

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 <

Sollte diese Publikation Links auf Webseiten Dritter enthalten,
so übernehmen wir für deren Inhalte keine Haftung,
da wir uns diese nicht zu eigen machen, sondern lediglich
auf deren Stand zum Zeitpunkt der Erstveröffentlichung verweisen.



Verlagsgruppe Random House FSC® N001967

5. Auflage
Originalausgabe 2019

Copyright © 2019 by Wilhelm Heyne Verlag, München,
in der Verlagsgruppe Random House GmbH,

Neumarkter Straße 28, 81673 München

Illustratoren: Sandra Schulze, Katharina Zweig

Redaktion: Heike Gronemeier

Umschlaggestaltung: Favoritbüro, München,

unter Verwendung eines Fotos von Leremy / Shutterstock

Herstellung: Helga Schörnig

Satz: Leingärtner, Nabburg

Druck und Bindung: CPI books GmbH, Leck

Printed in Germany

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 Urteilstkraft 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 Erden³ – auch wenn es sich für andere nicht sehr verlockend anhören mag, seine Wochenenden damit zu verbringen, knietief in riesigen Datensätzen 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 aus schöpfen 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 daten-getriebener 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: Algorithmus, Big Data und Computerintelligenz 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.

