

Finale mündliche Prüfung – Note 5.
Der Prüfling fällt auf die Knie und ruft verzweifelt:
»Oh Jesus, I am going to die!«

KAPITEL I

Wie begründet man eine Entscheidung?

Wissenschaftlerin zu sein ist mein Traumjob. Ich kann mir nichts Besseres vorstellen, als den ganzen Tag zu meinen Themen zu forschen und darüber zu schreiben. Andere Aspekte des Professorinnen-daseins hingegen gefallen mir deutlich weniger: So kommt es von Zeit zu Zeit vor, dass wir eine >finale Prüfung< abnehmen müssen. Das ist die allerletzte Wiederholung einer Prüfung, bei der der Prüfling tatsächlich seine Chance auf einen Abschluss verlieren kann. Es ist sogar noch etwas schlimmer: Werden Studierende an diesem Punkt aus dem Studium ausgeschlossen, können sie auch an anderen Universitäten in Deutschland nicht mehr Informatik studieren. Als ich an die TU Kaiserslautern berufen wurde, habe ich mich daher intensiv mit der Frage beschäftigt, warum wir eigentlich Noten geben und wie ich sie im Einzelfall rechtfertigen kann. Und natürlich war eine der wichtigsten Fragen: Warum lassen wir Studierende überhaupt durchfallen? Welchen Sinn hat das? Denn

»Warum haben Sie mich durchfallen lassen?«

Wie man eine Note begründet

Recht auf faire Entscheidungen,
Recht auf Begründungen

am Ende muss ich meine Notenentscheidung gegenüber dem Prüfling, aber auch gegenüber meinen Kollegen und Kolleginnen jeder-

Heute entscheiden auch
Maschinen über Menschen

zeit begründen können. Der Satz oben ist in einer meiner Prüfungen tatsächlich gefallen – und natürlich habe ich versucht, dem Prüfling zu erklären, wie ich die Entscheidung rechtfertige, denn es ist sein gutes Recht, das zu erfahren.

Und so ist es auch bei vielen anderen Entscheidungen: Bürgerinnen und Bürger haben ein Recht auf eine Begründung und auf Widerspruch. Wir alle haben ein Recht auf eine faire Behandlung, darauf, dass wir nicht nur wegen der Zugehörigkeit zu bestimmten Gruppen schlechter und andere allein wegen ihrer Zugehörigkeit zu einer anderen Gruppe besser behandelt werden – das steht schon im Grundgesetz.

Heute gibt es nun die ersten Maschinen, die menschliches Handeln bewerten. Die meisten dieser Maschinen bezeichnet man heute als »Künstliche Intelligenz« (kurz KI) und verweist auf ihre Erfolge

Ist die Technologie schon
so weit entwickelt, dass sie
verlässlich über Menschen
entscheiden kann?

beim Übersetzen oder bei der Produktempfehlung in Online-Shops. Basierend auf diesen Erfolgen hat sich in Teilen der Informatik und der Wirtschaft die Idee verbreitet, dass Computer sicherlich auch in der Lage sein werden, Menschen in

komplexeren Situationen als beim Einkauf zu bewerten. Diese Position finde ich nach jahrelanger Forschung in diesem Bereich nicht überzeugend. Damit bin ich nicht die Einzige. Trotzdem werden heute entscheidungsberechnende Systeme vielfältig eingesetzt: zum Beispiel Maschinen, die Notenvorschläge für von Schülerinnen und Schülern geschriebene Essays berechnen; Maschinen, die vor Gericht verwendet werden, um die Rückfalligkeitswahrscheinlichkeit von Kriminellen zu bewerten. Maschinen, die Bewerbungen danach bewerten, wie gut die Bewerberin auf den Job passt, oder vorhersagen, ob der Bewerber zukünftig erfolgreich mitarbeiten wird. Auch in Deutschland wurden schon die ersten Pilotprojekte gestartet. Gerade

im Personalwesen passiert es immer häufiger, dass solche Systeme verwendet werden; momentan berate ich dazu mit meinen Mitarbeitern zwei Firmen und eine Karriereplattform. Ich erwarte, dass wir in den nächsten Jahren viele dieser Systeme in Unternehmen, Schulen und Universitäten sowie in der Verwaltung sehen werden. Diese Systeme werden Entscheidungen über Menschen treffen und die betroffenen Personen werden damit nicht immer einverstanden sein. Daher brauchen wir für diese Entscheidungen gute Begründungen. Aber ist das möglich?

1.1 Können KI-Entscheidungen begründet werden?

KI-Systeme sind aus verschiedenen Gründen heute oft sogenannte *Blackbox-Modelle*, also Systeme, in die man nicht so einfach hineinschauen kann. Das liegt teilweise daran, dass die Systeme von Firmen gebaut werden, die sich nicht gern in die Karten sehen lassen. Zumindest bei einflussreichen Entscheidungen brauchen Menschen aber ein Recht darauf, zu verstehen, wieso die Maschine so entschieden hat, wie sie entschieden hat. Ein anderer Grund liegt aber in der Natur der Technologie: Die meisten Methoden der Künstlichen Intelligenz, die heute verwendet werden, sind von sich aus nicht vollständig einsehbar.

Maschinelle Entscheidungssysteme sind *Blackboxes*: undurchdringliche Maschinen.

Und darum geht es in diesem Buch: Um die Frage, unter welchen Bedingungen man Entscheidungen von Computern infrage stellen kann und welche Art von Antworten wir erwarten können. Weil das eine ziemlich komplexe Frage ist, nehme ich Sie in diesem Buch mit auf eine Reise. Eine Reise zu verschiedenen Personen und Geschehnissen, bei denen maschinelle Entscheidungen falsch waren, und bei denen diese Fehlentscheidungen teils nur absurde, teils aber gravierende, ja sogar tödliche Konsequenzen hatten. Inzwischen gibt

es eine ziemlich große Anzahl an Beispielen dafür – und die meisten sind vielschichtig. Bei manchen war wenigstens offensichtlich, dass die Maschine fehlerhaft entschieden hatte. Bei anderen konnte sogar

Wie wollen wir in Zukunft entscheiden?

geklärt werden, wie es dazu kommen konnte. Bei vielen ist es dagegen unklar. Diese Klarheit wäre aber notwendig, insbesondere, wenn wir die Perspektive vom Betroffenen

zum Verwender von automatisierten Entscheidungen wechseln: Wir müssen als Menschen besser verstehen, in welchen Bereichen Maschinen uns bei Entscheidungen unterstützen können. Wenn wir weder als Verwender noch als Betroffene befriedigende Erklärungen bekommen können, sollten wir nicht einzelne Entscheidungen, sondern die Verwendung solcher Systeme infrage stellen. Am Ende wollen wir wissen, wie es zu einer falschen Entscheidung kam: War's die KI?

Teil I startet mit dem Beispiel einer Frau, die von einer Maschine deutlich schlechter behandelt wurde als ihr Ehemann. Der ließ das nicht auf sich sitzen und mobilisierte mit seinen verärgerten Tweets auf Twitter Tausende seiner Follower. Um zu verstehen, warum er denkt, dass der betreffende Algorithmus sexistisch entscheide, gibt es in diesem Teil des Buches eine Einführung in *Das kleine ABC der Informatik*: **A** wie Algorithmus, **B** wie Big Data und **C** wie Computerintelligenz.

Im zweiten Teil erfahren Sie, was Menschen in Situationen passiert ist, in denen im Wesentlichen klar war, dass die Maschine eine falsche Entscheidung getroffen hatte. Aber wie konnte es dazu kommen? Die große Bandbreite der Erklärungen dafür zeigt, wie verwickelt die Auflösung sein kann. Das ist ähnlich wie bei einem »Whodunit«, einem klassischen Kriminalroman, bei dem der Mord am Anfang passiert und dann die Deduktion zum Täter oder zur Täterin führt. Aber anders als bei einer gut konzipierten Detektivgeschichte ist das reale Leben meistens vielschichtiger: Hier gibt es nicht nur einen Täter, sondern meistens mehrere Fehler von beteiligten Personen.

Schwieriger wird es natürlich bei automatisierten Entscheidungen, die nicht direkt als fehlerhaft erkannt werden können. Je schwieriger das ist, desto wichtiger ist es, dass wir den Prozess hinter der Entscheidung verstehen. Darum geht es in Teil III, der zeigen wird, dass es fundamentale Gründe dafür gibt, warum wir maschinelle Entscheidungen nicht verstehen können. Das ist ein gravierender Nachteil. Trotzdem bieten automatisierte Entscheidungen oft auch Vorteile. In Teil IV diskutiere ich daher, wie wir als Menschheit in Zukunft darüber entscheiden könnten, wann wir unsere Entscheidungen allein und wann wir sie mit maschineller Unterstützung treffen wollen.

Und warum gibt es in diesem Buch so viele Beispiele? Weil die Frage danach, wann wir automatisierte Entscheidungen über Menschen nutzen wollen, mit uns allen zu tun hat. Es ist eine gesellschaftliche Frage, in der jede und jeder gehört werden sollte. Und weil Fehlentscheidungen uns so in Rage bringen können – es trifft uns, wenn wir das Gefühl haben, ungerecht behandelt zu werden. Teil I beginnt daher mit der Geschichte des amerikanischen Softwareentwicklers David Heinemeier Hansson, der nach einer Computerentscheidung, die sich eigentlich nicht gegen ihn, sondern gegen seine Ehefrau gerichtet hatte, in den »Twitterkampf« gegen Apple zog. Es ist ein Beispiel dafür, welche starken Emotionen erzeugt werden können, wenn über Menschen mithilfe von Software ohne weitere Begründung und ohne Möglichkeit des Widerspruchs geurteilt wird. Und eine Geschichte mit einer interessanten Auflösung. Bühne frei!

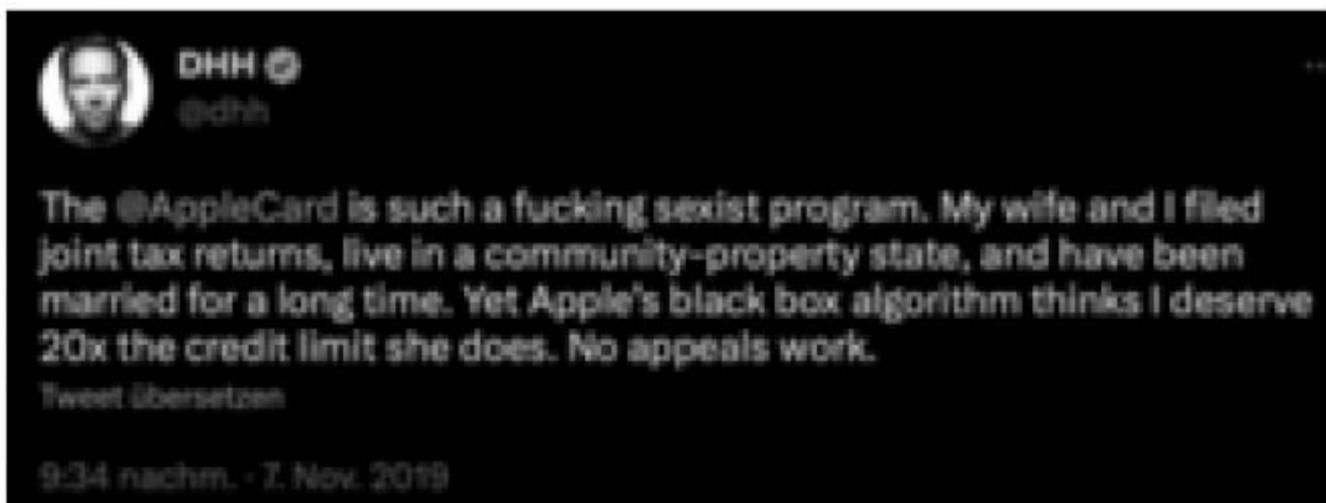
TEIL I

Wie Maschinen zu ihren Entscheidungen kommen

KAPITEL 2

Apples Pay Card: Sexismus pur?

Im November 2019 beantragten David Heinemeier Hansson und seine Frau Jamie jeweils eine Apple Card. Und weil Apple die Firma ist, die sie ist, ist die Beantragung der Kreditkarte sehr einfach und komplett digital: Man kann sie direkt über ein Apple-Gerät beantragen, und sobald die AGB anerkannt wurden, bekommt der Nutzer eine Entscheidung, ob er eine Karte bekommt und wo die Kreditlinie liegt. Das geht so blitzschnell, dass dahinter natürlich eine Software-Entscheidung steht. Und siehe da: Seine Frau Jamie Heinemeier Hansson bekam einen Kreditrahmen von um die 50 US-Dollar und ihr Ehemann David einen um die 1000 US-Dollar. Und damit begann der Ärger von David Heinemeier Hansson. Aber das kann er selbst Ihnen besser erzählen als ich:¹



Ich konnte nach diesen starken Worten einfach nicht anders, als mir Heinemeier Hansson als Ritter vorzustellen, der seiner Ehefrau und allen anderen Frauen beispringt:

The @AppleCard
is such a
* 7 8 9 0 *
Sexist program.



*Alter gallischer Fluch



... leben in einer
Zugewinnge-
meinschaft...



... und sind seit
Jahren
verheiratet.

Meine Frau und ich
zahlen gemeinsam
Steuern...



Und trotzdem denkt
Apples Blackbox-
Algorithmus, dass ich
20x mehr Kredit verdiene
als sie.



Widerspruch
zwecklos!





Wir sehen ihn also jetzt in seiner polierten Rüstung, das Schwert zur Verteidigung gezogen – und plötzlich geht sein Tweet viral. Denn David Heinemeier Hansson ist ein interessanter Typ – Rennfahrer, Softwareentwickler, Autonarr, dem auf Twitter viele Personen folgen: Über 11 000 Mal wird der Tweet am Ende weiterverbreitet, über 25 000 Mal klicken die Leser und Leserinnen auf »like«. Die Autorin Safiya Noble, die das Buch »Algorithms of Oppression« schrieb, twitterte beispielsweise: »This is an algorithm of oppression if I've ever seen one.« Völlig überraschend mischte sich dann auch Steve Wozniak in die Unterhaltung ein, der sich demütig wie folgt vorstellte: »Ich bin momentan Apple-Mitarbeiter und Gründer der Firma.« Er bestätigte, dass auch seine Ehefrau einen zehnfach geringeren Kreditrahmen erhalten hatte, obwohl sie all ihr Vermögen teilen würden, und ermahnte Apple, hier Verantwortung zu zeigen. Dieser Kommentar war eine gewichtige Unterstützung von Heinemeier Hanssons Anliegen.

Andere Kommentare dagegen erzürnten Heinemeier Hansson dagegen fast so sehr wie die Entscheidung des Algorithmus. Dazu gehören insbesondere die drei folgenden Argumentationslinien:

1. »Aber das ist ja gar nicht Apple selbst, sondern Goldman Sachs, also die Bank, die die Apple Pay Card in Zusammenarbeit ausstellt. Warum Apple an den Pranger stellen?«
2. »Hat deine Frau nicht vielleicht einfach eine schlechtere Kreditwürdigkeit als du?«
3. »Wieso denkst du, dass das sexistisch ist, könnte doch auch andere Gründe geben?«

Nun, all das lässt Heinemeier Hansson natürlich nicht auf sich sitzen. Das erste Argument ist tatsächlich keines: Denn wer nun genau den Algorithmus verantwortet, ist unerheblich für die Frage, ob er funktioniert, wie er soll. Das Beispiel ist damit auch eine Warnung für Unternehmen, die Software nutzen, um Entscheidungen automatisch zu berechnen: Für die Kunden und Kundinnen ist am Ende immer in erster Linie die Person verantwortlich, die den Algorithmus nutzt – unabhängig davon, wer ihn entwickelt hat. Das ist ein wichtiger Punkt, den ich später nochmals aufgreifen werde: Der Verwender eines KI-Systems für einen sozialen Prozess ist die Person, an die sich die Kunden und Kundinnen nachher wenden werden – unabhängig davon, ob er die eigentliche Entscheidungslogik dahinter verantwortet oder nicht.

Die zweite Frage bringt Heinemeier Hansson dazu, dass sich beide für schlappe 25 US-Dollar pro Monat (!) bei TransUnion anmelden, um ihren Kreditwürdigkeits-Score zu bekommen, den sogenannten FICO-Score – das ist im Wesentlichen dasselbe wie in Deutschland der Schufa-Score. Und, hey, Heinemeier Hansson ist immer noch auf 180: »Also melden wir uns natürlich beide wutentbrannt für verdammte 25 Dollar pro Monat bei der Krediteinsichtsschwachsinnabzocke namens TransUnion an. Vielleicht hat ja jemand die Identität meiner Frau gestohlen? Obwohl wir vorher schon einmal geprüft hatten, dass alles okay ist. Aber, ratet mal: IHRE KREDITWÜRDIGKEIT IST SOGAR HÖHER ALS MEINE!!!«²

Sein Ärger ist natürlich irgendwie nachvollziehbar, aber mir persönlich gefällt diese Frage der Twitternutzer nach ihrer Kreditwürdigkeit,

denn tatsächlich muss es in den Eingangsdaten, die eine Software bekommt, irgendeinen Unterschied bei den beiden Personen geben.

Denn zumindest darauf können wir uns bei Software verlassen: Wenn genau dieselben Daten hineingehen, dann kommt auch genau dasselbe dabei heraus.³

Die Frage danach, was sich bei den beiden Personen unterscheidet, ist also erst einmal in Ordnung und auch sehr wichtig – und dass es der FICO-Score sein könnte, liegt auf der Hand. Ärgerlich ist, dass Heinemeier Hansson und seine Frau nicht einfach direkt von Goldman Sachs Einsicht in die jeweils über sie verwendeten Eingangsdaten und deren Werte bekamen. Denn es sollte nicht so sein, dass die bewertete Person weder weiß, welche Daten relevant sind, noch ob diese Daten vielleicht fehlerhaft sind. Heinemeier Hansson und seine Frau sind jetzt nicht schlauer als zuvor: An dieser Stelle unterscheiden die beiden sich zwar, aber tendenziell würde man erwarten, dass der Kreditrahmen für seine Frau hätte höher ausfallen sollen als seiner.

Auch der dritte Kommentar stellt eine valide Frage, da wir oft mit »Sexismus« die absichtsvolle Schlechterstellung von Menschen wegen ihres Geschlechts meinen. Daneben gibt es aber die unbeabsichtigte Schlechterstellung, die ebenfalls manchmal mitgemeint ist: gerade in den USA unterscheidet man im Bürgerrechtsgesetz zwischen intendierter Diskriminierung (*disparate treatment*) und der Schlechterstellung von Gruppen (*disparate impact*) mit oder ohne Intention. Auch Letzteres ist laut *Title VI* des Bürgerrechtsgesetzes in den USA beispielsweise für alle Programme verboten, die staatliche Förderung erhalten.⁴ Um Schlechterstellung nachzuweisen, reicht es dabei aus, zu zeigen, dass eine Gruppe basierend auf einem Merkmal, das gesetzlich geschützt ist, statistisch signifikant schlechtergestellt ist. Die Frage, ob diese Schlechterstellung durch eine gewollte Diskriminierung passiert, oder durch eigentlich neutral formulierte Gesetze und

Prozesse geschieht, ist dabei erst einmal unerheblich – die staatliche Förderung wird gestrichen. Diese amerikanische Idee von *disparate impact*, der Schlechterstellung von Gruppen, ist es, die David und Jamie Heinemeier Hansson hier antreibt. Jamie Heinemeier Hansson schreibt in einem Blogpost, dass sie eigentlich Privatheit sehr schätzt und sie darum mit ihrem Mann beschlossen hat, dass er ihre Apple-Card-Geschichte erzählt. Als Millionärin ist es für sie nicht so relevant, dass ihr persönlicher Kreditrahmen erhöht wird.⁵ Es geht ihr um etwas anderes: »Es ist relevant für die Frau, die gerade ein Geschäft eröffnet in einer Welt, die immer noch denkt, dass Frauen weniger erfolgreich und kreditwürdig sind als Männer. Es ist relevant für die Ehefrau, die versucht, aus einer gewalttätigen Beziehung auszubrechen. Es ist relevant für Minderheiten, die durch systemische Vorurteile beeinträchtigt werden.«⁶ Aufgrund des hohen Medienechos wurde Jamie Heinemeier Hanssons Kreditrahmen schleunigst hochgesetzt – aber aus den genannten Gründen reichte ihr das nicht: »Das ist nicht einfach eine Geschichte über Sexismus und Blackbox-Kreditvergabealgorithmen, sondern darüber, dass Reiche immer irgendwie ihren Kopf durchsetzen. Gerechtigkeit für irgendeine reiche weiße Frau ist überhaupt keine Gerechtigkeit.«

Das Ehepaar bezieht sich hier also auf *disparate impact*, und daher lohnt es sich, die Bedeutung dieses Begriffes etwas näher anzugucken. Ein Handbuch zum Umgang mit dieser Art von statistischer Ungleichbehandlung gibt ein interessantes Beispiel für einen ungewollt schlechterstellenden Prozess: An einer Schule müssen alle Schüler und Schülerinnen, die zu spät kommen, zum Gespräch zur Schulleiterin. Damit verpassen sie natürlich noch mehr Unterricht als durch das Zuspätkommen allein. Es stellt sich nun heraus, dass eine Gruppe, die asiatisch-amerikanischen Kinder, viel öfter zur Schulleiterin muss als andere Kinder. Eine genauere Untersuchung ergibt, dass diese Kinder weiter weg von der Schule wohnen und selbst der früheste Schulbus oft zu spät zur Schule kommt. Da es ja eigentlich um einen Schutz des Unterrichts geht, der durch Disziplinlosigkeit entsteht, die Schüler und Schülerinnen für ihre Verspätung aber nichts

können, gilt das als *disparate impact*. Daher muss der Prozess laut amerikanischem Recht hier geändert werden: Zum Beispiel könnte die Regel für Kinder, die aufgrund einer Verspätung des Schulbusses zu spät kommen, nicht gelten, oder die Schule sorgt dafür, dass die Schulbusse früher losfahren.

Für uns in Deutschland ist diese statistische Sichtweise wie gesagt eher ungewohnt – illegale Diskriminierung bezeichnet bei uns die sachgrundlose Schlechter- oder Besserbehandlung von Personen aufgrund einiger gesetzlich geschützter Merkmale. Das entspricht eher dem *disparate treatment*. Zudem ist es eher ein Akt von Personen oder Institutionen: Man muss als Geschädigter den Verursacher nennen können und dessen Gesinnung wird dann geprüft. Im amerikanischen Recht reicht dagegen der Nachweis der Schlechterstellung aus – warum das so ist und wie man das abstellen kann, ist dann die Sache desjenigen, der bewusst oder unbewusst schlechterstellt. Eine Absicht muss nicht nachgewiesen werden.

Der statistische Aspekt, der unabhängig vom Nachweis einer Intention einfach berechnet werden kann, kommt daher aus meiner Sicht hauptsächlich aus den USA. Auch andere Diskussionen zeigen, dass das, was als Diskriminierung gilt und wie man es nachweisen kann, auch eine kulturelle Frage ist. Als Informatikerin liegt mir der statistische Ansatz natürlich – da können wir als Data Scientists leicht weiterhelfen. Ich persönlich finde die Frage nach der Intention zwar auch wichtig, aber zuerst sollte der Befund gemacht werden und für Abhilfe gesorgt werden: Für die Betroffenen sind die Auswirkungen ja zuerst einmal unabhängig davon, ob die Schlechterstellung absichtlich oder unabsichtlich geschieht. Und in diesem Sinne ist es auch interessant zu wissen, ob der Algorithmus hinter der Kreditwürdigkeitsbestimmung statistisch gesehen Frauen schlechterstellt als Männer, und damit sexistisch entscheidet.

Was Heinemeier Hansson aber am meisten erzürnt, sind die vielen Leute, die ihm erklären wollen, dass das halt die Art und Weise ist, in der Kredite vergeben werden. Er prangert an, dass es keine Stelle gibt,

bei der man nachfragen kann, was genau die Eingabe in diesen Algorithmus ist – und ob sich dabei Falschinformationen eingeschlichen haben könnten. Ihn verärgert, dass ihm und seiner Frau keiner sagen kann, wie genau man von den Eingabedaten zur Entscheidung kommt oder wo die Daten seiner Frau von seinen eigenen abweichen, um die Diskrepanz im Kreditrahmen zu erklären. Am Ende hatten sie insgesamt mit sechs Mitarbeitern von Apple bzw. Goldman Sachs gesprochen und keiner konnte ihnen erklären, wie genau die Entscheidung abläuft. Aber jeder sagte: »Natürlich diskriminieren wir niemanden! Das ist einfach nur der Algorithmus!« Mit »das ist einfach nur der Algorithmus« speist man einen Heinemeier Hansson auf jeden Fall nicht ab. Er fasst die Situation in einem weiteren Tweet zusammen:

»Es versteht also niemand DEN ALGORITHMUS. Niemand kann DEN ALGORITHMUS untersuchen oder überprüfen. Und trotzdem ist jeder, mit dem wir von Apple oder Goldman Sachs sprechen können, VÖLLIG ÜBERZEUGT, dass DER ALGORITHMUS weder voreingenommen ist noch in irgendeiner Form diskriminiert. Das ist eine großartige Zurschaustellung von kognitiver Dissonanz.«⁷

Und wissen Sie was? Er hat damit völlig recht: Wenn man nicht weiß, wie eine Entscheidung zustande kommt, kann man nicht einfach sagen: »Aber es ist doch einfach der Algorithmus.« Das ist so, als würde man eine Entscheidung damit begründen, dass diese Entscheidung einem klar definierten Prozess folgt – denn für nichts anderes steht in diesem Zusammenhang das Wort Algorithmus.

Gehen wir noch einmal zurück zu meinem Prüfling, der durchgefallen ist: Der Algorithmus, mit dem ich seine Note berechnet habe, bestand darin, die Noten für die verschiedenen Prüfungsthemen zusammenzurechnen und den Durchschnitt auszurechnen:

Schritt 1: Rechne die Zahlen zusammen.

Schritt 2: Teile durch die Anzahl der Zahlen.

Das ist doch keine Begründung für das Ergebnis »durchgefallen«, schon gar nicht in einer finalen Prüfung! Dass hinter einer Berechnung ein Algorithmus steht, ist offensichtlich keine Begründung dafür, dass das Resultat tatsächlich die Prüfungsleistung des Teilnehmers widerspiegelt. Vielmehr muss der Prozess selbst sinnvoll sein und auch im richtigen Moment verwendet werden. Der kurze Algorithmus zur Berechnung des Durchschnitts ist zum Beispiel definitiv korrekt – aber wenn ich in der Prüfung sehr lange ein Thema behandelt hätte, das der Student gut konnte, hätte man vielleicht dieses Thema doppelt gewichten sollen: Der Algorithmus selbst wäre richtig, aber für die Situation nicht angemessen. Daher kann der reine Verweis der Bank auf »den Algorithmus« nicht ausreichen: Es muss vielmehr klar sein, dass der Algorithmus in sich sinnvoll und der Situation angemessen ist. Genauso wenig kann die Tatsache, dass ein Kreditrahmen mit einem Algorithmus berechnet wird, eine Garantie dafür sein, dass das Resultat eine passende, stimmige Lösung für das Problem der Kreditwürdigkeitsbewertung ist. Und ganz bestimmt kann damit auch die Qualität der Entscheidung nicht beurteilt werden.

Ich werde auf diesen Fall zurückkommen, denn tatsächlich dauerte die offizielle Untersuchung fast anderthalb Jahre. In dieser Zeit konnten Menschen, denen unklar war, wie es zu der Einschätzung der Software kam, ihre Informationen einschicken und sie neu bewerten lassen. Die Überprüfung wurde dann manuell durchgeführt, was naturgemäß lange dauert. Daher bitte ich um etwas Geduld, bis ich Ihnen die Auswertung der Apple Pay Card/Goldman-Sachs-Affäre mitteile. Denn um zu verstehen, wie man die Qualität einer maschinellen Entscheidung bewerten und im Einzelfall infrage stellen kann, möchte ich erst einmal darstellen, wie Maschinen überhaupt Entscheidungen berechnen.⁸

Im ersten Teil des Buches »Wie Maschinen zu ihren Entscheidungen kommen«, beginne ich mit dem »kleinen ABC der Informatik« – ein kurzer Abriss über A wie **A**lgorithmus, B wie **B**ig Data, und C wie

Computerintelligenz. Wer das in aller Tiefe noch einmal nachlesen will, der findet mehr dazu in meinem Buch »Ein Algorithmus hat kein Taktgefühl«. Ich werde hier aber einen Punkt betonen, der für dieses Buch besonders wichtig ist: Dass hinter allen Algorithmen erst einmal ein Modell im Kopf seiner Entwicklerinnen und Entwickler steht, bevor die Maschine dann ebenfalls ein statistisches Modell berechnet, auf dessen Grundlage alle maschinellen Entscheidungen beruhen. Und diese beiden Modelle – das menschliche und das maschinelle – müssen wir in den meisten Fällen verstehen, um eine maschinell berechnete Entscheidung zu verstehen und ihr zu vertrauen.

KAPITEL 3

Das kleine ABC der Informatik

A wie Algorithmus In den Debatten um die Künstliche Intelligenz ist immer viel von »dem Algorithmus« die Rede und selten wird genau gesagt, was damit eigentlich gemeint ist. Dabei benutzen Sie alle im Alltag andauernd Algorithmen, es hat Ihnen nur niemand gesagt, dass Sie das tun. Ich spiel zum Beispiel sehr gerne Doppelkopf und freue mich jetzt schon darauf, das im Sommer mit meinem Bruder und meinen Eltern bei einem gemeinsamen Urlaub mal wieder zu tun. Der erste Schritt besteht bei allen Kartenspielen darin, die Karten aufzunehmen und für sich zu sortieren. Ich mache das, indem ich jede Karte einzeln aufnehme, und dann die nächste Karte immer schon an die richtige Stelle setze – danach sind die Karten dann definitiv so sortiert, wie ich das gerne hätte. Andere nehmen einfach alle Karten auf einmal auf und sortieren dann so lange um, bis alles seine Richtigkeit hat. Am Ende können beide Methoden dazu genutzt werden, um die Karten zu sortieren. Informatiker würden an dieser Stelle sagen: Beide Methoden lösen das »Sortierproblem«, sie sind Algorithmen für das Sortierproblem.

Algorithmiker, also Menschen, die Algorithmen entwerfen, sind Problemlöser. Ein Problem ist dabei eine allgemeine Beschreibung von einer Ist-Situation und einer Soll-Lösung. Die Ist-Situation besteht beim Sortierproblem darin, dass jemand eine Reihe von Dingen in irgendeiner Reihenfolge vorliegen hat. Die Soll-Lösung besteht darin, dass die Software eine Reihenfolge der Dinge nennen soll, sodass sie nach einem vorher festgelegten Kriterium sortiert sind.

Solche Probleme formulieren wir Algorithmiker sehr gerne ganz allgemein: Man kann man ja nicht nur Spielkarten sortieren, sondern auch Dokumente, z. B. nach Datum, oder Bücher nach ihren Autoren und Autorinnen. Kinder sortieren ihre Kuscheltiere vielleicht nach Flauschigkeit oder einfach nach Liebe. Sortieralgorithmen ist es ziemlich egal, was genau nach welchem Kriterium sortiert werden soll. Sie lösen das ganz allgemeine »Sortierproblem«: Wenn ich ihm eine Menge von Dingen gebe und sage, wann ein Ding »besser« (oder »aktueller« oder »flauschiger«) ist als das andere, garantiert mir ein Algorithmus, dass er diese Dinge danach in genau der gewünschten Reihenfolge angeordnet hat.

Ein Algorithmus ist dabei erst einmal einfach ein »Verfahren«. Aber nicht jedes Verfahren ist auch ein Algorithmus: Man könnte ja zum Beispiel die Spielkarten einfach in die Luft schmeißen, auf einen Haufen schieben und dann hoffen, dass die neue Reihenfolge zufällig die gewünschte ist – das ist ein Verfahren, aber kein Algorithmus. Denn offensichtlich könnte dabei einmal die richtige Sortierung herauskommen, aber es könnte auch passieren, dass man einfach Pech hat und die gewünschte Reihenfolge nie eintritt – auf jeden Fall nicht zu Lebzeiten der Person, die die Dinge gerne sortiert hätte.

Als »Algorithmus« bezeichnen wir in der Informatik daher nur solche Verfahren, die bewiesenermaßen in endlicher Zeit das Problem lösen. Wir müssen also beides beweisen: dass das Verfahren wirklich die richtige Lösung finden kann und dass es das auch wirklich tun wird.

Für das Sortierproblem gibt es etliche, sogar Dutzende Verfahren, die genau das leisten. Und viele davon haben Sie garantiert heute schon verwendet, ohne es zu wissen. Wenn Sie heute schon E-Mails gelesen haben, dann wurden diese für Sie sortiert, z. B. nach dem Datum, an dem Sie sie erhalten haben. Als Sie auf eine E-Mail geantwortet haben, wurde Ihre Antwort in viele kleine Päckchen unterteilt und über

das Internet an den Empfänger oder die Empfängerin gesendet – und dort zuerst sortiert und dann wieder zusammengesetzt. Ihre favorisierte Social-Media-Plattform hat für Sie Einträge von anderen bewertet und dann so sortiert, dass Sie möglichst lange auf der Plattform bleiben.

B wie Big Data Aber genau da fängt jetzt auch schon die Frage an: Die Sortierung an sich mag fehlerlos sein, aber wie genau kommen denn die Bewertungen auf den Social-Media-Plattformen oder die Bewertungen der Kreditwürdigkeit zustande? Und hier betreten wir nun den Bereich von Big Data und der »Computerintelligenz«: Gerade auf Social-Media-Plattformen werden den ganzen lieben Tag lang Informationen mitgeloggt, d. h. digital gespeichert: Was haben Sie angeklickt, wie lange war der Post auf dem Bildschirm zu sehen, haben Sie interagiert mit dem Post, also ihn beispielsweise weitergeleitet oder geliket? Über welche Vorschläge haben Sie drüberge-scrollt oder draufgeklickt, sind aber sofort wieder zurückgesprungen? All diese Informationen werden interpretiert als »Interesse« oder »Desinteresse«. Dabei muss natürlich ein Klick auf einen Link oder Post im Einzelfall nicht viel bedeuten: Vielleicht haben Sie sich verklickt oder wollten einem Freund etwas zeigen, das ihn interessiert. Oder Ihre Freundin hat Ihnen etwas geschickt, das Sie nur aus Höflichkeit ihr gegenüber anklicken. Die Idee ist aber, dass in der riesigen Datenmenge, die dabei entsteht, Sie doch meistens die Themen anklicken und sich mit den Posts beschäftigen, die Sie interessieren. Und das ist genau die Idee von »Big Data«: Daten zu verwenden, die quasi »nebenbei« anfallen, die im Einzelfall auch nicht unbedingt interessant sein müssen, aber in der Summe statistisch auswertbar sind und dann interessante Muster aufweisen könnten. Zusätzlich könnten diese Daten auch noch von verschiedenen Plattformen oder anderer Internetnutzung zusammengezogen werden.

Insgesamt versteht man unter »Big Data« die Nutzung von sehr großen Datenmengen, die statistisch ausgewertet werden, weil sie im Einzelfall fehlerhaft oder nicht sehr aussagekräftig sind. Die Aspekte der Vielfältigkeit der Datenquellen und der großen Geschwindigkeit, mit der die Daten anfallen, werden auch häufig noch genannt.

C wie Computerintelligenz Dass es immer mehr dieser Daten gibt, hat in den letzten 30 Jahren dazu geführt, dass die Methoden des sogenannten »maschinellen Lernens« immer populärer wurden. Das maschinelle Lernen gehört zur Künstlichen Intelligenz und führt zu den vielen Erfolgen, die es in den letzten Jahren gab: Zum Beispiel basieren Sprachassistenten wie Siri oder Alexa auf maschinellem Lernen, genauso wie das Übersetzerprogramm der deutschen Softwarefirma DeepL und der Chatbot ChatGPT, der Ende 2022 viel von sich reden machte. Der Begriff Künstliche Intelligenz ist dabei mehrfach besetzt: Auf der einen Seite meint man damit das Forschungsfeld, das Methoden entwickelt, um Computer Dinge machen zu lassen, für die Menschen Intelligenz brauchen würden. Auf der anderen Seite nennt man die Methoden selbst auch manchmal »Künstliche Intelligenz«. Und nicht zuletzt, als wäre nicht alles schon verwirrend genug, nennt man insbesondere auch die Software, die diese Methoden verwendet, ebenfalls »Künstliche Intelligenz«.

Das Forschungsfeld ist jedenfalls schon ziemlich alt: In den 1950ern waren einige Wissenschaftler sehr optimistisch, dass mit der beginnenden Digitalisierung die Computer schnell in der Lage sein sollten, menschliche Sprache, menschliche Sichtweisen und menschliches Denken zu verstehen. Die Methoden, die zuerst entwickelt wurden, waren dabei oft regel- und wissensbasiert, d. h., man versuchte, das Wissen der Welt in computerverdaulicher Weise darzustellen. Dieser Weg führte aber insbesondere im Bereich Sprachverständnis und Bilderkennung nicht zu den gewünschten Ergebnissen;