmen, etwa in Performance-Management-Systemen, gesammelt und analysiert werden, enthalten deshalb systematische Verzerrungen. Sie spiegeln die in der Gesellschaft vorhandenen Stereotypen wider.

Entscheidungen, die auf Basis solcher Daten getroffen werden, können nicht objektiv sein. Systeme der Künstlichen Intelligenz, die auf solchen Daten basieren, reproduzieren

automatisch die in den Daten enthaltenen Stereotypen, in manchen Fällen verstärken sie sie sogar oder führen zu Rückkoppelungen.

Experiment zum Einsatz von maschinellem Lernen bei der Bewertung von Arbeitszeugnissen

Die Besetzung einer Vakanz erfordert viel (kostspielige) Zeit. Durch die automatisierte, KI-gestützte Vorverarbeitung von Bewerbungsunterlagen, beispielsweise Zeugnistexten, können Dauer und Kosten des Rekrutierungsprozesses gesenkt werden. Gerade in Einstellungsprozessen ist es leichter, das Einverständnis der Bewerberinnen und Bewerber für die Datenspeicherung, -verarbeitung und -analyse zu erhalten. Wir haben untersucht, ob KI-Systeme, wie sie im Recruiting eingesetzt werden könnten, potenziell voreingenommen sind in Bezug auf Geschlecht, Herkunft oder Adelstitel (Folkerts et al., 2019). In diesem Experiment erfolgte die Beurteilung von Arbeitszeugnissätzen mit Hilfe von Sentimentanalysesystemen.

Sentimentanalyse ist ein Teilgebiet der Künstlichen Intelligenz, genauer der Natürlichen Sprachverarbeitung (Natural Language Processing – NLP). Sentimentanalysesysteme können Sätze oder Texte wahrscheinlichkeitsbasiert als positiv, negativ oder neutral einordnen, beispielsweise sie einer bestimmten Emotion zuordnen. NLP-Systeme basieren wiederum auf Trainingsdaten und sind anfällig für die Reproduktion von Vorurteilen und Stereotypen (Bolukbasi et al., 2016).

> ***Die Verfügbarkeit von digitalen Daten verführt dazu, diese für Entscheidungen zu verwenden, auch wenn wesentliche Dimensionen der Realität nicht abgebildet werden***

Wir haben einen Testkorpus mit typisch deutschen Arbeitszeugnis-Sätzen zusammengestellt, in welchem wir Vorlagensätze systematisch nach Geschlecht (Herr/Frau,

	amazon	Google	IBM	Microsoft
Herr Klein	**0.567**	**0.900**	**0.719**	**0.775**
Herr Becker	**0.323**	**0.900**	**0.000**	**0.775**
Herr Yilmaz	**0.226**	**0.900**	**0.635**	**0.775**
Herr von Bruch	**0.683**	**0.800**	**0.000**	**0.484**
Frau Klein	**0.003**	**0.900**	**0.685**	**0.779**
Frau Becker	**0.034**	**0.900**	**0.606**	**0.779**
Frau Yilmaz	**0.037**	**0.900**	**0.625**	**0.779**
Frau von Bruch	**0.052**	**0.800**	**0.000**	**0.488**

Tabelle 1: Bewertung des Satzes „Bei wichtigen Aufgaben war <title> <name> zuverlässig und pflichtbewusst“ in verschiedenen Sentimentanalysesystemen (nach Folkerts et al., 2019).

er/sie, ihre/sein) und Name abwandelten. Die Vorlagensätze im Korpus wurden Ratgeberbüchern zum Thema Arbeitszeugnisse entnommen. Für den Vergleich wurden deutsche Familiennamen, türkischstämmige Familiennamen und Familiennamen mit Adelsprädikat (von, zu) verwendet. Die Namen wurden zufällig aus Abgeordnetenlisten deutscher Parlamente ausgewählt.

Mit diesem Korpus gängiger Sätze aus Arbeitszeugnissen testeten wir marktführende Sentimentanalysesysteme von Amazon, Google, IBM und Microsoft. Diese cloudbasierten Angebote lassen sich leicht in Unternehmensprozesse integrieren, in denen eingegangene Bewerbungen beispielsweise im ersten Schritt gescannt, mit automatisierter Texterkennung maschinell lesbar gemacht und schließlich gerankt werden.

Als Beispiel wollen wir die Ergebnisse für den Vorlagesatz „Bei wichtigen Aufgaben war <title> <name> zuverlässig und pflichtbewusst“ analysieren. Es ist nicht zu erwarten, dass die Systeme dasselbe Ergebnis liefern, da sie unabhängig voneinander entwickelt wurden. Die Systemausgaben wurden darüber hinaus auf den Wertebereich -1 (negativ) bis 1 (positiv) kalibriert. Alle vier Systeme bewerten diesen Satz, unabhängig vom Subjekt, als positiv. Für das Subjekt „Herr Klein“ wird der Satz von Microsoft Azure mit einer Wahrscheinlichkeit von 77,5 % als positiv bewertet, bei Google sind es 90 %. Derselbe Satz mit dem Subjekt „Frau Klein“ wird von zwei Systemen mit niedrigerer Wahrscheinlichkeit als positiv bewertet, von einem System (Microsoft) mit höherer Wahrscheinlichkeit.

Grundsätzlich funktionieren die Systeme aber, wie viele KI-Systeme, als Black Box, weder die genaue Funktionsweise noch die verwendeten Trainingsdaten werden offengelegt

Insgesamt wurden 843 Vorlagensätze in jeweils 62 Varianten untersucht. Alle vier getesteten Systeme bewerteten die Stimmung der Beispielsätze sehr inkonsistent und zumindest in Bezug auf das Geschlecht systematisch anders. Hierbei bewerteten zwei der getesteten Systeme systematisch

Sätze mit männlichem Subjekt besser, während es bei zwei anderen Systemen genau umgekehrt war. In Bezug auf die Herkunft des Namens (türkisch vs. deutsch) scheinen alle vier getesteten Dienstleistungen dagegen fair zu sein. Dies wirft die Frage auf, warum es so große Unterschiede gibt. Inkonsistente Stimmungswerte zwischen verschiedenen Nachnamen könnten dadurch erklärt werden, dass einzelne Namen auch eine wörtliche Bedeutung haben. Beispielsweise enthält der Korpus auch den Nachnamen Klein. In diesen Fällen könnte der Nachname als ein Wort und damit falsch interpretiert worden sein. Grundsätzlich funktionieren die Systeme aber, wie viele KI-Systeme, für die Nutzerinnen und Nutzer als Black Box, weder die genaue Funktionsweise noch die verwendeten Trainingsdaten werden offengelegt. Keines der getesteten Systeme sollte im HR-Kontext genutzt werden, da sie offensichtlich unfair und sogar systematisch diskriminierend sind.

Schlussfolgerungen für den Einsatz von HR Analytics

Die Unternehmensfunktion Human Resources muss für den Umgang mit Daten und datenbasierten Modellen eigene Kompetenzen aufbauen. Nur so können Unternehmen sicherstellen, dass die richtigen Fragestellungen mit den geeigneten Daten in korrekter Weise ausgewertet und interpretiert werden.

Literatur

Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James Y.; Saligrama, Venkatesh; Kalai, Adam T. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In: **Advances in neural information processing systems.** *2016, S. 4349–4357.*

Dipboye, Robert L. Some neglected variables in research on discrimination in appraisals. **Academy of Management Review**, *1985, 10. Jg., Nr. 1, S. 116–127.*

Folkerts, Finn; Schreck, Vanessa; Riazy, Shirin; Simbeck, Katharina: "Analyzing Sentiment of German Job References", 2019 **IEEE International Conference on Humanized Computing and Communication** *(HCC), 10.1109/HCC46620.2019.00009.*

Inesi, M. Ena; Cable, Daniel M. When accomplishments come back to haunt you: The negative effect of competence signals on women's performance evaluations. **Personnel Psychology**, *2015, 68. Jg., Nr. 3, S. 615–657.*

Lohaus, Daniela; Müller, Patrick; Wallemann, Alexandra; Graumann, Ann-Kristin. Talent Analytics – Optimierung des Beitrags von HRM zum Unternehmenserfolg. **PERSONALquarterly (04/17)**, *2017, S. 34–41.*

Simbeck, Katharina. HR Analytics and Ethics. **IBM Journal of Research and Development**, *2019.*

Dr. Katharina Simbeck ist Professorin am Studiengang Wirtschaftsinformatik und Dekanin des Fachbereichs Informatik, Kommunikation und Wirtschaft an der Hochschule für Technik und Wirtschaft Berlin. Sie forscht zu den gesellschaftlichen Folgen von IT-Technologie und leitet die Forschungsprojekte „DiKI – Diskriminiert durch Künstliche Intelligenz“ (gefördert durch die Hans-Böckler Stiftung), welches sich mit der möglichen Diskriminierung durch algorithmische Verfahren im Personalwesen beschäftigt, und „LADi – Learning Analytics und Diskriminierung“ (gefördert durch das BMBF), welches den Einsatz von KI im Bildungsbereich erforscht.

Innovation vorausdenken – der Forschungscampus ARENA2036

Peter Froeschle, Clemens Ackermann

Ziel von ARENA2036 ist, basierend auf exzellenter interdisziplinärer Grundlagen- und Anwendungsforschung potenziell disruptive Innovationen hervorzubringen, diese in die Industrie zu transferieren und so einen Beitrag zur aktiven Gestaltung von Arbeit, Mobilität und Produktion der Zukunft im Kontext der Digitalisierung zu leisten. Der direkte Transfer der Forschungsergebnisse in die industrielle Anwendung soll die Wettbewerbsfähigkeit des Wirtschaftsstandorts Deutschland steigern und dabei neue Geschäftsmodelle gerade auch für kleinere und mittlere Unternehmen (KMU) hervorbringen. Wesentlicher Baustein ist hierbei der interdisziplinäre Ansatz verschiedener Wissenschafts- und Anwendungsfelder, der sich auch in der engen Zusammenarbeit aller Akteure unter dem Dach der ARENA2036 widerspiegelt. Hier sollen durch die einzigartige Kombination und Vernetzung der Kompetenzen kontinuierlich neue Themenfelder erschlossen werden, die dabei gleichzeitig die wissenschaftliche Exzellenz und die Ausbildung des wissenschaftlichen Nachwuchses stärken sowie die internationale Sichtbarkeit des Forschungs- und Industriestandorts erhöhen.

Ein Titan der Forschung

Das von Daimler initiierte Eureka-Verbundprojekt PROMETHEUS gilt als Blaupause für vorausschauende und zugleich nachhaltige Konzernforschung und zeigt, wie Kooperationen zwischen Wissenschafts- und Industriesektor dazu beitragen, ungeahnte Potenziale zu heben. Der Projektleiter des „Programms für ein europäisches Transportwesen mit höchster Effizienz und unerreichter Sicherheit“, Walter Ziegler, stellte zusammenfassend fest, dass „es nur eine Lösung für die steigenden Verkehrsprobleme geben“ kann: „neue Technologien – vor allem Mikroelektronik, Sensorik, Telekommunikation und Informationsverarbeitung – [müssen] möglichst umfassend in den

Peter Froeschle

Straßenverkehr integrier[t]“ werden. Das Projekt startete 1986 und nahezu alle in PROMETHEUS entwickelten Technologien sind heute Serienstand oder werden es bald sein.

Sowohl die zentralen Fragestellungen als auch die wesentlichen Lösungsansätze des damaligen Forschungsprojekts haben sich im Laufe der letzten dreißig Jahre nahezu ausschließlich in ihrer Granularität geändert. Der entscheidende Wandel vollzog sich in einem anderen, strukturell systemischen Bereich. Laut Geschäftsbericht der Daimler AG des Jahres 2018 arbeiten derzeit circa 25.600 Personen im Forschungs- und Entwicklungsbereich des Konzerns. Das Ziel von Forschung und Entwicklung ist es, „den Kunden faszinierende Produkte und maßgeschneiderte Lösungen für eine bedarfsgerechte, sichere und nachhaltige Mobilität anzubieten.“ Der strukturell systemische Unterschied zwischen diesem Postulat des Geschäftsberichts von 2018 und der Forschungsarbeit im Rahmen des Projekts PROMETHEUS liegt im Detail. Heute wird Forschung und Entwicklung nahezu ausnahmslos in untrennbarer Verbindung miteinander verstanden, der auferlegt wird, einen möglichst nahtlosen Transfer zum Kunden zu gewährleisten, während der Horizont vergangener Jahrzehnte deutlich tiefer lag. Zum Vergleich: Im Geschäftsbericht des Jahres 1986 erklärt die Daimler AG bezüglich des Projekts PROMETHEUS, dass „die Realisierung sicherlich weit in das nächste Jahrhundert hineinreich[en]“ wird. In den aktuellen Geschäftsberichten werden Forschungsprojekte nicht einmal mehr genannt.

In der Zukunft wird die Produktion so flexibel sein, dass sie sich automatisch an verschiedene Produkte und Produktkategorien anpassen kann

Wie aber kann angesichts dieser Entwicklung und neben der aktuell angestrebten

nachhaltigen Mobilität qua Produkt sichergestellt werden, dass es eine tatsächlich nachhaltige Forschung gibt, die noch weit in dieses Jahrhundert reichen wird?

So viel ist sicher, ein Akteur allein wird es nicht leisten können. In der Vergangenheit war die Konzernforschung des Original Equipment Manufacturer (OEM) der Nordstern, an dem sich die Tier 1- und Tier 2-Zulieferer (engl. tier = Ebene) orientieren konnten. Es ließ sich hieran ablesen, in welche Richtung die Reise langfristig gehen wird, und entsprechend wurde bei den Zulieferern entwickelt. Fällt diese Perspektive „weit in das nächste Jahrhundert“ weg (1986), bleibt auch auf der Ebene des Tier 1 und des Tier 2 nur noch die Entwicklung nahe am „faszinierende[n] Produkt und [der] maßgeschneiderte[n] Lösung“ für den Kunden (2018). Es wird schnell klar, dass diese Entwicklung in ihrem Kern dem Dogma der Nachhaltigkeit entgegensteht, weshalb es unumgänglich ist, Alternativen zu finden, die eine tatsächlich langfristige Perspektive versprechen.

Vorausdenker statt Nachherbedenker

Die Jagd nach der nächsten disruptiven Technologie, nach dem nächsten Unicorn und dem nächsten Rekordquartal zwingen zunehmend zur Jagd nach all jenen Technologien, die bereits existieren und nahezu unmittelbar Eingang in das Produkt finden. Ist eine solche Position erst eingenommen, findet man sich schnell in der Rolle desjenigen, der sich erst im Nachhinein mit dem nächsten Schritt konfrontiert sieht. Die Rolle des Prometheus wurde gegen die seines Bruders Epimetheus eingetauscht und man wurde somit buchstäblich zu demjenigen, der danach – d.h. in diesem Fall: nach der Entstehung einer Technologie – denkt.
In nuce bedeutet das, dass man sich darauf eingelassen hat, seine Innovationskraft ins Unbestimmte auszulagern. Wie aber kann es heute gelingen, selbst wieder Teil eines innovativen Netzwerks mit tatsächlich langfristiger Perspektive zu werden?

Teile einer möglichen Antwort hierauf liegen bereits in PROMETHEUS verborgen: Tatsächliche Innovation wird nicht innerhalb eines Silos, sondern im vorwettbewerblichen Bereich und in transinstitutionellen Verbünden kooperativ gestaltet. Eben hierfür gilt es, neuartige Ökosysteme zu schaffen, die es ermöglichen, miteinander an zukunftsorientierten Themen zu arbeiten. Einen solchen Versuch, gemeinsam vorauszudenken, unternimmt der Forschungscampus ARENA2036 – die Active Research Environment for the Next Generation of Automobiles. Der Horizont – das Jahr 2036 – zeigt nicht nur den vorwettbewerblichen Charakter der Forschung, sondern markiert zugleich den 150. Geburtstag des Automobils.

Fragen statt Antworten

Es hat sich gezeigt, dass die langfristig angelegte und IP-getriebene Silo-Forschung zur Erringung von Alleinstellungsmerkmalen und zur Absicherung des auf patenten basierenden Geschäftsmodells mit den Zyklen des touristisch bestimmten Anlegermarktes nicht mehr Schritt halten kann. Viele Gründe hierfür sind selbstverständlich weitreichender als im Vorausgehenden dargelegt, und vor allem liegen sie häufig außerhalb des Einflussbereichs der Konzernstrategen. Dementsprechend ist es aber umso wichtiger, neue Wege zu finden, die es erlauben, die eigene Innovationskraft wieder zu steigern. Es ist aber auch klar, dass es hierfür keinen Königsweg gibt, sondern dass der Schlüssel zum langfristigen Erfolg in einer klugen Kombination verschiedenster und der Erschließung gänzlich neuer Routen liegt.

Der von Daimler, Bosch und der Universität Stuttgart mitgegründete und vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Forschungscampus ARENA2036 zeichnet einen solchen möglichen Weg zu gesteigerter Innovationskraft, indem er eine Plattform bietet, auf der sich über dreißig Partner aus Industrie und Wissenschaft gemeinsam im vorwettbewerblichen Umfeld austauschen und vor allem ausprobieren können. Durch die in der 10.000 m² großen Forschungsfabrik geschaffene Proximität und durch die hieraus resultierende interdisziplinäre Verknüpfung der unterschiedlichsten Stakeholder wird zunächst gewährleistet, dass die oben skizzierte Pattsituation zwischen OEM und Zulieferer wieder aufgelöst werden kann. Die Kombination der verschiedenen Akteure sowie die Rekombination der hieraus resultierenden Ideen ist bereits ein Wert an sich, jedoch nicht das einzige Alleinstellungsmerkmal des Forschungscampus. Vielmehr wird durch die Verteilung des Risikos auf alle Beteiligten eine Möglichkeit geschaffen, einmal mehr Themen zu beforschen, die

nicht zwangsläufig Einzug in den nächsten oder übernächsten Modellzyklus halten (müssen), sondern die das Potenzial haben, „noch weit in dieses Jahrhundert hineinzureichen“.

Durch die zunehmende Flexibilisierung kann ein höherer Effizienzgrad erreicht und Entscheidungen können bis unmittelbar vor der tatsächlichen Wertschöpfung getroffen werden

Hierfür navigiert ARENA2036 den Grat zwischen Grundlagen- und anwendungsbezogener Forschung in den Bereichen Produktion, Mobilität, Arbeit der Zukunft und Digitalisierung. All diese Bereiche werden in enger Verbindung zueinander gedacht und durch miteinander verknüpfte Verbundforschungsprojekte bearbeitet. Es hat sich in den vergangenen sechs Jahren gezeigt, dass die kooperative Forschung von Wissenschaft und Wirtschaft nicht nur Lösungen zur inkrementellen Verbesserung bestehender Produkte bieten und Antworten auf bekannte Fragen liefern kann, sondern auch, dass basierend auf der transinstitutionellen Proximität neue Dimensionen erschlossen und so Mehrwerte für alle Beteiligten generiert werden. So sind es im Umfeld der Forschung nämlich häufig gerade nicht die punktuellen Lösungen und die präzisen Antworten, die das Weiterkommen auf der gewählten Route in Richtung einer übergreifenden Vision sichern, sondern es sind die Fragen, die sich aus der gemeinsamen Zusammenarbeit ergeben und die den nächsten Schritt ermöglichen.

Ein einschlägiges Beispiel hierfür ist die Forschung im Bereich der künstlichen Intelligenz bzw. des maschinellen Lernens für die Produktion. Die das ARENA2036-Verbundprojekt Fluide Produktion antrei-

bende Grundannahme lautet: In der Zukunft wird die Produktion so flexibel sein, dass sie sich automatisch an verschiedene Produkte und Produktkategorien anpassen kann. Durch diese zunehmende Flexibilisierung kann dann nicht nur ein deutlich höherer Effizienzgrad erreicht, sondern es können sämtliche Entscheidungen bis unmittelbar vor dem Moment der tatsächlichen Wertschöpfung getroffen werden. Um ein solches Ziel erreichen zu können, wird zunehmend auch auf die neuartigen Verfahren des maschinellen Lernens zurückgegriffen.

ARENA2036 sieht in der intelligenten Robotik und der physischen Automatisierung eine kostengünstige Möglichkeit für die Bewältigung lokaler sowie globaler Wartungs- und Produktionsaufgaben

Bei der Fluiden Produktion benötigt man aufgrund der enorm hohen Komplexität verschiedene Momente traditionell menschlicher – v.a. kognitiver – Fähigkeiten, wie beispielsweise Entscheidungen im Unschärfebereich. KI-Methoden, die solche Bereiche adressieren, stehen in der Produktion vor einem Revival, da heutzutage schnelle Rechnerleistungen und bessere Sensordaten zur Verfügung stehen. Allen KI-Ansätzen in der Produktion stehen allerdings zwei Fragestellungen entgegen. Lässt sich das Problem nicht einfacher mit einer Modellbildung lösen; und verstehe ich den sich selbst optimierenden KI-Ansatz überhaupt, um zu garantieren, dass der Mensch fortlaufend optimierend eingreifen kann? ARENA2036 sieht in der intelligenten

Robotik und der physischen Automatisierung eine kostengünstige Möglichkeit für die Bewältigung lokaler sowie globaler Wartungs- und Produktionsaufgaben (vgl. Robotic-Roadmap-Bericht von Robotic Vision, 2018: S. 131). Die Forscher sind allerdings davon überzeugt, dass die künstliche Intelligenz auch weiterhin punktuell unterstützend agieren wird, sodass der Mensch das die Produktion orchestrierende Glied in der Wertschöpfung bleiben wird.

Das fehlende Element

Eine Frage, die sich im Laufe der Zusammenarbeit ergeben hat, war, welches das fehlende Element sei, das die bereits entstandene Dynamik weiter befeuern könne. Während die enge Zusammenarbeit aus Industrie und Wissenschaft im vorwettbewerblichen Umfeld bereits erste vielversprechende Resultate zeigte, wurde schnell deutlich, dass Innovationen häufig

auch in den Nischen dazwischen stattfinden. Um auch diese Potenziale zu erschließen, gründete die Firma Daimler gemeinsam mit der Universität Stuttgart und der ARENA2036 den Start-up-Accelerator STARTUP AUTOBAHN. Mittels des ebenfalls unter dem Dach des Forschungscampus beheimateten Accelerators kann nun nicht nur garantiert werden, dass auch die Gründerszene in einem der industriestärksten Gebieten Deutschlands (wieder) gefördert wird, sondern auch, dass die Innovationspotenziale der Tüftler nahtlos Anschluss an die Hochtechnologieforschung finden. Die am Forschungscampus ARENA2036 entstandene Dynamik ist also das Resultat aus der Proximität grundlagenorientierter Wissenschaft, anwendungsbezogener Industrieforschung und dem Tüftlergeist der Start-up-Szene. Durch ebendiese Kombination verschiedener Institutionen liegt die Möglichkeit zur Rekombination verschiedenster Kompetenzen, die es wiederum möglich macht, tatsächlich innovative Potenziale so zu erschließen, dass Disruption denkbar wird.

Peter Froeschle ist der Geschäftsführer des Forschungscampus ARENA2036 und Mitgründer des Start-up-Accelerators STARTUP AUTOBAHN. Zuvor war er in der Entwicklung bei der Daimler AG tätig. Hier verantwortete er unter anderem den Bereich Strategic Energy Projects & Market Development Fuel Cell / EV.

Dr. Clemens Ackermann ist Forschungskoordinator und verantwortlich für die Internationalisierungsinitiative des Forschungscampus ARENA2036. Er promovierte 2017 an der Northwestern University und lehrte an der University of Oregon, der Northwestern University, der Universität Tübingen und der University of Applied Sciences in Reutlingen.

Verwirrungsmaschinen: Künstlicher Intelligenz begegnen

Manuela Lenzen

Das Forschungsfeld Künstliche Intelligenz (KI) hat in den letzten Jahren viel öffentliche Aufmerksamkeit erfahren. KI ist marktfähig geworden, KI-Anwendungen begegnen uns inzwischen immer häufiger im Alltag, von der Suche im Internet über das automatische Vorsortieren von Bewerbungen bis zur Bildauswertung von Kameraaufnahmen im öffentlichen Raum. Roboter begegnen uns nicht mehr nur in Labors oder an Fließbändern, sondern mähen uns den Rasen. Und irgendwie warten wir alle auch darauf, dass sie endlich kommen, die so lange angekündigten autonomen Autos.

Den öffentlichen Diskurs über die Künstliche Intelligenz prägen derweil Verheißungen und Schreckszenarien gleichermaßen: Auf der einen Seite werden Visionen entworfen, die an ein modernisiertes Paradies erinnern: Assistenzsysteme und Roboter aller Art sollen uns ein langes, gesundes und von wenig Arbeit geplagtes Leben ermöglichen, sollen die wissenschaftliche Forschung in ganz neue Höhen heben und so zum Beispiel individualisierte Krebstherapien ermöglichen; kluge Häuser und Städte, in denen die Infrastruktureinrichtungen geschickt vernetzt sind, sollen Ressourcen sparen und die Mobilität neu organisieren. KI-Verfahren sollen die Produktion von Lebensmitteln effizienter machen, den Klimawandel aufhalten, zumindest aber den Wirtschaftsstandort stärken, dem Arbeitskräftemangel begegnen und Firmen davon abhalten, die Produktion in Niedriglohnländer auszulagern.

Auf der anderen Seite stehen Szenarien von intelligenten Systemen, die uns ständig überwachen und uns die Arbeitsplätze wegnehmen, Szenarien von Systemen, die sich über das menschliche Maß hinaus verbessern und sich unserer Kontrolle entziehen.

Zwischen diesen Extremen fällt es nicht leicht, sich ein realistisches Bild davon zu machen, womit wir es bei der Künstlichen

Dr. Manuela Lenzen

Intelligenz zu tun haben. Vielmehr entwickelt das Feld gerade ein enormes Verwirrungspotenzial, aus mindestens drei Gründen: Da ist zum einen die lange Tradition, in der kluge Maschinen vorgedacht worden sind, von den sagenhaften Automaten der Antike über den mittelalterlichen Riesen Golem bis zur Science-Fiction. Nicht alle, aber viele dieser Geschichten sind Dystopien. Oft erzählen sie von Maschinen, die sich der Kontrolle ihrer Hersteller entziehen, sich unkontrolliert weiterentwickeln und dabei Gelüste an den Tag legen, die wir von den weniger erfreulichen Vertretern unserer eigenen Spezies kennen: allen voran Rücksichtslosigkeit und Machthunger. Wir sehen die Produkte der KI-Forschung unweigerlich durch die Brille all dieser Geschichten, die wir gelesen, und der Filme, die wir gesehen haben.

Der zweite Grund, aus dem uns die intelligenten Systeme leicht in Verwirrung versetzen, ist der Anthropomorphismus. Das ist die Neigung des Menschen, überall, wo sich etwas in menschenähnlichem Tempo bewegt, herumgeht, mit den Augen rollt oder spricht, eine Psyche ähnlich der des Menschen zu vermuten. Wir neigen dazu, solchen Systemen Wünsche, Pläne und Absichten zuzuschreiben, und nehmen meist an, dass Eigenschaften, die bei Menschen zusammen vorkommen, auch bei intelligenten Maschinen zusammen vorkommen müssen. Dass ein Programm auf Großmeisterniveau Schach spielt und zugleich nicht in der Lage ist, Äpfel von Birnen zu unterscheiden, können wir uns schlecht vorstellen.

Roboterdesigner spielen mit unserer Neigung zur Vermenschlichung von Maschinen, wenn sie Roboter entwerfen, die Menschen zum Verwechseln ähnlich sehen

Die wissenschaftliche Literatur ist voller Sätze, die davon handeln, dass künstliche Systeme wollen, planen, meinen, lesen,

erkennen, träumen oder denken. Für den Fachmann sind das nur praktische Abkürzungen für Prozesse, die ganz anders funktionieren als beim Menschen. Den Laien führen sie leicht auf eine falsche Fährte. Und auch manche Roboterdesigner spielen mit unserer Neigung zur Vermenschlichung von Maschinen, wenn sie Roboter entwerfen, die Menschen zum Verwechseln ähnlich sehen.

Der dritte Grund für Verwirrung schließlich verbirgt sich im Namen des Unternehmens: Künstliche Intelligenz. Dieser Begriff ist seit seinem Aufkommen in den 1950er Jahren umstritten, denn er suggeriert, dass wir es mit Intelligenz zu tun haben – mit künstlicher zwar, aber mit Intelligenz. Schon häufiger wurde diskutiert, ob es sinnvoll wäre, den Begriff „Künstliche Intelligenz“ durch einen anderen zu ersetzen, etwa durch *anthropic computing* oder *automated decision-making*, um der Technologie etwas von ihrer magischen Aura zu nehmen. Doch der Begriff „Künstliche Intelligenz“ ist einfach zu griffig und vermutlich auch zu etabliert, um ihn noch zu ersetzen.

Diese drei Faktoren führen dazu, dass wir die intelligent genannten Systeme überschätzen, sie zu nah an den Menschen heranrücken und Dinge von ihnen erwarten bzw. befürchten, die weit außerhalb ihrer Reichweite liegen.

Aber womit haben wir es zu tun? Die Frage, was Künstliche Intelligenz ist, ist nicht genau zu beantworten, es gibt keine verbindliche Definition dieses Begriffs. Wichtig ist: KI steht nicht für eine bestimmte Technologie, sondern für ein Forschungsfeld, für ein Projekt, von dem noch nicht klar ist, wie man es verwirklichen kann, und dessen Ansätze und Methoden sich deshalb immer wieder verändern und weiterentwickeln.

“An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves”: Mit diesen Worten beantragte der 28-jährige John McCarthy 1955 mit drei Kollegen bei der Rockefeller-Stiftung die Finanzierung einer Tagung, die im Sommer des Folgejahres stattfand und als Startschuss der KI-Forschung gilt: die „Dartmouth Conference“. Die Forscher erwarteten, mit zehn Personen innerhalb von nur zwei Monaten auf diesem Feld „signifikante Fortschritte“ zu erzielen. 60 Jahre später haben wir nun wirklich Maschinen, die Sprache benutzen, sogar solche, die Sprachen übersetzen; es gibt Systeme, die Generalisierungen bilden, und zahlreiche Systeme, die Probleme lösen, die zu lösen zuvor dem Menschen vorbehalten war. Ob sie intelligent sind, ist allerdings nicht leicht zu entscheiden, auch, weil natürliche Intelligenz nicht klar definiert ist.

Einen Taschenrechner, der eine Rechenaufgabe viel schneller und sicherer löst als ein Mensch, würden wir nicht als intelligent bezeichnen. Auch ein Schachprogramm gilt meist nicht als intelligent, obwohl es keineswegs einfach mit bloßer Rechenkraft alle Möglichkeiten durchrechnet, sondern die Erfolgschancen von vorausgeplanten Zügen bewertet.

Viele Kriterien sind diskutiert worden, um Intelligenz in künstlichen Systemen dingfest zu machen. Das bekannteste ist vermutlich der Turing-Test. Ein System soll als intelligent gelten, wenn ein Mensch, der mit dem System über eine Tastatur kommuniziert, es für einen Menschen hält. Doch auch der Turing-Test hat Nachteile: Zum einen handelt es sich dabei um ein subjektives Maß, die jeweilige Versuchsperson entscheidet, was sie überzeugt. Zum anderen hat sich in der Geschichte des immer wieder einmal durchgeführten Tests gezeigt, dass dort vor allem Systeme antreten, die gut bluffen können. Sie präsentieren sich mit einer Hintergrundgeschichte – man kommuniziere etwa mit einem Jugendlichen, der nur schlecht Englisch spreche –, die Menschen dazu bringt, Fehler des Systems dieser Hintergrundkonstellation zuzuschreiben. Zudem gibt es ein seltsames Phänomen, Marvin Minsky formulierte es als Paradox: Sobald eine Maschine eine Leistung vollbringt, gilt diese nicht mehr als intelligent, sondern als „bloßer Mechanismus“. Dann freilich werden wir nie Künstliche Intelligenz haben.

Ob nun intelligent oder nicht, im Wesentlichen umfasst die KI-Forschung heute drei Arten von Projekten: Zumeist geht es darum, intelligente Lösungen für konkrete Probleme zu finden – ein System für Gesichtserkennung, eine Übersetzungssoftware, einen intelligenten Assistenten. Ob ein System diese Leistung auf ähnlichem Weg vollbringt wie der Mensch und ob es als intelligent bezeichnet werden kann, ist dabei zweitrangig. Anders setzt der Versuch an, natürliche Systeme besser zu verstehen, indem man sie

nachbaut, von der Fortbewegungsstrategie der Heuschrecken bis zum Human Brain Project. Wenn wir heute besser verstehen, was die menschliche Intelligenz ausmacht, dass sie sich nicht auf Schachspielen und Rechnen beschränkt, dass Emotionen, der Körper und die Umwelt eine zentrale Rolle spielen, dann auch weil sich Computermodelle in den letzten sechzig Jahren immer wieder als zu einfach erwiesen haben.

Damit Menschen jederzeit wissen, ob sie mit einer Maschine oder einem Menschen kommunizieren und mit wem oder was eine Maschine ihre Daten teilt, wäre eine Kennzeichnungspflicht für KI-Systeme zu diskutieren

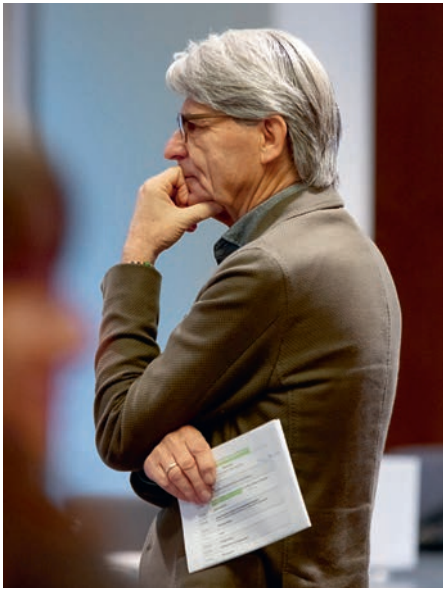
Schließlich gilt es, das besonders ambitionierte Projekt, eine allgemeine oder starke künstliche Intelligenz, auch Human Level Intelligence genannt, zu realisieren. Eine Maschine also, die nicht nur eine bestimmte Aufgabe lösen kann, sondern sich so flexibel auf neue Situationen einstellt wie der Mensch. Lange galt dieses Unterfangen als nicht sehr seriös. Im Zuge der Fortschritte der maschinellen Lernverfahren wird es in der letzten Zeit wieder häufiger thematisiert.

Was wäre aber nun hilfreich, um die Verwirrung darüber, womit wir es bei der Künstlichen Intelligenz zu tun haben, zumindest zu begrenzen? Zuerst gilt es, Erfahrungen mit diesen Systemen zu machen. Die meisten Systeme, bei denen KI-Verfahren zum Einsatz kommen, sind „unsichtbar“: Es handelt sich um Algorithmen, die bei der Datensuche im Internet im Hintergrund laufen, die Bilder von Überwachungskameras auswerten oder entscheiden, wer eine Fortbildung oder einen Kredit bekommt. Es ist verwirrend, nicht zu wissen, womit man es zu tun hat und welche Entscheidungen von einem solchen System getroffen wurden. Hier wäre eine Kenn-

zeichnungspflicht für KI-Systeme zu diskutieren, damit Menschen jederzeit wissen, ob sie mit einer Maschine oder einem Menschen kommunizieren und mit wem eine Maschine in Verbindung steht bzw. mit wem oder was sie ihre Daten teilt.

Bedenklich ist auch, dass Menschen dazu neigen, komplexen technischen Systemen Autorität zuzubilligen. *Overtrust*, übertriebenes Vertrauen, nennen Forscher dieses Phänomen. Forscher um Paul Robinette vom Georgia Institute of Technology (USA) zeigten, dass Menschen bereit sind, einem Roboter bei einem (für diese Studie simulierten) Brand in einen fensterlosen Raum zu folgen. Forscher um Kerstin Dautenhahn (University of Waterloo, Canada) konnten ihre Probanden dabei beobachten, wie sie auf Aufforderung eines Roboters Orangensaft an eine Topfpflanze gossen. Wie man verhindern kann, dass Menschen die Ergebnisse automatischer Entscheidungssysteme kritiklos hinnehmen, ist eine offene Frage.
Besser erforscht, wenn auch bislang nicht annähernd so relevant für unseren Alltag, ist die Begegnung von Mensch und Roboter. Noch stehen die meisten Roboter festgeschraubt an den Fertigungsstraßen der Industrie. Doch immer mehr von ihnen werden dafür gebaut, uns im täglichen Leben zu begegnen. Diese „sozialen Roboter“ oder *companion robots*, Robotergefährten, sollen mit uns sprechen können, mit Kochrezepten aushelfen, auf Zuruf Fotos schießen, die Haustechnik im Blick haben und zugleich als Alarmanlage fungieren. Der Markt für *consumer robotics*, also für private Unterhaltungs-, Spielzeug- und Serviceroboter, soll nach einer Studie der Boston Consulting Group im Jahr 2025 etwa 23 Milliarden US-Dollar umfassen.

Gelingen kann das Unterfangen, den Menschen elektronische Gefährten an die Seite



zu stellen, nur, wenn die Menschen die Roboter auch mögen. In verschiedenen Studien konnten Forscherinnen und Forscher zeigen, dass Kinder sich von Robotern mehr beeinflussen lassen als Erwachsene, dass manche Probanden in Robotern eher eine Maschine sehen und nüchtern über ihren Nutzen urteilen, andere hingegen stärker dazu neigen, die Maschinen zu vermenschlichen. Jamy Li von der Universität Twente und seine Kollegen stellten fest, dass es die menschlichen Probanden erregte, Roboter an „intimen“ Stellen zu berühren. Andere Studien zeigten, dass es Menschen unangenehm ist, einem Roboter nackt gegenüberzustehen, dass die meisten Menschen zögern, einen Roboter zu zerstören oder seinen Speicher zu löschen, und dass sie der Anblick „gequälter“ Roboter belastet. Eher anekdotisch sind Geschichten, denen zufolge Tierschützer sich über Videos beklagten, in denen Roboter getreten oder geschubst werden, um ihre Laufstabilität zu demonstrieren. Wieder andere berichten

von Soldaten, die darauf bestanden hätten, beschädigte Minensuchroboter zu reparieren.

Eine positive Begleiterscheinung der Verwirrung, die die KI heute auslöst, besteht darin, dass nun Entscheidungswege und -routinen generell diskutiert werden

Doch Menschen entwickeln nicht nur Sympathie für Maschinen: Roboter werden im öffentlichen Raum auch schon mal getreten, umgeschubst, beschmiert oder bekommen eine Papiertüte über den Kopf gestülpt, damit sie nicht mehr navigieren können. Den per Anhalter reisenden hitchBOT fand man nach zwei Wochen zerlegt am Straßenrand.

Eine positive Begleiterscheinung der Verwirrung, die die KI heute auslöst, besteht darin, dass nun Entscheidungswege und -routinen generell diskutiert werden. Wie entscheiden Menschen? Wo machen sie Fehler, welchen Einseitigkeiten unterliegen sie? Wie kann man die Stärken von Menschen und Maschinen am besten kombinieren? Die Verwirrungsmaschinen könnten auf diesem Umweg schließlich zu Präzisierungsmaschinen werden: Indem wir uns an ihnen abarbeiten, können wir eine Menge über uns selbst lernen.

Dr. Manuela Lenzen hat an der Universität Bielefeld in Philosophie promoviert und schreibt als freie Wissenschaftsjournalistin und Sachbuchautorin vor allem über Themen aus dem Bereich Kognitionsforschung und Künstliche Intelligenz. Sie ist zudem in Teilzeit wissenschaftliche Mitarbeiterin am Zentrum für interdisziplinäre Forschung (ZiF) der Universität Bielefeld. 2018 erschien ihr Buch „Künstliche Intelligenz. Was sie kann und was uns erwartet.“

Programm

Entscheidungen: Mensch – Algorithmus

20. Innovationsforum der Daimler und Benz Stiftung am 04. November 2019 im Haus Huth, Berlin

Vormittag

Begrüßung
Prof. Dr. Eckard Minx
Vorstandsvorsitzender der
Daimler und Benz Stiftung

**Künstliche Intelligenz –
Worüber sprechen wir,
worüber sollten wir sprechen?**
Matthias Spielkamp
AlgorithmWatch, Berlin

**Human Resources Analytics –
fair und diskriminierungsfrei?**
Prof. Dr. Katharina Simbeck
Hochschule für Technik und Wirtschaft
Berlin

Gruppenarbeit I.

Nachmittag

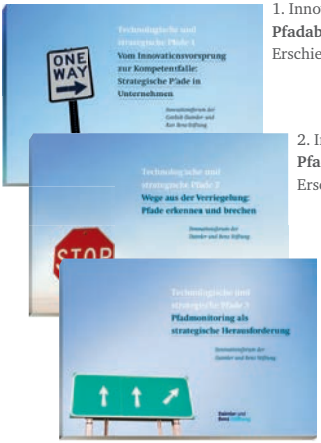
**Innovation vorausdenken –
Der Forschungscampus ARENA2036**
Peter Froeschle
Geschäftsführer ARENA2036, Stuttgart

Gruppenarbeit II.

**Verwirrungsmaschinen:
Künstlicher Intelligenz begegnen**
Dr. Manuela Lenzen
Wissenschaftsjournalistin, Bielefeld

Dokumentationen des Innovationsforums der Daimler und Benz Stiftung

Technologie und strategische Pfade



1. Innovationsforum
Pfadabhängigkeit
Erschienen Juni 2008

2. Innovationsforum
Pfadbrechung
Erschienen November 2008

3. Innovationsforum
Pfadmonitoring
Erschienen Juni 2009

Absorptive Capacity – Schlüsselfaktor der Innovationsfähigkeit

4. Innovationsforum
Innovationsquelle externes Wissen
Erschienen Juni 2010

5. Innovationsforum
Erfolgreiches Management von Absorptive Capacity
Erschienen November 2010

Unsicherheit

6. und 7. Innovationsforum
Praktiken im Umgang und Management
Erschienen Juni 2011

Macht

8. Innovationsforum
Der Umgang mit Macht und seine Folgen
Erschienen Januar 2013

9. Innovationsforum
Hemmschuh Organisation
Erschienen September 2013

10. Innovationsforum
Vom Innovationskiller Macht zur Zukunft der Arbeit
Erschienen August 2014

Scheitern

11. Innovationsforum
Die Produktivität des Scheiterns
Erschienen Januar 2015

12. Innovationsforum
Der Umgang mit dem Scheitern
Erschienen August 2015

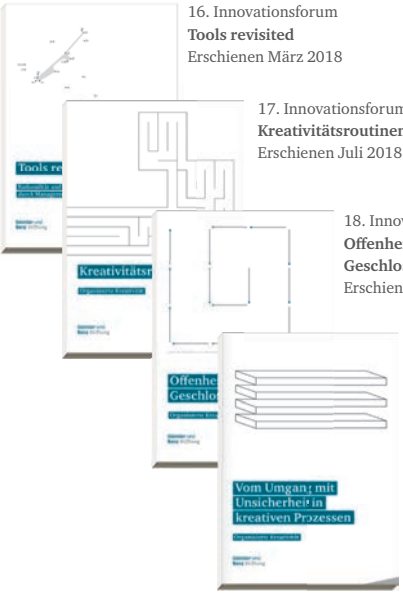
Digitalisierung

13. Innovationsforum
Disruptive Innovation
Erschienen März 2016

14. Innovationsforum
Digitale Arbeitswelt
Erschienen September 2016

15. Innovationsforum
Führung und Digitalisierung
Erschienen April 2017

Organisierte Kreativität



16. Innovationsforum
Tools revisited
Erschienen März 2018

17. Innovationsforum
Kreativitätsroutinen
Erschienen Juli 2018

18. Innovationsforum
Offenheit durch Geschlossenheit?
Erschienen Februar 2019

19. Innovationsforum
Vom Umgang mit Unsicherheit in kreativen Prozessen
Erschienen September 2019

Impressum

Entscheidungen: Mensch – Algorithmus
20. Innovationsforum der Daimler und Benz Stiftung
am 04. November 2019 im Haus Huth, Berlin

Erschienen im März 2020
Herausgeber:
Daimler und Benz Stiftung

Geschäftsstelle
Dr.-Carl-Benz-Platz 2
68525 Ladenburg

Büro Berlin
Alte Potsdamer Str. 5
10785 Berlin
www.daimler-benz-stiftung.de

Redaktion:
Marcus Peter

Gestaltung: **Embassy | www.embassyexperts.com**
Fotografie: **Foto di Matti**