MAXIMILIAN KASY

DIE MITTEL DER PRÄDIKTION

Wie KI wirklich funktioniert (und wer davon profitiert)

SUHRKAMP

SV

MAXIMILIAN KASY

DIE MITTEL DER PRÄDIKTION

Wie KI wirklich funktioniert (und wer davon profitiert)

Aus dem Englischen von Andreas Wirthensohn

Suhrkamp

Die Originalausgabe erschien 2025 unter dem Titel
The Means of Prediction. How AI Really Works (and Who Benefits)
bei The University of Chicago Press, Chicago, Illinois, USA.

Erste Auflage 2026
Deutsche Erstausgabe

Umschlaggestaltung: Rothfos & Gabler, Hamburg
Umschlagabbildung: ZinetroN/Shutterstock
Satz: Satz-Offizin Hümmer GmbH, Waldbüttelbrunn
Druck: CPI books GmbH, Leck
Printed in the EU
ISBN 978-3-518-58856-7

Suhrkamp Verlag GmbH
Torstraße 44, 10119 Berlin
info@suhrkamp.de
www.suhrkamp.de

Für Sanaz

Inhalt

Vorwort

Über künstliche Intelligenz gibt es jede Menge guter Bücher. Einige von ihnen erklären ausführlich und detailliert, wie die Technologie, die Statistik und die Computerwissenschaft der KI funktionieren. Andere konzentrieren sich auf eines der vielen Probleme, die KI mit sich bringen könnte: auf Algorithmen, die nicht wie versprochen funktionieren, Algorithmen, die diskriminieren, Algorithmen, die sich gegen ihre menschlichen Meisterinnen wenden, und Algorithmen, die menschliche Arbeitskraft automatisieren und überflüssig machen. Wieder andere erörtern die problematischen Grundlagen, auf denen KI basiert, von der Überwachung von Internetnutzern und der Ausbeutung von Clickworkerinnen bis hin zur Umweltzerstörung durch Rechenzentren und Rohstoffgewinnung. All das sind wichtige Themen.

Was bisher jedoch fehlte, war ein einheitlicher Rahmen, um zu verstehen, wie KI in einer von Macht und Ungleichheit geprägten Gesellschaft wirksam werden wird. Das vorliegende Buch will einen solchen liefern. Inmitten all der hektischen Debatten über technische Details, neue Möglichkeiten und gesellschaftliche Probleme vertrete ich die These, dass das zentrale Thema, das alle Probleme der KI in sich vereint, die Wahl der Ziele ist, die KI verfolgt, und die Frage, wer diese Ziele kontrolliert. Bestimmt wird die Kontrolle über diese Ziele durch die Kontrolle über die Ressourcen, die für die Entwicklung von KI erforderlich sind – Daten, Recheninfrastruktur, technisches Know-how und Energie. Ich nenne diese Ressourcen die Mittel der Prädiktion oder kurz: die Prädiktionsmittel (*means of prediction*).

Wer bin ich, dass ich dieses Buch schreibe? Ich bin derzeit Profes-

sor für Wirtschaftswissenschaften an der Universität Oxford, wo ich Doktoranden in der Theorie maschinellen Lernens unterrichte und die Gruppe für maschinelles Lernen und Wirtschaft koordiniere. Ich habe Mathematik und Statistik sowie Wirtschaftswissenschaften studiert, und ein Großteil meiner Forschung beschäftigt sich mit Methodenfragen. Auf diesen Hintergrund stütze ich mich, wenn ich auf den folgenden Seiten den aktuellen Stand der KI in den Blick nehme.

In einigen meiner Forschungsarbeiten versuche ich, über die technischen Fragen der Statistik und des maschinellen Lernens hinauszugehen und diese Bereiche in ihrem sozialen, politischen und ökonomischen Kontext zu verstehen. Auf einem anderen Forschungsfeld beschäftige ich mich mit wirtschaftlicher Ungleichheit und damit, was die Politik tun kann, um allen ein erfülltes und sicheres Leben zu ermöglichen. Ich untersuche Pilotprogramme zur Arbeitsplatzgarantie und zum Grundeinkommen.

In meinen Augen haben Forschende die Pflicht, zu einer Gesellschaft beizutragen, in der wir gemeinsam über unsere Zukunft debattieren und entscheiden, anstatt Fragen von Technologie und Politik allein Technokraten und Expertinnen zu überlassen. In diesem Sinne verstehen sich meine Ausführungen als Beitrag zu einer breiten öffentlichen Debatte über die Zukunft der KI.

Alle Bücher bauen auf den Ideen, den Erkenntnissen und der Arbeit unzähliger Menschen auf, denen ihre Verfasserinnen viel zu verdanken haben. Entsprechend habe auch ich bei der Vorbereitung und beim Schreiben von der Kreativität, der Kritik, den Vorschlägen, den Diskussionen und der Forschungsunterstützung zahlreicher Freunde, Mitautorinnen, Kollegen und Studierender profitiert, darunter die folgenden (in alphabetischer Reihenfolge):

Alberto Abadie, Rediet Abebe, Daron Acemoğlu, Isaiah Andrews, Johanna Barop, Stefano Caria, Nicolò Cesa-Bianchi, Gary Chamberlain, Roberto Colomboni, Ellora Derenoncourt, Binta Zahra Diop, Pirmin Fessler, Susann Fiedler, Alex Frankel, Carlos Gonzalez Perez,

Verena Halsmayer, Ian Jewitt, Jeremy Large, Lukas Lehner, Gregory Levy, Peter Lindner, Carrie Love, Lester Mackey, Gerhard Meszaros, Sanaz Mobasseri, Christopher Muller, Suresh Naidu, Harald Oberhauser, Dietmar Offenhuber, Walter Palmetshofer, Daniela Platsch, Carina Prunkl, Simon Quinn, Alvaro Ramos-Chaves, Anja Sautmann, Frederik Schwerter, Jann Spiess, Alexander Teytelboym, Martin Weidner, Ashia Wilson, Noam Yuchtman und Chad Zimmerman.

Teil I:
Einleitung

Haben Sie Angst vor künstlicher Intelligenz? Das sollten Sie – sofern man einigen populären Erzählungen über die Bedrohung durch KI und den künftigen Konflikt zwischen Mensch und Maschine Glauben schenken darf.

Diesen Erzählungen zufolge wird die KI übermenschliche Fähigkeiten erlangen und damit beginnen, sich selbst zu verbessern. Sie wird die Menschheit im Namen der Selbsterhaltung bedrohen und schließlich zu einer existenziellen Gefahr für sie werden. Diese Geschichten, die von Film, Literatur, Industrie und Wissenschaft erzählt werden, rühren an unsere tiefsten und fundamentalsten Ängste. Wir haben Angst, unsere Lebensgrundlage zu verlieren und in die Armut abzurutschen. Wir haben Angst, unsere Autonomie zu verlieren und von unverständlichen und bösartigen Akteuren kontrolliert zu werden, denen unser Schicksal gleichgültig ist. Wir haben Angst, unser Leben zu verlieren. Und – was noch schlimmer ist – wir fürchten, dass das Überleben der Menschen, die wir lieben, und das Überleben der Menschheit insgesamt bedroht sein könnte. Darüber hinaus fürchten wir die künstliche Intelligenz als eine unergründliche Kraft, die unausweichlich auf uns zukommt und deren Auswirkungen offenbar niemand kontrollieren kann.

Diese Geschichten über KI, die Geschichten über den existenziellen Konflikt zwischen Mensch und Maschine, werden in Hollywood und im Silicon Valley immer wieder erzählt. Aber diese Geschichten helfen uns nicht, gute Entscheidungen zu treffen – weder auf dem Feld der Technologie noch in der Politik. Sie erwecken vielmehr den Eindruck, als gäbe es nur eine mögliche Richtung für die Entwicklung von KI und als könne die Gesellschaft nichts dagegen tun. Überdies verschleiern diese Geschichten, wer bei der Entwicklung von KI gewinnt und wer verliert. Damit wird verhindert, dass sich die öffentliche Debatte auf die wirklichen Probleme konzentriert, und es hält die Menschen davon ab, etwas dagegen zu unternehmen. Und dieses Nichtstun dient den Interessen derjenigen, die davon profitieren, dass die Dinge so bleiben, wie sie sind.

Dieses Buch bietet eine andere Perspektive auf KI und Gesellschaft. Im Gegensatz zu den populären Geschichten ist der Fortschritt der KI nicht schicksalhaft gegeben, sondern vielmehr ein Produkt menschlicher Entscheidungen. Die wichtigsten Konflikte bestehen nicht zwischen Mensch und Maschine, sondern zwischen verschiedenen Menschen. Die Antwort auf diese Konflikte ist eine gemeinsame demokratische Kontrolle der KI sowie der Ziele, die sie verfolgt: Diejenigen, die von algorithmischen Entscheidungen betroffen sind, müssen ein Mitspracherecht bei diesen Entscheidungen haben.

Um die Grundlage für eine solche andere Perspektive zu schaffen, entwickle ich in diesem Buch zunächst eine unvoreingenommene Sichtweise auf KI, so wie sie von Expertinnen für maschinelles Lernen gedacht und verstanden wird. Dabei zeigt sich, wo die Grenzen von KI liegen und wie KI so gestaltet werden kann, dass sie allen Menschen etwas bringt. Dazu müssen wir die Geschichten, die wir uns über Mensch und Maschine erzählen, revidieren.

1. Die Geschichte von Mensch und Maschine

Im Filmklassiker *2001: A Space Odyssey*, der 1968 in die Kinos kam, ist das Raumschiff *Discovery One* auf dem Weg zum Jupiter. Ausgestattet ist dieses Raumschiff mit einem Bordcomputer namens HAL 9000, der sich im Laufe der Zeit zu einem tödlichen Gegner der Astronauten an Bord entwickelt. Nach einem scheinbaren Fehler des Computers versuchen mehrere Besatzungsmitglieder, HAL abzuschalten, doch zum Schutz der geheimen Mission des Raumschiffs tötet dieser die Besatzungsmitglieder. Schließlich gelingt es dem Astronauten Dave Bowman, HAL zu deaktivieren, auch weil er die verzweifelten Bitten des Computers, damit aufzuhören, ignoriert.

In *The Terminator* (1984) wird der Konflikt zwischen den Menschen und einer sich selbst erhaltenden KI noch etwas dramatischer, er wird zu einer Frage des Überlebens der gesamten menschlichen Spezies.

Viele dieser Motive tauchen auch in anderen Filmen wie *The Matrix* (1999), *I, Robot* (2004), *Transcendence* (2014), *Ex Machina* (2015), *M3gan* (2022) oder *The Creator* (2023) auf. Sie spiegeln eine besondere Angst vor der KI wider, die in jüngster Zeit von bekannten Persönlichkeiten aus der Technologiebranche noch verstärkt wird: dass wir auf einen Konflikt zwischen Mensch und Maschine zusteuern. So vertrat Elon Musk auf dem KI-Gipfel in Bletchley Park die Ansicht, dass KI »eine der größten Bedrohungen für die Menschheit« sei und dass wir zum ersten Mal »mit etwas konfrontiert sind, das viel intelligenter sein wird als wir«. Und in den Augen von Sam Altman von OpenAI könnte generative KI das Ende der menschlichen

Zivilisation herbeiführen; für ihn stellt KI ein Risiko dar, das mit Atomkriegen und globalen Pandemien vergleichbar ist.

Auch in der akademischen Welt hat diese Geschichte einigen Widerhall gefunden. Der Philosoph Nick Bostrom hat sich ausführlich mit den existenziellen Risiken der KI für die Menschheit und einer möglichen Intelligenzexplosion beschäftigt, bei der sich die KI immer weiter verbessert, sobald sie menschliches Niveau erreicht hat. Der Informatiker Stuart Russell hat zusammen mit seinen Mitarbeitern am Center for Human-Compatible Artificial Intelligence an der University of California, Berkeley, das sogenannte *alignment problem* in den Blick genommen, das heißt die Frage, wie sich maschinelle Ziele mit menschlichen Zielen in Einklang bringen lassen.

Eine andere dystopische Geschichte, die fast ebenso beängstigend ist, besagt, dass die KI uns zwar nicht umbringen, aber die menschliche Arbeitskraft überflüssig machen wird, was unweigerlich zu Massenarbeitslosigkeit und sozialen Unruhen führt. In einem Bericht von Goldman Sachs aus dem Jahr 2023 wird beispielsweise behauptet, generative KI könnte in Europa und den Vereinigten Staaten dreihundert Millionen Vollzeitbeschäftigte ersetzen.

Die Geschichten, die in Hollywood und im Silicon Valley erzählt werden, handeln meist von einem heroischen Konflikt zwischen einem Menschen (in der Regel ist es ein Mann) und einer Maschine – Dave Bowman und HAL 9000 in *2001*, Kyle Reese und der Terminator, Nathan und Ava in *Ex Machina* oder Sam Altman und die durch KI verursachte Auslöschung der Menschheit. In der akademischen Version der Geschichte, wie sie von Computerwissenschaftlern erzählt wird, stehen in der Regel ebenfalls ein Mensch und eine Maschine im Mittelpunkt, ob es nun um ein Problem mit der Werteausrichtung (*value-alignment*) der Maschine (das heißt ein falsch definiertes Ziel) geht oder um eine Voreingenommenheit (*bias*) der Maschine in Bezug auf ihr Ziel.

2. Was die alte Geschichte nicht erzählt

Dieses Buch befasst sich mit den Schlüsselfragen, die in der Geschichte von Mensch und Maschine außer Acht gelassen werden: Technologie ist kein Schicksal. Genauso wie Menschen eine Technologie erschaffen, entscheiden Menschen, wie sie eingesetzt wird und welchen Interessen sie dient. Diese Entscheidungen werden wieder und wieder neu getroffen, wenn KI entwickelt und eingesetzt wird. Darüber hinaus ist KI letztlich gar nicht so kompliziert. Wie KI funktioniert, kann jeder nachvollziehen. Der eigentliche Konflikt besteht nicht zwischen Mensch und Maschine, sondern zwischen den verschiedenen Mitgliedern der Gesellschaft. Und die Antwort auf die verschiedenen Risiken und nachteiligen Folgen von KI ist die öffentliche Kontrolle der KI-Ziele mit demokratischen Mitteln.

Im Kern ist KI eine automatisierte Entscheidungsfindung mittels *Optimierung*. Das heißt, dass KI-Algorithmen darauf ausgelegt sind, ein messbares Ziel so groß wie möglich zu machen. Solche Algorithmen könnten zum Beispiel die Anzahl der Klicks auf eine Anzeige maximieren. KI erfordert daher, dass jemand das Ziel – die *Belohnung* – auswählt, das optimiert werden soll. Jemand muss also buchstäblich in seinen Computer tippen: »Das ist das Maß der Belohnung, das uns wichtig ist.«

Die entscheidende Frage ist demnach, wer die Ziele von KI-Systemen bestimmen darf. Nun leben wir in einer kapitalistischen Gesellschaft, und in einer solchen Gesellschaft werden die Ziele der KI in der Regel von den Kapitalbesitzern festgelegt. Die Kapitaleigner kontrollieren die *Prädiktionsmittel*, die für die Entwicklung von KI benötigt werden – Daten, Recheninfrastruktur, technisches Know-how

und Energie. Ganz allgemein werden die Ziele der KI von denjenigen bestimmt, die über gesellschaftliche Macht verfügen, sei es in der Strafjustiz, im Bildungswesen, in der Medizin oder in den Geheimpolizeien autokratischer Überwachungsstaaten.

Ein Bereich, in dem KI in der Gesellschaft eingesetzt wird, ist der Arbeitsplatz. In robotergesteuerten Amazon-Lagern und bei der algorithmischen Lenkung von Uber-Fahrern kommt KI genauso zum Einsatz wie bei der Auswahl von Bewerberinnen durch große Unternehmen. Auch in Bereichen außerhalb des Arbeitsplatzes wird KI eingesetzt, beispielsweise bei der Filterung und Auswahl von Facebook-Feeds und Google-Suchergebnissen mit dem Ziel, die Zahl der Anzeigenklicks zu maximieren. Ein dritter Bereich ist die prognosebasierte Polizeiarbeit (*predictive policing*) und die Inhaftierung von Angeklagten, die auf der Grundlage prognostizierter Straftaten, die sie noch gar nicht begangen haben, vor Gericht kommen. Am verheerendsten ist vielleicht, dass KI auch in der Kriegsführung eingesetzt wird; so wird sie beispielsweise seit 2023 verwendet, um zu entscheiden, welche Wohnhäuser in Gaza bombardiert werden.

Natürlich haben zahlreiche Forscher und Kritikerinnen vor den Gefahren des Einsatzes von KI in diesen wichtigen Bereichen gewarnt. Joy Buolamwini, Informatikerin am MIT Media Lab, hat sich ausführlich mit den Gefahren von ungenauen und rassistisch voreingenommenen Gesichtserkennungssystemen befasst. Ruha Benjamin, Soziologin in Princeton, hat betont, dass KI bestehende soziale Ungleichheiten in Bereichen wie Bildung, Beschäftigung, Strafjustiz und Gesundheitsversorgung replizieren und verstärken kann. In ähnlicher Weise warnte Timnit Gebru, eine Informatikerin, die eine Zeit lang bei Google arbeitete, vor den Gefahren großer Sprachmodelle, die sich wie stochastische Papageien verhalten, die Sprachmuster wiederholen, ohne sie zu verstehen, und dabei die in ihren Trainingsdaten enthaltenen Verzerrungen replizieren. Meredith Whittaker, derzeitige Präsidentin der Signal Foundation, hat die politische Ökonomie der Technologiebranche kritisiert, in der KI von mächtigen Akteuren

auf eine Art und Weise eingesetzt wird, die eine Marginalisierung verstärken kann. Kate Crawford, Professorin an der University of Southern California und Mitbegründerin des AI Now Institute, hat darauf hingewiesen, dass KI in ihrem Wesenskern eine extraktive und ausbeuterische Industrie ist.

Angesichts dieser sich überschneidenden Kritiken, die sich jeweils auf einen speziellen Aspekt und Fallstrick der KI konzentrieren, ist es nicht ganz einfach, eine systematische Sichtweise auf KI in der Gesellschaft zu formulieren. Eine mögliche vereinheitlichende Perspektive bietet die Informatik. Informatiker sind darauf trainiert, die meisten Probleme als Optimierungsprobleme zu betrachten. In diesem Zusammenhang geht es bei der Optimierung darum, die Entscheidung zu finden, die eine gegebene Belohnung so groß wie möglich macht, und zwar bei begrenzten Rechenressourcen und begrenzten Daten.

Die Perspektive der Informatik hat einen Großteil des öffentlichen Diskurses über KI-Sicherheit und KI-Ethik geprägt, vor allem in Bezug auf Themen wie Fairness oder Werteausrichtung: »Wenn etwas falsch läuft, muss ein Optimierungsfehler vorliegen.« So gesehen, besteht das Problem schlicht darin, dass eine Handlung gewählt wurde, die das vorgegebene Ziel nicht maximal erreicht. Diese Sichtweise trifft jedoch in den meisten Fällen nicht den Kern des Problems, weil sie sich nicht mit der Wahl des Ziels selbst befasst.

Ich vertrete die Ansicht, dass das zentrale Problem nicht Optimierungsfehler, sondern Interessenkonflikte über die Kontrolle der KI-Ziele sind. Wenn KI Menschen Schaden zufügt, liegt das Problem meist nicht darin, dass ein Algorithmus nicht perfekt optimiert hat. Das Problem ist vielmehr, dass das vom Algorithmus optimierte Ziel gut für die Menschen ist, die die Mittel der Prädiktion kontrollieren – Menschen wie Jeff Bezos, Gründer und ehemaliger CEO von Amazon, und Mark Zuckerberg, Gründer und CEO von Meta –, aber nicht gut für den Rest der Gesellschaft.

Diese Erkenntnis verändert die Art und Weise, wie wir über mögliche Lösungen für die Probleme der KI nachdenken sollten. Wie