Katharina Zweig

EIN ALGORITHMUS HAT KEIN TAKTGEFÜHL

Katharina Zweig

EIN ALGORITHMUS HAT KEIN TAKTGEFÜHL

Wo künstliche Intelligenz sich irrt,
warum uns das betrifft und
was wir dagegen tun können

HEYNE

Penguin Random House Verlagsgruppe FSC® N001967

6. Auflage
Originalausgabe 2019

Illustratoren: Sandra Schulze, Katharina Zweig
Redaktion: Heike Gronemeier
Umschlaggestaltung: Favoritbüro, München,
unter Verwendung eines Fotos von Leremy / Shutterstock
Herstellung: Helga Schörnig
Satz: satz-bau Leingärtner, Nabburg
Druck und Bindung: CPI books GmbH, Leck
Printed in the EU
ISBN: 978-3-453-20730-1

www.heyne.de

Gewidmet meiner Mutter, die mich das Lehren lehrte
und mir ihre Schreibe vererbt hat.

Inhalt

Vorwort 9

TEIL I Der Werkzeugkoffer 17

Kapitel 1 Robo-Richter mit schlechtem Urteilsvermögen 19

Kapitel 2 Die Faktenfabriken der Naturwissenschaften 31

TEIL II Das kleine ABC der Informatik 47

Kapitel 3 Algorithmen – Handlungsanweisungen für Computer 49

Kapitel 4 Big Data und Data Mining 83

Kapitel 5 Computerintelligenz 125

Kapitel 6 Maschinelles Lernen versus Mensch (2:0) 181

Kapitel 7 Alphabetisierung geglückt? 201

TEIL III Der Weg zu besseren Entscheidungen – mit und ohne Maschinen 203

Kapitel 8 Algorithmen, Diskriminierung und Ideologie 205

Kapitel 9 Wie man die Kontrolle behält 233

Kapitel 10 Wer will eigentlich die Maschine als Entscheider über Menschen? 253

Kapitel 11 Die Sache mit der starken KI 267

Schlusswort 281

Ein Dank zum Schluss 287

Anhang 289

Anmerkungen 289

Glossar 313

Bildnachweis 319

Vorwort

Das Wichtigste an diesem Buch sind Sie, meine lieben Leserinnen und Leser! Denn künstliche Intelligenz – oder kurz KI – wird überall Einzug halten und Entscheidungen über uns, mit uns und für uns treffen. Um diese Entscheidungen so gut wie möglich zu treffen, müssen wir alle darüber nachdenken, was gute Entscheidungen eigentlich sind – und ob der Computer sie an unserer statt treffen kann. Und dafür steige ich mit Ihnen in den Maschinenraum hinter diesem Ansatz. Dort können Sie sehen, wie viele Handgriffe wir Informatiker:innen und sogenannte Data Scientists in Wirklichkeit ausführen, um aus Daten Entscheidungen zu wringen. Und hier kommen Sie ins Spiel – denn an diesen Stellen geht es um die Frage, wie **Sie** entscheiden würden. Denn die Gesellschaft sollte den Maschinen nur dann wichtige Entscheidungen überlassen, wenn sie darauf vertrauen kann, dass sie nach unseren kulturellen und moralischen Maßstäben handeln. Daher will ich mit diesem Buch vor allen Dingen eins: Sie ermächtigen! Ihnen das Gefühl des Kontrollverlusts nehmen, das viele beschleicht, wenn es um Algorithmen geht. Ihnen die notwendigen Begriffe erklären und aufzeigen, wo und wie Sie sich einmischen können. Sie aufrütteln, damit Sie mit uns Informatiker:innen, mit der Politik und Ihren Arbeitgebern über den Sinn und Unsinn von künstlicher Intelligenz diskutieren können.

Und warum wird die künstliche Intelligenz überall Einzug halten? Zum einen, weil sie uns die lästigen, immer wiederkehrenden Teile der Arbeit abnehmen kann und damit Prozesse effizienter macht. Zum anderen sehe ich im Moment aber auch die Tendenz,

dass künstliche Intelligenz dazu genutzt werden soll, Entscheidungen über Menschen zu treffen. Indem beispielsweise aus bestimmten Daten herausgelesen wird, ob eine Bewerbung zu einem Vorstellungsgespräch führen sollte, ob eine Person fit genug für ein Studium ist oder ob jemand vielleicht terroristische Neigungen hat.

Doch wie konnte es so weit kommen, dass manche überhaupt erwägen, Maschinen für die besseren Richter über Menschen zu halten? Nun, zuallererst können Computer natürlich Datenmengen bewältigen, die Menschen nicht mehr analysieren können. Wichtiger scheint mir aber, dass es im Moment nicht weit her ist mit unserem Vertrauen in die Urteilskraft des Menschen. Nicht erst seitdem Daniel Kahneman 2002 für seine Forschung zur Irrationalität des Menschen und 2017 dann Richard Thaler für seine Idee des »Nudgings«[1] mit dem Nobelpreis geehrt wurden, nehmen wir die Menschheit in ihrer Gesamtheit als irrational wahr, als manipulierbar, subjektiv und voreingenommen. Dabei ist natürlich der jeweils andere immer wesentlich irrationaler als man selbst,[2] insbesondere, wenn er oder sie uns in unserer eigenen Individualität und Komplexität völlig falsch beurteilt! Wir hoffen daher darauf, dass die unbestechlichen Maschinen objektivere Entscheidungen treffen können, dass sie mit ein wenig »Magie« Muster und Regeln im menschlichen Verhalten entdecken, die den Experten entgangen sind, und die damit für sicherere Prognosen sorgen.

Woher kommt diese Hoffnung? In den letzten Jahren haben Entwicklerteams aus aller Welt gezeigt, wie gut und schnell Computer mithilfe künstlicher Intelligenz Aufgaben lösen, die noch vor zwei Jahrzehnten als große Herausforderungen galten: Die Maschinen schaffen es, täglich Milliarden von Webseiten zu durchforsten und uns die besten Ergebnisse für unsere Suchanfragen zu präsentieren; sie erkennen halbverdeckte Radfahrer und Fußgänger auf Kamerabildern und können deren nächste Bewegungen recht zuverlässig vorhersagen; sie haben im Schach und dem asiatischen Brettspiel »Go« sogar die jeweiligen Weltmeister geschlagen. Ist es da nicht naheliegend, dass sie die Gesellschaft auch dabei unterstützen könnten,

faire Urteile über Menschen zu treffen? Oder sollten die Maschinen diese Urteile einfach gleich selbst fällen?

Manche versprechen sich davon, dass die Entscheidungen dadurch objektiver werden – das ist an vielen Stellen auch nötig! Eines der Länder, in denen heute schon algorithmische Entscheidungssysteme wichtige menschliche Entscheidungen vorbereiten, sind die USA. In einem Land, das 20 Prozent aller weltweit offiziell gemeldeten Gefangenen beherbergt und in dem Afroamerikaner ein circa sechsfach höheres Risiko haben, inhaftiert zu werden als Weiße, wünscht man sich Systeme, die jeglichen latenten Rassismus vermeiden. Und das möglichst, ohne deutlich mehr Geld dafür aufwenden zu müssen. Dies führte zur Einführung von sogenannten »Rückfälligkeitsvorhersagealgorithmen«, die eine Einschätzung darüber abgeben, wie stark rückfallgefährdet eine schon früher straffällig gewordene Person sei. Diese Systeme basieren auf einer automatischen Analyse der Eigenschaften von bekannten Kriminellen, die oft bei denen zu finden sind, die rückfällig werden, und selten bei denen, die es nicht werden. Es hat mich sehr erschüttert, dass wir in unserer Forschung zeigen konnten, dass eines dieser vielfach verwendeten algorithmischen Entscheidungssysteme dabei bis zu 30 Prozent, in Bezug auf schwere Straftaten sogar bis zu 75 Prozent Fehlurteile (!) produziert. Das bedeutet: Von allen Personen, die der Algorithmus in eine Hochrisikogruppe für Rückfälligkeit steckt, werden bei einfachen Straftaten drei von zehn Personen **nicht** rückfällig, und bei der Vorhersage einer schweren Straftat begeht tatsächlich nur jeder vierte von ihnen eine solche Tat. Ein einfaches Raten, das die allgemeinen Rückfälligkeitswahrscheinlichkeiten berücksichtigt, wäre nur wenig schlechter gewesen, hätte aber wenigstens den Vorteil gehabt, dass man sich des »reinen Ratens« bewusst gewesen wäre.

Was also geht schief, wenn Maschinen den Menschen bewerten? Als Wissenschaftlerin mit einem sehr interdisziplinären Lebenslauf betrachte ich die Aus- und Nebenwirkungen von Software unter einer besonderen Perspektive: Die der Sozioinformatik. Die Sozio-

informatik ist ein junges Teilgebiet der Informatik, das Methoden und Ansätze aus der Psychologie, Soziologie, den Wirtschaftswissenschaften, der statistischen Physik und natürlich der Informatik nutzt. Wir gehen dabei davon aus, dass die Interaktionen von Technik und Programmierern einerseits und die von Nutzern und Software andererseits nur verstanden werden können, wenn sie als Gesamtsystem betrachtet werden. Solche Systeme nennen wir »sozio-technische Systeme«.

Konkret forsche ich seit über 15 Jahren dazu, wie und wann wir mit Computern unsere komplexe Welt besser verstehen können und zwar mithilfe des sogenannten Data Minings, der »Nutzbarmachung von Daten«. Damit gehöre ich zu den Menschen mit dem sexiest Job auf Erden[3] – auch wenn es sich für andere nicht sehr verlockend anhören mag, seine Wochenenden damit zu verbringen, knietief in riesigen Datenmengen zu stehen und diese mithilfe von statistischen Methoden nach aufregenden Zusammenhängen zu durchkämmen. Für mich ist es das auf jeden Fall! Zu Beginn meiner Karriere war ich allerdings erst einmal nur eine reine Nutzerin dieser Methoden. Immer unsicher, ob ich dieses oder jenes Verfahren überhaupt anwenden dürfte und ob die Ergebnisse wirklich aussagekräftig sein würden. Das lag daran, dass ich nach dem Abitur zuerst einen typischerweise mathematikfernen Studiengang gewählt hatte, die Biochemie. Hier bekamen wir Grundlagenwissen in Biologie, Medizin, Physik und Chemie – aber keine einzige Stunde in Statistik. Die Hoffnung war wohl, dass das Wissen durch reine Diffusion in unsere Köpfe fließen würde, wenn wir nur genügend Experimente nachkochten.

Später studierte ich noch Bioinformatik, einen damals ganz neuen Studiengang, der das Design und die Anwendung von Methoden zur Untersuchung der damals in immer größeren Mengen anfallenden Biodaten lehrte. Auch hier fehlte allerdings die Statistik. Und in keinem der beiden Studiengänge wurden wir in Wissenschaftstheorie unterrichtet – eine völlig unverständliche und gefährliche Lücke im Lehrplan fast aller naturwissenschaftlichen Studiengänge, die Fakten produzieren wollen und sollen.

Und so ist es nicht verwunderlich, dass viele Informatiker und Ingenieure sich zu sicher sind, dass die Methoden die reine und objektive Wahrheit aus den Daten holen, und insbesondere im Data Mining und im maschinellen Lernen, der Grundlage für künstliche Intelligenz, das Heil bei der Lösung aller komplexen Probleme sehen. Denn wer nicht weiß, dass er nur mit Modellen hantiert und niemals endgültige Gewissheit erlangen kann, der schwingt sich schnell zu Aussagen wie den folgenden auf: »Stellen Sie sich eine Welt vor, in der Sie das maximale Potenzial jeder Sekunde Ihres Lebens ausschöpfen könnten. Ein solches Leben wäre produktiv, effizient und einflussreich. Sie werden (schlussendlich) Superkräfte haben – und viel mehr Freizeit. Vielleicht würden manche diese Welt auch als ein bisschen langweilig ansehen – solche, die gerne unberechenbare Risiken eingehen. Ganz sicher aber nicht alle diejenigen Organisationen, die Profit machen wollen. Diese Organisationen geben schon heute Millionen für Manager aus, die nur dazu da sind, um mit Risiken umzugehen. Und wenn es irgendetwas da draußen gibt, das Sie darin unterstützt, gleichzeitig Ihre Arbeitsschritte optimiert und die Profite maximiert, dann sollten Sie es definitiv kennenlernen. Dieses Hilfsmittel ist die Welt der analytischen Vorhersagen.«[4]

Und das war nur die Einleitung zu einem kurzweiligen Lehrbuch zum Thema! Ernster wird es dann schon, wenn Firmen für ihre Data-Mining-Software im Bereich »Vorhersage der Leistung von Arbeitnehmer:innen« mit den Worten werben:

»(…) am Ende sind die Möglichkeiten zur Vorhersage im Wesentlichen unbegrenzt, wenn nur genügend gute Daten zur Verfügung stehen. (…) Lassen Sie uns die Gefühle aus dem Bewerbungsprozess nehmen und sie durch einen daten-getriebenen Ansatz ersetzen!«[5] Aber das ist ein gutes Stichwort, denn auch bei Ihnen möchte sich jemand bewerben, nämlich »KAI«. Er möchte Ihr völlig datengetriebener Buchbegleiter werden. KAI ist eine künstliche Intelligenz (KI, englisch *artificial intelligence, AI)* und noch ein bisschen schwer von Begriff, wenn es darum geht, die Menschen wirklich zu verstehen. Er gibt sich aber redlich Mühe!

Das ist KAI, eine Künstliche Intelligenz oder KI – im Englischen spricht man auch von AI für *artificial intelligence*. KAI möchte sich bei Ihnen als Buchbegleiter bewerben. Er kann ganz schön viel, ist aber auch noch ein bisschen naiv. Seien Sie also bitte nett zu ihm!

Nach diesen beiden Zitaten ahnen Sie sicher schon, dass ich vor übermäßigem Vertrauen in KAI warnen möchte. Wann wir uns auf die Ergebnisse maschinellen Lernens nicht allzu leichtfertig verlassen sollten, werde ich Ihnen in diesem Buch aufzeigen. Es ist auf der anderen Seite aber auch wichtig, zu verstehen, welche enormen Chancen im Data Mining liegen, also im Aufbereiten von Daten durch Algorithmen. Daher werde ich konkrete Vorschläge machen, wo solche algorithmischen Entscheidungssysteme aus technischen oder gesellschaftlichen Überlegungen heraus nicht zulässig sind. Für die Fälle, wo es möglich ist, sie entscheiden zu lassen, werde ich konkret zeigen, wann wir ihnen dabei auf die Finger sehen müssen. Dazu mache ich Vorschläge, wie sie so entwickelt, kontrolliert und reguliert werden könnten, dass sie wirklich bestmöglich entscheiden.

Dieses Buch gibt Ihnen also die notwendigen Informationen, um zu verstehen, wie Computer zu Richtern über Menschen werden,

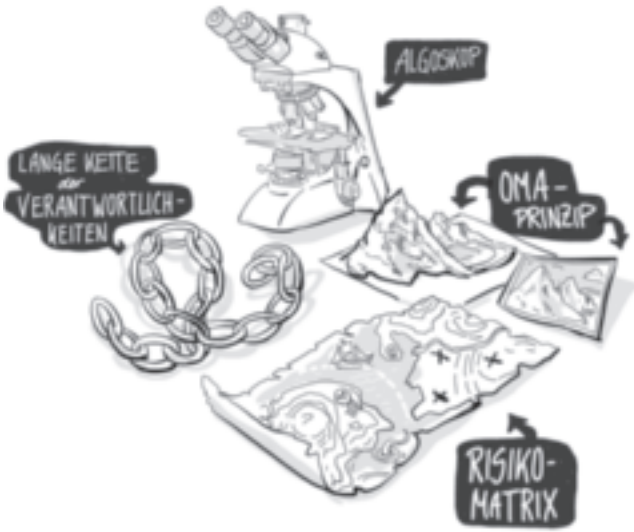
warum sie das im Moment oft nicht gut machen, und auch, wie wir sie verbessern können. Ich werde aber insbesondere auch diskutieren, wo wir sie gar nicht erst einsetzen sollten, um zu verhindern, dass wir mit vermeintlicher Objektivität und scheinbarer Gewissheit falsche Urteile über unsere Mitmenschen fällen.

Das Buch besteht aus drei Teilen: Im ersten Teil des Buches präsentiere ich Ihnen die naturwissenschaftliche Methode der Erkenntnisgewinnung und stelle Ihnen schon mal Ihren Werkzeugkoffer für die Gestaltung von KI-Systemen vor. Im zweiten Teil geht es in den Maschinenraum, wo ich Ihnen das kleine ABC der Informatik erkläre: **A**lgorithmus, **B**ig Data und **C**omputerintelligenz und wie sie zusammenhängen. Im dritten Teil geht es dann konkret um die Frage, wo die Ethik in den Rechner kommt und wie man diesen Prozess bestmöglich gestaltet.

Dies Buch soll Ihnen das Werkzeug an die Hand geben, damit Sie sich einmischen können. Damit wir als Gesellschaft bessere Entscheidungen treffen können – sowohl mit als auch ohne Maschinen.

TEIL I

Der Werkzeugkoffer



Wer sich mit »künstlicher Intelligenz« anlegen will, der braucht die richtigen Tools. Wenn Ihre Arbeitgeberin oder der Staat planen, ein algorithmisches Entscheidungssystem einzusetzen, werden Sie in Zukunft erst mal Ihren Werkzeuggürtel mit den in diesem Buch beschriebenen vier Werkzeugen bestücken. Und dann klappern Sie methodisch die möglichen Fallstricke ab – oder geben gleich Entwarnung – denn nicht alles, was gefährlich aussieht, ist es auch.

KAPITEL 1

Robo-Richter mit schlechtem Urteilsvermögen

Es war nicht das erste Mal, dass ich fassungslos vor den Ergebnissen unserer wissenschaftlichen Nachforschungen saß – aber vermutlich das eindrücklichste Mal. Zusammen mit meinem Doktoranden Tobias Krafft hatte ich mir die Vorhersagen einer speziellen Software angesehen, die in den USA in Gerichtssälen eingesetzt wird. Und wir waren entsetzt, wie schlecht diese von einem Staat genutzten Vorhersagen sind, die in einem so wichtigen Vorgang genutzt werden. Die Idee hinter der Nutzung von Algorithmen zur Vorhersage, ob eine Person eine Straftat begehen wird, erinnert an den Film »Minority Report«. Tom Cruise spielt darin einen Polizisten, der durch die Zusammenarbeit mit hellseherisch begabten »Precogs« Personen identifizieren kann, die in Zukunft Straftaten begehen werden. Noch bevor es dazu kommt, kann Cruise die potenziellen Straftäter in Gewahrsam nehmen. Diese bizarre Geschichte, die der berühmte Science-Fiction-Autor Philipp K. Dick schon 1956 als Kurzgeschichte entwickelte, ist Realität geworden – nur leider fehlt die Präzision der vorhersagenden Maschinerie.

Im Gegensatz zum Film kann die vorhersagende Software die eigentliche Tat natürlich nicht »sehen« oder gar den genauen Zeitpunkt wissen. Stattdessen bekommt sie über alle Kriminelle, die sie bewerten soll, grundlegende Informationen: Wie oft sie schon verhaftet worden sind, welche Arten von Straftaten sie bisher begangen haben, dazu Informationen über ihr Alter und ihr Geschlecht.

Daraus berechnet der Computer einen »Risikoscore«, den Sie sich vorstellen können wie die Schadensfreiheitsklasse in einer Autoversicherung: Dort sind Personen mit hohem Risiko in einer Klasse zusammengefasst, solche mit niedrigem Risiko in einer anderen. Wenn eine Person in eine solche Klasse einsortiert wird, passiert etwas Merkwürdiges: Obwohl sie selbst (noch) gar nichts gemacht hat, wird sie so behandelt, wie die, die schon früher in diese Klasse einsortiert wurden. Waren diese Personen in viele Unfälle verwickelt, bezahlt man mehr. Waren sie in wenige Unfälle verwickelt, bezahlt man weniger. Wieviel man zahlt, ist bei der ersten Einstufung also nicht vom persönlichen zukünftigen Verhalten abhängig, sondern nur davon, wem man ähnelt und wie sich diese anderen in der Vergangenheit verhalten haben. Bei Autoversicherungen wird so das finanzielle Risiko auf alle Personen in derselben Klasse verteilt.

Aber wie soll das Verfahren bei der Frage nach zukünftigen Straftaten funktionieren? Nun, das Prinzip ist erst einmal dasselbe: Der Rechner sucht diejenigen Eigenschaften, die bei rückfälligen Kriminellen häufig sind und selten bei solchen, die in der Gesellschaft wieder Fuß fassen. Diese Eigenschaften bestimmen dann das Risiko einer Person. Im Autoversicherungsbeispiel sind diese risikobestimmenden Eigenschaften das Alter der Fahrer und die Anzahl der durchgehend unfallfreien Beitragsjahre. Das muss man nicht fair finden – es ist sicherlich auch unterkomplex. Wäre es nicht beispielsweise gerechter, einen Persönlichkeitstest durchzuführen und danach zu entscheiden, wer in welche Klasse kommt?

Es ist natürlich der Effizienz geschuldet, dass die Einstufung anhand sehr einfacher und leicht zu messender Eigenschaften geschieht. Das Verfahren ist aber insofern gerecht, als dass alle Fahrerinnen und Fahrer, die mit 18 Jahren ihren Führerschein bekommen, genau gleich starten und ihre spätere Klassifikation nur von ihrem eigenen Fahrverhalten abhängig ist und nicht mehr von dem ihrer Generation.

Das kann man vom Einstufungsverfahren der von uns untersuchten Rückfälligkeitsvorhersagesoftware namens COMPAS nicht behaupten: Neben den oben genannten Informationen über bisherige Straftaten wird in einem Fragebogen zum Beispiel auch abgefragt, ob Eltern und Geschwister straffällig wurden oder ob die Eltern schon früh geschieden waren. Das sind Umstände, die ein Individuum zwar prägen mögen, aber von ihm weder zu verantworten noch zu ändern sind.[6] Basierend auf allen Eigenschaften, welche die Software-Firma für relevant hielt, wird eine kriminelle Person nun bewertet und in eine Risikokategorie eingestuft: Sind dort viele eingestuft, die in der Vergangenheit rückfällig wurden, geht die Software auch bei dieser Person davon aus, dass sie rückfällig werden wird.

Der Bewertungsalgorithmus wird damit beworben, dass circa 70 Prozent seiner Entscheidungen richtig seien.[7] Das allein fanden Tobias und ich schon beunruhigend niedrig für eine Software, die

von einer staatlichen Stelle vor Gericht eingesetzt wird. In der Medizin würde eine solch geringe Prozentzahl auch tatsächlich als nicht genügend angesehen werden. Aber nun lagen Ergebnisse vor uns, die belegten, wie viele Personen aus der höchsten Risikokategorie tatsächlich rückfällig geworden sind: Es waren zwar etwas mehr als 70 Prozent bei allgemeinen Straftaten, aber nur um die 25 Prozent bei denjenigen, bei denen gewalttätige Straftaten vorhergesagt wurden. Das heißt, dass nur jeder Vierte, der oder die mit einem kaum zu ignorierenden Alarmsignal als anfällig für eine weitere schwere Gewalttat versehen wird, auch tatsächlich wieder eine solche Tat begeht. Zudem zeigten andere Kollegen, dass auch Laien eine solche Vorhersage mit im Wesentlichen derselben Qualität treffen können.[8]

Ich habe die letzten drei Jahre damit verbracht, zu verstehen, warum irgendjemand so schlecht vorhersagende Algorithmen verwenden wollte und wieso Regierungen sie in Auftrag geben oder kaufen. Und natürlich wollte ich die Königsfrage lösen, wie wir bessere Software erstellen können und ob es vielleicht Situationen gibt, in denen Algorithmen über Menschen grundsätzlich nicht entscheiden sollten. Aber hat das etwas mit Ihnen zu tun, liebe Leserinnen und Leser? Ist das nicht alles so technisch, dass Sie dabei einfach keinen Gestaltungsspielraum haben? Ihre und meine gemeinsame Erfahrung in den letzten Jahren ist eher, dass wir keine Chance haben, die Algorithmen zu verändern, die unser Leben mitbestimmen: Von Google über Facebook zu Amazon – alles ist verwirrend und zu weit weg vom Alltag. Wir Individuen, aber selbst die Gesellschaft als Ganzes, Deutschland und vielleicht sogar Europa, scheinen nahezu ohnmächtig gegenüber diesen transatlantischen Algorithmen. Das Gefühl des Kontrollverlustes liegt aber nicht nur in der Tatsache verankert, dass diese und andere Firmen sich global immer dort ansiedeln, wo die angenehmsten Regeln herrschen. Es liegt auch an der Technik selbst. Sie wird häufig präsentiert als eine objektive Methode, die aus Daten Entscheidungen generiert. Ja, geradezu als ein Ansatz, der aus Daten die **WAHRHEIT** extrahiert. Die einzige

Entscheidung, die einem unter diesen Umständen übriggeblieben zu sein scheint, ist binär: Wollen wir Algorithmen, die über uns bestimmen, oder wollen wir sie nicht? Verweigern wir uns dieser gesamten Digitalisierung oder opfern wir unsere persönlichen Daten für die vielen neuen Dienste?

Für die algorithmischen Entscheidungssysteme, denen wir in den nächsten Jahren begegnen werden, gibt es glücklicherweise nicht nur dieses Entweder-oder. Bei diesen Systemen werden Sie sich einmischen können – und sollten es auch tun. Denn diese algorithmischen Entscheidungssysteme werden von Ihrem Arbeitgeber eingesetzt werden, von Ihrer Ausbildungsstätte, Ihrem Versicherer oder vom Staat. Und bei jedem dieser Einsätze haben Sie als Arbeitnehmer:in, als Schüler:in oder Student:in, als Verbraucher:in oder Bürger:in einen Hebel: Sie können dagegen Widerspruch einlegen und sich in den Entwicklungsprozess einbringen. Diese Erkenntnis hilft aber nur bei dem einen der beiden genannten Probleme: Die Ansprechpartner sind hier vor Ort und damit greifbar.

Aber gibt es inhaltlich wirklich Punkte, an denen **Sie** sich einbringen können? Dazu ist es fundamental wichtig, zu verstehen, wie die Maschinerie hinter der künstlichen Intelligenz und dabei insbesondere die des sogenannten **maschinellen Lernens** funktioniert. Und das ist eher so ein Daniel-Düsentrieb-Prozess des Herumdokterns. Der Prozess ist weit weniger objektiv und selbstgesteuert, als Sie es vermuten würden. Die daraus resultierende Maschine zur Entscheidungsfindung ist an vielen Stellen justierbar – und an manchen nur mit Bindfäden zusammengehalten. Daher ist es auch so wichtig, dass manche dieser Maschinen eng überwacht werden – und zwar auf dieser Maschinenraumebene.

Abbildung 1: Der Prozess der maschinellen Entscheidungsfindung ist das Ergebnis von professionellem Herumtüfteln an Daten und Maschinerie.

In ihrer Gesamtheit werden wir als Gesellschaft auf die möglichen Einsätze von künstlicher Intelligenz umfassend arbeits-, sozial- und bildungspolitisch reagieren müssen. Aber in diesem Buch geht es um die Frage danach, wie wir sie gestalten wollen, wann wir diese Maschinen überwachen müssen, wo das möglich ist und wo wir sie gar nicht einsetzen sollten.

Und tatsächlich brauchen wir Datenwissenschaftler Sie als Arbeitnehmer:in, Verbraucher:in und Bürger:in dazu im Maschinenraum. Für Ihren Einsatz dort stelle ich Ihnen in diesem Buch einen Werkzeugkoffer zusammen, den ich Ihnen im Folgenden kurz vorstelle.

Ausgerüstet mit den Werkzeugen, die ich in den folgenden Kapiteln detailliert beschreibe, können Sie dann erkennen, ob Sie a) überhaupt ranmüssen; b) wo Sie ansetzen können; und c) welche Konsequenzen Ihre Einschätzung für den kontrollierten Einsatz der Maschine hat. Denn natürlich müssen Sie sich nicht überall einmischen. Für die Entscheidung, ob Sie sich einmischen sollten, möchte ich Ihnen ein erstes Instrument in die Hand drücken: das **Algoskop**. Es filtert diejenigen Systeme heraus, um die man sich prinzipiell kümmern muss.

Sind das alle Systeme, die künstliche Intelligenz verwenden? Über diese Fragen haben sich in den letzten Jahren viele Personen Gedanken gemacht. 2013 machten Viktor Mayer-Schöneberger und Kenneth Cukier in ihrem Buch »Big Data – Die Revolution, die unser Leben verändern wird« den Vorschlag, einen generellen Algorithmen-TÜV einzuführen. Das ist aus verschiedenen Gründen in dieser Form weder sinnvoll noch notwendig, wie ich später zeigen werde. Insbesondere müssen aber nicht alle algorithmischen Entscheidungssysteme auf den Prüfstand. Im Wesentlichen sind es **nur die Systeme**, die

- **über Menschen entscheiden** oder
- **über Ressourcen, die Menschen betreffen,** oder die
- **solche Entscheidungen treffen, die die gesellschaftlichen Teilhabemöglichkeiten von Personen ändern**,

die einer Regulierung und Kontrolle ihrer inneren Mechanik bedürfen.[9] Es handelt sich damit nur um kleinen Teil aller möglichen »Algorithmen«. Diese Fokussierung auf ethisch relevante algorithmische Entscheidungssysteme nenne ich das **Algoskop**. Warum es im Wesentlichen nur diese Systeme sind, die verstärkt kontrolliert und reguliert werden müssen, erkläre ich ausführlich in Teil II und III dieses Buches.

Abbildung 2: Das **Algoskop** benennt, um welche Art von Software wir uns verstärkt kümmern müssen: um algorithmische Systeme, die unmittelbar über Menschen entscheiden oder Entscheidungen fällen, die Menschen mittelbar betreffen.

Das heißt: Systeme, die entscheiden, ob eine Schraube defekt ist und vom Produktionsband gepustet werden sollte, fallen nicht darunter. Ein System, das punktgenau Dünger auf einem Acker ausbringt, fällt ebenfalls nicht darunter. Ein autonomes Auto, das im Zweifelsfall in Unfälle verwickelt sein könnte, dagegen natürlich schon. Systeme, die einfach nur Bilder erkennen oder Sprache übersetzen, gehören eher nicht dazu – es sei denn, sie sind in autonomen Autos verbaut, wo sie wiederum zu Unfällen beitragen. Definitiv dazu gehören KI-Systeme im medizinischen Bereich. Darunter wiederum solche weniger, die uns freiverkäufliche Produkte empfehlen als solche, die über Therapien entscheiden.

Wenn also der KI-Alarm ertönt, schauen Sie erst einmal, **was** das System entscheiden soll. Wenn es weder direkt noch indirekt um das menschliche Wohlbefinden geht, dürfen Sie zurück in den Pausenraum.

Wenn es aber doch um das menschliche Wohl geht, dann ist die Qualität der Entscheidungen durch die Maschine von den folgenden Erfolgsfaktoren abhängig:

- von der Qualität und Quantität der eingehenden Daten,
- von den grundlegenden Annahmen über die Natur der Fragestellung
- und davon, was die Gesellschaft eigentlich für eine »gute« Entscheidung hält.

Dieser letzte Punkt, die Frage nach der »guten Entscheidung«, ist aus Informatiksicht ein »Modell für eine gute Entscheidung« – die Philosophie würde es eine »Moral« nennen. Damit Algorithmen eine solche Moral befolgen können, muss für die Maschine »messbar« gemacht werden, wie sehr eine Entscheidung dieser Moral entspricht. Nur dann kann der Computer versuchen, Entscheidungen zu optimieren. Das ist aber gar nicht so einfach. Wenn eine Software genutzt werden soll, um Kinder Schulen so zuzuteilen, dass ihre Schulwege möglichst kurz sind: Soll der Schulweg dann im Durchschnitt klein sein? Oder für kein Kind ein bestimmtes Maximum überschreiten? Diese Entscheidung, wie nachher die Güte eines algorithmischen Ergebnisses bewertet werden soll, ermöglicht eine

Messung, wie gut eine algorithmische Lösung ist. Diese **Messbarmachung** nennen wir »Operationalisierung«.

Neben dieser Entscheidung steht die Frage, was genau der Computer als Informationen bekommt, um die Länge des Schulweges zu berechnen: Werden dabei ideale Fahrzeiten zugrunde gelegt oder reale? Fußwege zur Bushaltestelle mitberücksichtigt? Diese Entscheidungen nennen wir das »Modell des Problems«, das vom Computer gelöst werden soll.

Damit die Ergebnisse der Datenverarbeitung der vorher festgelegten Moral folgen, müssen also die dafür notwendigen Messbarmachungen (**O**perationalisierungen), das **M**odell des Problems und der **A**lgorithmus zusammenpassen: Das ist das **OMA-Prinzip**, und damit Ihr zweites Werkzeug. Was es damit genau auf sich hat, und wie man das OMA-Prinzip handhabt, werde ich Ihnen anhand vieler Beispiele, beginnend in Kapitel 2, vorstellen.

Doch selbst das OMA-Prinzip ist noch nicht ausreichend, um zu beurteilen, ob und wann Maschinen einen Teil menschlicher Entscheidungen übernehmen können. Dazu ist es zudem notwendig, deren Rolle im Gesamtprozess zu betrachten.

Die nächste Abbildung zeigt, wie lang der Prozess der Entwicklung und des Einsatzes von **algorithmischen Entscheidungssystemen** ist. Diesen Prozess werde ich im Buch Stück für Stück erklären – seine Länge ist vor allen Dingen deshalb ein Problem, weil dabei die Verantwortung für einzelne Entscheidungen auf so viele Schultern verteilt wird, dass es nachher schwierig ist, sie bei einer Person zu verorten. Für den Moment ist für Sie aber erst einmal wichtig, dass es darin **nur wenige Stellen** gibt, an der technisches Wissen notwendig ist. In jeder Phase gibt es aber Aspekte, bei denen Sie auch mitreden können und sollten. Diese Darstellung nenne ich die **lange Kette der Verantwortlichkeiten**.[10] An ihr entlang und um sie herum entwickeln sich die Themen des Buches. Und mit dieser langen Kette der Verantwortlichkeiten haben Sie nun das dritte Werkzeug in Ihrem Koffer, denn sie zeigt, wo man hingucken muss.

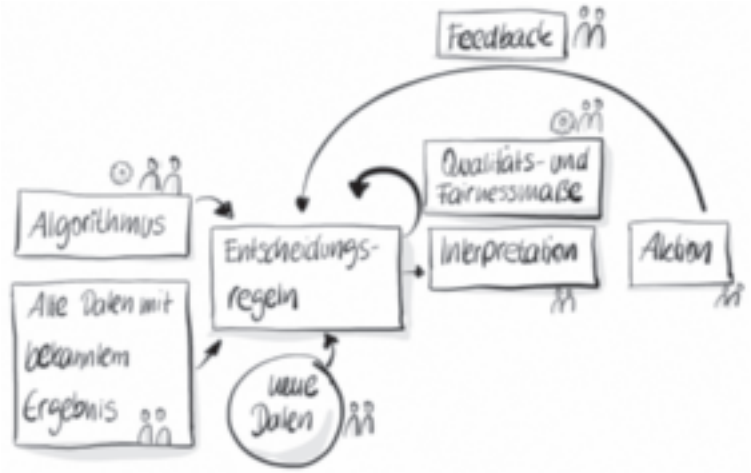


Abbildung 3: Lange Kette der Verantwortlichkeiten. Nur an zwei Stellen ist auch technisches Wissen notwendig – an allen Stellen aber können und sollten Sie sich einmischen. Die einzelnen Schritte des Prozesses und was dabei jeweils schiefgehen kann, werden in den nächsten Kapiteln detailliert erklärt. Ein Zahnrad deutet an, dass bei den an dieser Stelle notwendigen Entscheidungen (auch) technisches Wissen notwendig ist. Die beiden Personen deuten an, dass an diesen Stellen gesunder Menschenverstand ausreicht und gesellschaftlicher Diskurs notwendig ist.

Wie stark man nun eine Maschine, die Entscheidungen berechnet, überwachen sollte, hängt im Wesentlichen davon ab, wieviel Schaden sie verursachen und wie gut man sich dagegen wehren kann. Dazu biete ich Ihnen als viertes Werkzeug eine **Messung der Regulierungsnotwendigkeit** an, die mit verschiedenen Kontrollmaßnahmen verknüpft ist. Dieses Werkzeug erläutere ich an ein paar Beispielen, nachdem ich Ihnen den Maschinenraum gezeigt habe.

So, mit diesen Instrumenten ist der Werkzeugkoffer komplett, und sobald Sie deren Anwendung näher kennengelernt haben, können Sie mit diesen Instrumenten bestimmen, wo sich Ihr Einsatz lohnt.

Ich starte meine Führung durch den Maschinenraum der künstlichen Intelligenz mit einem Ausflug in die Labore der Naturwissen-